

MACQUARIE UNIVERSITY

DOCTORAL THESIS

---

**Virtual Acoustic Environments for the  
Evaluation of Hearing Devices**

---

Christos Oreinos  
Diplomate Engineer

A thesis submitted in fulfilment of the requirements  
for the degree of Doctor of Philosophy at the

National Acoustic Laboratories

&

Department of Linguistics | Faculty of Human Sciences

© **Christos Oreinos**

“Virtual Acoustic Environments for the Evaluation of Hearing Devices”

Department of Linguistics | Faculty of Human Sciences | Macquarie University

National Acoustic Laboratories

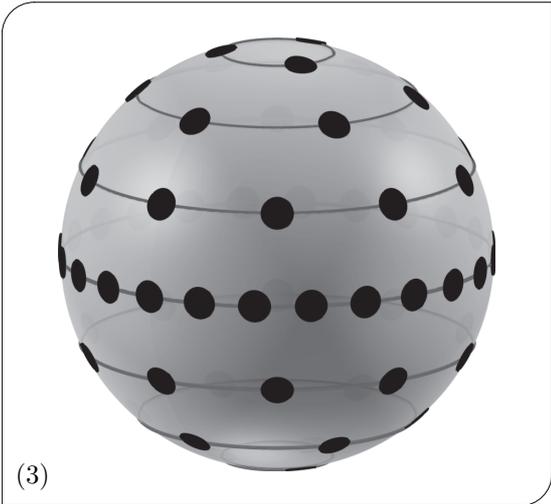
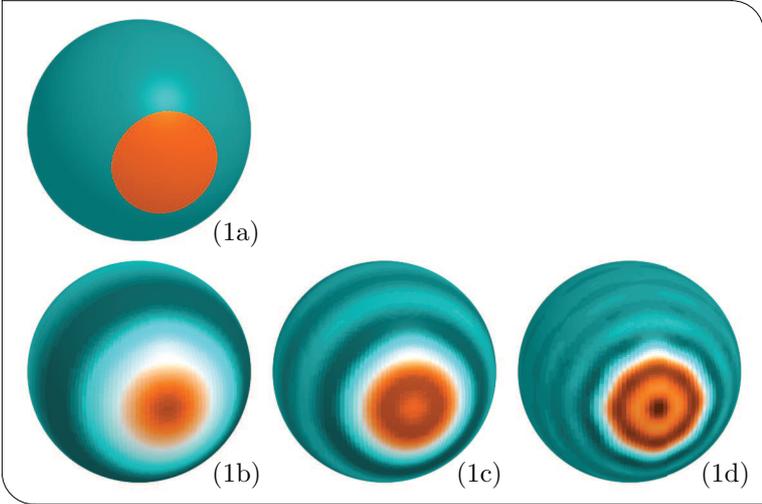
**Supervisors:**

Dr. Jörg M. Buchholz

Dr. Jorge Mejia

Sydney, Australia

March 2015



---

Images on previous page:

- (1a) A discontinuous distribution on the sphere.
- (1b) 4<sup>th</sup> order SHF expansion of above distribution. Note the Gibbs oscillations near the discontinuity.
- (1c) 10<sup>th</sup> order SHF expansion of above distribution. — ” —.
- (1d) 20<sup>th</sup> order SHF expansion of above distribution. — ” —.
- (2) 3D panning functions of a 4<sup>th</sup>/7<sup>th</sup> mixed-order Ambisonics system [artistic colour shading].
- (3) The 62-ch *simulated* National Acoustic Laboratories microphone array.
- (4) The 41-ch National Acoustic Laboratories loudspeaker array.
- (5) Hearing aid satellites used for the speech-in-noise test of Chapter 6.
- (6) Room setup used for the speech-in-noise test of Chapter 6.

# Declaration of Authorship

I, Christos Oreinos, declare that this thesis titled: “*Virtual Acoustic Environments for the Evaluation of Hearing Devices*” and the work presented in it are my own. Specifically, I confirm that:

- This work was conducted exclusively while in candidature for a PhD degree at Macquarie University.
- This thesis has been submitted solely to Macquarie University for consideration for the doctoral degree.
- Where I have cited or quoted the published work of others, this is always clearly attributed. With the exception of such referencing, this PhD thesis is entirely my own work.
- My principle supervisor, Dr. Jörg M. Buchholz, has assisted me in improving the applied methods, in better interpreting the results and in improving the structure and writing style of the manuscripts that have been submitted for publication. Credit has accordingly been given in the Acknowledgement section.
- I have additionally acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have identified my contribution.
- Ethics approval for the experiment described in Chapter 6 has been obtained from the Australian Hearing Human Research Ethics Committee, bearing the approval number: AHHREC2014-3. The signed approval can be found in Appendix C.

Signed: \_\_\_\_\_



Date: \_\_\_\_\_

20 March 2015



# *Abstract*

---

Listening tests are essential for optimising and evaluating novel signal processing concepts for hearing devices. When they are performed inside the laboratory they are highly controlled, but typically lack ecological validity. On the other hand, field-studies provide high ecological validity, but are difficult to control, time-consuming and costly. Even though field-studies may ultimately be required, there still is a necessity for more realistic laboratory-based listening tests.

This thesis focuses on creating and validating realistic virtual sound environments (VSEs), primarily based on the method of higher-order Ambisonics (HOA), for testing hearing aids (HAs). Particular emphasis has been given on how the VSE limitations affect the output of multi-microphone directional HAs. To systematically and objectively analyse the errors incurred by such VSEs, we developed and verified a framework that consists of: (1) the simulation of a reverberant acoustic scene; (2) the coding of that scene using HOA; (3) the reconstruction of the VSE via a loudspeaker array; and (4) the evaluation of the SNR benefit and output pressure of directional HAs worn on an acoustic manikin placed inside the VSE. A listening experiment was additionally designed to test the simulation results. For that purpose, a real-room ‘cocktail-party’ acoustic scene was created and acoustically modelled. Ambisonics was then applied to reproduce that scene inside a loudspeaker array. Listening tests involving hearing-impaired subjects, fitted with directional HAs, were next conducted inside the real and the virtual sound environments. The intelligibility and acceptable-noise-level benefits of the directional HAs were thereby estimated.

The above evaluation illustrated that most of the real-environment properties can be faithfully reconstructed by the considered VSEs. Moreover, directional HA algorithms were shown to behave and adapt in a similar way as in real environments. However, sound-field reconstruction deviations were noticeable. For the case of highly directional schemes, such as the considered bilateral beamformer, such deviations may result in a reduced benefit estimation. Hence, we can conclude that although the described VSEs illustrated a reduced sensitivity in demonstrating real-life benefits, they are nonetheless suitable for evaluating directional HA processing schemes.

«Μα η δυσκολία στάθηκε πάντα στη ζωή ο μέγας ερεθισμός που ξυπνάει και κεντρίζει όλες μας τις ορμές, καλές και κακές, για να πηδήξουμε το εμπόδιο που υψώθηκε ξαφνικά μπροστά μας· κι έτσι, επιστρατεύοντας όλες μας τις δυνάμεις, που αλλιώς θα ύπνωναν ή θα ενεργούσαν σχόρπια κι ανόρεχτα, φτάνουμε κάποτε πολύ μακρύτερα απ' ότι ελπίζαμε.»

—Νίκος Καζαντζάκης, Αναφορά στο Γκρέκο

*“Hardship, however, has always been life’s stimulant, awakening and goading all our impulses, both good and bad, in order to make us overleap the obstacle which has suddenly risen before us. Thus, by mobilising all our forces, which otherwise would have remained asleep or acted in a wayward and languishing manner, we finally reach a point much further than we had hoped.”*

—Nikos Kazantzakis, *Report to Greco*

# *Acknowledgements*

---

During the past three-and-a-half years I have gained some of the most enriching and life-changing experiences to date. This is the place to thank the people that have contributed, in one way or another, in those adventures of mine.

Starting off, I want to express my deepest gratitude to Jörg Buchholz, my principle supervisor. Jörg put confidence in my skills as well as my motives to make a career shift, back in 2011, with a view to pursuing a research path. He supported me and finally admitted me as his first PhD student at his new position at the National Acoustic Laboratories in Sydney, Australia. Furthermore, Jörg guided my first steps in research with enthusiasm and patience by devoting ample time during both scheduled and *ad hoc* meetings. His contribution has also been instrumental in improving the applied research methods, in better interpreting the results and in enhancing the structure and writing style of this thesis. It has been a great pleasure to cooperate with Jörg on a day-to-day basis, even when he ruthlessly suppressed my –perhaps failed– attempts at embellishing research writing with elements from literature. (Jörg, forgive me for reverting to my personally-preferred style of writing for this section!)

I also wish to extend my sincere thanks to Jorge Mejia, my associate supervisor. Jorge has been a source of constant influence and guidance in attaining hands-on skills relevant to fitting, measuring and recording procedures with hearing aids, to name but a few. Thanks to Jorge’s experience and, above all, his will to transmit know-how to me during our regular meetings, I acquired valuable new practical skills which benefited my PhD work and, at the same, nourished my deeply application- and technical-oriented engineering nature. Certainly, I am also thankful to Jorge who shared with me the open code for his proprietary bilateral beamformer algorithm, which has been employed in various parts of this work.

Credit is further due to my NAL colleagues for their camaraderie and support towards my work; namely, to Cong-Van for our lengthy and fruitful discussions on signal processing and algorithm-implementation, to Margot for carrying out the data collection in a rather intricate listening test, to Fabrice for inspiring me to dedicate my best efforts on my research endeavour, and to Adam for indoctrinating me to the world of  $\text{\LaTeX}$ .

On a more pragmatic level, I should sincerely acknowledge the HEARing CRC and Macquarie University for the funding they have provided me with. Ultimately, it has been this funding that empowered me to accomplish the present work.

I also wish to bestow my gratitude to Nicolas-Alexander Tatlas who encouraged my initial thoughts on seeking a PhD path and, at the same time, assuaged my concerns about embarking on a doctoral programme while being past my thirties. Nicolas was also the person who introduced me to Jörg, and so was the catalyst for everything that followed.

Unusual as it might appear for a thesis of engineering orientation, I would herein like to express my gratefulness towards two of the most notable Greek authors, namely “Nikos Kazantzákis” (Νίκος Καζαντζάκης, 1883 – 1957) and “Aléxandros Papadiamántis” (Αλέξανδρος Παπαδιαμάντης, 1851 – 1911). Their monumental work, characterised by a spiritual restlessness, has unceasingly accompanied me during all this period of investigating the implications of sound waves, further assisting me to sustain a strong bond, on a cultural and lyrical level, with my homeland.

Στο σημείο αυτό αρμόζει να ευχαριστήσω τους γονείς μου· όχι μόνο γιατί ενθάρρυναν κι υποστήριξαν την εκπόνηση της παρούσας διατριβής, αλλά γιατί σε αυτούς οφείλω, εν πολλοίς, την εξ απαλών ονύχων σφυρηλάτηση του φιλοπερίεργου χαρακτήρα, καθώς και της ορθολογικής σκέψης που πιστεύω ότι με χαρακτηρίζουν.

Reaching the highly sought after “last but not least” paragraph, I desire to wholeheartedly thank my partner Yanyan for her contribution in this work. Yanyan has supported and encouraged me during moments of fatigue, discontent and even exasperation (one can perhaps imagine that those moments were not all too scarce). She has sympathised and stood by me all the while I have been devoting long hours of work during weekdays, weekends and holidays. But most of all, she has been the pillar of our family, assiduously attending to the needs of our son Nicholas during the day and, often enough, during the night. Completing this PhD work without compromising the family-life balance should be primarily attributed to her. 延延深深感谢你为我和小尼克所做的一切。也非常感谢你容忍我周日午后的“阴郁”心情！

# Contents

<b>Declaration of Authorship</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>Acknowledgements</b>	<b>ix</b>
<b>Contents</b>	<b>xi</b>
<b>List of Publications</b>	<b>xv</b>
<b>Abbreviations</b>	<b>xvii</b>
<b>Symbols</b>	<b>xix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Context of thesis . . . . .	1
1.1.1 Evaluation of hearing aids inside sound-treated booths and rooms	2
1.1.2 Binaural technology for evaluation of hearing aids via headphones	4
1.1.3 The R-SPACE sound system . . . . .	5
1.1.4 Loudspeaker-based virtual sound environments . . . . .	5
1.2 Aims of the thesis . . . . .	6
1.3 Thesis organisation . . . . .	7
<b>2 The theory of higher-order Ambisonics</b>	<b>11</b>
2.1 Fourier-Bessel series expansion . . . . .	11
2.2 Plane-wave assumption – Decoding to a loudspeaker array . . . . .	13
2.3 Spherical-wave expansion . . . . .	18
2.4 Sound field reconstruction . . . . .	19
2.5 HOA directivity plots . . . . .	22
2.6 Modified ( $\max-r_E$ ) decoding . . . . .	23
2.7 Field encoding with a microphone array (Model-based encoding) . . . . .	25
2.8 ‘Shape-matching’ encoding method (Measurement-based encoding) . . . . .	29
2.9 Comparison of ‘shape-matching’ and ‘Ambisonics’ encoding methods . . . . .	30
2.10 Mixed-Order Ambisonics . . . . .	34
<b>3 Measurement of a full 3D set of HRTFs for in-ear and hearing aid microphones on a Head and Torso Simulator (HATS)</b>	<b>37</b>
3.1 Introduction . . . . .	38
3.2 Methods . . . . .	39

3.2.1	Measurement Setup . . . . .	39
3.2.2	Spherical head model realization . . . . .	42
3.2.3	Combination of the head model with the measured data . . . . .	44
3.3	Results . . . . .	46
3.4	Summary and conclusions . . . . .	53
<b>4</b>	<b>Objective analysis of Ambisonics for hearing aid applications: Effect of listener's head, room reverberation, and directional microphones</b>	<b>55</b>
4.1	Introduction . . . . .	56
4.2	Higher-Order Ambisonics (HOA) basics . . . . .	59
4.2.1	Fourier-Bessel series expansion . . . . .	59
4.2.2	Decoding to a loudspeaker array . . . . .	60
4.2.3	Modified ( $\max-r_E$ ) decoding . . . . .	62
4.2.4	Field encoding with a microphone array . . . . .	63
4.2.5	Plane-wave assumption & near-field effects . . . . .	66
4.3	Methods . . . . .	66
4.3.1	Acoustic scene simulation . . . . .	67
4.3.1.1	ODEON room model . . . . .	67
4.3.1.2	Rendering to auxiliary sound sources . . . . .	70
4.3.2	Sound field coding . . . . .	71
4.3.2.1	HOA microphone array considerations . . . . .	72
4.3.3	Playback and field synthesis . . . . .	72
4.3.3.1	HOA reconstruction errors . . . . .	73
4.3.3.2	Loudspeaker array configuration . . . . .	74
4.3.3.3	HOA system noise due to microphone array . . . . .	75
4.3.4	Directional hearing aid microphone and SNR benefit . . . . .	76
4.4	Simulation results . . . . .	78
4.4.1	Microphone array noise . . . . .	78
4.4.2	Anechoic virtual source reconstruction . . . . .	79
4.4.2.1	HOA pressure errors . . . . .	79
4.4.2.2	Directional HA-microphone HOA errors . . . . .	85
4.4.3	Reverberant virtual source reconstruction . . . . .	86
4.4.3.1	HOA pressure errors . . . . .	86
4.4.3.2	Directional HA-microphone HOA errors . . . . .	88
4.5	Discussion . . . . .	90
4.5.1	HOA system noise . . . . .	90
4.5.2	HOA errors for anechoic source reconstruction . . . . .	91
4.5.2.1	HOA pressure errors . . . . .	91
4.5.2.2	Directional HA-microphone HOA errors . . . . .	92
4.5.3	HOA errors for reverberant source reconstruction . . . . .	92
4.5.3.1	HOA pressure errors . . . . .	92
4.5.3.2	Directional HA-microphone HOA errors . . . . .	94
4.6	Conclusions . . . . .	94
<b>5</b>	<b>Effect of higher-order Ambisonics on evaluating beamformer benefit in realistic acoustic environments</b>	<b>97</b>
5.1	Introduction . . . . .	98

5.2	Methods . . . . .	98
5.2.1	Higher-order Ambisonics (HOA) . . . . .	99
5.2.2	Hearing aid beamforming . . . . .	101
5.3	Performance Analysis . . . . .	102
5.4	Discussion . . . . .	104
<b>6</b>	<b>An objective and subjective validation of applied loudspeaker-based virtual sound environments used for directional hearing aid testing</b>	<b>107</b>
6.1	Introduction . . . . .	108
6.2	Methods . . . . .	110
6.2.1	Acoustic scene generation . . . . .	110
6.2.1.1	Real-room environment . . . . .	110
6.2.1.2	Direct room acoustics model-based sound rendering . . . . .	112
6.2.1.3	Mixed-Order Ambisonics sound-field reconstruction . . . . .	115
6.2.2	Directional hearing aid processing . . . . .	116
6.2.3	Objective evaluation . . . . .	117
6.2.3.1	Pressure reconstruction errors . . . . .	117
6.2.3.2	SNR benefit errors . . . . .	118
6.2.4	Subjective evaluation . . . . .	119
6.2.4.1	Test subjects . . . . .	119
6.2.4.2	Hearing aid fitting . . . . .	120
6.2.4.3	Speech intelligibility . . . . .	120
6.2.4.4	Acceptable noise level . . . . .	121
6.3	Results . . . . .	122
6.3.1	Objective analysis . . . . .	122
6.3.1.1	Pressure reconstruction errors . . . . .	122
6.3.1.2	SNR benefit errors . . . . .	125
6.3.2	Subjective measures . . . . .	126
6.3.2.1	Cardioid microphone performance . . . . .	126
6.3.2.2	Beamformer benefit on speech intelligibility . . . . .	128
6.3.2.3	Beamformer benefit on acceptable noise level . . . . .	130
6.4	Discussion . . . . .	130
6.4.1	Objective analysis . . . . .	130
6.4.1.1	Pressure reconstruction errors . . . . .	130
6.4.1.2	SNR benefit errors . . . . .	132
6.4.2	Subjective evaluation . . . . .	133
6.4.2.1	Cardioid microphone performance . . . . .	133
6.4.2.2	Beamformer benefit on speech intelligibility . . . . .	134
6.4.2.3	Beamformer benefit on acceptable noise level . . . . .	135
6.5	Conclusion . . . . .	136
<b>7</b>	<b>Concluding discussion</b>	<b>139</b>
7.1	Summary of findings . . . . .	139
7.2	Limitations of this work . . . . .	143
7.3	Recommendations and prospects for future work . . . . .	144
<b>A</b>	<b>Employed HOA equation formulation</b>	<b>147</b>

<b>B Co-authored work</b>	<b>151</b>
<b>C Ethics application approval</b>	<b>153</b>
<b>D Practical considerations for the microphone array simulation applied in Chapter 6</b>	<b>155</b>
<b>Bibliography</b>	<b>157</b>

# List of Publications

## Peer-reviewed journal and conference-proceeding papers

- [1] Oreinos, C. and Buchholz, J. M. (2015), “Objective Analysis of Ambisonics for Hearing Aid Applications: Effect of Listener’s Head, Room Reverberation, and Directional Microphones”, Manuscript accepted for publication at the J. Acoust. Soc. Am.
- [2] Oreinos, C. and Buchholz, J. M. (2013), “Measurement of a Full 3D set of HRTFs for In-Ear and Hearing Aid Microphones on a Head and Torso Simulator (HATS),” *Acta Acust. United Ac.* 99, 836–844.
- [3] Oreinos, C. and Buchholz, J. M. (2014), “Validation of Realistic Acoustic Environments for Listening Tests Using Directional Hearing Aids”, in *International Workshop on Acoustic Signal Enhancement, IWAENC* (Antibes – Juan les Pins).
- [4] Oreinos, C., Buchholz, J. M., and Mejia, J. (2013), “Effect of Higher-Order Ambisonics on Evaluating Beamformer Benefit in Realistic Acoustic Environments”, in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA* (New Paltz).
- [5] Weller, T., Buchholz, J.M., and Oreinos, C. (2014), “Frequency Dependent Regularization of a Mixed-Order Ambisonics Encoding System”, in *55<sup>th</sup> Conference of the Audio Engineering Society on Spatial Audio*.

## Papers submitted to peer-reviewed journals for publication

- [1] Oreinos, C. and Buchholz, J. M. (2014), “An Objective and Subjective Validation of Applied Loudspeaker-Based Virtual Sound Environments Used for Directional Hearing Aid Testing”, Manuscript prepared for submission to journal.

## Conference-proceeding papers (non-peer-reviewed)

- [1] Oreinos, C. and Buchholz, J. (2013), “Objective Analysis of HOA Sound-Field Reproduction for Hearing Aid Applications,” in *Proceedings of 21<sup>st</sup> International Conference on Acoustics, ICA* (Montréal).



# Abbreviations

2D	Two-dimensional (i.e., horizontal plane) sound field reproduction
4FAHL	Four-frequency average hearing loss
ADM	Adaptive directional microphone
AMB	Ambisonics
ANL	Acceptable noise level
BBF	Bilateral hearing aid beamformer
BNL	Background noise level
BTE-TF	Behind-the-ear (hearing aid)-related transfer functions
CRD	Cardioid hearing aid microphone
DFT	Discrete Fourier transform
DS	Direct signal
EDT	Early decay time
EIN	Equivalent input noise
ER	Early reflections
ERB	Equivalent rectangular bandwidth
fen	Function
FFT	Fast Fourier transform
HA	Hearing aid
HATS	Head and torso simulator
HI	Hearing impaired
HOA	Higher-order Ambisonics
HRIR	Head-related impulse response
HRTF	Head-related transfer function
ILD	Interaural level difference
IQR	Interquartile range
ITD	Interaural time difference
JND	Just-noticeable difference
LoRA	Loudspeaker-based room auralisation
L/S	Loudspeaker
Lt	Late (diffuse) part of room impulse response
mic	Microphone
MCL	Most comfortable level
MOA	Mixed-order Ambisonics
NAL	National Acoustic Laboratories
NMSE	Normalised mean-square error

n.d.	no date (relevant to citations lacking a date)
PSD	Power spectral density
REF	reference sound-field coding
RIR	Room impulse response
RMS	Root mean square
SHF	Spherical harmonic function
SiN	Speech-in-noise
SM	Shape-matching
SNR	Signal-to-noise ratio
SPL	Sound pressure level
SRTn	Speech reception threshold in noise
STFT	Short-time Fourier transform
TF	Transfer function
TSVD	Truncated singular value decomposition
VBAP	Vector base amplitude panning
VSE	Virtual sound environment
WFS	Wave field synthesis
WNG	White noise gain

# Symbols

$b(i, k)$	ADM mixing gain of front- and back-facing cardioids
$\underline{b}(i, k)$	Optimum, adaptive, ADM mixing gain
$\mathbf{b}$	Vector of HOA components $B_{mn}^\sigma$
$\hat{\mathbf{b}}$	Vector of least-squares solutions for HOA components
$\tilde{\mathbf{b}}$	Vector of regularised least-squares solutions $\tilde{B}_{mn}^\sigma$
$B(f)$	Single-microphone equivalent input noise PSD
$\bar{B}(f)$	Single-microphone EIN integrated in one-ERB-wide bands
$B_{mn}^\sigma$	HOA components (coefficients)
$\tilde{B}_{mn}^\sigma$	Regularised least-squares solutions for HOA components
$c$	Speed of sound ( $c = 343$ m/s)
$c_{i_{\text{front}}}(n)$	Output of $i$ -th filter-bank channel of front-facing cardioid HA mic
$C_{\text{front}}(f_i)$	DFT of $c_{i_{\text{front}}}$
$c_{i_{\text{back}}}(n)$	Output of $i$ -th filter-bank channel of back-facing cardioid HA mic
$C_{\text{back}}(f_i)$	DFT of $c_{i_{\text{back}}}$
$\mathbf{C}$	Matrix of SHFs sampled at the array loudspeaker positions
$d$	Index of point used for calculation of HOA microphone directivities
$d_{\text{mic}}$	Distance between the 2 HA microphones
$D$	Number of points for calculation of HOA microphone directivities
$\mathbf{D}$	HOA basic decoding matrix
$\mathbf{D}_m$	HOA max- $r_E$ decoding matrix
$\text{DI}(f)$	Directivity index
$e$	Natural logarithm base
$\vec{E}$	Sound-field energy vector
$\mathcal{E}$	SHF orthonormality error
$\mathcal{E}_{\text{HOA}}(f)$	HOA sound-field pressure reconstruction errors
$\mathcal{E}_S^{\text{VSE}}[n, f]$	Short-time spectral errors due to the considered VSE
$\mathcal{E}_{\Delta\text{SNR}}^{\text{VSE}}[n, f]$	Short-time SNR benefit error due to the considered VSE
$\mathbf{E} \left[ \tilde{\mathbf{E}}, \tilde{\mathbf{E}}_{\text{AMB}} \right]$	‘Ambisonics’ encoding matrix [regularised solution]
$\mathbf{E}_{\text{SM}} \left[ \tilde{\mathbf{E}}_{\text{SM}} \right]$	‘Shape-matching’ encoding matrix [regularised solution]
$f$	Frequency
$f_{\text{al}}$	Spatial aliasing frequency of spherical microphone array
$f_{\text{HOA}}$	HOA cut-off frequency
$f_s$	Sampling frequency
$f_x$	Crossover frequency of filter combining basic/max- $r_E$ decoding

$g_m$	$m$ -th degree max- $r_E$ decoding gain
$g\Delta(if)$	DFT of aligning time delay $\Delta$
$G$	Number of sources employed in the ‘Shape-matching’ optimisation
$G(\gamma)$	HOA panning function (loudspeaker-gain directivity)
$\mathcal{G}_{\text{idl}}(d, u, f)$	HOA microphone array ideal directivity
$\mathcal{G}_{\text{rec}}(d, u, f)$	HOA microphone array reconstructed directivity
$h_m^{(2)}(x)$	$m$ -th order spherical Hankel function of the 2 <sup>nd</sup> kind
$h[n]$	Impulse response
$h_u^{\text{DE}}[n]$	Impulse response of $u$ -th auxiliary source relevant to DS and ER
$h_v^{\text{Lt}}[n]$	Impulse response of $v$ -th auxiliary source relevant to Late
$H_{\text{eq},m}(kR)$	Inverse of $m$ -th degree rigid sphere mode strength
$\tilde{H}_{\text{eq},m}(kR)$	Regularised inverse of $m$ -th degree rigid sphere mode strength
$\mathbf{H}_{\text{eq},3\text{D}}(kR)$	Vector of inverse-equalisation functions $H_{\text{eq},m}(kR)$
$H_{\text{HOA}}(f)$	Ear response due to the HOA coding of the reference sound field
$H_{\text{Ref}}(f)$	Ear response due to the reference sound field
$H_{\text{TF}}(f, \theta, \delta)$	TF from source at $(\theta, \delta)$ to the considered ear of either the HATS or a transparent head
$H_u^{\text{aux/DE}}(f)$	Normalised frequency response of DS and ER auxiliary sources
$H_v^{\text{aux/Lt}}(f)$	Normalised frequency response of Late auxiliary sources
$i$	Imaginary unit
$I$	Number of quadrature points for the calculation of $\text{DI}(f)$ ; Number of observation points on a sphere for the calculation of $\epsilon_{\text{Enc}}$
$\mathbf{I}_J$	Identity matrix with dimensions $[J \times J]$
$j_m(x)$	$m$ -th order spherical Bessel function of 1 <sup>st</sup> kind
$k$	Wavenumber; Virtual source index
$\mathbf{k}$	Wave-vector pointing at direction of wave incidence; $\mathbf{k} = (k, \theta_k, \delta_k)$
$L$	Number of loudspeakers comprising the reproduction array
$m$	Degree of SHFs and Legendre polynomials ( <i>Literature usually refers to the descriptor <math>m</math> as the ‘order’ of the Bessel-family functions</i> )
$M$	HOA system order
$M_{3\text{D}}, M_{2\text{D}}$	MOA periphonic and horizontal-only orders
$n$	Order of SHFs and Legendre polynomials; Time index of discrete-time signal
$N_{\text{DE}}$	Total number of DS and ER auxiliary sources in LoRA
$N_{\text{FFT}}$	Length of the fast Fourier transform
$N_{\text{Lt}}$	Number of Late auxiliary sources in LoRA
$N_{\text{HOA}}(f)$	HOA system noise PSD
$N_{\text{T}}$	ODEON image-source/ray-tracing transition order
$p$	Acoustic pressure; p-value of statistical test
$p_R(\theta_q, \delta_q)$	Acoustic pressure at the $q$ -th sensor of the mic array (radius: $R$ )
$\tilde{p}$	Pressure derived from the series expansion using $\tilde{\mathbf{b}}$

$\mathbf{p}_R$	Vector of pressures at all array mics due to a given sound field
$P_m(x)$	Legendre polynomial of $m$ -th degree
$P_{mn}(x)$	Associated Legendre polynomial of $m$ -th degree, $n$ -th order
$\mathbf{P}_G$	Matrix of $Q$ microphone pressures due to $G$ sources
$Q$	Number of microphones comprising the recording array
$Q1, Q3$	First (25%) and third (75%) quartile of distribution
$\mathcal{Q}$	Acoustic source directivity
$r$	Radial distance from origin of coordinate system; (Pearson product-moment) correlation coefficient
$\mathbf{r}$	Position vector $\mathbf{r} = (r, \theta, \delta)$
$r_c$	Critical distance in reverberant space
$r_e$	Transparent head radius (ear distance from centre of coordinates)
$r_E$	Norm of the sound field energy vector $\vec{E}$
$r_o$	Radius of reproduction area of interest (sweet spot)
$r_s, r_{src}$	Single source radial distance
$\mathbf{r}_s$	Position vector of single source $\mathbf{r}_s = (r_s, \theta_s, \delta_s)$
$R, R_Q$	Radius of spherical microphone array
$R_L$	Radius of spherical loudspeaker array
$RT_{60}$	Reverberation time (unspecified evaluation range)
$s_l$	Gain ( <i>filter</i> in the general case) relevant to the $l$ -th loudspeaker
$\mathbf{s}$	Vector of loudspeaker filters $s_l$
$\hat{\mathbf{s}}$	Vector of least-squares solutions for loudspeaker filters
$\tilde{\mathbf{s}}$	Vector of solutions for L/S filters derived by using $\tilde{\mathbf{E}}$
$\tilde{s}_{u,l}^{DE}(f)$	Regularised filter of $l$ -th L/S relevant to $u$ -th DS/ER aux. source
$\tilde{s}_{v,l}^{Lt}(f)$	Regularised filter of $l$ -th L/S relevant to $v$ -th Late aux. source
$S[n, f]$	Short-time power spectral density
$S_l^{HOA}(f)$	L/S response relevant to HOA coding of the LoRA sound field
$S_u^{Ref/DE}(f)$	L/S response relevant to reference DS and ER auxiliary sources
$S_v^{Ref/Lt}(f)$	L/S response relevant to reference Late auxiliary sources
$SNR_c(f)$	Long-term SNR at the output of a cardioid HA microphone
$SNR_o(f)$	Long-term SNR at the output of an omnidirectional HA microphone
$SNR_o^I$	Long-term, intelligibility-weighted $SNR_o(f)$
$t(df)$	t-value of a t-test with $df$ degrees of freedom
$T_{20}, T_{30}$	Reverberation times derived by extrapolating the 5 dB to 25 dB or 5 dB to 35 dB, respectively, decay below the initial level
$T_{EDT}$	Early decay time
$\tilde{\mathbf{T}}$	Matrix of TFs from mics to L/S (combined encoding/decoding)
$U = (M+1)^2$	Number of SHF components in a HOA system of order $M$
$\vec{V}$	Sound field velocity vector
$W_m(kR)$	Rigid sphere mode strength
$W_L(i, k)$	BBF adaptive gains for the $i$ -th channel and $k$ -th block; left HA
$W_R(i, k)$	As above for the right HA
$\mathbf{W}_{3D}(kR)$	Vector of rigid sphere mode strengths $W_m(kR)$
$\mathcal{W}$	Wronskian determinant

$y_i(n)$	Output of $i$ -th filter-bank channel of hearing aid ADM
$y_m(z)$	$m$ -th order spherical Bessel function of 2 <sup>nd</sup> kind (Neumann fcn)
$Y_{mn}^\sigma(\theta, \delta)$	Real-valued spherical harmonic function of $m$ -th degree, $n$ -th order
$\mathbf{Y}$	Matrix of SHFs sampled at the array microphone positions
$\mathbf{Y}_G$	Matrix of SHFs sampled at $G$ positions (refer to ‘Shape-matching’)

Greek symbols (Ελληνικά σύμβολα):

$\alpha$	Spherical head model radius; Directivity parameter of first-order directional microphone
$\alpha_q, \alpha_i$	Quadrature weights relevant to $q$ -th and $i$ -th point on a sphere
$\gamma$	Angle between a loudspeaker at $(\theta_l, \delta_l)$ and a source at $(\theta_k, \delta_k)$
$\delta$	Elevation angle
$\delta_e$	Transparent-head ear elevation angle
$\delta_k$	Virtual source elevation angle
$\delta_l$	Loudspeaker elevation angle
$\delta_s, \delta_{\text{src}}$	Single source elevation angle
$\Delta$	Delay employed to time-align the spherical head model with the HRIR measurement
$\Delta\mathcal{G}_{\text{RMS}}(u, f)$	RMS error of reconstructed HOA mic directivities
$\Delta\mathcal{G}_{\text{sRMS}}(u, f)$	Signed RMS error of reconstructed HOA mic directivities
$\Delta\mathcal{P}(u, f)$	Difference between ideal/reconstructed HOA mic directivity power
$\Delta\text{SNR}[n, f]$	Short-term SNR benefit provided by directional microphone
$\Delta\text{SNR}(f)$	Long-term SNR benefit provided by directional microphone
$\Delta\text{SNR}^{\text{I}}$	Long-term, intelligibility-weighted $\Delta\text{SNR}(f)$
$\Delta T_{\text{EDT}}$	Early decay time misestimation in room acoustics model
$\epsilon$	Normalised truncation error (normalised mean-square error)
$\epsilon_{\text{Enc}}(kr_e, \theta_k, \delta_k)$	Pressure NMSE due to HOA encoding; relevant to a source at $(\theta_k, \delta_k)$
$\theta$	Azimuth angle (‘right-hand’ convention)
$\theta_e$	Transparent-head ear azimuth angle (‘right-hand’ convention)
$\theta_k$	Virtual source azimuth angle (‘right-hand’ convention)
$\theta_l$	Loudspeaker azimuth angle (‘right-hand’ convention)
$\theta_s, \theta_{\text{src}}$	Single source azimuth angle (‘right-hand’ convention)
$\lambda, \lambda_{\text{AMB}}$	‘Ambisonics’ encoding regularisation parameter
$\lambda_{\text{SM}}$	‘Shape-matching’ encoding regularisation parameter
$\tau$	Rear microphone time delay of first-order directional mic
$\hat{\tau}$	Time-of-arrival (TOA) of HRIR
$\tau_{\text{ST}}$	Short-time Fourier transform (sliding) window length
$\omega$	Angular frequency

*Dedicated to my family;  
my partner Yanyan and my son Nicholas.*



# Chapter 1

## Introduction

### 1.1 Context of thesis

A large number of research projects in psychoacoustics, audiology, and hearing devices technology are currently undertaken in laboratories. These projects aim to provide new insights on hearing and its impairment, including the management of hearing loss using hearing aids (HAs). The outcomes of this research need to be thoroughly evaluated in a realistic context to ensure that the theories that are formulated and tested in the laboratory also apply in the complex acoustical environments commonly encountered in everyday life.

Specifically in the field of signal enhancement strategies, implemented on HAs, researchers have been commonly evaluating novel schemes using traditional listening tests. Such tests are typically performed either in a controlled but, at the same time, oversimplified manner (e.g., laboratory setups typically employing 2-8 loudspeakers in a horizontal ring), or in an uncontrolled and lacking repeatability but, at the same time, realistic manner (e.g., field studies inside real-life environments). However, a number of researchers (Ricketts, 2000; Compton-Conley *et al.*, 2004; Cord *et al.*, 2004) have reported significant discrepancies between the performance of even simple directional HA schemes when tested in artificial laboratory settings and the benefits observed in real-life scenes.

Moreover, a plethora of anecdotal observations supports this inconsistency between laboratory test results and real-life benefit, highlighting the need to develop and establish ecologically-valid listening experiment configurations. As Luts *et al.* (2010) epigrammatically put forth: *“The reason for the limited application of digital signal enhancement techniques in commercial hearing instruments is twofold. First, few signal processing schemes have been developed and sufficiently evaluated for real-world applications [...]”*.

This PhD thesis aims to validate a particular type of virtual sound environment (VSE) which is based on the coding scheme of higher-order Ambisonics (HOA). The validation is performed by estimating a number of objective (e.g., the sound-field pressure and the SNR at the ears and HA microphones) and subjective measures (e.g., the speech intelligibility and the acceptable noise level) inside a certain VSE and then comparing

them to relevant estimations derived inside a reference complex acoustic environment. The first part of the work considers simulated virtual and realistic sound environments while the second part examines actual real-room scenes and VSEs reconstructed inside a real loudspeaker array.

The following sections provide a synopsis of the listening test configurations employed in the literature to evaluate the performance of various HA processing strategies. The overview begins with the most common and least realistic configuration of listening tests conducted inside sound-treated booths fitted with loudspeakers. Then, it progresses to listening tests administered via headphones. This transition is probably the first step towards introducing some realism into laboratory testing. The next section summarises a recording and loudspeaker-reproduction sound system (R-SPACE) that aims to reconstruct the realism of complex acoustic environments inside the laboratory. The overview finally paves the way to consider periphonic (i.e., covering the whole three-dimensional, 3D, space) loudspeaker-based VSEs as the closest step proposed to reconstructing real-life acoustic environments.

### **1.1.1 Evaluation of hearing aids inside sound-treated booths and rooms**

One of the most common ways to evaluate HAs is by conducting listening tests inside sound-treated booths. In such a testing paradigm, hearing-impaired subjects are fitted with HAs featuring the signal enhancement algorithm(s) being tested and are seated inside the test booth. Commonly, a small number of loudspeakers are employed to imitate an acoustic environment comprising a target (usually spoken words or sentences) and some interfering sound sources. Sound material that has been previously recorded inside a reverberant or an anechoic environment is next presented via the loudspeakers and a series of listening tests (e.g., speech intelligibility, localisation, etc.) are administered. In that way the perceptual benefit of the given algorithms, *as derived inside the laboratory environment*, can be estimated.

However, the above laboratory-test configuration cannot adequately predict the real world performance of HA signal-enhancing schemes such as directional microphones. This occurs because the applied acoustic sound-fields commonly fail, and often do not even attempt, to reconstruct the complexity of real-world listening conditions, mainly because they: (1) only provide very few locations (loudspeakers) where sound sources can be placed, (2) cannot control the sound-field diffuseness (reverberation) in a flexible manner so as to create a range of different acoustic scenes, (3) do not usually feature the means to reproduce elevated or distant sources, and (4) cannot approximate dynamic scenes comprising moving sources. Of course, even when laboratory-test configurations

are elaborate enough to recreate simple acoustic scenes that could occur, for example, in a domestic environment, they lack the flexibility to present different acoustic conditions so that the under-testing signal-enhancing scheme can be evaluated in a range of scenarios.

A number of researchers have demonstrated, either explicitly or implicitly, the deficiencies related to such simplistic listening tests conducted inside sound-treated booths. For example, [Desloge \*et al.\* \(1997\)](#) conducted speech reception threshold (SRT) tests with HAs inside an acoustically-live room (reverberation time  $RT_{60} = 0.6$  s) using a nearby frontal target loudspeaker and 4 more distant loudspeakers reproducing the interfering signals (distractors). Counter-intuitively enough, while testing two binaural directional algorithms, some of the SRT improvement results exceeded the theoretical intelligibility-weighted SNR benefit values. This outcome suggests that the chosen source configuration probably constituted a best-case scenario that favoured the tested directional algorithms. Obviously, such best-case scenarios are in stark contrast to the real-life complex acoustic environments usually encountered. Although testing in such best-case scenarios could be useful in terms of estimating an upper limit of the benefit that a processing scheme could offer, it should necessarily be complemented with tests conducted in a wide range of more complex environments in order to yield more general conclusions.

In another study using similar methods, [Ricketts \(2000\)](#) evaluated the directional benefit and performance of static directional HA microphones for four noise configurations inside two different rooms ( $RT_{60} = 0.63$  s and  $RT_{60} = 1.10$  s). The author's conclusions highlight that both the noise configuration and the room type significantly affect the estimated directional benefit of the HAs. This observation indicates that solely relying on testing inside a single, or at best inside a few such artificial environments proves inadequate.

In the same manner, [Cord \*et al.\* \(2004\)](#) investigated whether the commonly employed laboratory testing configurations can adequately predict the real-world performance of directional HA microphones and concluded that the improvement of speech intelligibility, as derived through commonly used listening tests, is a poor predictor of the benefit experienced in everyday life. The same research group ([Cord \*et al.\*, 2002](#)) had previously reached the same conclusion after having conducted a field study.

[Maj \*et al.\* \(2006\)](#) performed listening tests with HAs inside a room using four different distractor-source configurations also verifying the well-documented fact that the environment affects the way signal enhancement algorithms, in their case an adaptive beamformer, perform.

In a more extensive study, conducted across 4 different laboratories, [Luts \*et al.\* \(2010\)](#) evaluated the benefit of a number of HA signal enhancement algorithms and found some of them to be sensitive to the testing environment characteristics (e.g., number of sources and amount of reverberation). They also found a number of the algorithms to be more robust in varying acoustic environment settings. These findings point to the need for evaluating HA algorithms using versatile laboratory testing environments so that a wide range of different real-life acoustic scenes can be reconstructed and presented to the hearing subjects.

### 1.1.2 Binaural technology for evaluation of hearing aids via headphones

In order to bridge the gap between artificial laboratory scenes, such as those described above, and real-world environments, some researchers ([Rychtáriková \*et al.\*, 2009](#); [Völk and Fastl, 2010](#); [Fels, 2013](#)) have employed binaural technology to conduct listening tests via headphones. For example, [Van den Bogaert \*et al.\* \(2008, 2009\)](#) evaluated three binaural HA noise-reduction algorithms using reverberant signals presented to listeners through headphones. The binaural signals were derived by off-line processing of the reconstructed HA microphone input signals. The input signals were, in turn, derived by convolving anechoic signals with impulse responses measured from horizontal-plane loudspeakers, placed inside an acoustically-live room ( $RT_{60} = 0.61$  s), to the microphones of HAs worn on an acoustic manikin.

Likewise, [Mueller \*et al.\* \(2012\)](#) employed headphone signal presentation to conduct a series of localisation tests. The headphone signals were generated by processing HA microphone signals derived: (1) via real-room measurements, and (2) via room acoustics simulation. The authors considered four different acoustic environments and three different HA processing strategies. The outcomes of comparing the real to the simulated environments, both presented binaurally, prove binaural technology to be a reliable tool to evaluate sound localization with HAs.

However, although binaural signal presentation via headphones is capable of recreating the auditory cues of complex real-life environments, it exhibits a number of real-life limitations that constrain its theoretical capacity in an applied context:

- (1) It ideally requires that individual head-related transfer functions (HRTFs) to the HA microphones are measured for every listener.
- (2) It precludes head movements, unless an elaborate head-tracking system is applied.
- (3) It fails to accurately reproduce the significant, especially at low frequencies, acoustic path contribution (leakage) of the individual's ear-HA device.
- (4) It precludes the use of off-the-shelf or the subject's individual HAs for the testing.

### 1.1.3 The R-SPACE sound system

The R-SPACE sound system (Revit *et al.*, 2002; Compton-Conley *et al.*, 2004; Revit *et al.*, 2007) has been one of the first attempts to record real-world acoustic environments and then reproduce them inside a laboratory for the purpose of testing HAs. The system employs an horizontal-plane ring array of 8 highly directional (shotgun) microphones positioned with their diaphragms on a circle with a radius of 0.6 m. The microphones are evenly distributed with an angular spacing of  $45^\circ$  and point radially outward from the array centre. The recordings are subsequently fed to an array of 8 loudspeakers, configured in the same manner as the microphone array. The R-SPACE virtual sound system has been validated using a series of perceptual measures, such as the acoustic scene ‘realism’ and the SRT improvement provided by first- and higher-order directional HA microphones, as presented by the above authors.

The operation principle of the R-SPACE system resembles the combined encoding and decoding operations of a 2D (height-invariant) Ambisonics system, as explained by Moreau *et al.* (2006, Section 2.1.2.). However, that encoding-decoding principle is not strictly satisfied by the R-SPACE system since the directional microphones are not coincident at a reference point but rather capture the sound-field at the same radial distance at which the reproduction loudspeakers are situated. Although that system has yielded promising results in terms of capturing and reproducing complex acoustic scenes, it nevertheless exhibits a number of limitations:

- (1) It is inevitably limited in terms of frequency bandwidth and spatial accuracy due to the small number of employed microphones and loudspeakers; those factors, however, were not addressed by Revit *et al.* (2002), Revit *et al.* (2007) or Compton-Conley *et al.* (2004).
- (2) It can only control horizontal-component reconstruction. The elevated sources of the original sound-field are merely mapped to the horizontal loudspeakers by means of the microphone directivity.
- (3) It is partly validated by relying on perceptual attributes of the acoustic scenes (e.g., scene realism), which is ill-suited in the case of HA testing, given that the multi-microphone HA processing significantly differs from the human auditory processing.

### 1.1.4 Loudspeaker-based virtual sound environments

Building up on the principle of the R-SPACE system, it is suggested that using a loudspeaker-based virtual sound environment (VSE) that attempts to realistically reconstruct a complex acoustic scene appears to be a viable candidate for conducting

ecologically-valid listening tests with hearing devices (Minnaar *et al.*, 2010; Seeber *et al.*, 2010). Such a VSE can be generated by either acoustically simulating or by recording a real-life environment, followed by a decoding operation to generate appropriate loudspeaker signals. The main methods used to derive loudspeaker signals, in accordance with the previous paradigm, are: (1) vector base amplitude panning – VBAP (Pulkki, 2001), (2) wave field synthesis – WFS (Berkhout *et al.*, 1993), and (3) higher-order Ambisonics – HOA (Daniel, 2001; Poletti, 2005). The first method aims at recreating faithful perception attributes for the reconstructed virtual sources while the remaining two aim at physically reconstructing the sound field inside a defined region.

For the purpose of testing HAs inside a VSE, this work focuses on a sound-field reconstructing method. This choice was made because the HA processing differs significantly from the human auditory processing, particularly when the HAs feature multi-microphone directional algorithms. Hence, solely relying on reconstructing the perceptual attributes, e.g., using the VBAP method, is expected to influence the HA processing in an uncontrolled way. Among the remaining candidates of HOA and WFS, the former is herein preferred over the latter due to a number of advantages (Daniel *et al.*, 2003; Spors and Ahrens, 2008) that it offers for the described application; namely: (1) it consistently optimises the field reconstruction at a centred region (sweet spot), which is well-suited for the case of single-listener hearing tests; (2) it is a scalable format, i.e., the scene recording is completely decoupled from the reproduction, which can be performed on a number of different loudspeaker arrays, making it thus possible to exchange generated environments and to compare test results across laboratories. It should be mentioned that a number of researchers (Spors and Ahrens, 2008; Ahrens and Spors, 2009; Ahrens *et al.*, 2010) refrain from favouring either sound-field reconstruction technique over the other, following a detailed analysis of their time-domain and spectral properties. Although these results provide valuable and novel insights on HOA and WFS, the centre-oriented advantage of HOA prevails when it comes to the application of individualised listening tests.

## 1.2 Aims of the thesis

The present PhD thesis aims to achieve two main goals:

- (1) Establish a versatile framework to objectively validate the limitations of a HOA-based VSE for testing hearing aids. Such a framework is intended to:

- a) Consider the entire path of signal flow, from the initial, real or simulated, acoustic environment, to the final loudspeaker signals determined via the HOA coding.
  - b) Analyse the effect of reverberation existing in real-life acoustic environments.
  - c) Take into account the presence of a listener inside the reproduction loudspeaker array.
  - d) Be able to convert loudspeaker signals to in-ear and HA microphone signals, in a flexible way, independent of the specific geometry of the loudspeaker array.
  - e) Evaluate the output of a number of static and adaptive spatially-selective HA algorithms driven by the HA microphone signals generated above.
  - f) Ultimately provide a rule of thumb concerning the required HOA order that would ensure a sufficiently accurate VSE for HA testing.
- (2) Subjectively verify the outcomes of the objective analysis of loudspeaker-based sound systems for testing HAs. For that purpose, a real reverberant multi-talker environment will be created and reproduced using two different VSE methods; a direct, model-based reproduction of the real environment; and a HOA-based reproduction of the previously modelled acoustic environment. The different VSEs will be evaluated and compared to the real environment by:
- a) Measuring the intelligibility benefit offered by directional HA algorithms to hearing-impaired listeners positioned inside all three considered environments.
  - b) Applying a number of objective measures to the output of the directional HAs placed on an acoustic manikin inside all three considered environments.

### **1.3 Thesis organisation**

The work presented in this thesis is structured as follows:

Chapter 2 presents the basic theoretical background of higher-order Ambisonics (HOA). Starting from basic concepts, such as the expansion of a sound-field's pressure in spherical harmonic functions (SHFs), it details the reproduction of virtual sources using loudspeaker arrays and then presents the fundamental principles of using a spherical microphone array to capture generic sound fields. The chapter concludes with an analysis and theoretical validation of the 'shape-matching' operation, a relatively novel method for deriving HOA components from a microphone array. Mixed-order Ambisonics (MOA), a specific SHF selection scheme, is also briefly described. Inevitably, there is some overlap between the content of Chapter 2 and the shorter theory sections of the journal publication-style chapters (most notably that of Chapter 4). The reader who studies

Chapter 2 in detail may consider skipping the relevant theory sections of the following chapters.

Chapter 3 presents the methodology employed to measure and validate a dense 3D set of head-related transfer functions (HRTFs) to the in-ear and behind-the-ear (BTE) hearing aid microphones of a head and torso Simulator (HATS). We were motivated to measure such a set of HRTFs by the need to simulate dense loudspeaker arrays and image sources of arbitrary locations. Simulating such virtual sources proves a valuable tool in evaluating spatial signal processing algorithms (e.g., directional microphones, beamformers) in hearing devices as well as in evaluating sound-field reconstruction methods.

Chapter 4 constitutes the backbone of this work, presenting the main findings relevant to employing HOA to reconstruct realistic, complex acoustic scenes. Specifically, the paradigm of a listener fitted with directional HAs and positioned at the centre of the HOA reproduction loudspeaker array is examined. Moreover, Chapter 4 investigates how room reverberation affects the HOA pressure reconstruction errors. It demonstrates that including the room reverberation effectively reduces the HOA reconstruction errors, thus extending the usable frequency range of the system.

Chapter 5 expands the work presented in Chapter 4 by analysing how two adaptive directional HA strategies perform while operating inside a HOA-reconstructed reverberant environment. The aim of this short chapter is twofold: (1) it studies the performance of more complex directional strategies, such as a first-order adaptive directional microphone and a state of the art adaptive bilateral beamformer; and (2) it considers the short-time behaviour of both observable quantities (e.g., SNR benefit) and internal system quantities (e.g., adaptive gains of directional algorithms). The results imply that the HOA coding preserves, in its region of operation, the correct phase and timing of the initial sound field.

Chapter 6 describes a listening experiment aimed to validate the above theoretical observations. The experiment starts with setting a realistic ‘cocktail party’ acoustic scene inside a real-room. The acoustic environment is then modelled and presented either directly or via MOA to the loudspeaker array available at the National Acoustic Laboratories (NAL). Speech intelligibility and acceptable noise level are measured on 18 aided hearing-impaired (HI) subjects inside the real and the two virtual sound environments. The HI subjects are fitted with HAs featuring two directional algorithms; a static cardioid array and NAL’s proprietary adaptive bilateral beamformer. The objective analysis of the HA directional outputs inside all three acoustic environments shows that the VSEs incur considerable quasi-random short-time errors both in the pressure and the SNR benefit, which, nevertheless and most importantly, average to minor only values. The subjective analysis demonstrates that the general behaviour and trends in

---

the data are preserved in both VSEs for the two considered directional HA processing methods. However, the achievable benefits are slightly reduced in the direct model-based VSE, and further reduced in the MOA-based VSE. The results indicate that both VSEs can be employed to evaluate directional HAs, but the sensitivity of the applied test measures, in particular for the MOA-based VSE, might be noticeably reduced when compared to the real world case.



# Chapter 2

## The theory of higher-order Ambisonics

### 2.1 Fourier-Bessel series expansion

Ambisonics, or more accurately Higher-Order Ambisonics (HOA) since the term Ambisonics on its own refers only to a first order system (Gerzon, 1973; Bamford and Vanderkooy, 1995), is a mathematical concept that permits the decomposition and resynthesis of sound fields. It is based on the representation of acoustic waves in space as weighted sums of spherical harmonics, an orthogonal system that can be used to transform any square-integrable function on the unit sphere (Williams, 1999). Such a spherical harmonic expansion, also termed a ‘Fourier-Bessel series’ (being one particular kind of the generalised Fourier series), results into an equivalent description of the function in terms of its spatial spectrum, in the same way that a Fourier series results into an equivalent description of periodic functions in terms of their frequency spectrum (i.e., weighted sum of sines and cosines, or complex exponentials). In the context of acoustics, the use of spherical harmonics provides a convenient tool to tackle the problem of sound field reconstruction (Fazi and Nelson, 2007).

Adopting the notation used by Daniel (2001) and Moreau *et al.* (2006), i.e., the use of *real-valued* spherical harmonic functions (SHF) and a spherical coordinate system where elevation  $\delta$  is measured from the horizontal plane and azimuth  $\theta$  increases counter-clockwise from the x-axis to the y-axis as seen from positive  $z$ -axis, the pressure inside a source-free region  $\Omega_{r_L}$  can be written as:

$$p(kr, \theta, \delta) = \sum_{m=0}^{\infty} i^m j_m(kr) \sum_{n=0}^m \sum_{\sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \delta), \quad (2.1)$$

where  $i = \sqrt{-1}$  is the imaginary unit,  $j_m(kr)$  the spherical Bessel function of degree  $m$ ,  $k$  the wavenumber,  $B_{mn}^{\sigma}$  the expansion coefficients or simply HOA components, and  $Y_{mn}^{\sigma}(\theta, \delta)$  the real-valued spherical harmonic functions (SHFs) of degree  $m$  and order  $n$ ,

defined as:

$$Y_{mn}^{\sigma}(\theta, \delta) = \sqrt{(2m+1)(2-\delta_{0,n})} \frac{(m-n)!}{(m+n)!} P_{mn}(\sin \delta) \cdot \begin{cases} \cos n\theta & \text{if } \sigma = +1 \\ \sin n\theta & \text{if } \sigma = -1 \text{ (ignored if } n = 0) \end{cases}, \quad (2.2)$$

where  $\delta_{0,n}$  is the Kronecker delta, equal to 1 for  $n = 0$  and 0 elsewhere, and  $P_{mn}$  the associated Legendre functions (as defined for example by [Moreau \(2006\)](#), i.e., dropping the  $(-1)^m$  term used by [Williams \(1999\)](#) and other researchers).

Figures 2.1 and 2.2 show two different visual representations of the SHFs. The first one (Fig. 2.1) depicts the SHFs as azimuth- and elevation-dependent densities on the unit sphere, with the red colour showing positive values and the blue colour showing negative values. The nodal lines are also shown in the plots. SHFs up to degree  $m = 5$  are plotted, with part of the  $\sigma = -1$  functions omitted due to space considerations. As can be confirmed graphically, the  $Y_{mn}^{-1}$  values result from the relevant  $Y_{mn}^{+1}$  values with a simple azimuth rotation (yaw) of  $\alpha = \pi/2$ . The second way of visualising the SHFs is by making use of directivity balloons. Hence, Fig. 2.2 depicts the SHFs as  $\mathbf{r} = \left( |Y_{mn}^{\sigma}(\theta, \delta)|, \theta, \delta \right)$ , with  $\theta \in [0, 2\pi)$  and  $\delta \in [-\pi/2, \pi/2]$ . The red and blue colours show positive and negative values as before.

Equations (2.1) and (2.2) may differ to other formulations commonly found in literature. The observed discrepancies are due to the:

- (1) different normalizations applied to the spherical harmonic functions: standard normalization ensuring orthonormality of SHFs, or unit power normalization (used in this study), or Schmidt semi-normalization);
- (2) use of complex-valued SHFs by a big part of the research community;
- (3) different assumptions for the time dependence:  $e^{+i\omega t}$  (used in this study) or  $e^{-i\omega t}$ ;
- (4) use of different conventions for the spherical coordinate system: azimuth  $\theta$  measured either clockwise or counter-clockwise;  $\delta$  representing either the elevation or the inclination.

More detailed considerations and some proofs of equivalence between different notations can be found in [Appendix A](#).

As a consequence of considering a wave time-dependence of  $e^{+i\omega t}$ , a plane wave of the form  $e^{i\mathbf{k}\cdot\mathbf{r}}$  represents an incident wave. Therefore, the wave vector  $\mathbf{k}$  shows the direction of incidence rather than the direction of propagation.

Each term of the pressure expansion series of Eq. (2.1) consists of a sum of SHFs of degree  $m$  and order  $n$ , which expresses a dependence with azimuth  $\theta$  and elevation  $\delta$ , as seen in Figs. 2.1 and 2.2, weighted per degree by the spherical Bessel functions, which describe the radial dependence. For small arguments  $kr$  (low frequencies and small distances from the centre of coordinates), the values of the radial weights  $j_m(kr)$  decrease with increasing degree  $m$  (Moreau *et al.*, 2006; Williams, 1999). This permits the truncation of the infinite sum to a degree  $m = M$ , referred to in literature as the order of the HOA system. Associated with the  $m = M$  truncation of the sum are  $(M + 1)^2$  SHF terms  $Y_{mn}^\sigma(\theta, \delta)$  since there are  $(2m + 1)$  order terms for each degree  $m$ . This truncated pressure expansion series constitutes the building block of the HOA sound field reconstruction technique.

## 2.2 Plane-wave assumption – Decoding to a loudspeaker array

HOA reproduction is commonly (but not always; e.g., see Daniel and Moreau, 2004; Poletti, 2005; Favrot and Buchholz, 2010a) based on the plane wave expansion. Plane waves –being solutions to the homogeneous Helmholtz equation– form a convenient basis (leading to simpler equation formulation) to represent random sound fields. Loudspeakers can be considered as plane-wave sources at realistic array radius distances (e.g.,  $r \sim 2$  m) when looking at a small portion of the wave-front in the vicinity of a human head centred in the array.

A unit-amplitude plane wave can be expanded in spherical harmonic terms as (Moreau, 2006):

$$e^{i\mathbf{k}\cdot\mathbf{r}} = \sum_{m=0}^{\infty} i^m j_m(kr) \sum_{n=0}^m \sum_{\sigma=\pm 1} Y_{mn}^\sigma(\theta_k, \delta_k) Y_{mn}^\sigma(\theta, \delta), \quad (2.3)$$

where the wave vector  $\mathbf{k}$  is expressed in spherical coordinates as  $\mathbf{k} = (k, \theta_k, \delta_k)$  and the observation vector  $\mathbf{r}$  is expressed as  $\mathbf{r} = (r, \theta, \delta)$ . Direct comparison of Eq. (2.1) with the above expansion reveals that the HOA components of a unit-amplitude plane wave arriving from position  $(\theta_k, \delta_k)$  equal the SHFs sampled at the direction of the wave arrival:

$$B_{mn}^\sigma \Big|_{\text{pl.wave from } (\theta_k, \delta_k)} = Y_{mn}^\sigma(\theta_k, \delta_k). \quad (2.4)$$

Truncating the infinite sum of Eq. (2.3) to the HOA order  $m = M$ , as reasoned at the end of Section 2.1, incurs a spatial band-limiting error commonly quantified by the

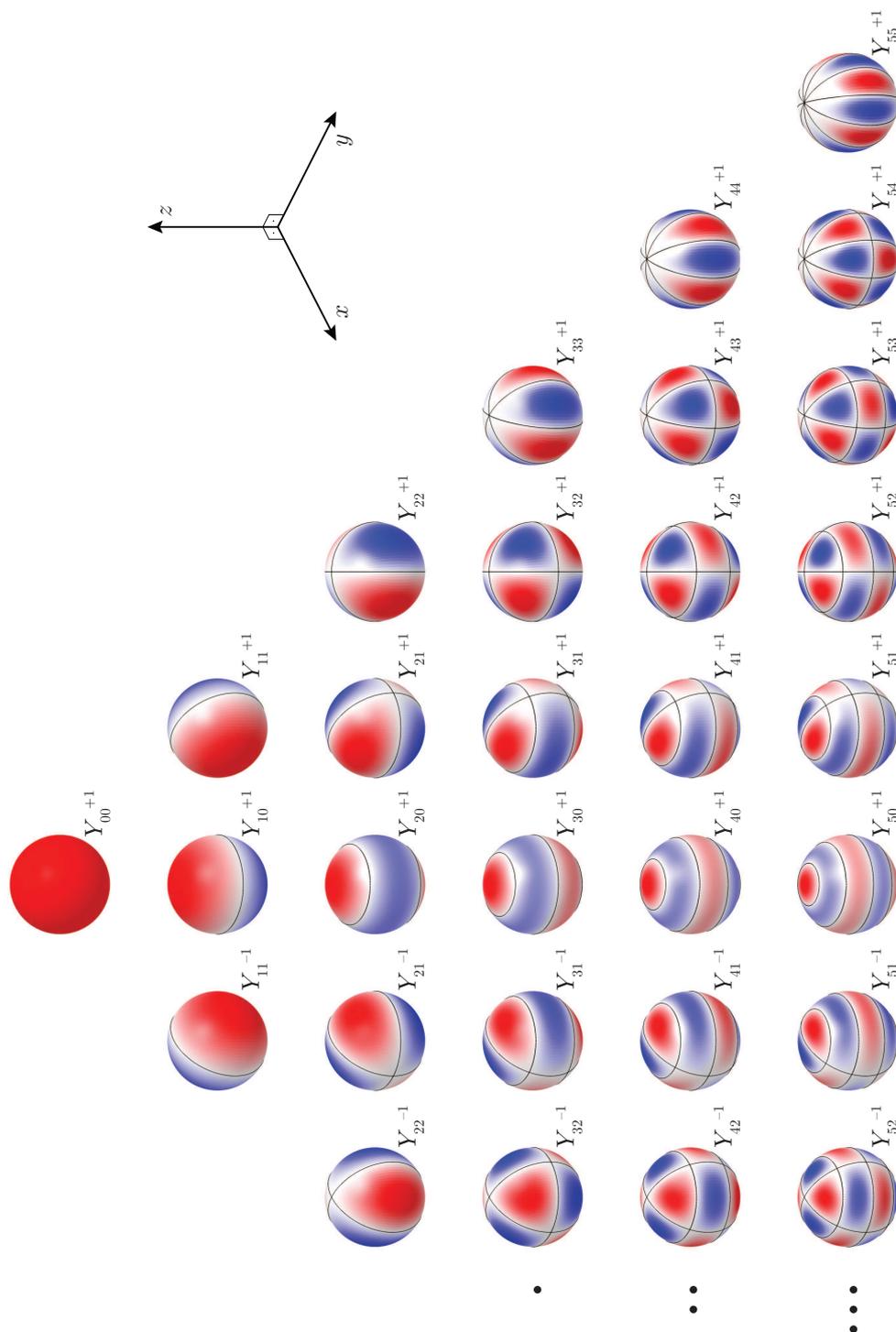


FIGURE 2.1: SHFs of  $m = 1 \dots 5$  plotted as densities on the sphere. Values of  $Y_{m,n}^{-1}$  for  $n > 2$  are omitted due to space considerations. Red and blue colours depict positive and negative values respectively. The longitudinal and latitudinal circles illustrate the nodal lines (zeros of SHFs).

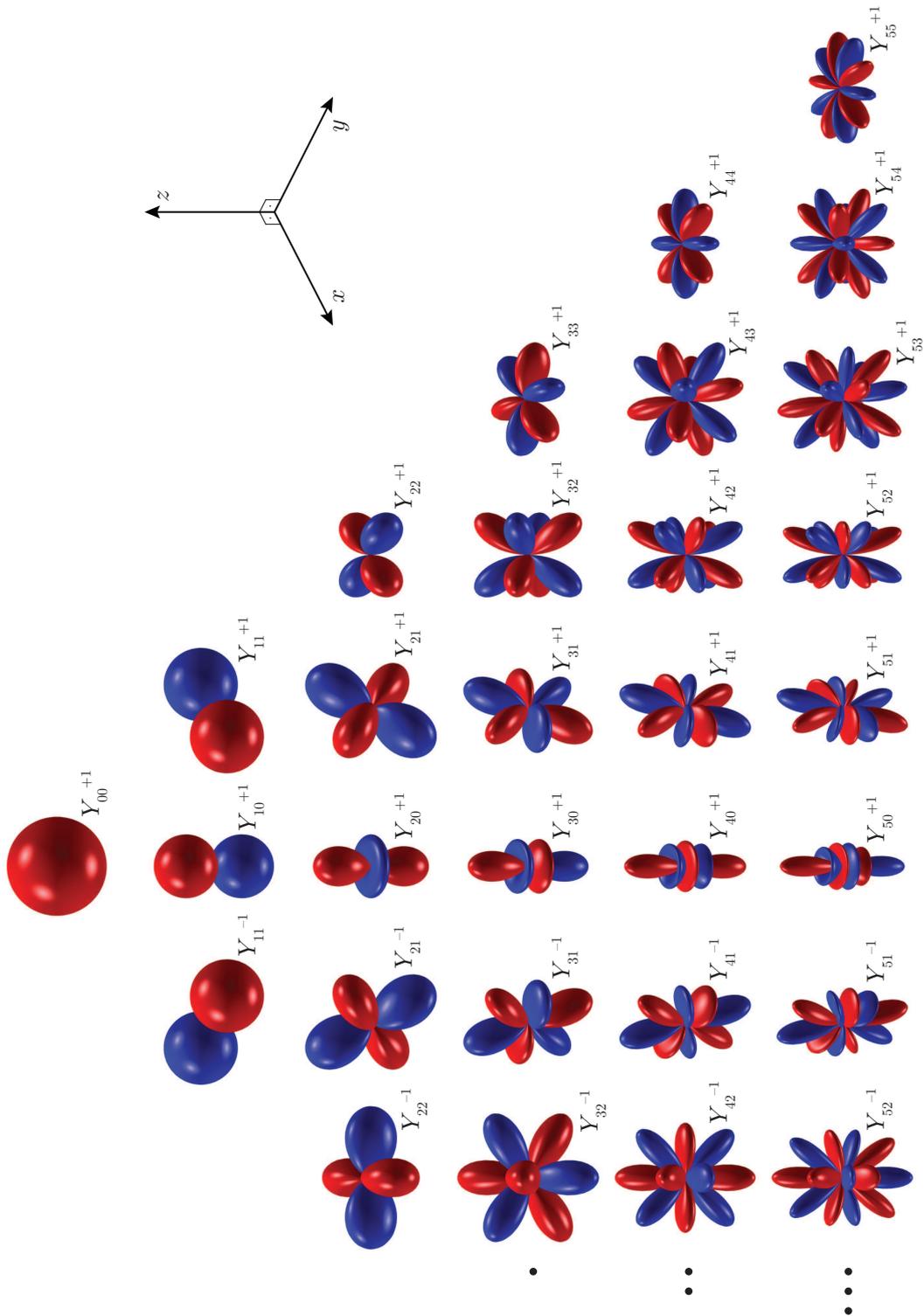


FIGURE 2.2: Similar to Fig. 2.1 with the SHFs here plotted as directivity balloons  $\mathbf{r} = (|Y_{m\sigma}^{\sigma}(\theta, \delta)|, \theta, \delta)$ .

normalised truncation error (Ward and Abhayapala, 2001):

$$\epsilon(kr) = \frac{\int_{\Omega} |p(kr, \theta, \delta) - \check{p}(kr, \theta, \delta)|^2 d\Omega}{\int_{\Omega} |p(kr, \theta, \delta)|^2 d\Omega}, \quad (2.5)$$

where  $\check{p}$  denotes the pressure given by the truncated series. The integration is performed on the unit sphere  $\Omega$ . A rule of thumb can be deduced for the product  $kr$  so that the error of Eq. (2.5) is sufficiently low for most practical applications (Ward and Abhayapala, 2001; Moreau *et al.*, 2006) as below:

$$M = \lceil kr_o \rceil, \quad (2.6)$$

where  $r_o$  is the radius of the reproduction region of interest and  $\lceil \cdot \rceil$  is the ceiling function rounding its argument upwards to the next integer value. Equation (2.6) ensures that  $\epsilon(kr) < 4\%$  ( $-14$  dB) for all points with  $r \leq r_o$ . The above error criterion is based on a free-space propagation condition (i.e., no head/torso or other diffracting object present inside the loudspeaker reproduction array). Solving Eq. (2.6) for the (implied) maximum frequency yields:

$$f_{\text{HOA}} \leq \frac{c}{2\pi r_o} M, \quad (2.7)$$

where  $c = 343$  m/s represents the speed of sound at usual room conditions (i.e., dry air at  $20^\circ\text{C}$  at sea level).

The HOA decoding operation can be formulated as follows; given a spherical array of  $L$  loudspeakers placed at angles  $(\theta_l, \delta_l)$  and emitting plane waves weighted by the gains  $s_l$ , we need to derive the optimum values for those gains in order to recreate a desired unit-amplitude plane wave field, described by the HOA components of Eq. (2.4), in an area inside the loudspeaker array. Expanding the plane waves emitted by the loudspeakers according to Eq. (2.3) and equating their superposition to the desired plane-wave field (wave vector  $\mathbf{k}$ ) yields:

$$\sum_{l=1}^L s_l \sum_{m=0}^M \sum_{n=0}^m \sum_{\sigma=\pm 1} Y_{mn}^{\sigma}(\theta_l, \delta_l) = \sum_{m=0}^M \sum_{n=0}^m \sum_{\sigma=\pm 1} Y_{mn}^{\sigma}(\theta_k, \delta_k). \quad (2.8)$$

Generalizing to the case of an arbitrary field (i.e., superposition of plane waves emanating from different positions  $\mathbf{k}$ ) characterised by the HOA coefficients  $B_{mn}^{\sigma}$  (weighted sum of  $Y_{mn}^{\sigma}(\theta_k, \delta_k)$  SHFs) and expressing the above nested sums in matrix form, yields the below compact equation:

$$\mathbf{C}\mathbf{s} = \mathbf{b}, \quad (2.9)$$

where the  $[(M+1)^2 \times L]$  matrix  $\mathbf{C}$  contains the sampled, by the loudspeakers, SHFs  $Y_{mn}^\sigma(\theta_l, \delta_l)$ , the  $[L \times 1]$  vector  $\mathbf{s}$  contains the loudspeaker gains  $s_l$ , and the  $[(M+1)^2 \times 1]$  vector  $\mathbf{b}$  contains the HOA components  $B_{mn}^\sigma$  characterising the sound field to be reconstructed.

Assuming that the number of loudspeakers is at least equal to the number of HOA components, i.e.,  $L \geq (M+1)^2$  for 3D reproduction, the solution of the above under-determined system (fewer constraining equations [defined by total number of SHFs] than number of unknowns [number of loudspeaker gains]) can be derived as:

$$\hat{\mathbf{s}} = \mathbf{C}^T (\mathbf{C}\mathbf{C}^T)^{-1} \mathbf{b} = \text{pinv}(\mathbf{C}) \mathbf{b} = \mathbf{D}\mathbf{b}, \quad (2.10)$$

where the hat ( $\hat{\cdot}$ ) denotes the least-squares solution, the superscript  $(\cdot)^T$  symbolises the matrix transpose (this study uses real-valued SHFs such that the conjugate transpose is equivalent to a simple transpose),  $(\cdot)^{-1}$  denotes matrix inversion, and  $\text{pinv}(\cdot)$  is the Moore-Penrose pseudoinverse operator. The above solution to the system is unique among the possibly infinite solutions in that it minimises the loudspeaker energy  $\|\hat{\mathbf{s}}\|^2$ . The pseudoinverse matrix  $\text{pinv}(\mathbf{C})$  is typically called the decoding matrix  $\mathbf{D}$ .

Generally,  $\mathbf{C}$  can be ill-conditioned, requiring the application of regularization techniques (Poletti, 2005) to solve the linear system of Eq. (2.9). Alternatively, if quasi-regular loudspeaker setups (Fliege and Maier, 1999; Fliege, 2007) are employed, then the matrix  $\mathbf{C}$  results in being well-behaved as for example shown by Duraiswami *et al.* (2005). The ill-conditioning of  $\mathbf{C}$  is due to the discrete sampling of the SHFs by the loudspeaker which can violate the fundamental orthonormality property (see Sneeuw (1994); Li and North (1997) and Zotter (2009, p. 69) for a rigorous analysis of discrete sampling on the sphere). A measure of the orthonormality error of the sampled SHFs is (Moreau *et al.*, 2006):

$$\mathcal{E} = \mathbf{I}_U - \frac{1}{L} \mathbf{C}\mathbf{C}^T, \quad (2.11)$$

where  $\mathbf{I}_U$  is the  $[U \times U]$  identity matrix, with  $U = (M+1)^2$ .

The Moore-Penrose pseudoinverse, as used in Eq. (2.10), provides the least-squares solution provided  $\mathbf{C}$  is of full row rank (i.e., right invertible). For loudspeaker setups that violate the discrete orthonormality condition of SHFs, this condition is generally not met and the formulation of  $\text{pinv}(\mathbf{C})$  as  $\mathbf{C}^T (\mathbf{C}\mathbf{C}^T)^{-1}$  can potentially result in unstable solutions. In this work, however, the pseudoinverse is computed using MATLAB's `pinv` function which employs a robust singular-value decomposition method (Mathworks, 2013), discarding the singular values that are lower than some threshold [method referred to as 'truncated singular-value decomposition, TSVD' in Granados *et al.* (2013)].

In that way, stable –albeit potentially sub-optimum (Solvang, 2008; Trevino *et al.*, 2010)– solutions for the loudspeaker gains can be derived even if the system is ill-conditioned.

The above loudspeaker gain decoding method is termed ‘mode-matching’ since it is obtained by requiring that the modal decomposition of the loudspeaker sound field matches that of the desired field (Poletti *et al.*, 2010). An alternative procedure is based on the concept of a continuous spherical monopole source distribution  $\mu(r)$ , sampled by the loudspeakers. That ‘simple source’ (Poletti, 2005; Poletti *et al.*, 2010) approach uses quadrature weights to approximate the integral of the source distribution with a discrete sum (Fliege and Maier, 1999), thus omitting one matrix inversion. Although computationally more efficient, it fails to compensate for the non-orthogonality of the sampled SHFs. For that reason, this work employs the mode-matching decoding method.

For the most recent research on the topic of decoding to loudspeaker arrays, the interested reader is referred to the works of Epain *et al.* (2014); Heller and Benjamin (2014).

### 2.3 Spherical-wave expansion

In the interest of completeness, the expansion of a spherical wave, created by a point source at  $(r_s, \theta_s, \delta_s)$ , is herein presented:

$$r_s e^{ikr_s} \frac{e^{-ik\|\mathbf{r}-\mathbf{r}_s\|}}{\|\mathbf{r}-\mathbf{r}_s\|} = -ikr_s e^{ikr_s} \sum_{m=0}^{\infty} j_m(kr) h_m^{(2)}(kr_s) \sum_{n=0}^m \sum_{\sigma=\pm 1} Y_{mn}^{\sigma}(\theta_s, \delta_s) Y_{mn}^{\sigma}(\theta, \delta), \quad (2.12)$$

where  $h_m^{(2)}(kR)$  denotes the  $m$ –th degree spherical Hankel function of the second kind. The employed normalisation  $r_s e^{ikr_s}$  ensures a unit-amplitude and zero-phase pressure at the centre of coordinates ( $\mathbf{r} = \mathbf{0}$ ). Direct comparison of Eq. (2.1) with the above expansion yields the HOA components of a spherical wave created by a point source at  $(r_s, \theta_s, \delta_s)$  and normalised at the centre of coordinates:

$$B_{mn}^{\sigma} \Big|_{\text{sph. wave from } (r_s, \theta_s, \delta_s)} = i^{-(m+1)} k r_s e^{ikr_s} h_m^{(2)}(kr_s) Y_{mn}^{\sigma}(\theta_s, \delta_s). \quad (2.13)$$

As the distance between the point-source ( $r_s$ ) and the observation point ( $r$ ) increases, the spherical wave description reduces to that of the plane wave as seen below:

$$r_s e^{ikr_s} \frac{e^{-ik\|\mathbf{r}-\mathbf{r}_s\|}}{\|\mathbf{r}-\mathbf{r}_s\|} \rightarrow \frac{1}{1 - \frac{\widehat{\mathbf{r}_s} \cdot \mathbf{r}}{r_s}} e^{i\mathbf{k} \cdot \mathbf{r}} \rightarrow e^{i\mathbf{k} \cdot \mathbf{r}}, \text{ for } r_s \gg r, \quad (2.14)$$

where  $\mathbf{k} = k\hat{\mathbf{r}}_s$  and  $(\hat{\cdot})$  denotes the unit vector of its argument. The following vector norm approximation for  $r_s \gg r$  has been used to derive the above result:

$$\begin{aligned} \|\mathbf{r} - \mathbf{r}_s\| &= \sqrt{r^2 + r_s^2 - 2\mathbf{r} \cdot \mathbf{r}_s} = r_s \sqrt{\left(\frac{r}{r_s}\right)^2 + 1 - 2\frac{\mathbf{r} \cdot \hat{\mathbf{r}}_s}{r_s}} \\ &\rightarrow r_s \sqrt{1 - 2\frac{\mathbf{r} \cdot \hat{\mathbf{r}}_s}{r_s}} \rightarrow r_s \left(1 - \frac{\mathbf{r} \cdot \hat{\mathbf{r}}_s}{r_s}\right) = r_s - \mathbf{r} \cdot \hat{\mathbf{r}}_s, \end{aligned} \quad (2.15)$$

where the small argument approximation  $\sqrt{1+x} \approx 1 + \frac{1}{2}x$  has been used on the last line.

In order to derive the limit of the Eq. (2.13) for  $r_s \gg r$ , the large argument approximation of the  $h_m^{(2)}(z)$  functions –derived using the first term of Eq. 10.1.17, p.439 in [Abramowitz and Stegun \(1972\)](#)– is employed:

$$\lim_{z \rightarrow +\infty} h_m^{(2)}(z) = i^{m+1} \frac{e^{-iz}}{z}. \quad (2.16)$$

With the above result, it is easily confirmed that Eq. (2.13) reduces to Eq. (2.4), i.e.:

$$\lim_{r_s \rightarrow +\infty} B_{mn}^\sigma \Big|_{\text{sph.wave from } (r_s, \theta_s, \delta_s)} = B_{mn}^\sigma \Big|_{\text{pl.wave from } (\theta_s, \delta_s)}. \quad (2.17)$$

Practically, when the observation point lies on a centred human head, such that  $r \sim 0.09$  m, the spherical waves emitted from point sources lying at  $r_s > \sim 2$  m can be safely considered as plane waves in the vicinity of the head.

## 2.4 Sound field reconstruction

To illustrate the wave reconstruction principle using a HOA system, we consider an  $L = 100$  quasi-regular ([Fliege, 2007](#)) spherical loudspeaker array decoding a plane wave incident from  $(\theta_k, \delta_k) = (+60^\circ, 0^\circ)$ , employing an  $M = 3$  and  $M = 7$  order successively. The loudspeaker gains are derived from Eq. (2.10). Although an  $L_3 = 16$  and  $L_7 = 64$  loudspeaker array would suffice for the decoding of a 3<sup>rd</sup> and 7<sup>th</sup> order HOA system respectively (according to the  $L \geq (M+1)^2$  rule), the orthonormality error of quasi-regular configurations of that size –computed via Eq. (2.11)– was significant. The employed 100 channels array ensured  $\mathcal{E} < 0.05$  for all terms up to  $M = 7$  order and was used to decode both HOA systems for comparability of results. With the loudspeaker gains known, the HOA-reconstructed field can be computed as the superposition of  $L$  plane waves, with amplitudes equal to  $s_l$  ( $l = 1, \dots, L$ ), incident from the loudspeaker positions.

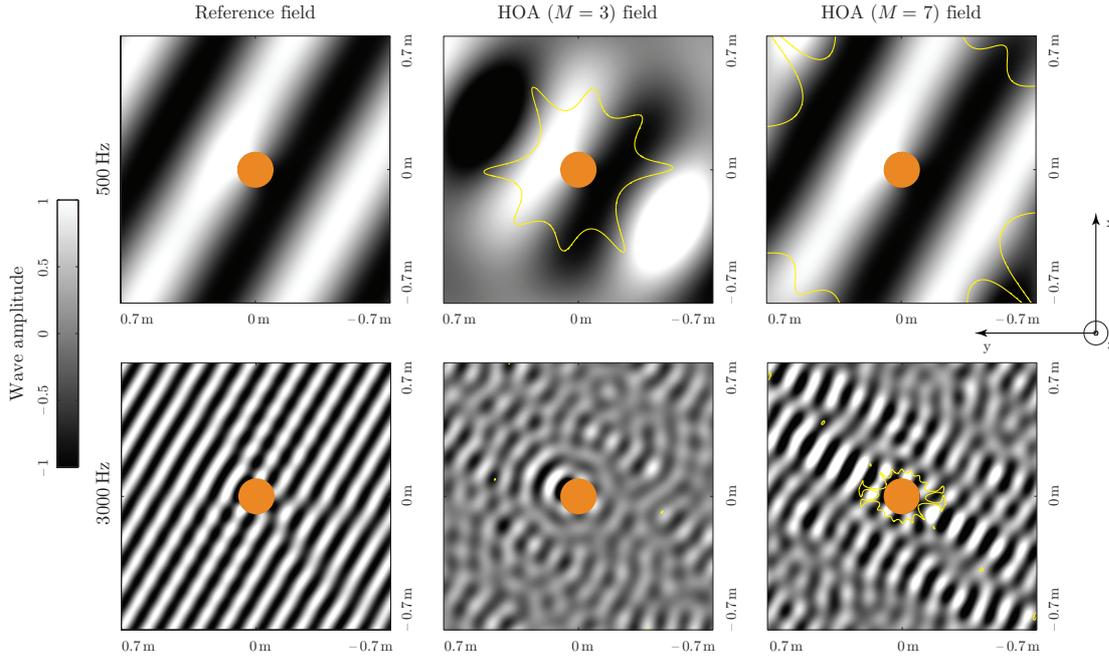


FIGURE 2.3: Reconstruction of a unit-amplitude plane wave from  $(+60^\circ, 0^\circ)$  (first column) using an  $M = 3$  (second column) and  $M = 7$  (third column) HOA system. Two different frequencies are examined: 500 Hz (top row) and 3000 Hz (bottom row). The middle (orange) circles represent a head, modelled as a rigid sphere of radius  $a = 0.09$  m. The (yellow) contour isolines mark the regions where the reconstruction error is  $\leq -14$  dB.

Commonly in literature wave reconstruction surface plots assume free-space propagation. However, in line with one of the aims of this work, the effect of a listener's head (herein modelled by a rigid sphere of radius  $a = 0.09$  m) is taken into account while creating the pressure surface plots. It should be emphasised that including a listener (acoustic manikin) inside the reconstructed sound field yields increased HOA errors at lower frequencies (compared to the free-space case that omits the listener). This observation is illustrated in much detail in Section 4.4.2.1 of Chapter 4.

Fig. 2.3 shows the instantaneous pressure due to two reference unit-amplitude plane waves of frequencies 500 Hz and 3000 Hz as well as the instantaneous pressure resulting from their HOA sound field reconstruction. The central (orange) circle in the plots represents the head-modelling sphere and is plotted to scale with the wave-fronts and the axes labelling. The isoline contours delimit regions that satisfy the rule of thumb  $\epsilon(kr) < 4\%$  ( $-14$  dB). The wave-front plots highlight the fundamental HOA property that pressure reconstruction errors increase with increasing frequency and decreasing HOA order, for a given distance  $r_o$  from the centre of the loudspeaker array.

Regarding the effect of the number of the loudspeakers [provided that  $L \geq (M + 1)^2$ ] on the overall HOA pressure reconstruction errors, Solvang (2008) concluded that for a 2D HOA system (i.e., horizontal circular loudspeaker array reconstructing fields with

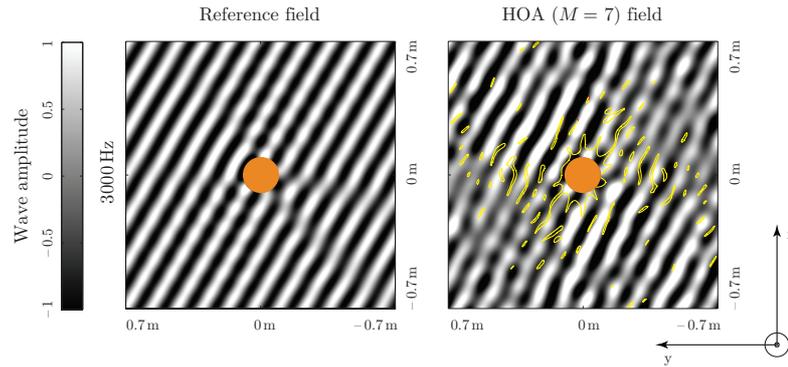


FIGURE 2.4: As Fig. 2.3 using an  $L = 64$  loudspeaker array. Only the 3000 Hz wave and an  $M = 7$  order HOA system are considered.

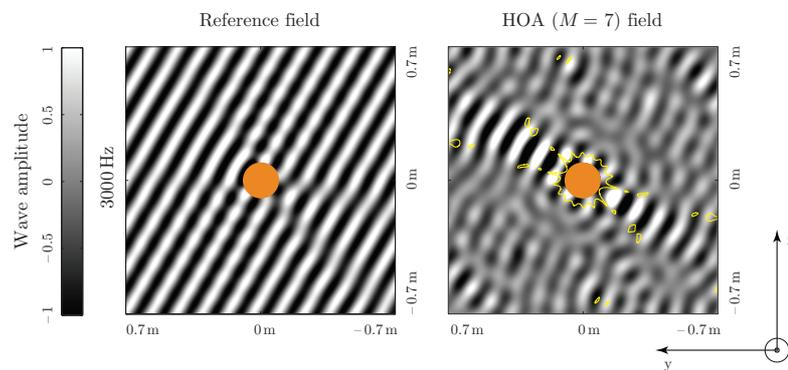


FIGURE 2.5: As Fig. 2.3 using an  $L = 196$  loudspeaker array. Only the 3000 Hz wave and an  $M = 7$  order HOA system are considered.

no height information), the use of more than the necessary [i.e.,  $(M + 1)^2$ ] number of loudspeakers impairs the pressure reconstruction *outside the sweet-spot* with radius  $r_o$ . However, as the author concludes: “*utilizing more loudspeakers can decrease the reproduction error in the near perfect reproduction region*”, i.e., when  $r \approx r_o$ . Similar observations have been discussed by the author with various researchers. In this work, without delving into a detailed analytical solution, we plot the pressure reconstruction wave-fronts, similar to Fig. 2.3 but only for the 3000 Hz and  $M = 7$  case, employing:

- (1) the minimum required  $L = (7 + 1)^2 = 64$  loudspeakers arranged in a quasi-regular setup (Fig. 2.4);
- (2) a denser quasi-regular setup of  $L = 196$  loudspeakers (Fig. 2.5).

The reconstructed wave-fronts and the range of the ‘accurate-enough’ regions graphically suggest that the smallest loudspeaker array ( $L = 64$ ) yields the smallest pressure error further away from the centred hard sphere (Fig. 2.4), while the largest loudspeaker array ( $L = 196$ ) effectively smears the wave-fronts further away from the centred hard sphere (Fig. 2.5). However, in the vicinity of the hard sphere, the  $L = 64$  array seems to perform worse, followed by the two larger arrays ( $L = 100$  and  $L = 196$ ). Hence, for

a centred single listener, it appears that a dense loudspeaker array is preferable for a ‘best-case’ validation of HOA systems of different orders.

Another real-world consideration that should be mentioned is that increasing the number of loudspeakers leads to a higher density of reflections off their front baffle. Such secondary radiation is expected to impair the reconstructed sound-field in an uncontrollable manner.

## 2.5 HOA directivity plots

To further understand the inherent directional properties of a HOA system, we consider a dense-enough loudspeaker array (i.e.,  $L \gg (M + 1)^2$ ) that *satisfies the discrete orthonormality property*. In this case the solution for the loudspeaker gains can be interpreted as a panning function  $G$  (Neukom, 2007), given by (Daniel, 2001):

$$G(\gamma) = \frac{1}{L} \sum_{m=0}^M (2m + 1) P_m(\cos \gamma), \quad (2.18)$$

where  $P_m$  is the Legendre polynomial of degree  $m$  and  $\gamma$  is the angle between each loudspeaker position  $(\theta_l, \delta_l)$  and the virtual plane-wave source direction  $(\theta_k, \delta_k)$ , such that:

$$\cos \gamma = \hat{\mathbf{r}}_l \cdot \hat{\mathbf{k}} = \sin \delta_l \sin \delta_k + \cos \delta_l \cos \delta_k \cos(\theta_l - \theta_k), \quad l = 1, \dots, L. \quad (2.19)$$

where  $\hat{\mathbf{r}}_l \cdot \hat{\mathbf{k}}$  represents the inner product between the unit position vector of the  $l$ -th loudspeaker,  $\hat{\mathbf{r}}_l = (1, \theta_l, \delta_l)$ , and the unit wave vector,  $\hat{\mathbf{k}} = (1, \theta_k, \delta_k)$ .

Fig. 2.6 shows the 3D plot of  $G(\gamma)$ , for HOA systems of order  $M = 3$  and  $M = 7$  respectively. The employed virtual loudspeaker array  $(\theta_l, \delta_l)$ , with  $l = 1 \dots L$ , does not necessarily preserve the discrete orthonormality condition and are solely used for the purpose of producing smooth plots for the panning functions  $G(\gamma)$  for two different HOA system orders ( $M = 3$  and  $M = 7$ ). A virtual plane-wave source is considered at  $(\theta_k, \delta_k) = (0^\circ, 0^\circ)$  so that the main lobe faces the positive  $x$ -axis.

The HOA directivity plots, or panning functions, are instantly understood as depicting the loudspeaker gain dependence with azimuth and elevation for a certain virtual plane-wave source (incidence direction indicated by the direction where the maximum lobe of  $G(\gamma)$  points to). Alternatively, the HOA directivity plots can be interpreted via a projection paradigm, where the  $l$ -th loudspeaker is conceptually fed with the signal captured by a centred virtual microphone having a directivity of  $G(\gamma)$ , with the main

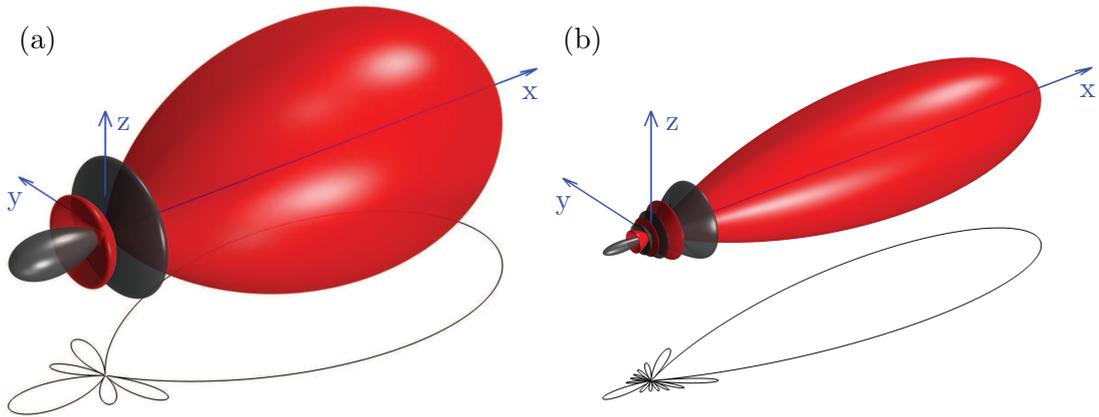


FIGURE 2.6: Directivity plots (HOA panning function)  $G(\gamma)$  of an: (a)  $M = 3$ , and (b)  $M = 7$  system. The virtual plane wave was considered to impinge from  $(\theta_k, \delta_k) = (0^\circ, 0^\circ)$  where the main lobe points to. The red and grey lobes represent positive and negative loudspeaker gains respectively. The contour plots below the 3D surfaces show the directivity pattern at the horizontal  $xy$ -plane.

lobe of the directivity pointing to the  $l$ -th loudspeaker (Moreau *et al.*, 2006; Li *et al.*, 2004b).

Figure 2.6 illustrates that a HOA decoding system becomes more spatially selective with increasing order  $M$ . At the limit of  $M \rightarrow \infty$ , Eq. (2.18) reduces to a Dirac function pointing to the position of the virtual plane-wave source, which results, as expected, in a perfect reconstruction of the field.

## 2.6 Modified (max- $r_E$ ) decoding

Since HOA cannot control the field at frequencies above the estimation of Eq. (2.7), it would be beneficial to modify the decoded loudspeaker signals so that most of the energy is emitted by loudspeakers that are close to the estimated direction of the virtual wave. This is effectively achieved by the ‘max- $r_E$ ’ alternative decoding method, proposed by Daniel *et al.* (1998). The max- $r_E$  decoding aims at maximizing the norm  $r_E$  of the energy vector  $\vec{E}$  proposed by Gerzon (1992). Although the energy vector  $\vec{E}$  does not directly correspond to a physical property of the system, Daniel suggested (Daniel *et al.*, 1998; Daniel, 2001) that its norm  $r_E$  is related to the concentration of the high frequency energy in the direction of  $\vec{E}$ . Therefore a solution that maximises the  $r_E$ , results in perceptually focusing the high frequency energy towards the expected direction. (Daniel *et al.*, 1998) also shows that it improves the reconstruction of Interaural Level Differences (ILDs).

However, such a solution reduces the norm  $r_V$  of the velocity vector  $\vec{V}$  (Gerzon, 1992; Daniel, 2001) which translates into a degradation of the low frequency reconstruction.

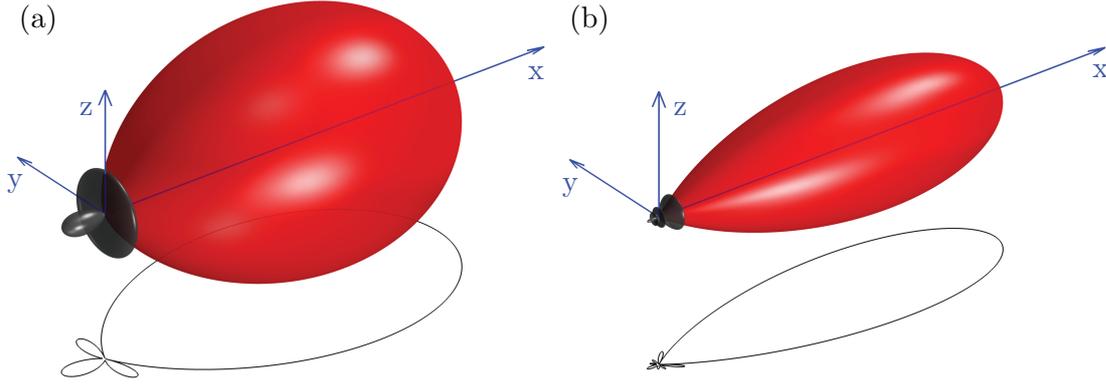


FIGURE 2.7: Same as Fig. 2.6 using the max- $r_E$  decoding.

Hence, this modified decoding should only be used above a frequency threshold, as is for example  $f_{\text{HOA}}$  of Eq. (2.7).

The max- $r_E$  decoding involves the use of order-dependent gains  $g_m$ , which weigh the decoding matrix  $\mathbf{D}$ , effectively tapering the contribution of higher order components:

$$\mathbf{D}_m = \mathbf{D} \text{diag} \left[ g_0 \ g_1 \ g_1 \ g_1 \ \cdots \ \underbrace{g_m \cdots g_m}_{2m+1} \ g_M \cdots g_M \right], \quad (2.20)$$

where the  $\text{diag}[\cdot]$  function returns a square  $[U \times U]$ , diagonal matrix, with  $U = (M + 1)^2$ , whose main diagonal elements are equal to the input argument vector. The gains  $g_m$ , as well as their derivation, are given by Daniel (2001, p. 183).

The modified decoding matrix  $\mathbf{D}_m$  results in a concentration of the energy towards the expected direction, by reducing the secondary lobes at the expense of broadening the main one as can be observed by comparing Fig. 2.7 with Fig. 2.6. The below equation –modified version of Eq. (2.18), as derived by Daniel *et al.* (2003)– was used to plot the directivity patterns of Fig. 2.7:

$$G(\gamma) = \frac{1}{L} \sum_{m=0}^M (2m + 1) g_m P_m(\cos \gamma), \quad (2.21)$$

For the sake of completeness, the interested reader can refer to Epain *et al.* (2014) and Heller and Benjamin (2014) for optimum methods of decoding –applying the max- $r_E$  concept– HOA components to irregular loudspeaker setups.

## 2.7 Field encoding with a microphone array (Model-based encoding)

Until now it has been assumed that the positions of the considered plane-wave virtual sources are known so that Eq. (2.4) can be used to calculate the HOA components  $B_{mn}^\sigma$  needed by Eq. (2.10) to compute the loudspeaker gains  $s_l$  (HOA decoding operation). However, in real-life scenarios, the positions of the sources are not known and instead need to be estimated. A spherical microphone array is usually employed for that purpose, performing the so-called HOA encoding of the acoustic scene.

In this work we consider the most frequently used, due to its robustness (Meyer, 2001; Meyer and Elko, 2002; Rafaely, 2005), implementation of a rigid sphere of radius  $R$  using  $Q$  flush-mounted omnidirectional microphones. To illustrate the encoding process, we first assume a theoretical, continuous pressure spherical microphone of radius  $R$ . The pressure on its surface, due to an impinging wave characterised by the HOA components  $B_{mn}^\sigma$ , can be calculated (Moreau *et al.*, 2006; Williams, 1999; Duda and Martens, 1998) using the series expansion:

$$p(r = R, \theta, \delta) = \sum_{m=0}^{\infty} W_m(kR) \sum_{n=0}^m \sum_{\sigma=\pm 1} B_{mn}^\sigma Y_{mn}^\sigma(\theta, \delta), \quad (2.22)$$

with the rigid sphere mode strength term  $W_m(kR)$  defined by:

$$\begin{aligned} W_m(kR) &= i^m \left( j_m(kR) - \frac{j'_m(kR)}{h_m^{(2)'}(kR)} h_m^{(2)}(kR) \right) \\ &= \frac{i^{m-1}}{(kR)^2 h_m^{(2)'}(kR)}, \end{aligned} \quad (2.23)$$

where the prime  $(\cdot)'$  denotes the derivative of the spherical Bessel and Hankel functions with respect to their argument  $kR$ . In order to simplify the first line of the above equation we have employed the definition  $h_m^{(2)}(z) = j_m(z) - iy_m(z)$ , where  $y_m(z)$  are the  $m$ -th degree spherical Bessel functions of the second kind –also known as Neumann functions–, along with the Wronskian of the spherical Bessel functions [see Eq. 10.1.6, p.437 in Abramowitz and Stegun (1972)]:

$$\mathcal{W} \{ j_m(z), y_m(z) \} = j_m(z) y'_m(z) - j'_m(z) y_m(z) = z^{-2}. \quad (2.24)$$

Given the pressure on the continuous sensing sphere, expressed by Eq. (2.22), and using the orthonormality property of the spherical harmonic functions (Williams, 1999), the

unknown HOA components, which fully characterise the field, can be resolved as:

$$B_{mn}^\sigma = \frac{1}{W_m(kR)} \int_{\Omega} p_R(\theta, \delta) Y_{mn}^\sigma(\theta, \delta) d\Omega, \quad (2.25)$$

provided  $W_m(kR) \neq 0$ . The pressure on the rigid sphere, expressed by Eq. (2.22), is referred to as  $p_R(\theta, \delta)$  in the above equation for brevity of notation.

Since a continuous sensing sphere is not realizable, a finite number of  $Q$  microphones, as described above, are employed to sample the pressure at discrete points. Hence, the integral of Eq. (2.25) can be approximated by a quadrature:

$$B_{mn}^\sigma \approx \frac{1}{W_m(kR)} \sum_{q=1}^Q \alpha_q p_R(\theta_q, \delta_q) Y_{mn}^\sigma(\theta_q, \delta_q), \quad (2.26)$$

with the quadrature weights  $\alpha_q$ , number  $Q$  and exact locations of the microphones  $(\theta_q, \delta_q)$  derived such that the orthonormality of the sampled, at the microphone points, SHFs property is preserved (Rafaely, 2005).

In a practical context, the microphone diaphragms are not dimensionless points but rather occupy some surface on the array sphere. Thus their output is not exactly equal to  $p_R(\theta_q, \delta_q)$  but, instead, the average of the pressure on their diaphragm. Meyer and Elko (2002) refer to that phenomenon as ‘modal low-pass filtering’. Although not vigorously investigated, this phenomenon is not expected to affect the lower degree components which are of interest in this work. Hence, the sensors of the considered spherical microphone arrays will be assumed as being ideal, point pressure-sensors.

An alternative method to estimate the HOA components from the microphone signals that does not require a quadrature can be formulated exploiting the matrix notation in a similar way as presented in Section 2.2. Sampling Eq. (2.22) at the microphone positions  $(\theta_q, \delta_q)$  and converting the summations to matrix multiplications, the following compact notation (Moreau *et al.*, 2006) can be derived:

$$\mathbf{Y} \text{diag}[\mathbf{W}_{3D}(kR)] \mathbf{b} = \mathbf{p}_R, \quad (2.27)$$

where the  $[Q \times (M+1)^2]$  matrix  $\mathbf{Y}$  contains the sampled by the microphones SHFs  $Y_{mn}^\sigma(\theta_q, \delta_q)$ , the  $[(M+1)^2 \times (M+1)^2]$  diagonal matrix  $\text{diag}[\mathbf{W}_{3D}(kR)]$  contains the inverse rigid sphere mode strength terms, the  $[(M+1)^2 \times 1]$  vector  $\mathbf{b}$  contains the HOA components  $B_{mn}^\sigma$  to be estimated and the  $[Q \times 1]$  vector  $\mathbf{p}_R$  contains the microphone pressure signals  $p_R(\theta_q, \delta_q)$ . The diagonal mode strength matrix is formed using the

following vector passed as argument to the  $\text{diag}[\cdot]$  function:

$$\mathbf{W}_{3\text{D}} = \begin{bmatrix} W_0 & W_1 & W_1 & W_1 & \cdots & \underbrace{W_m \cdots W_m}_{2m+1} & W_M \cdots W_M \end{bmatrix}, \quad (2.28)$$

where the  $kR$  dependence has been dropped for conciseness. All elements of the aforementioned matrices are relevant to a *single frequency point*. Hence, characterising a broadband sound field necessitates the consideration of Eq. (2.27) *separately* for every frequency of interest.

Assuming  $Q \geq (M+1)^2$ , the least-squares solution of the under-determined linear system of Eq. (2.27) can be written (Moreau *et al.*, 2006; Poletti, 2005) as:

$$\begin{aligned} \hat{\mathbf{b}} &= \text{diag}[\mathbf{W}_{3\text{D}}(kR)]^{-1} \mathbf{Y}^T (\mathbf{Y}\mathbf{Y}^T)^{-1} \mathbf{p}_R \\ &= \text{diag}[\mathbf{H}_{\text{eq},3\text{D}}(kR)] \text{pinv}(\mathbf{Y}) \mathbf{p}_R \\ &= \mathbf{E} \mathbf{p}_R, \end{aligned} \quad (2.29)$$

where  $\mathbf{E}$  denotes the encoding matrix. The equalization diagonal matrix  $\text{diag}[\mathbf{H}_{\text{eq},3\text{D}}(kR)]$  has its diagonal elements  $H_{\text{eq},m}(kR)$  formed by inverting the terms of Eq. (2.28) one-by-one.

In the above equation, vector  $\hat{\mathbf{b}}$  contains only the HOA components  $B_{mn}^\sigma$  up to the order of interest  $M$ . Jin *et al.* (2014), however, formulate the minimisation problem to include higher-order HOA components and force the encoding matrix  $\mathbf{E}$  to minimise those higher-order components so that they do not pollute (i.e., via aliasing) the useful lower order components.

Inverting the mode strength terms poses implementation problems due to the very low value of high-degree ( $m$ ) functions  $W_m(kR)$  at low frequencies (Rafaely, 2004; Baumgartner *et al.*, 2011). Increasing the radius  $R$  of the microphone shifts the resulting large gains to lower frequencies, at the expense of shifting spatial aliasing problems to lower frequencies (Moreau *et al.*, 2006; Rafaely, 2005; Rafaely *et al.*, 2007; Abhayapala and Ward, 2002). Estimating the HOA components at low frequencies, without having to resort to a large radius  $R$  design, is commonly treated by applying Tikhonov regularization (Moreau *et al.*, 2006; Moreau, 2006; Poletti, 2005; Zotkin *et al.*, 2010) to the solution of Eq. (2.29). Effectively, the equalization terms are modified so that:

$$\tilde{H}_{\text{eq},m}(kR) = \frac{|W_m(kR)|^2}{|W_m(kR)|^2 + \lambda^2} \cdot \frac{1}{W_m(kR)}, \quad (2.30)$$

where the parameter  $\lambda$  controls the amount of regularization, with higher values yielding a more regularised solution. The parameter  $\lambda$  tapers off the high magnitudes of the

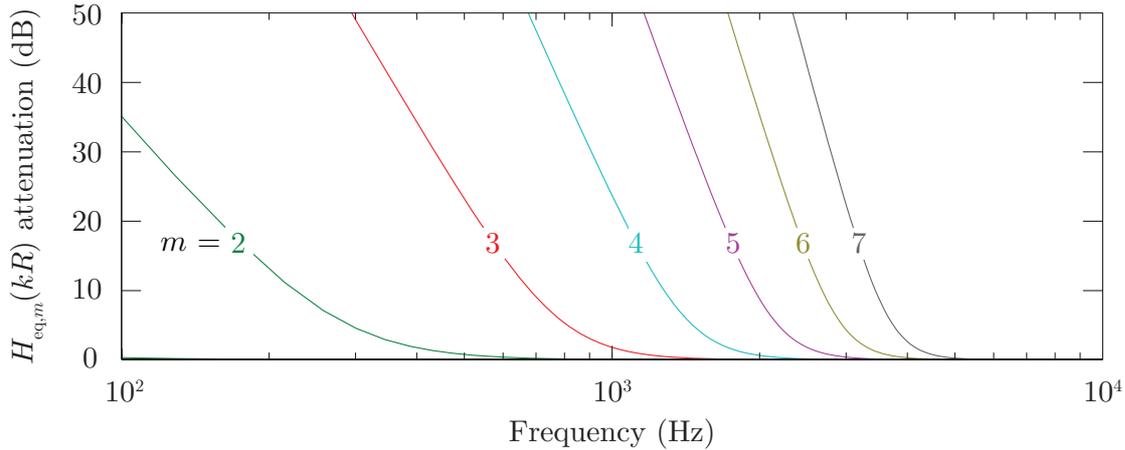


FIGURE 2.8: Attenuation of the radial equalization filters  $H_{\text{eq},m}(kR)$ ,  $m = 0..7$  relevant to a spherical microphone array of radius  $R = 0.06$  m, due to a realistic regularization value of  $\lambda = 0.01$ .

inverse of the mode strengths (radial equalization filters) at low frequencies (see [Epain and Daniel, 2008](#), Section 3) so that the microphone self-noise is not excessively boosted. Fig. 2.8 illustrates how regularization attenuates the radial equalization filters of the  $m$ -th degree by showing the ratio  $H_{\text{eq},m}(kR)/\tilde{H}_{\text{eq},m}(kR)$ . A number of researchers [[Moreau et al. 2006](#), Eq. (34); [Jin et al. 2014](#), Eq. (15)] provide rules to set  $\lambda$  according to the accepted amplified noise level. However, this work takes another approach by examining the actual noise, due to the encoding of the microphone array signals, at the ears of a dummy listener (see Sections 4.3.3.3 and 4.4.1).

After applying the regularised equalization filters to Eq. (2.29), the encoding matrix is modified to:

$$\tilde{\mathbf{E}} = \text{diag}[\tilde{\mathbf{H}}_{\text{eq},3\text{D}}(kR)] \text{pinv}(\mathbf{Y}). \quad (2.31)$$

The operation of employing the regularised encoding matrix  $\tilde{\mathbf{E}}$  to yield the HOA components  $\tilde{\mathbf{b}} = [\tilde{B}_{mn}^\sigma]$  from the microphone pressure signals  $\mathbf{p}_R$  using Eq. (2.29) will be referred to as ‘Ambisonics’ encoding.

Interested readers can refer to [Gauthier et al. \(2011\)](#) for a more advanced treatment of regularisation, including the presentation of a novel relevant approach.

An alternative way of spherical microphone array encoding is formulated in the theory of beamforming or plane-wave decomposition ([Meyer and Elko, 2002](#); [Rafaely, 2004](#); [Li and Duraiswami, 2007](#); [Zotkin et al., 2010](#)). According to these formulations weights are applied to the microphones in order to create a given directional response (usually a truncated Dirac) pointing towards a look-out direction and then signals mainly incident from that direction are picked up. Hence, the field can be spatially sampled by

performing a number of such beamforming operations on a given grid of look-out directions. On the contrary, the Ambisonics approach, described above, is aiming, through a least-squares optimization, at estimating the set of HOA components  $\tilde{B}_{mn}^\sigma$  that best explains the whole captured sound field.

## 2.8 ‘Shape-matching’ encoding method (Measurement-based encoding)

An alternative encoding method, termed ‘shape-matching’ (SM) has been validated and employed in this study. SM was first reported by [Moreau *et al.* 2006, Eq. (46)] and then used by Marschall *et al.* (2012), while other researchers (Farina *et al.*, 2010; Zaunschirm and Zotter, 2014; Jin *et al.*, 2014) have also employed similar measurement-based encoding techniques. The SM encoding operation aims at minimising the error between the desired SHFs  $Y_{mn}^\sigma(\theta_g, \theta_g)$ , sampled at the direction  $(\theta_g, \theta_g)$  of an incident plane wave, and the directivity patterns reproduced by the array. The minimisation should be performed over  $g = 1 \dots G$  incident plane waves evenly distributed on the unit sphere.

Essentially, SM can be understood as performing  $G$  encoding operations for  $G$  incident plane waves and then minimising the error between the derived and ideal  $B_{mn}^\sigma$  components, with the ideal  $B_{mn}^\sigma$  being equal to the SHFs sampled at the plane-wave incidence directions. In matrix formulation:

$$\underset{\mathbf{E}_{\text{SM}}}{\text{minimise}} \quad \left\| \mathbf{E}_{\text{SM}} \mathbf{P}_G - \mathbf{Y}_G \right\|_F^2, \quad (2.32)$$

where  $\mathbf{E}_{\text{SM}}$  is the  $[U \times Q]$  SM encoding matrix,  $\mathbf{P}_G$  the  $[Q \times G]$  matrix of all  $Q$ -mic pressures due to the assumed  $G$  incident plane waves,  $\mathbf{Y}_G$  is the  $[U \times G]$  matrix with the  $U = (M + 1)^2$  SHFs sampled at the incident wave positions, and  $\|\cdot\|_F$  is the Frobenius norm (Golub and Loan, 1996, §2.3). The above problem can be converted to a standard least-squares problem by vectorizing  $\mathbf{Y}_G$  and using the Kronecker product (Horn and Johnson, 1994, §4). Considering a regularised formulation for the minimisation problem, as in (Boyd and Vandenberghe, 2004, §6.3.2), leads to:

$$\underset{\mathbf{E}_{\text{SM}}}{\text{minimise}} \quad \left\| (\mathbf{P}_G^H \otimes \mathbf{I}_U) \text{vec}(\mathbf{E}_{\text{SM}}) - \text{vec}(\mathbf{Y}_G) \right\|_2^2 + \lambda_{\text{SM}} \left\| \text{vec}(\mathbf{E}_{\text{SM}}) \right\|_2^2, \quad (2.33)$$

where  $(\cdot)^H$  denotes the Hermitian transpose,  $\text{vec}(\cdot)$  stacks the columns of its matrix argument on top of one another, and  $\otimes$  denotes the Kronecker product. The above problem can be expressed in the form:  $\underset{\mathbf{x}}{\text{minimise}} \quad \left\| \mathbf{A}\mathbf{x} - \mathbf{b} \right\|_2^2 + \lambda \left\| \mathbf{x} \right\|_2^2$ , with:  $\mathbf{A} = \mathbf{P}_G^H \otimes \mathbf{I}_U$ ,

$\mathbf{x} = \text{vec}(\mathbf{E}_{\text{SM}})$  and  $\mathbf{b} = \text{vec}(\mathbf{Y}_G)$ . The solution can thus be expressed as:

$$\begin{aligned}
\mathbf{x} &= (\mathbf{A}^H \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^H \mathbf{b} \Rightarrow \\
\text{vec}(\tilde{\mathbf{E}}_{\text{SM}}) &= [(\mathbf{P}_G \otimes \mathbf{I}_U) (\mathbf{P}_G^H \otimes \mathbf{I}_U) + \lambda_{\text{SM}} \mathbf{I}_{UQ}]^{-1} (\mathbf{P}_G \otimes \mathbf{I}_U) \text{vec}(\mathbf{Y}_G) \Rightarrow \\
\text{vec}(\tilde{\mathbf{E}}_{\text{SM}}) &= [(\mathbf{P}_G \mathbf{P}_G^H \otimes \mathbf{I}_U) + \lambda_{\text{SM}} \mathbf{I}_{UQ}]^{-1} \text{vec}(\mathbf{Y}_G \mathbf{P}_G^H) \Rightarrow \\
[(\mathbf{P}_G \mathbf{P}_G^H \otimes \mathbf{I}_U) + \lambda_{\text{SM}} \mathbf{I}_{UQ}] \text{vec}(\tilde{\mathbf{E}}_{\text{SM}}) &= \text{vec}(\mathbf{Y}_G \mathbf{P}_G^H) \Rightarrow \\
\text{vec}(\tilde{\mathbf{E}}_{\text{SM}} \mathbf{P}_G \mathbf{P}_G^H + \lambda_{\text{SM}} \tilde{\mathbf{E}}_{\text{SM}}) &= \text{vec}(\mathbf{Y}_G \mathbf{P}_G^H). \tag{2.34}
\end{aligned}$$

Finally, the last equation provides the solution for the regularised shape-matching encoding matrix as seen below:

$$\tilde{\mathbf{E}}_{\text{SM}} = \mathbf{Y}_G \mathbf{P}_G^H (\mathbf{P}_G \mathbf{P}_G^H + \lambda_{\text{SM}} \mathbf{I}_Q)^{-1}. \tag{2.35}$$

Jin *et al.* (2014) give a slightly different formulation for the above problem and also extend the optimisation order  $M$  to higher values in order to better control the spatial aliasing products. Considering such higher order SHF components was herein deemed unnecessary, provided that a dense-enough grid (i.e., large value of  $G$ ) was used for the minimisation.

## 2.9 Comparison of ‘shape-matching’ and ‘Ambisonics’ encoding methods

The shape-matching encoding method offers the obvious advantage that it takes into considerations the sensitivity and phase mismatches of the microphones as well as any positioning errors on the sphere. That occurs because it employs a set of measured pressures ( $\mathbf{P}_G$ ), from  $G$  source positions to all  $Q$  microphones, to compute the encoding matrix  $\tilde{\mathbf{E}}_{\text{SM}}$ . On the contrary, the Ambisonics encoding matrix  $\tilde{\mathbf{E}}$ , hereafter referred to as  $\tilde{\mathbf{E}}_{\text{AMB}}$  for disambiguation, relies on the exact knowledge of the array parameters (response and positioning of microphones) and on the assumption that the hard-sphere model accurately expresses the pressure on the non-zero surface sensors of the real spherical microphone array.

Apart from the above practical advantage, the shape-matching encoding method is herein shown to better control the spatial aliasing (i.e., large erroneous directivity lobes) of the microphone array at high frequencies. This is presumed to result from the penalisation of the deviations between the reconstructed directivities and the ideal SHFs in the minimisation problem formulation of Eq. (2.33). In turn, better control over spatial

aliasing results in lower pressure reconstruction errors. Similar observations were also reported by [Jin et al. \(2014\)](#). In that work, the control of the higher frequency aliasing products is attained by forcing the encoding matrix to decompose the SHF components up to the order of interest  $M$  and at the same time to minimise the power of SHF components of higher-than- $M$  orders. The interested reader is encouraged to seek more enlightening details in the Sections II.B and IV.B of [Jin et al. \(2014\)](#).

To quantify the pressure errors on the head of a listener (modelled by a rigid sphere) incurred by the HOA-encoded sound field, we use the normalised mean-square error (NMSE). For that purpose, the simulation framework outlined in the following steps is considered:

- the  $Q = 62$ ,  $R = 0.05$  m NAL microphone array configuration [see Fig. 6.3(b)] is employed, assuming perfectly matched and positioned sensors, operating at the HOA order  $M = 5$ ;
- a grid of  $G = 400$  quasi-regularly ([Fliege, 2007](#)) positioned virtual sources is employed to derive the shape-matching encoding matrix  $\tilde{\mathbf{E}}_{\text{SM}}$ , according to Eq. (2.35), with a regularisation parameter of  $\lambda_{\text{SM}} = 0.4$ ;
- the Ambisonics encoding matrix  $\tilde{\mathbf{E}}_{\text{AMB}}$  is derived via Eq. (2.31) with a regularisation parameter of  $\lambda_{\text{AMB}} = 0.005$ ;
- the choice of  $\lambda_{\text{SM}}$  and  $\lambda_{\text{AMB}}$  yields a good compromise between microphone noise and pressure reconstruction errors ([Oreinos and Buchholz, 2014](#)), while ensuring *equal HOA system noise power for both encoding strategies*;
- a set of  $K = 100$  quasi-regularly ([Fliege, 2007](#)) positioned virtual sources are considered. The generated pressures on the microphone array sensors are calculated using the rigid-sphere model solution via Eqs. (2.22) and (2.23);
- the HOA components  $\tilde{B}_{mn}^{\sigma}$  are derived via the encoding operation of Eq. (2.29). Both encoding matrices  $\tilde{\mathbf{E}}_{\text{SM}}$  and  $\tilde{\mathbf{E}}_{\text{AMB}}$  are applied in succession to yield two sets of HOA components;
- the ideal HOA components  $B_{mn}^{\sigma}$  of the spatially low-passed sound field (series expansion truncated at  $m = M$ ), due to all  $K$  virtual sources, are computed using Eq. (2.4);
- a grid of  $I = 400$  quasi-regularly ([Fliege, 2007](#)) positioned observation points lying on a hard sphere of radius  $r_e = 0.09$  m is considered;
- the pressure values on all  $i = 1 \dots I$  points of the hard sphere –due to the  $k$ –th virtual source– are calculated [Eqs. (2.22) and (2.23)] using the ideal HOA components  $B_{mn}^{\sigma}$ , thus yielding the reference pressure values  $p(kr_e, \theta_i, \delta_i, \theta_k, \delta_k)$ ;
- the pressure values on all  $i = 1 \dots I$  points of the hard sphere –due to the  $k$ –th

virtual source— are additionally calculated [Eqs. (2.22) and (2.23)] using the derived (from the encoding) HOA components  $\tilde{B}_{mn}^\sigma$ , thus yielding the pressure values  $\tilde{p}(kr_e, \theta_i, \delta_i, \theta_k, \delta_k)$ .

Finally, the resulting NMSE values for each  $k = 1 \dots K$  virtual source were calculated as:

$$\epsilon_{\text{Enc}}(kr_e, \theta_k, \delta_k) = 10 \log_{10} \frac{\sum_{i=1}^I \alpha_i \left| p(kr_e, \theta_i, \delta_i, \theta_k, \delta_k) - \tilde{p}(kr_e, \theta_i, \delta_i, \theta_k, \delta_k) \right|^2}{\sum_{i=1}^I \alpha_i \left| p(kr_e, \theta_i, \delta_i, \theta_k, \delta_k) \right|^2}, \quad (2.36)$$

with  $\alpha_i$  being the quadrature weights associated with the chosen grid of  $I$  quasi-regular positioned observation points on the hard sphere. It is important to note that the reference pressure values  $p(kr_e, \theta_i, \delta_i, \theta_k, \delta_k)$ , used for the  $\epsilon_{\text{Enc}}$  calculation above, are derived from the *truncated series expansion* so that if the microphone array were to be ‘transparent’ signal-processing-wise, the NMSE would be zero ( $-\infty$  dB).

The median and interquartile range (IQR) of  $\epsilon_{\text{Enc}}(kr_e, \theta_k, \delta_k)$ , with  $k = 1 \dots K$ , are shown in Fig. 2.9 for both shape-matching and Ambisonics encoding strategies. The plot illustrates that above the spatial aliasing cut-off frequency [estimated as  $f_{\text{al}} = 7.4$  kHz using Moreau *et al.* 2006, Eq.(29)] the pressure errors are significantly reduced when using the shape-matching encoding. It should be noted that even small reductions of the *normalised* error  $\epsilon_{\text{Enc}}$  above 0 dB are translated into significantly lower pressure errors. Oreinos and Buchholz (2014) present a perceptually-driven comparison of the two encoding methods, by examining the pressure reconstruction errors at the in-ear microphones of an acoustic manikin (HATS) lying inside the HOA sound field, further supporting the above observations.

Alternatively, the two encoding strategies can be compared via the error that they incur on the ‘reconstructed directivities’. The reconstructed directivities can be understood as follows; given a plane-wave incident from direction  $(\theta_d, \delta_d)$  and impinging on a spherical microphone error, the encoded components  $\tilde{B}_{mn}^\sigma$  [due to source from:  $(\theta_d, \delta_d)$ ] are expected to approximate the SHFs sampled at the wave incidence position, i.e.,  $Y_{mn}^\sigma(\theta_d, \delta_d)$ , when inside the HOA system’s region of operation. Considering a dense grid of  $d = 1 \dots D$  such plane-wave sources (distributed on the sphere) encoded in succession by the microphone array, the set of components  $\tilde{B}_{mn}^\sigma$  [due to all sources from:  $(\theta_d, \delta_d)$ ] are termed ‘reconstructed directivities’, while the set of all SHFs  $Y_{mn}^\sigma(\theta_d, \delta_d)$  are termed ‘ideal directivities’. Employing compact matrix notation as in Eq. (2.32),

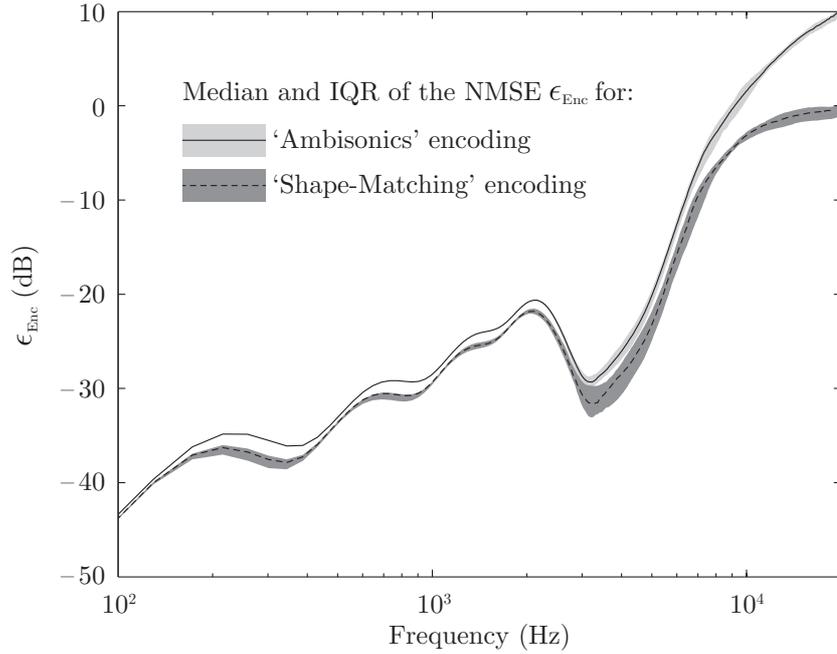


FIGURE 2.9: Normalised mean-square error incurred by the shape-matching and Ambisonics encoding, considering the NAL spherical microphone array ( $Q = 62$ ,  $R = 0.05$  m). The chosen regularisation values,  $\lambda_{\text{SM}}$  and  $\lambda_{\text{AMB}}$ , yield the same HOA system noise power for both encoding strategies. The solid and dashed lines represent the median value while the grey-shaded regions show the interquartile range, IQR, of the NMSE.

the shape of the reconstructed and ideal directivities can be respectively expressed as:

$$\begin{aligned} \mathcal{G}_{\text{rec}}(d, u, f) &= \text{Re} \left\{ \sum_{q=1}^Q [\tilde{\mathbf{E}}]_{u,q} [\mathbf{P}_D]_{q,d} \right\} \\ \mathcal{G}_{\text{idl}}(d, u, f) &= [\mathbf{Y}_D]_{u,d}, \end{aligned} \quad (2.37)$$

with the notation  $[\mathbf{A}]_{i,j}$  denoting the entry in the  $i$ -th row and  $j$ -th column of a matrix  $\mathbf{A}$ , and  $\tilde{\mathbf{E}}$  representing either the shape-matching encoding matrix  $\tilde{\mathbf{E}}_{\text{SM}}$  or the Ambisonics encoding matrix  $\tilde{\mathbf{E}}_{\text{AMB}}$ . The above directivities are functions of the  $d = 1 \dots D$  considered plane-wave sources, the  $u = 1 \dots U$   $[=(M+1)^2]$  SHF components and frequency  $f$ . Note that in the case of reconstructed directivities, the encoding operation gives rise to erroneous imaginary parts for the  $\tilde{B}_{mn}^\sigma$  components at frequencies outside the operating region of the microphone array. Only the real part  $\text{Re}\{\cdot\}$  is herein taken into account.

The RMS error between those directivities can thus be derived as:

$$\Delta \mathcal{G}_{\text{RMS}}(u, f) = \frac{1}{D} \sqrt{\sum_{d=1}^D [\mathcal{G}_{\text{rec}}(d, u, f) - \mathcal{G}_{\text{idl}}(d, u, f)]^2}. \quad (2.38)$$

The above RMS error only shows how far the reconstructed directivity shapes deviate

from the ideal ones but does not provide any information about the sign of the error. That is, no information is provided on whether  $\mathcal{G}_{\text{rec}}(d, u, f)$  is shrunk or enlarged compared to  $\mathcal{G}_{\text{idl}}(d, u, f)$ . For that reason, the difference of directivity total power (difference of squares of directivities) is calculated as:

$$\Delta\mathcal{P}(u, f) = \sum_{d=1}^D \left[ \mathcal{G}_{\text{rec}}(d, u, f) \right]^2 - \sum_{d=1}^D \left[ \mathcal{G}_{\text{idl}}(d, u, f) \right]^2. \quad (2.39)$$

Hence, signed RMS error values can be derived as:

$$\Delta\mathcal{G}_{\text{sRMS}}(u, f) = \text{sgn} \left( \Delta\mathcal{P}(u, f) \right) \Delta\mathcal{G}_{\text{RMS}}(u, f), \quad (2.40)$$

with  $\text{sgn}(\cdot)$  representing the sign function returning +1 if its argument is positive, -1 if negative and 0 if zero.

The above signed RMS error  $\Delta\mathcal{G}_{\text{sRMS}}(u, f)$  is derived by employing the considered microphone array configuration ( $Q = 62$ -mic array and  $\lambda_{\text{SM}} = 0.4$  or  $\lambda_{\text{AMB}} = 0.005$ ) and a dense grid of  $D = 2030$  points quasi-regularly placed on the sphere (Burkardt, 2010). Both encoding strategies are considered, with the shape-matching optimisation problem of Eq. (2.32) solved for  $G = 400$  quasi-regularly (Fliege, 2007) positioned sources, as reported above.

Figure 2.10 shows  $\Delta\mathcal{G}_{\text{sRMS}}(u, f)$ , for both encoding strategies, as surface plots, with the abscissa representing the SHF index  $u$  and the ordinate representing frequency  $f$ . The plot illustrates that the higher order SHF components can only be reconstructed with low errors (white areas) at limited frequency regions. At low frequencies, the regularisation turns off higher-order components in order to avoid excessive noise amplification, while at high frequencies, the microphone spatial aliasing causes the directivity components of all orders to break. Moreover, Fig. 2.10a illustrates that at high frequencies the SM encoding causes the reconstructed directivities to shrink (blue coloured errors) whereas Fig. 2.10b shows the high frequency AMB encoding directivities to be enlarged (red coloured errors). Shape-matching encoding, thus, controls the high frequency reconstructed directivities so that no excessive energy is fed to the reproduction loudspeaker array after the decoding. That results to lower pressure reconstruction errors, as can be seen in Fig. 2.9 and in Oreinos and Buchholz (2014, Fig. 6).

## 2.10 Mixed-Order Ambisonics

The traditional periphonic (i.e., 3D) coding of Higher-Order Ambisonics results in an area of accurate reconstruction (sweet spot) that does not depend on the position of

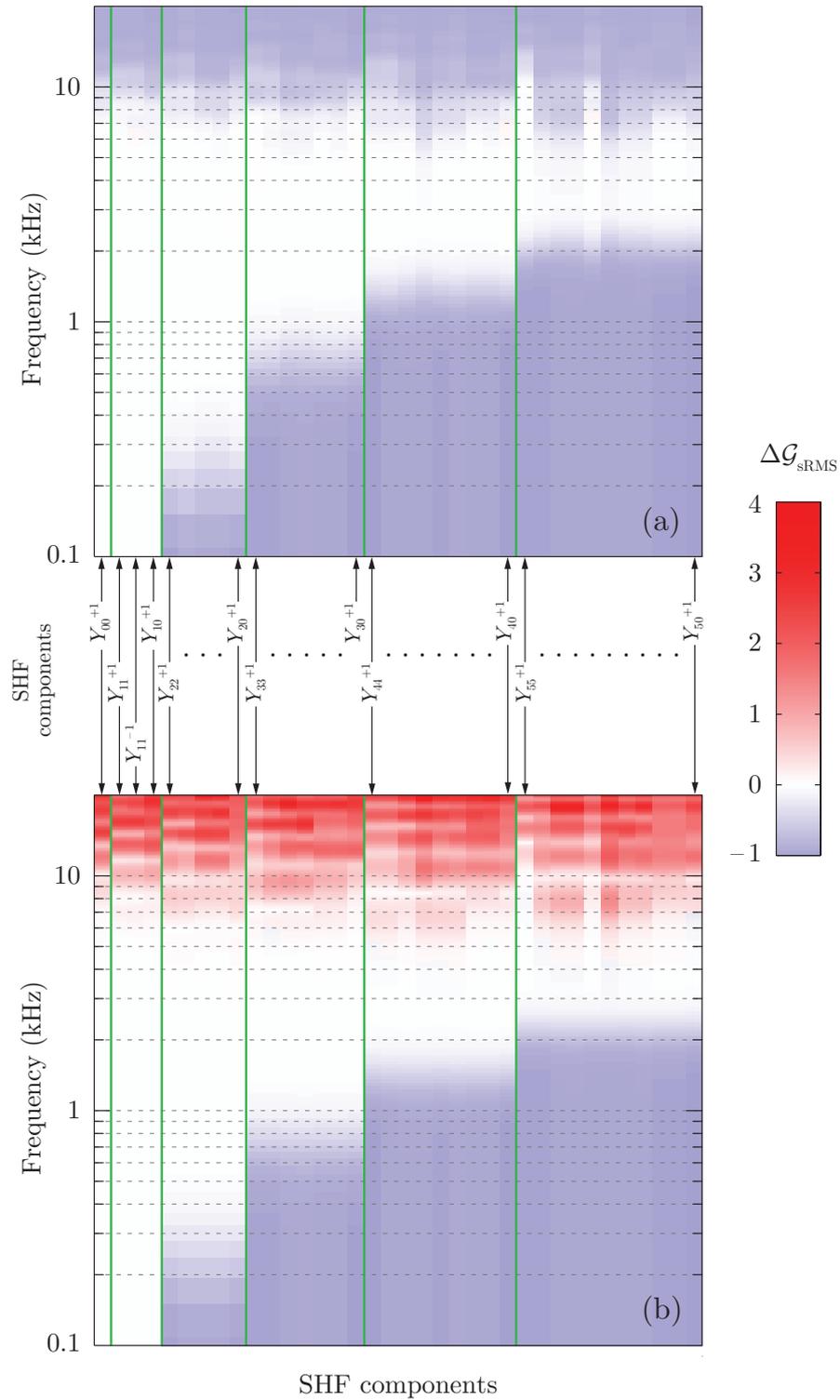


FIGURE 2.10: Signed RMS error of reconstructed directivities employing the  $Q = 62$ ,  $R = 0.05$  m NAL microphone array. All  $U = 36$  directivity components ( $M = 5$ ) are shown as columns of the surface plots, with the frequency dependence shown as rows. Plot (a) is relevant to shape-matching and plot (b) is relevant to Ambisonics encoding.

the coded virtual sources, provided that the employed microphone and loudspeaker arrays exhibit a quasi-regular geometry. Additionally, as shown in Chapter 3, the above property is subject to a free-field propagation (i.e., no listener inside the loudspeaker array).

Mixed-Order Ambisonics (MOA) is an alternative coding scheme that favours the reconstruction of horizontal-plane virtual sources (Favrot *et al.*, 2011; Käsbaach *et al.*, 2011; Favrot and Marschall, 2012; Marschall *et al.*, 2012). The applied MOA scheme employs two orders  $M_{3D}$ ,  $M_{2D}$ , relevant to periphonic (3D) and horizontal-plane-only (2D) reproduction respectively, according to the following pressure expansion:

$$\begin{aligned}
 p(kr, \theta, \delta) \simeq & \sum_{m=0}^{M_{3D}} i^m j_m(kr) \sum_{n=0}^m \sum_{\sigma=\pm 1} B_{mn}^\sigma Y_{mn}^\sigma(\theta, \delta) \\
 & + \sum_{m=M_{3D}+1}^{M_{2D}} i^m j_m(kr) \sum_{\sigma=\pm 1} B_{mm}^\sigma Y_{mm}^\sigma(\theta, \delta).
 \end{aligned} \tag{2.41}$$

The expansion of Eq. (2.41) uses the following  $U = (M_{3D}+1)^2 + 2(M_{2D}-M_{3D})$  SHF components:

$$\begin{aligned}
 & \left\{ Y_{00}^{+1}, Y_{10}^{+1}, Y_{11}^{\pm 1}, \dots, Y_{M_{3D}(M_{3D}-1)}^{\pm 1}, Y_{M_{3D}M_{3D}}^{\pm 1} \right\}, \\
 & \left\{ Y_{(M_{3D}+1)(M_{3D}+1)}^{\pm 1}, Y_{(M_{3D}+2)(M_{3D}+2)}^{\pm 1}, \dots, Y_{M_{2D}M_{2D}}^{\pm 1} \right\}.
 \end{aligned} \tag{2.42}$$

Although the MOA SHF selection scheme of Eq. 2.42 improves the horizontal-plane source reconstruction, it does not increase the accuracy of the system to that of an  $M = M_{2D}$  horizontal-plane-only (2D) HOA system. That would require the use of all SHFs that satisfy  $m+n = 2u$ ,  $u \in \mathbb{Z}^+$ , i.e., those components that do not exhibit a nodal line on the horizontal plane (see Figs. 2.1 and 2.2). In turn, estimating those components would necessitate the use of a microphone array with  $Q \geq (M_{3D}+1)^2$  sensors. Same argumentation would call for at least  $L = Q$  loudspeakers to ensure controlled playback. However, choosing only the sectoral SHFs  $Y_{mm}^\sigma(\theta, \delta)$  for degrees  $M_{3D} < m \leq M_{2D}$  necessitates only an increased density of microphones (and loudspeakers) on the equator, while at the same time providing a significant benefit over the baseline periphonic (3D) HOA system of  $M = M_{3D}$ . More elaborate MOA-selection schemes can be found in the work of Travis (2009).

## Chapter 3

# Measurement of a full 3D set of HRTFs for in-ear and hearing aid microphones on a Head and Torso Simulator (HATS)\*

### Abstract

The accurate reproduction of acoustic real-world environments is becoming of increasing importance in hearing device research and development. It is thereby often required to accurately predict the sound pressure at the microphones of a hearing device in a simulated or recorded acoustic environment. For that reason, an extensive set of head-related transfer functions (HRTFs) was measured in free-field with a pair of behind-the-ear (BTE) hearing aids placed on a Head and Torso Simulator (HATS). Transfer functions to the in-ear HATS microphones were also measured. A spherical head model was applied to extend the usable frequency range towards low frequencies. Special care was given to preserve the phase properties of the measurements so that the HRTFs could be widely used in phase-sensitive technical applications, including the evaluation of spatial signal processing algorithms (i.e., directional microphones, beamformers) in hearing devices and the evaluation of sound field resynthesis methods. The extended HRTF set can also be used for research in psychoacoustics. It is available for download at: [http://www.nal.gov.au/download/HATS\\_BTE\\_hrirDatabase.zip](http://www.nal.gov.au/download/HATS_BTE_hrirDatabase.zip).

---

\*Chapter published as:

Oreinos, C. and Buchholz, J. M. (2013), "Measurement of a full 3D set of HRTFs for in-ear and hearing aid microphones on a Head and Torso Simulator (HATS)", *Acta Acust. United Ac.* **99**, 836–844.

### 3.1 Introduction

Head-related transfer functions (HRTFs) contain all the spectral, temporal, and spatial information that is available in a given acoustic environment. The HRTFs in their free-field form are calculated as the ratio of the Fourier transform of the sound pressure at a point in the ear canal of a subject to the sound pressure that would have been measured at the centre of the head, with the subject not being present. They contain the diffraction effect of the head and torso as well as the resonances and scattering effects of the pinna. Given that HRTFs are determined by the detailed size and shape of the head, torso, and in particular ears, they are highly individual. This is demonstrated, for example, in localization experiments, where listeners' performance decreases significantly when non-individual HRTFs are used, in particular along the cones of confusion (Wenzel *et al.*, 1993). However, non-individual HRTFs, as for example measured on a standardised head and torso simulator (HATS), are often used in applications where it is impractical to measure individual HRTFs, such as virtual auditory spaces and computer games. Non-individual HRTFs can also be considered sufficient when analysing and verifying the general effect of multi-channel loudspeaker playback methods, where listener-specific effects are not of primary interest. In particular in the latter application, the measurement of a full 3D HRTF data set is required. By applying such a 3D HRTF data set, it is then possible to transform the loudspeaker output signals of basically any multi-channel playback system [e.g., using higher-order Ambisonics (Moreau *et al.*, 2006; Ward and Abhayapala, 2001; Epain *et al.*, 2010)] into ear signals. Although a large amount of 3D HRTF data sets are publicly available (e.g., Gardner and Martin, 1995; Bovbjerg *et al.*, 2000; Algazi *et al.*, 2001; Grassi *et al.*, 2003; Nakado *et al.*, 2008; Carpentier *et al.*, 2014), very limited HRTF data that additionally contain responses to the microphones of hearing aids (HA) fitted to a listener's (or HATS') ears are available (e.g., Kayser *et al.*, 2009; Acoustic Research Institute, n.d.). Such HTRF sets are essential for the research and development of hearing devices, the evaluation of hearing aid algorithms, and the verification of multi-channel loudspeaker systems for recreating realistic acoustic scenes (e.g., Minnaar *et al.*, 2010) aimed at hearing aid testing.

When multi-microphone signal processing (or enhancement) techniques are considered in hearing devices, the HRTFs need to provide accurate phase and amplitude information. Particularly when considering delay-and-subtract directional HA processing (Kates, 2008), even very small phase and amplitude errors between the two HA microphones are critical at low frequencies due to the differentiation and subsequent equalization of the high-pass 6dB/octave roll-off. Such accuracy was difficult to achieve here at low frequencies, by the measurements alone, for two reasons: (a) because a small measurement loudspeaker, with poor low-frequency response, had to be used due to weight

restrictions of moving equipment and (b) because the used anechoic chamber was not rated anechoic down to the required frequencies. To remedy that, the measured HRTFs were extended towards low frequencies by applying a spherical head model.

Although spherical head models have been previously used for this purpose ([Algazi \*et al.\*, 2002](#)), the applied methods for the combination of the model with the measurements have been evaluated mainly perceptually ([Kistler and Wightman, 1992](#)). The signal processing in hearing devices is very different from the auditory signal processing, and in particular for the case of multi-microphone signal processing techniques, it is much more sensitive to small amplitude and phase variations. Hence, different (or modified) techniques are required for combining measured and modelled HRTFs. The low-frequency extension methods developed and verified in this study are important for any HRTF measurement where the available anechoic chamber has a limited frequency bandwidth or when only semi-anechoic or even reverberant chambers are available and impulse response truncation methods need to be applied to remove the room artefacts from the measured HRTFs.

The present study comprises:

- The measurement of an extensive 6-channel HRTF data set for 1784 source locations covering the full 3D space using a Brüel & Kjær type 4128C HATS with two behind-the-ear (BTE) hearing aids, featuring two microphones each, fitted above the left and right pinna.
- The application of a spherical head model to extend the usable frequency range of the measured HRTFs towards low frequencies.
- A method to combine measured and modelled HRTFs that is adequate for multi-microphone hearing aid signal processing techniques. The applicability of this method is verified by considering directivity plots of an example hyper-cardioid microphone that is realised by combining the signals from the two microphones of either BTE hearing aid.

## 3.2 Methods

### 3.2.1 Measurement Setup

HRTFs were measured in the anechoic chamber of the Auditory Neuroscience Laboratory, Department of Physiology, University of Sydney. The chamber is 3.5 m long, 4.6 m wide and 2.4 m high and is rated as anechoic above 300 Hz. It contains a motor controlled

semicircular hoop with a radius of 1.2 m, which supports an Audience A3 loudspeaker. The hoop is fully controlled by computer software with an angular precision of less than  $0.1^\circ$  and is calibrated before the start of each set of measurements.

The measurements of the HRTFs were performed on the 4128C head and torso simulator (HATS) manufactured by Brüel & Kjær. The HATS complies with the ANSI standard for manikins for simulated in-situ airborne acoustic measurements (ANSI S3.36-1985 [R2006]) and the ITU-T P.58 recommendation for HATS telephonometry. It comes pre-fitted with 4158C/4159C type occluded ear simulators, which were connected to a pair of 2610 Brüel & Kjær measuring amplifiers. For these HRTFs measurements, a pair of –intentionally– slightly asymmetric pinnae (B&K models DZ-9763 & DZ-9764) were used. Two behind-the-ear (BTE) hearing aids were placed and secured, with sticky tape, above the HATS’ left and right pinnae [Fig. 3.1(a)]. The hearing aids were provided by Phonak and featured cables connected to their microphones and in-built receiver (not used for this set of measurements). The signals picked up by the microphones were sent via balanced cables to a purpose-built pre-amplifier designed by Phonak. It must be noted that these hearing aid “satellites” did not perform any signal processing. Finally, the six channels (two in-ear plus two BTE microphones for each side) were fed to an RME Fireface 400 sound-card connected to a PC running MATLAB. Transfer functions were measured using a logarithmic sweep of 1 s duration, sampled at 44.1 kHz with 24-bit resolution.

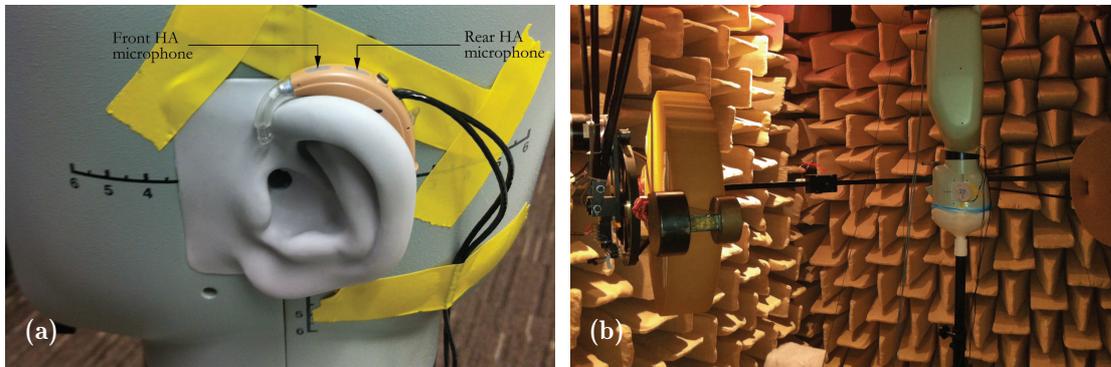


FIGURE 3.1: The acoustic manikin used for the measurements: (a) left ear of the HATS showing the hearing aid satellite with its two microphones; (b) the HATS standing upside down on a tripod and held in place by a purpose-built plastic fitting.

An important issue that had to be resolved was the fact that the lowest elevation angle where the hoop could be positioned at was  $\delta = -55^\circ$ . In order to work around this technical restriction the HATS had to be mounted upside down supported by its head, facing the front as in the upright placement, for some of the measurements. For that purpose, it had to be placed on a custom-built mount that was shaped around the top

TABLE 3.1: The coordinates of the 1784 points on the sphere where the HRTFs and BTE transfer functions were measured. Resolution of elevation range:  $5^\circ$ . Number of measurements:  $N$

Elevation range	Azimuth resolution	$N$
$[-40^\circ, 40^\circ]$	$5^\circ$	1224
$(-60^\circ, -40^\circ) \cup (40^\circ, 60^\circ)$	$8^\circ$	270
$(-70^\circ, -60^\circ) \cup [60^\circ, 70^\circ)$	$10^\circ$	144
$(-80^\circ, -70^\circ) \cup [70^\circ, 80^\circ)$	$15^\circ$	96
$(-90^\circ, -80^\circ) \cup [80^\circ, 90^\circ)$	$30^\circ$	48
$-90^\circ, 90^\circ$	$360^\circ$	2

of the head using a thermoplastic material (polymorph pellets). Fig. 3.1(b) shows the setup used for the upside down mounting of the HATS. For those measurements, the source (loudspeaker) coordinates (azimuth  $\theta$  and elevation  $\delta$ ) were referenced to the upside down HATS according to the transformations:  $\delta' = -\delta$  and  $\theta' = -\theta$ , so that the role of the ears remained consistent. The use of asymmetric pinnae for the left and right ears of the HATS dictated this choice. If both pinnae were exactly the same, another option would have been to transform the coordinates according to  $\delta' = -\delta$  and  $\theta' = \theta$ , while exchanging the roles of left and right side microphones (in-ear and BTE).

In both measurement conditions (standing upright and upside down) the accurate alignment of the HATS was facilitated by using two laser beam pointers mounted on the hoop at an arc distance of  $\pi/2$ . Correct alignment in all three axes could be attained by targeting the laser beams to the ear canals and the tip of the nose of the HATS. That resulted in effectively aligning the centre of the interaural axis with the centre of the hoop, which is the centre of the coordinate system against which the measurements were referenced. An observation that will prove useful in Section 3.2.2 is that the centre of the interaural axis does not coincide with the centre of the HATS' head.

In order to further verify the correct alignment of the setup, all loudspeaker locations with elevation  $-55^\circ \leq \delta \leq -40^\circ$  were measured with the HATS standing both upright and upside down. The corresponding transfer functions were compared during the measurement process and were found to match very well (absolute magnitude error between the upright and upside down HRTF measurements, averaged across all available overlapping locations,  $-55^\circ \leq \delta \leq -40^\circ$  and  $0^\circ \leq \theta < 360^\circ$ ,  $< 1$  dB up to at least 4 – 5 kHz and  $< 2$  dB up to at least 10 kHz), confirming the centred placement of the HATS in both conditions. Table 3.1 summarises all 1784 measurement positions.

HRTFs are typically normalised by the free-field response measured with the same loudspeaker and microphones as used in the HRTF measurements, but with the microphones placed in the origin of the applied coordinate system (here the centre of the hoop) and

the head (HATS) removed (Blauert, 1997). Unfortunately, due to time restrictions on the use of the facility, the loudspeaker response could not be measured with the in-ear and the BTE microphones in a free-space configuration. Instead, the loudspeaker magnitude response was equalised, with a 512 tap FIR filter, using a measurement provided by our collaborators, and the magnitude responses of the BTE microphones were equalised, with a 64 tap FIR filter, after being measured inside a hearing aid test box. The frequency response of the HATS ear simulators (4158C & 4159C) were already calibrated up to 20 kHz, on purchase of the product, so no further correction was necessary. The gain necessary to compensate for the in-ear microphones/amplifiers sensitivity was estimated by comparing the measured  $(0^\circ, 0^\circ)$  HRTF to the calibration chart provided by B&K, for the specific HATS that was used, as well as by comparing the measured HRTFs to the ITU-T P.58 recommendation (“Head and torso simulator for telephony”) values for HRTFs at  $0^\circ, 90^\circ, 180^\circ, 270^\circ$  on the horizontal plane. The sensitivity of the four BTE microphones/amplifiers was compensated for by minimizing the RMS error between their individual magnitude responses, averaged over all source directions, and the corresponding averaged in-ear microphone responses in the frequency range of 300 – 600 Hz.

### 3.2.2 Spherical head model realization

In this work, a spherical head model was applied to extend the usable frequency range of the HRTFs, i.e., to extrapolate information to lower frequencies, while preserving the sensitive phase information of the HA microphones. This model approximates the human head by a rigid sphere and as such, it does not take into account the acoustic resonances and scattering by the pinnae, at high frequencies, as well as the shadowing effects of the torso. Although it is an idealised model, it has been suggested that it can be used to significantly improve the low frequency accuracy of measured HRTFs (Kistler and Wightman, 1992; Duda and Martens, 1998; Brungart and Rabinowitz, 1999). Using the Fourier-Bessel series expansion of the wave equation solution for the incident (interior problem) and scattered (exterior problem) sound field created by a single point source positioned at  $(r_s, \theta_s, \delta_s)$  and imposing the boundary condition of zero total radial velocity on the rigid sphere of radius  $r = a$ , the pressure on its surface can be computed as:

$$p(r = a, \theta, \delta) = \sum_{m=0}^{\infty} \frac{i^{-1}}{(ka)^2 h_m^{(2)'}(ka)} \frac{h_m^{(2)}(kr_s)}{h_0^{(2)}(kr_s)} (2m + 1) P_m(\cos \gamma), \quad (3.1)$$

where  $h_m^{(2)}$  is the  $m$ -th order spherical Hankel function of the second kind,  $h_m^{(2)'}$  its derivative and  $\gamma$  is the angle between the vector of the point source and that of the

TABLE 3.2: Spherical head model parameters.

Parameter	Value
Head radius	0.1 m
Ear canal entrance locations	L: $(100^\circ, -11^\circ)$ R: $(-100^\circ, -11^\circ)$
Front BTE microphone locations	L: $(104^\circ, 10^\circ)$ R: $(-104^\circ, 10^\circ)$
Rear BTE microphone locations	L: $(109^\circ, 10^\circ)$ R: $(-109^\circ, 10^\circ)$
“Point” source distance	1.2 m

observation point.  $P_m$  is the  $m$ -th degree Legendre polynomial defined as:

$$P_m(x) = \frac{1}{2^m m!} \frac{d^m}{dx^m} (x^2 - 1)^m. \quad (3.2)$$

Eq. 3.1 assumes waves normalised so that they have unit amplitude and zero phase at the origin, having an implied  $e^{+i\omega t}$  time dependence.

The spherical head model response was computed using the parameters shown in Table 3.2. Given the non-spherical nature of the HATS’ head as well as the existence of the torso, which was disregarded by the applied spherical head model, the parameters of Table 3.2 were adjusted in a more heuristic, rather than rigorous, way. The head-related values were chosen by comparison to their locations on the HATS while iteratively minimising the RMS error of the modelled and measured ILDs and ITDs. The sound source distance was taken to be exactly the radius of the hoop where the speaker was mounted. The resulting mean absolute errors, averaged across all 1784 measurement points, were  $|\overline{\text{ILD}_{\text{meas}}^{\text{dB}} - \text{ILD}_{\text{mod}}^{\text{dB}}}| < 0.9 \text{ dB}$  and  $|\overline{\text{ITD}_{\text{meas}} - \text{ITD}_{\text{mod}}}| < 46 \mu\text{s}$ , where the ILDs were estimated at the region 300 – 600 Hz. A formal optimization was impractical due to the nature of the problem, wherein a spherical model was employed to describe a flattened (at the ear sides) ellipsoidal structure. However, the applied methodology reinserts, at a later stage, the correct timing to the model (Section 3.2.3 and Fig. 3.6), circumventing in this manner the need to apply much more complicated prolate spheroid models (Adelman *et al.*, 2014).

It should be mentioned that alternative, more complex approaches, such as the snowman model described by Algazi *et al.* (2002), exist and are expected to provide a more accurate representation of the low frequency response of the HRTFs. However, the spherical head model already resulted in small mean absolute ITD and ILD errors and thus, such elaborate models were not further considered here.

In order to maximise the computational efficiency, Eq. (3.1) was computed using a set of recursive formulas as described by Duda and Martens (1998) instead of being directly

computed via the analytic expression. Given that the origin of the coordinate system used in the above spherical head model is defined in the centre of the spherical head and thus is different from the one used in the HRTF measurements (see Section 3.2.1), the source positions (radius  $r'_s$ , azimuth  $\theta'_s$ , elevation  $\delta'_s$ ) to be used for the spherical head model computation had to be derived from the source positions of the HRTF data set according to:

$$\begin{aligned}\theta'_s &= \arctan\left(\frac{y}{x + x_{ic}}\right), \\ \delta'_s &= \arctan\left(\frac{z + z_{ic}}{(x + x_{ic})^2 + y^2}\right), \\ r'_s &= \sqrt{(x + x_{ic})^2 + y^2 + (z + z_{ic})^2},\end{aligned}\tag{3.3}$$

where  $[x, y, z] = [r_s \cos \theta_s \cos \delta_s, r_s \sin \theta_s \cos \delta_s, r_s \sin \delta_s]$  is the position vector of a given source, referenced to the centre of the hoop and  $[x_{ic}, 0, z_{ic}]$  is the position vector of the centre of the hoop (centre of the interaural axis) referenced to the centre of the HATS head. This change of coordinate systems is necessary so that, for example, the angle  $\gamma$  between the vector of the left ear ( $100^\circ, -11^\circ$ ), as referenced to the HATS' centre system, and the leftmost source point, whose coordinates expressed in the system centred at the midpoint of the interaural axis are ( $90^\circ, 0^\circ$ ), is correctly calculated to be zero.

One-sided pressure spectra were calculated with the spherical head model for each of the 6 microphone locations (see Table 3.2) at 4097 ( $N_{\text{FFT}} = 8192$ ) equidistant frequency points between 0 Hz and  $f_s/2$ . The corresponding real-valued head-related impulse responses (HRIRs) were then derived after first ‘‘mirroring’’ the resulting spectra to negative frequencies by applying the complex conjugate transformation:  $H(-i\omega) = H^*(i\omega)$ , and then applying an inverse (discrete) Fourier transform ( $N_{\text{FFT}} = 8192$ ). Note that the  $N_{\text{FFT}}$  was chosen long enough to avoid time-aliasing after the successive time shifting and filtering of the impulses.

### 3.2.3 Combination of the head model with the measured data

A primary question that needed to be addressed was how to combine the measured HRTFs with the transfer functions derived from the spherical head model at low frequencies. [Algazi \*et al.\* \(2002\)](#) have proposed cross-fading the magnitudes of the model and the data (linearly in dB) while either keeping the entire phase response of the model, or re-inserting the estimated time of arrival of the data impulses to the minimum-phase inverse-DFT of the combined (cross-faded) magnitude responses. The latter method, called minimum-phase reconstruction ([Kistler and Wightman, 1992](#)), has been perceptually verified for sound localization. However, the measured transfer functions of this

HRTF set will be used as inputs to hearing aids whose processing is very different from human auditory processing. In particular, multi-microphone signal enhancement techniques (e.g., directional microphones and adaptive beamformers) are highly sensitive to very small phase and amplitude variations. Hence, the HRTFs must preserve as accurately as possible phase and amplitude relationships, especially between the individual BTE hearing aid microphone pairs. In this regard the method proposed by [Algazi \*et al.\* \(2002\)](#) did not provide satisfactory results and needed to be modified.

The combination method pursued in this study consists of time aligning the model's impulse responses to the corresponding measured impulse responses and then combining them using "crossover" type low-pass and high-pass filters with cut-off frequencies in the regions of 400 – 500 Hz. First, the time of arrival (TOA) for all 1784 measured and modelled HRIRs was estimated separately for the in-ear and BTE microphones according to the method proposed by [Nam \*et al.\* \(2008\)](#) as the time instant of the maximum peak of the absolute of the cross-correlation function between a given HRIR and its minimum-phase version:

$$\hat{\tau} = \arg \max_{\tau} \left\{ \left| \sum_n h[n - \tau] h_{\text{mp}}[n] \right| \right\}. \quad (3.4)$$

Since the BTE hearing aid microphones were only spaced apart by  $d_{\text{mic}} = 0.009$  m, their maximum TOA difference across all source locations was about 26  $\mu\text{s}$ . Considering that the applied sampling frequency was  $f_s = 44.1$  kHz and, thus, the resolution of the applied TOA estimation method was only  $T = 1/f_s \approx 22.7$   $\mu\text{s}$ , the HRIRs had to be up-sampled (a factor of 100 was chosen) before the cross-correlation operation of Eq. 3.4 was performed.

An alignment delay was then calculated from the estimated TOA of the modelled HRIRs,  $\hat{\tau}_{\text{mod}}$ , and the corresponding measured HRIRs,  $\hat{\tau}_{\text{meas}}$ , given by:

$$\Delta = \frac{\hat{\tau}_{\text{mod}} - \hat{\tau}_{\text{meas}}}{100}. \quad (3.5)$$

To preserve even the smallest phase differences between each BTE microphone pair, the average alignment delay for each microphone pair was always applied to both microphones. In this way any (small) TOA estimation error only affected the absolute timing of the BTE microphone pair signals, but the inter-microphone phase differences provided by the spherical head model were preserved. Since the combination of the measured and modelled HRTFs was performed in the frequency domain, the derived alignment delays

$\Delta$  were transformed into complex gain factors:

$$g\Delta(\mathrm{i}f) = e^{-\mathrm{i}2\pi f\Delta/f_s}, \quad (3.6)$$

which were then multiplied with the corresponding modelled HRTFs. No further amplitude adjustments were required than those already described in Section 3.2.1.

Inspired by the design of loudspeaker crossovers, 256-taps long linear-phase FIR filters were used to realise the crossover between measured and time-aligned modelled HRTFs that resembled the magnitude spectrum of a complementary pair of asymmetric Linkwitz-Riley filters. An 8<sup>th</sup> order low-pass filter with a cut-off frequency of 500 Hz and a 4<sup>th</sup> order high-pass filter with a cut-off frequency of 400 Hz were chosen. The 4<sup>th</sup> order HP filter slope combined with the roll-off of the measurements effectively resulting in a higher order HP response, almost complimentary to the 8<sup>th</sup> order LP filter. The different slopes and cut-off frequencies of the filters was chosen to: (1) apply a small gain-peak of 1 – 2 dB around 300 – 500 Hz and in turn to provide a smoother transition between measured and modelled HRTFs and (2) ensure good agreement of the combined HRTF responses with data from the ITU-T P.58 recommendation and the HATS' own calibration chart. This filter design approach in the frequency domain controlled the steepness of the slopes of the cross-over filters and thereby limited the temporal ringing of the filters in the time domain. Choosing a linear-phase filter design minimised any phase (or comb-filter) effects that could otherwise be introduced by adding the low-pass and high-pass HRTF components. The filtering was done in the frequency domain, as mentioned above, and an impulse response was computed taking the inverse DFT (as in Section 3.2.2) of the combined modelled/measured conjugate-mirrored transfer function.

The final impulse responses of the combined HRTFs were then truncated to 256 samples using a one-sided Tukey window with a tapered-to-constant-section-ratio of 1/5. Special care was taken in the truncation process to maintain the very short pre-ringing that was introduced by the linear phase crossover filtering process. Finally, the impulse responses were saved separately to MATLAB files (.mat extension), with every file having the impulse responses of all six microphones relevant to a given source position. The final HRTF data set is publicly available at: [http://www.nal.gov.au/download/HATS\\_BTE\\_hrirDatabase.zip](http://www.nal.gov.au/download/HATS_BTE_hrirDatabase.zip).

### 3.3 Results

Fig. 3.2 shows the derived (0°, 0°) HRTF response (measurement combined with the spherical head model), along with the corresponding ITU-T P.58 recommendation's

range of values, highlighting the compliance with this standard. The HRTFs at  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$  on the horizontal plane were also compared to the relevant ITU-T P.58 values (not shown here) and were similarly found to comply.

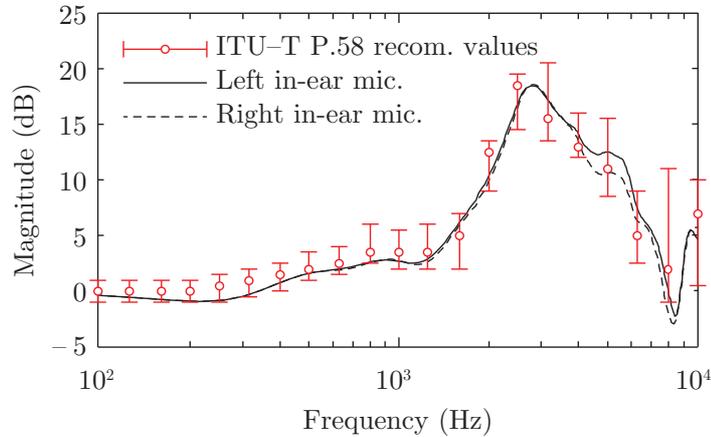


FIGURE 3.2: Processed  $(0^\circ, 0^\circ)$  HRTF magnitude response, extended to low frequencies using the spherical head model, compared against the values of the ITU-T P.58 recommendation.

The magnitude spectra of the measured, modelled, and combined HRTFs are plotted in Fig. 3.3 for an example source direction of  $(60^\circ, -30^\circ)$ . Responses for the left ear are shown in the left figures and for the right ear in the right figures. The top figures refer to the front microphone of the BTE hearing aids and the bottom figures refer to the in-ear microphones. In the top panels of Fig. 3.3 it can be seen that the overall trend of the modelled HRTFs matches with the measured HRTF of the BTE microphone. In contrast, the measured in-ear responses deviate significantly from the model at higher frequencies. This deviation is mainly due to the ear canal resonances, the most prominent of which is at around 2.5 kHz, as well as the effect of the pinnae. At low frequencies the application of the spherical head model increases the usable bandwidth and removes the low frequency dips and peaks, which are due to the measurement chamber (which is not anechoic below 300 Hz) room modes.

When comparing the details contained in the modelled and measured spectra, in particular at frequencies above about 1 kHz, clear differences can be observed. These differences can be even more pronounced at other source positions, especially at the contralateral side where more extensive rippling occurs. However, these discrepancies are expected because the shape of the head of the HATS is not strictly spherical and the presence of the torso and pinnae has not been considered in the spherical head model. Since the model is only applied to extend the usable frequency range below about 300 Hz, the mentioned discrepancies are not significant and can be neglected.

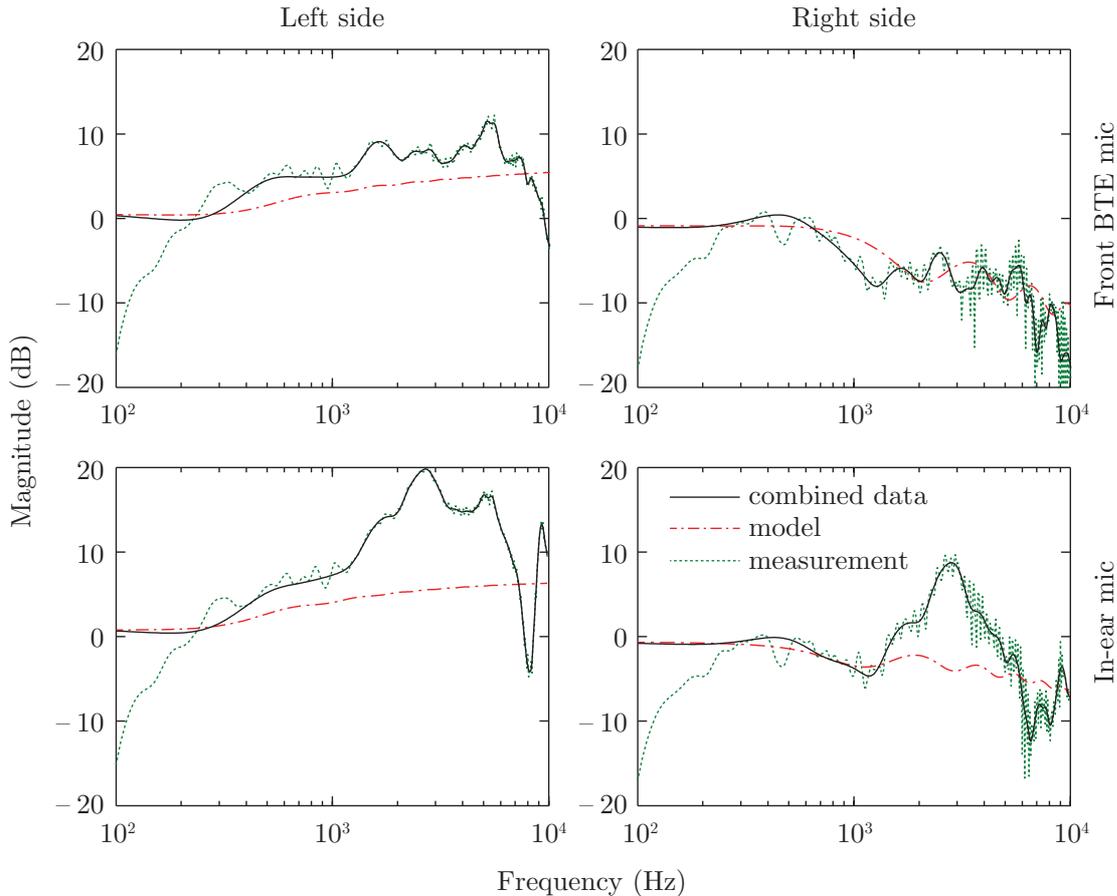


FIGURE 3.3: Spectrum magnitude of HRTF measurements (dashed curve), spherical head model (dash-dotted curve) and final combined and windowed responses (solid curve) relevant to the source direction ( $60^\circ, -30^\circ$ ). The left and right panels show left and right ear microphone responses, respectively. The top curves refer to the front BTE microphones and the bottom curves to the in-ear microphones.

Fig. 3.4 shows the HRTF magnitudes of the left in-ear and front HA microphones for all available source positions on the horizontal plane. The contour plots of the in-ear responses can be qualitatively compared to similar plots found in literature (Bovbjerg *et al.*, 2000; Algazi *et al.*, 2001; Brungart and Rabinowitz, 1999; Duraiswami *et al.*, 2004). However, a quantitative one-to-one comparison is not meaningful unless one is comparing measurements from the same manikin, fitted with the same pinnae and ear canal and simulator and driven by a similar measuring loudspeaker at the same distance. Comparing Fig. 3.4(a) and (b), it is evident that the ear-canal resonances at  $\sim 2.5$  kHz and  $\sim 5$  kHz are eliminated in the BTE responses. Additionally, some of the high frequency fine structure appears smoothed and attenuated due to the absence of pinna reflections.

As a next step to validate the combined HRTFs, the Interaural Level Differences (ILDs) versus frequency are plotted in Fig. 3.5 for sources located on the horizontal plane. ILDs were calculated as the magnitude-difference between the left and right ear's HRTFs. A

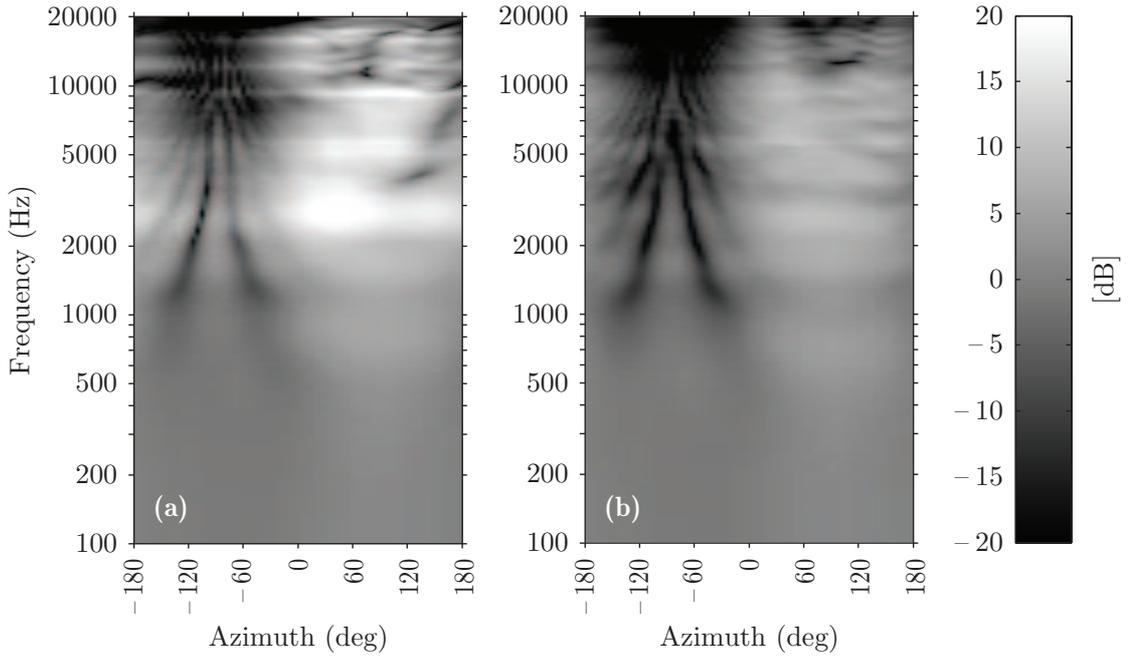


FIGURE 3.4: Contour plots of the magnitude of HRTFs to the: (a) in-ear and (b) hearing aid (front) microphone of HATS’ left side. The abscissa of the plots depicts the source direction on the horizontal plane, with azimuth  $\theta \in [-180^\circ, 180^\circ]$ .

number of observations can be made from these plots. First, due to the different pinnae used on the measured HATS the ILD for  $\theta = 0^\circ$  is slightly different from zero at high frequencies. A second observation is that for positive azimuths close to  $90^\circ$  (source on-axis with left ear) a positive ILD, which takes substantial values at high frequencies, is observed, while an almost symmetrical picture up to about 5 kHz is seen for negative azimuths (above  $\sim 5$  kHz the used pinnae cause deviations to the symmetric nature of the ILDs). The theoretical bright spots at  $\pm 90^\circ$  are also confirmed on that plot (i.e., a source at  $90^\circ$  creates a bright spot at  $180^\circ$  around the perimeter of the sphere, that is at  $-90^\circ$ , where the contralateral ear lies). As mentioned above for the HRTF magnitude case, it is not straightforward to compare the ILD contours from another data set, unless it is ensured that all geometrical, structural and acoustical properties of the setup are the same. The ILDs measured at the BTE-microphones (not shown here) look very similar to the in-ear ILDs of Fig. 3.5 at low frequencies, but above about 5 kHz are significantly smaller due to absence of the pinna and ear canal effects.

The broadband ITD of the measured HRTFs as well as the modelled HRTFs before and after TOA alignment (see Section 3.2.3) are shown in Fig. 3.6 for the in-ear microphones as a function of source azimuth. The ITDs were calculated as the TOA difference between the left and right ear microphones using Eq. 3.4. The ITDs for the measurement (solid line) as well as the TOA-aligned model (dashed line) match very well, confirming the accuracy of the TOA-alignment procedure described in Section 3.2.3. Moreover, these

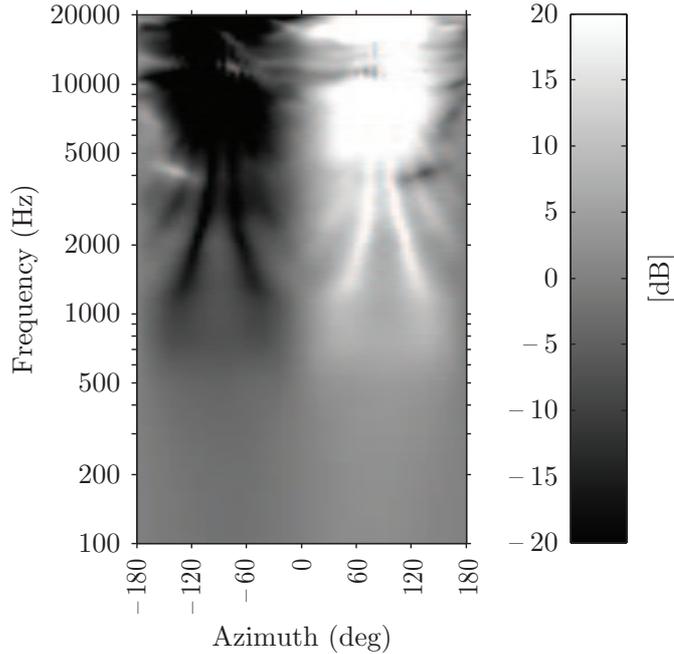


FIGURE 3.5: Interaural level differences between left and right in-ear microphones ( $H_L - H_R$ ), versus frequency, relevant to positions lying on the horizontal plane with azimuth  $\theta \in [-180^\circ, 180^\circ]$ .

ITDs exhibit the typical ITD shape measured in humans as seen for example in [Nam \*et al.\* \(2008\)](#) or [Busson \*et al.\* \(2005\)](#). The ITDs for the modelled HRTFs before TOA-alignment (dashed-dotted line) show a slightly less “peaky” behaviour than the ITDs for the measured HRTFs, but otherwise provide a very good fit and thus confirm the applicability of the parameters given in [Table 3.2](#) for the calculation of the spherical head model responses. As expected, the ITDs estimated from the HA microphones (not shown here) look almost the same as those of [Fig. 3.6](#).

In order to verify the accurate representation of phase (or timing) of the combined HRTFs, in particular between the front and back microphones of the BTE hearing aids, 3D directivity plots of a hyper-cardioid and an omnidirectional microphone were realised and shown in [Fig. 3.7](#) for three example frequencies, 300 Hz, 450 Hz and 1600 Hz. The radial distance of each directivity plot surface node represents the magnitude of the directional microphone output response (in dB) due to a point-source located at that node’s azimuth and elevation at a distance of  $r = 1.2\text{m}$  (distance between loudspeaker and HATS’ head centre). The hyper-cardioid output was generated by applying delay and subtraction beamforming ([Kates, 2008](#)) to the HRTFs of the front and back microphones of the left BTE hearing aid. Such an approach for creating directional microphones is highly sensitive to phase and amplitude errors, in particular at low frequencies. If the phase difference between the microphones is slightly disrupted (as was the case in the measured HRTFs at low frequencies), the directivity plots get severely distorted. The

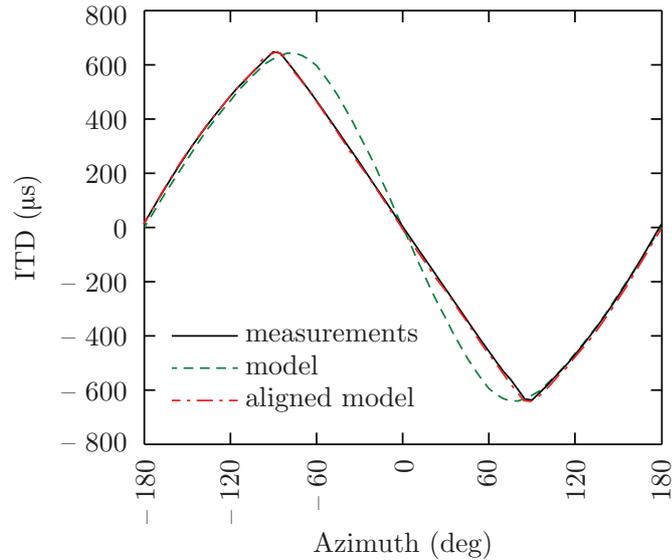


FIGURE 3.6: Interaural time differences between left and right in-ear microphones relevant to positions lying on the horizontal plane with azimuth  $\theta \in [-180^\circ, 180^\circ]$ . The ITD estimates of the measurements and the model, before and after time alignment, are plotted.

bottom polar plots of Fig. 3.7 show the hyper-cardioid directivity attained at low frequencies (300 Hz) to be smooth and resembling that at higher frequencies (1600 Hz). Thus the low frequency region where the model dominates the combined response produces a directional result that resembles that at higher frequencies where the measured data dominate. Moreover, the directivity plots at 450 Hz demonstrate that the correct magnitude and phase information of the HRTFs is also retained in the middle of the crossover region (Section 3.2.3).

Note that the presented directivity patterns differ from the well-known ideal hyper-cardioid polar plot (Kates, 2008) due to the presence of the head. The head shadowing effectively creates a higher-order directional behaviour, as compared to the first order ideal free-field patterns. In order to illustrate the head shadow effect on the microphone directionality, the directivity patterns of an omnidirectional BTE microphone on the head (i.e., only considering the front microphone) are shown in the top figures of Fig. 3.7 at 300 Hz, 450 Hz and 1600 Hz. They are very similar to the directivity patterns shown by Kates (2008), further demonstrating the validity of the measured HRTFs.

Lastly, Fig. 3.8 shows the SNR, averaged across all six microphone channels (which had very similar values) and all measurement positions (solid line). A noise estimate was formed using the last 256 samples of the 2048-sample long unprocessed measurement data. A signal estimate (in fact signal + noise) was formed using the preceding 1792 samples. The estimated SNR does not strictly depict the true microphone – amplifier – A/D SNR but describes the effective “SNR” after averaging and de-convolving the

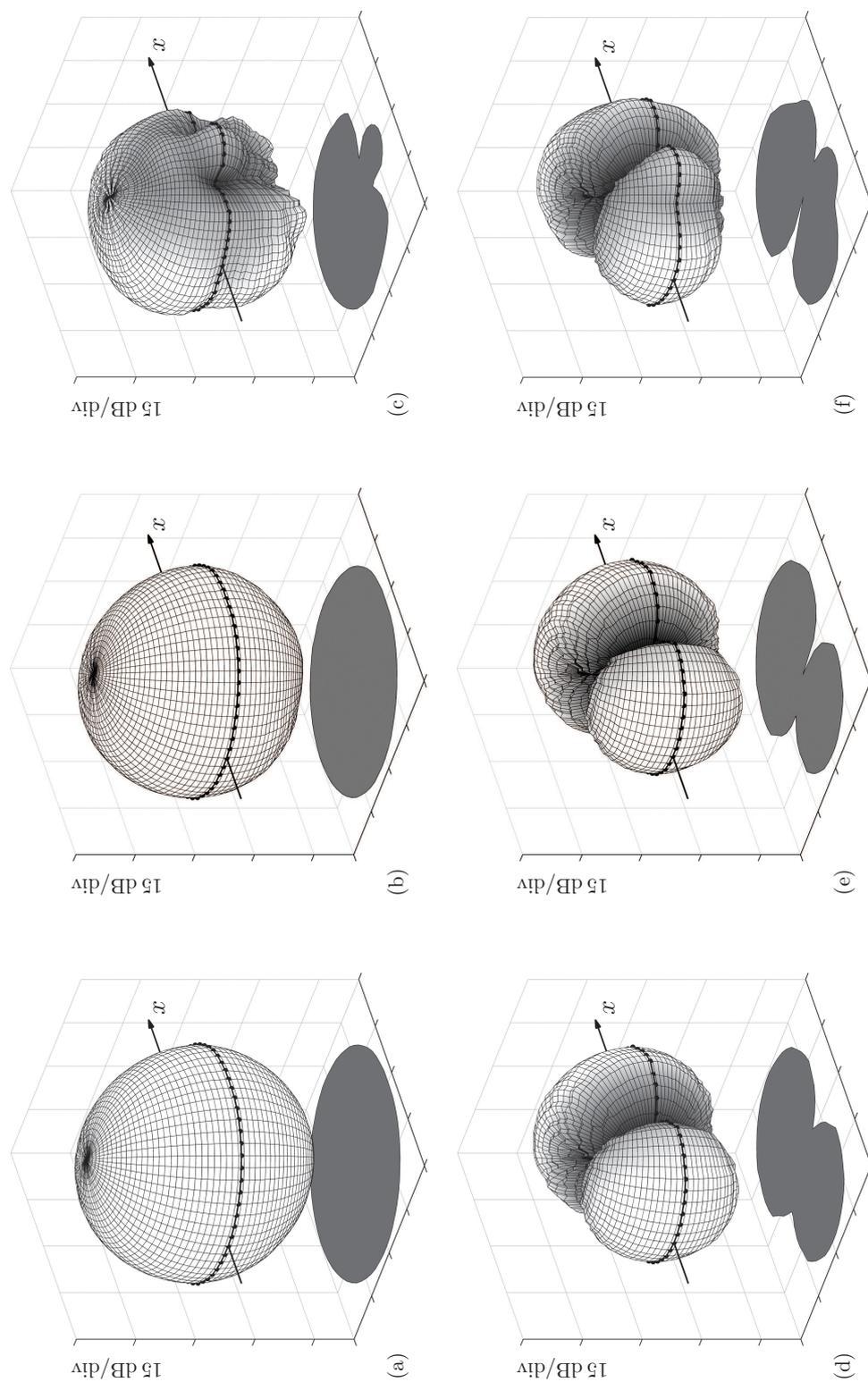


FIGURE 3.7: Directivity patterns of the left BTE microphone pair (worn on the HATS) for an omnidirectional configuration (i.e., using only the signals of the front BTE microphone) at (a) 300 Hz, (b) 450 Hz, (c) 1600 Hz and a hyper-cardioid configuration at (d) 300 Hz, (e) 450 Hz, (f) 1600 Hz. The 2D shadow plots illustrate the directivity plots of the 3D patterns on the horizontal plane.

recorded sine sweeps. The frequency-dependent SNR of the measurement with the worst broadband SNR, occurring at the position  $(30^\circ, -85^\circ)$ , is also plotted as a worst-case scenario. It can be seen that the SNR at low frequencies (300 – 400 Hz) is significantly lower than the values at mid and high frequencies. Especially if the 6dB/octave roll-off of a delay-and-subtract directional HA microphone is considered, the output SNR becomes negative at low frequencies. This confirms again the necessity of applying the spherical head model for a “phase-preserving” extrapolation of the magnitude to low frequencies.

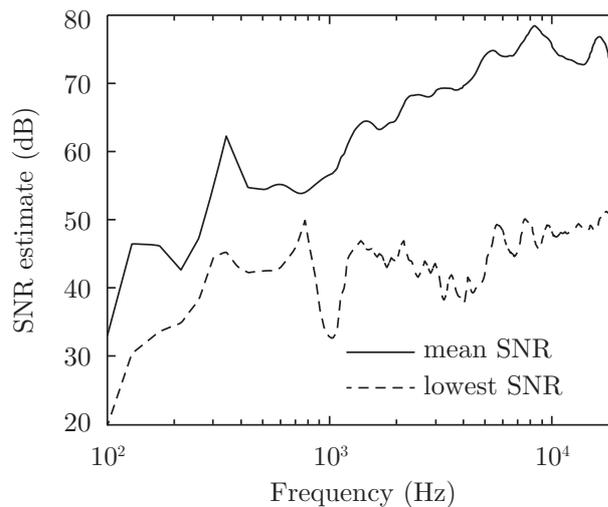


FIGURE 3.8: The frequency-dependent mean SNR of all microphone channels averaged across all positions (solid curve) along with the SNR of the measurement with the lowest broadband SNR (dotted curve).

### 3.4 Summary and conclusions

The procedure for the measurement of an extensive set of HRTFs from 1748 directions covering the whole sphere to the 2 in-ear and 4 BTE hearing aid microphones on a HATS was presented. A spherical head model was applied to extend the usable frequency range of the measured HRTFs towards low frequencies. A method was also proposed to combine the measurements with the model responses while maintaining the very sensitive amplitude and phase information carried by the HRTFs, which is particularly critical between the front and back hearing aid microphones. The applicability of this approach as well as the validity of the entire HRTF data set, especially when used for hearing aid applications, was confirmed by analysing magnitude spectra of the combined HRTFs, ITDs, ILDs and directivity plots of hyper-cardioid and omni-directional microphones.

The developed HRTF data set covering the full sphere at a high resolution can be used in a wide range of applications ranging from recreating virtual auditory spaces to

research and development with hearing devices, evaluation of hearing aid algorithms, including psychoacoustic aspects, and verification of multi-channel loudspeaker systems for recreating realistic acoustic scenes. The existence of measurements from both in-ear and BTE microphones permits realistic simulations of aided scenarios where sound reaches the hearing aids of the wearer but at the same time leaks through to the eardrum (this “acoustic path” will be affected by the fitting of the hearing aid, which may include ear-moulds with different vent sizes).

## **Acknowledgements**

The authors acknowledge the financial support of the HEARing CRC, established and supported under the Cooperative Research Centres Program – an initiative of the Australian Government. In addition, we would like to thank Johahn Leung, Heather C. Kelly, and Prof. Simon Carlile for letting us use the Auditory Neuroscience Laboratory, Department of Physiology, University of Sydney, as well as two anonymous reviewers for their helpful suggestions.

## Chapter 4

# Objective analysis of Ambisonics for hearing aid applications: Effect of listener's head, room reverberation, and directional microphones\*

### Abstract

Recently, an increased interest has been demonstrated in evaluating hearing aids (HAs) inside controlled, but at the same time realistic sound environments. A promising candidate that employs loudspeakers for realizing such sound environments is the listener-centred method of higher-order Ambisonics (HOA). Although the accuracy of HOA sound field reconstruction has been widely studied, it remains unclear to what extent the results can be generalised when: (1) a listener wearing HAs that may feature multi-microphone enhancement algorithms is considered inside the reconstructed sound field and (2) reverberant sound fields are recorded and reconstructed. For the purpose of objectively validating HOA for listening tests involving HAs, a framework was developed that simulates the entire path of sounds presented in a modelled room, picked up by a HOA microphone array, decoded to a loudspeaker array, and finally received at the ears and HA microphones of a dummy listener fitted with HAs. Reproduction errors at the ear signals and at the output of a cardioid HA microphone were analysed for different anechoic and reverberant scenes. It was found that the diffuse reverberation reduces the considered time-averaged HOA reconstruction errors which, depending on the considered application, suggests that reverberation can increase the usable frequency range of a HOA system.

---

\*An edited version of this Chapter has been accepted for publication at the J. Acoust. Soc. Am.

## 4.1 Introduction<sup>†</sup>

A large number of hearing research projects in psychoacoustics, audiology, and hearing devices technology are currently undertaken in laboratories, aiming at providing new insights on hearing loss, its prevention, and the required intervention. The outcomes of this research and development need to be thoroughly evaluated in realistic conditions to ensure that the theories that are formulated and tested in the laboratory also apply in the complex acoustical scenarios encountered in everyday life.

Currently, traditional listening tests involving hearing aids (HAs) are typically performed either in a controlled manner, but under over-simplifying assumptions (e.g., laboratory setups typically using 2-8 horizontal-plane loudspeakers), or in realistic environments, but with limited control or repeatability (e.g., field studies). [Cord \*et al.\* \(2004\)](#) examined the advantage of directional hearing aid microphones as measured in the laboratory (4 loudspeakers at  $0^\circ, 90^\circ, 180^\circ, 270^\circ$  inside a treated booth with floor size:  $3.0 \times 2.5$  m) and found it to be a poor predictor of the user benefit in everyday life. [Ricketts \(2000\)](#); [Revit \*et al.\* \(2007\)](#) and [Compton-Conley \*et al.\* \(2004\)](#) also demonstrated that the testing environment significantly affects the estimated HA benefit. A plethora of anecdotal observations also supports this discrepancy between laboratory test results and real-life benefit, highlighting the need to develop and establish controlled realistic environments for more ecologically-valid listening experiments.

In order to bridge the gap between artificial laboratory scenes and real-world environments, researchers ([Rychtáriková \*et al.\*, 2009](#); [Völk and Fastl, 2010](#); [Mueller \*et al.\*, 2012](#)) have employed binaural technology to present sounds via headphones. This method ideally requires that individual head-related transfer functions (HRTFs) are measured for every listener. However: (1) it precludes head movements, unless a head-tracking system is applied; (2) it disregards the significant, especially at low frequencies, acoustic path contribution (leakage) of the ear-HA system; and (3) it precludes the use of the subject's own or off-the-shelf HAs.

Using a loudspeaker-based virtual sound environment (VSE) to create a realistic 3D audio scene appears to be a more flexible candidate for listening tests with HAs ([Minnaar \*et al.\*, 2010, 2013](#)). The “Simulated Open-Field Environment” (SOFE) system ([Seeber \*et al.\*, 2010](#); [Kerber and Seeber, 2013](#)) and the sound reproduction system described by [Favrot and Buchholz \(2010b\)](#), for example, have been such attempts to induce realism into laboratory testing. Loudspeaker-based VSEs can be generated either by acoustically simulating, or by recording a given environment and then decoding the scene information to appropriate loudspeaker signals. The main methods used to derive loudspeaker signals

---

<sup>†</sup>The concepts of the first part of the Introduction have also been presented in Section 1.1.

from simulated or recorded acoustic information are: (1) direct mapping of sources to single loudspeakers (Seeber *et al.*, 2010), (2) vector base amplitude panning – VBAP (Pulkki, 2001), (3) wave field synthesis – WFS (Berkhout *et al.*, 1993), and (4) higher-order Ambisonics – HOA (Daniel, 2001; Poletti, 2005). The first two methods aim at recreating faithful perception attributes for the reconstructed virtual sources while the remaining two aim at physically reconstructing the sound field inside a defined region.

Even though any of these sound reproduction techniques may be considered for testing hearing aids, this study solely investigates the accuracy and limitations of HOA. The rationale for favouring a method that aims at physically reconstructing the sound field is that the HA processing –particularly when involving multi-microphone techniques such as directional microphones and beamformers– considerably differs from the human auditory processing. For example, for frequencies where HOA cannot control the sound-field any more, the human hearing system creates some kind of hearing impression through masking and summing localisation. It is unclear, though, how the processing of hearing aids would function in such circumstances. Hence, a method that relies only on the reconstruction of perceptual attributes may influence the HA processing in a non-controlled way. Moreover, HOA is chosen over WFS due to a number of advantages it offers (Daniel *et al.*, 2003) for the given application where a single listener is seated in the centre of the loudspeaker array: (1) it aims at optimizing the sound-field reconstruction only at the centred region (sweet spot) rather than a whole area, and (2) it is a scalable format, thus the field encoding is completely decoupled from the decoding, which can be easily executed on almost any loudspeaker array, thereby simplifying the sharing of stimuli between laboratories. However, the temporal behaviour of HOA, particularly outside the reproduction sweet spot and at high frequencies, is characterised by artefacts that may perceptually be more disturbing than those incurred by WFS (Ahrens and Spors, 2009; Ahrens *et al.*, 2010).

Despite the advantages of HOA over other methods, its accuracy is physically limited when practical loudspeaker and microphone systems are applied (Ward and Abhayapala, 2001; Daniel, 2001; Poletti, 2005; Moreau *et al.*, 2006; Epain *et al.*, 2010). Most prominently, HOA exhibits pressure errors in the reconstructed sound-field that increase with frequency and distance from the centre of the loudspeaker array. Both the sound-field reproduction and recording operations contribute to the aforementioned errors through a number of well-studied mechanisms. On the reproduction (decoding) side, errors are due to two main procedures (Poletti, 2005; Spors and Ahrens, 2008; Zotter *et al.*, 2009): (1) the truncation of the infinite-component series describing the pressure inside a space, and (2) the discrete sampling of the reproduction source distribution. On the recording (encoding) side, errors are again due to two distinct procedures Meyer and Elko (2004); Rafaely (2005); Li and Duraiswami (2007): (1) the discrete sampling of the sound-field

on the recording microphone array, and (2) the need to apply considerable amplification for practical small-sized ( $R \sim 0.06$  m) arrays, which, in turn, produces detrimental system noise.

To better understand the effect of HOA on the processing of HAs (or other hearing devices), the sound signals picked up by these devices have to be analysed. Inevitably, this analysis needs to include the effect of the listener's head inside the reproduced sound field. Even though most of the HOA limitations are very well understood in free-field propagation conditions, only very few studies have taken into account the effect of the listener's head (Daniel, 2001; Epain *et al.*, 2010). Hence, a major contribution of the present study is that it provides such a detailed analysis. This is accomplished by employing non-individual HRTFs to map the signals that are presented to an example loudspeaker array to the signals at a dummy listener's ears and HA microphones.

Moreover, since most HAs employ directional microphones to improve the signal-to-noise ratio in noisy conditions, the present work analyses how HOA affects their the output. Directional microphones exhibit higher sensitivity to small amplitude and phase variations than the auditory system and thus, are expected to show an increased sensitivity to HOA reconstruction errors. Sections 4.4.2.2 and 4.4.3.2 investigate this aspect by analysing the output of an example directional HA microphone. It should be emphasised that this study solely investigates the effect of HOA on the output of a HA that includes a static directional microphone processing. This provides a crucial first step before one can investigate the effect of HOA on other, more advanced, HA signal processing methods, such as non-linear amplification (or wide dynamic range compression), noise reduction, de-reverberation, or (bilateral) adaptive beamformers (for an overview of typical HA features refer to Kates, 2008).

Concerning the nature of the reconstructed sound field, the existing studies have analysed and optimised HOA coding for single, anechoic sound sources, such as ideal point sources or plane-wave sources. To the best knowledge of the authors, no study has taken into account the effect of multiple sound sources and room reverberation in a thorough and analytical manner. Hence, another key contribution of this paper is that it presents a systematic evaluation of HOA when employed to reconstruct realistic reverberant sound-fields (Section 4.4.3). Computer-based room simulation techniques are employed here to simulate the sound field inside an example office space (Section 4.3.1). Following, the errors introduced by the HOA encoding and decoding operations are analysed. The results (discussed in Section 4.5.3) suggest that the reverberation mitigates part of the HOA limitations, effectively extending its useful frequency range beyond the reported in literature (Ward and Abhayapala, 2001; Moreau *et al.*, 2006) limits, which are derived for idealised anechoic sound fields.

Finally, the present work proposes a practical method to assess how the noise inherent in a microphone array limits the accuracy of HOA. Thereby, Sections 4.3.2.1 and 4.4.1 compare the total HOA system noise produced at a listener's ears to the threshold of hearing. This analysis extends established performance measures, such as the white noise gain – WNG (Rafaely, 2005; Favrot and Marschall, 2012; Jin *et al.*, 2014), which do not provide sufficient information on the audibility of the total noise produced by a given HOA microphone array.

Apart from the above objective measures employed to characterise the errors introduced by the HOA sound-field coding, a number of perceptual measures, such as the subjective sound coloration, the apparent source width, and the localization accuracy, could have been examined. However, it is unclear how the limitations revealed in these basic perceptual measures translate into real-life performance changes of aided hearing impaired listeners in terms of speech intelligibility, particularly when complex (i.e., reverberant, multi-source) environments are reconstructed with HOA. Hence, even though these perceptual aspects are highly relevant and ultimately need to be investigated in the future, they are considered as being out of the scope of the present study.

The remaining chapter is organised as follows: Section 4.2 provides a summary of the fundamental HOA signal processing that is applied throughout the entire study. All methods that are applied to systematically analyse the effect of HOA on directional HAs operating in anechoic, as well as reverberant, environments are described in Section 4.3. That section presents details on the applied acoustic room modelling techniques, on the sound field coding and binaural resynthesis, on the directional microphone implementation, on the applied performance measures as well as on the choice of appropriate reference conditions. The derived results and HOA limitations are presented in Section 4.4 and discussed in Section 4.5. The study is finally concluded with a brief summary of the findings in Section 4.6.

## 4.2 Higher-Order Ambisonics (HOA) basics<sup>‡</sup>

### 4.2.1 Fourier-Bessel series expansion

Higher-Order Ambisonics (HOA) is a mathematical concept that permits the decomposition and resynthesis of sound fields. It is based on the representation of acoustic waves in space as weighted sums of spherical harmonics, an orthogonal system that can be used to expand any square-integrable function on the unit sphere (Williams, 1999).

---

<sup>‡</sup>The reader who studied the theory of HOA in Chapter 2 can safely skip this section.

The use of spherical harmonics provides a convenient tool to approach the problem of sound field reconstruction (Fazi and Nelson, 2007).

In the following, the notation used by Daniel (2001) and Moreau *et al.* (2006) is adopted, i.e., the study uses *real-valued* spherical harmonic functions (SHF) and a spherical coordinate system where elevation  $\delta$  is measured from the horizontal plane and azimuth  $\theta$  increases counter-clockwise, as seen from positive  $z$ -axis. The pressure at a point  $\mathbf{r} = (r, \theta, \delta)$  inside a source-free region  $\Omega_{r_L}$  can thus be written as:

$$p(kr, \theta, \delta) = \sum_{m=0}^{\infty} i^m j_m(kr) \sum_{n=0}^m \sum_{\sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \delta), \quad (4.1)$$

where  $i = \sqrt{-1}$  is the imaginary unit,  $j_m(kr)$  the spherical Bessel functions of degree  $m$ ,  $k$  the wavenumber,  $B_{mn}^{\sigma}$  the expansion coefficients or simply HOA components, and  $Y_{mn}^{\sigma}(\theta, \delta)$  the real-valued spherical harmonic functions (SHFs) of degree  $m$  and order  $n$ , defined as:

$$Y_{mn}^{\sigma}(\theta, \delta) = \sqrt{(2m+1)(2-\delta_{0,n}) \frac{(m-n)!}{(m+n)!}} P_{mn}(\sin \delta) \cdot \begin{cases} \cos n\theta & \text{if } \sigma = +1 \\ \sin n\theta & \text{if } \sigma = -1 \text{ (ignored if } n = 0) \end{cases}, \quad (4.2)$$

where  $\delta_{0,n}$  is the Kronecker delta, which equals 1 for  $n = 0$  and 0 elsewhere, and  $P_{mn}$  the associated Legendre functions (as defined for example by Moreau (2006), i.e., dropping the  $(-1)^m$  term used by Williams (1999) and other researchers). In this study, the spherical harmonic functions are normalised to have unit power. Additionally, an implicit time dependence of  $e^{+i\omega t}$  is assumed for the wave solution, such that a plane wave of the form  $e^{i\mathbf{k}\cdot\mathbf{r}}$  represents an incident wave. Therefore, the wave vector  $\mathbf{k}$  shows the incidence rather than the propagation direction.

## 4.2.2 Decoding to a loudspeaker array

HOA reproduction is commonly based on the plane wave expansion. A unit-amplitude plane wave characterised by the wave vector  $\mathbf{k} = (k, \theta_k, \delta_k)$ , can be expressed (Moreau, 2006) in the form of Eq. (4.1) with HOA components given by the SHFs sampled at the direction of the wave incidence:

$$B_{mn}^{\sigma} = Y_{mn}^{\sigma}(\theta_k, \delta_k). \quad (4.3)$$

For small arguments  $kr$ , the values of the radial weights  $j_m(kr)$  in Eq. (4.1) decrease with large degrees  $m$  (Moreau *et al.*, 2006; Williams, 1999). This permits the truncation of the infinite sum to a degree  $m = M$ , referred to in literature as the order of the HOA system. The infinite series truncation incurs a spatial band-limiting error commonly quantified in literature by the normalised truncation error (Ward and Abhayapala, 2001). Assuming that for most practical applications the normalised truncation error satisfies:  $\epsilon(kr) < 4\% = -14$  dB, the above authors deduce a rule of thumb for the product  $kr$ :

$$M \geq \lceil kr_o \rceil, \quad (4.4)$$

where  $r_o$  is the radius of the reproduction area of interest and  $\lceil \cdot \rceil$  is the ceiling function rounding its argument upwards to the next integer value. Based on Eq. (4.4) an upper cut-off frequency for accurate HOA reproduction is commonly defined as:

$$f_{\text{HOA}} \leq \frac{c}{2\pi r_o} M, \quad (4.5)$$

where  $c = 343$  m/s represents the speed of sound at usual room conditions.

The HOA reproduction operation can be formulated as follows: given a spherical array of  $L$  loudspeakers placed at angles  $(\theta_l, \delta_l)$  and emitting plane waves weighted by the gains  $s_l$ , we need to derive the optimum values for those gains in order to resynthesise a desired sound field characterised by the HOA components  $B_{mn}^\sigma$  (effectively a weighted sum of  $Y_{mn}^\sigma(\theta_k, \delta_k)$  SHFs describing a superposition of plane waves emanating from different positions  $\mathbf{k}$ ). Using a compact matrix notation to express the linear-equations system that describes the above operation (Daniel *et al.*, 2003; Poletti, 2005) yields:

$$\mathbf{C}\mathbf{s} = \mathbf{b}, \quad (4.6)$$

where the  $[U \times L]$  matrix  $\mathbf{C}$  contains the  $U = (M+1)^2$  SHFs  $Y_{mn}^\sigma(\theta_l, \delta_l)$  as sampled by the loudspeakers, the  $[L \times 1]$  vector  $\mathbf{s}$  contains the loudspeaker gains  $s_l$ , and the  $[(M+1)^2 \times 1]$  vector  $\mathbf{b}$  contains the HOA components  $B_{mn}^\sigma$  characterising the field to be reconstructed.

Assuming that the number of loudspeakers  $L$  is at least equal to the number of HOA components ( $L \geq U$  for 3D reproduction) the solution of the above underdetermined system (i.e., fewer constraining equations [defined by total number of SHFs] than number of unknowns [number of loudspeaker gains]) can be derived as:

$$\hat{\mathbf{s}} = \mathbf{C}^T (\mathbf{C}\mathbf{C}^T)^{-1} \mathbf{b} = \text{pinv}(\mathbf{C}) \mathbf{b} = \mathbf{D}\mathbf{b}, \quad (4.7)$$

where the hat ( $\hat{\cdot}$ ) denotes the least-squares solution, the superscript  $(\cdot)^T$  symbolises the

matrix transpose (this study uses real-valued SHFs such that the conjugate transpose is equivalent to a simple transpose),  $(\cdot)^{-1}$  denotes matrix inversion, and  $\text{pinv}(\cdot)$  is the Moore-Penrose pseudoinverse operator. The above solution to the system is unique among the possibly infinite solutions in that it minimises the total loudspeaker energy  $\|\hat{\mathbf{s}}\|^2$ . The pseudoinverse matrix  $\text{pinv}(\mathbf{C})$  is typically called the decoding matrix  $\mathbf{D}$ .

Generally,  $\mathbf{C}$  can be ill-conditioned, requiring the application of regularization techniques (Poletti, 2005) to solve Eq. (4.6). Alternatively, if quasi-regular loudspeaker setups (Fliege and Maier, 1999; Fliege, 2007) are employed, then the matrix  $\mathbf{C}$  results in being well-behaved as for example shown by Duraiswami *et al.* (2005). The ill-conditioning of  $\mathbf{C}$  is due to the discrete way the loudspeakers sample the SHFs. This discretization can violate the orthonormality property; see Li and North (1997) and Sneeuw (1994) for a rigorous analysis of discrete sampling on the sphere.

The Moore-Penrose pseudoinverse, as used in Eq. (4.7), provides the least-squares solution, given  $\mathbf{C}$  is of full row rank (i.e., right invertible). In the general case this condition is not guaranteed and the formulation of  $\text{pinv}(\mathbf{C})$  as  $\mathbf{C}^T (\mathbf{C}\mathbf{C}^T)^{-1}$  can potentially result in unstable solutions. This study, however, computes the pseudoinverse using MATLAB's `pinv` function which employs a robust singular-value decomposition method (Gregorčič, 2001), with the singular values that are lower than a given threshold being treated as zero (see 'truncated singular-value decomposition – TSVD' method in Granados *et al.*, 2013). In that way, stable –albeit potentially sub-optimal (Solvang, 2008; Trevino *et al.*, 2010)– solutions for the loudspeaker gains can be derived even if the system is ill-conditioned.

### 4.2.3 Modified (max- $r_E$ ) decoding

Since HOA does not control the sound field at frequencies above the estimated frequency cut-off given in Eq. (4.5), it is beneficial to modify the decoded loudspeaker signals so that most of the energy is emitted by loudspeakers that are close to the expected direction of the virtual sound. This is effectively the goal of the 'max- $r_E$ ' modified decoding method proposed by Daniel *et al.* (1998). Formally defined, the max- $r_E$  decoding aims at maximizing the norm  $r_E$  of the energy vector  $\vec{E}$  proposed by Gerzon (1992). Although the energy vector  $\vec{E}$  does not directly correspond to a physical property of the system, Daniel suggested (Daniel *et al.*, 1998; Daniel, 2001) that its norm  $r_E$  is related to the concentration of the high frequency energy in the direction of  $\vec{E}$ . Therefore a solution that maximises  $r_E$ , results in perceptually focusing the high frequency energy towards the expected direction. Daniel *et al.* (1998) also showed that this method improves the reconstruction of Interaural Level Differences (ILDs). Moreover, Frank (2009) illustrated

that the max- $r_E$  proves advantageous for more accurate localisation of phantom (virtual) sources.

However, such a solution reduces the norm  $r_V$  of the velocity vector  $\vec{V}$  (Gerzon, 1992; Daniel, 2001) which translates into a degradation of the sound-field reconstruction at low frequencies. Hence, this modified decoding should only be used above a frequency threshold  $f_x$ , usually lying close to the threshold  $f_{HOA}$ .

The max- $r_E$  decoding involves the use of order-dependent gains  $g_m$ , which weigh the elements of the decoding matrix  $\mathbf{D}$ , effectively tapering the contribution of higher order components:

$$\mathbf{D}_m = \mathbf{D} \text{diag} \left[ g_0 \ g_1 \ g_1 \ g_1 \ \cdots \ \underbrace{g_m \ \cdots \ g_m}_{2m+1} \ g_M \ \cdots \ g_M \right], \quad (4.8)$$

where the  $\text{diag}[\cdot]$  function returns a square  $[U \times U]$  diagonal matrix whose main diagonal elements are equal to the input argument vector. The gain values  $g_m$ , along with the way they are derived, are given by Daniel (2001, p. 183).

It should be mentioned that alternative methods exist that can improve the accuracy of the sound reproduction above the cut-off frequency  $f_{HOA}$ . For example, the loudspeaker signals can be equalised so that the generated sound-field intensity, averaged at a large set of points –either in free-field or on a rigid sphere–, exhibits a flat frequency response (Epain and Jin, 2014). This method was not pursued in the present work since our aim was to validate the range where the HOA coding can fully control the sound-field, without resorting to techniques tuned to a particular configuration.

For more recent advances on decoding to irregular loudspeaker arrays, the interested reader is referred to the works of Trevino *et al.* (2010, 2011); Epain *et al.* (2014); Heller and Benjamin (2014).

#### 4.2.4 Field encoding with a microphone array

Until now it has been assumed that the positions of the considered plane-wave virtual sources are known so that Eq. (4.3) can be used to derive the HOA components  $B_{mn}^\sigma$  required for Eq. (4.7) to yield the loudspeaker gains  $s_l$  (HOA decoding operation). However, in real-life scenarios, the positions of the sources are unknown, hence the components  $B_{mn}^\sigma$  need to be estimated. A spherical microphone array is usually employed for that purpose, performing the so-called HOA encoding of the acoustic scene.

In this study we consider the most frequently used robust implementation (Poletti, 2005) of a rigid sphere of radius  $R$  using  $Q$  flush-mounted omnidirectional microphones.

To illustrate the encoding process, we first assume a theoretical, continuous pressure spherical microphone of radius  $R$ . The pressure on its surface, due to an impinging wave characterised by the HOA components  $B_{mn}^\sigma$ , can be calculated (Moreau *et al.*, 2006; Williams, 1999; Duda and Martens, 1998) using the series expansion:

$$p(r = R, \theta, \delta) = \sum_{m=0}^{\infty} W_m(kR) \sum_{n=0}^m \sum_{\sigma=\pm 1} B_{mn}^\sigma Y_{mn}^\sigma(\theta, \delta), \quad (4.9)$$

with the rigid sphere mode strength term  $W_m(kR)$  defined by:

$$\begin{aligned} W_m(kR) &= i^m \left( j_m(kR) - \frac{j'_m(kR)}{h_m^{(2)'}(kR)} h_m^{(2)}(kR) \right) \\ &= \frac{i^{m-1}}{(kR)^2 h_m^{(2)'}(kR)}, \end{aligned} \quad (4.10)$$

where  $h_m^{(2)}(kR)$  is the  $m$ -th degree spherical Hankel function of the second kind and the prime  $(\cdot)'$  denotes the derivative of the spherical Bessel and Hankel functions with respect to their argument  $kR$ .

Given the pressures on all  $Q$  pressure sensors on the rigid sphere, the sound-field HOA components can be estimated using a quadrature approximation of the inverse spherical Fourier transform integral (Rafaely, 2005). An alternative method to estimate the HOA components from the microphone signals that does not require a quadrature can be formulated exploiting the matrix notation in a similar way as presented in Section 4.2.2. Sampling Eq. (4.9) at the microphone positions  $(\theta_q, \delta_q)$  and converting the summations to matrix multiplications, the following compact notation (Moreau *et al.*, 2006) can be derived:

$$\mathbf{Y} \text{diag}[\mathbf{W}_{3D}(kR)] \mathbf{b} = \mathbf{p}_R, \quad (4.11)$$

where the  $[Q \times U]$  matrix  $\mathbf{Y}$  contains the SHFs  $Y_{mn}^\sigma(\theta_q, \delta_q)$  as sampled by the microphones, the  $[U \times U]$  diagonal matrix  $\text{diag}[\mathbf{W}_{3D}(kR)]$  contains the inverse rigid-sphere mode-strength terms, the  $[U \times 1]$  vector  $\mathbf{b}$  contains the HOA components  $B_{mn}^\sigma$  to be estimated and the  $[Q \times 1]$  vector  $\mathbf{p}_R$  contains the microphone pressure signals  $p_R(\theta_q, \delta_q)$ . The diagonal mode strength matrix is formed using the following vector passed as argument to the  $\text{diag}[\cdot]$  function:

$$\mathbf{W}_{3D} = \left[ W_0 \ W_1 \ W_1 \ W_1 \ \cdots \ \underbrace{W_m \cdots W_m}_{2m+1} \ W_M \cdots W_M \right], \quad (4.12)$$

where the  $kR$  dependence has been dropped for conciseness. All elements of the aforementioned matrices are relevant to a *single frequency point*. Hence, characterising a

broadband sound field necessitates the consideration of Eq. (4.11) *separately* for every frequency of interest.

Assuming  $Q \geq U$ , the least-squares solution of the underdetermined linear-equation system of Eq. (4.11) can be written (Moreau *et al.*, 2006; Poletti, 2005) as:

$$\begin{aligned}\hat{\mathbf{b}} &= \text{diag}[\mathbf{W}_{3\text{D}}(kR)]^{-1} \mathbf{Y}^T (\mathbf{Y}\mathbf{Y}^T)^{-1} \mathbf{p}_R \\ &= \text{diag}[\mathbf{H}_{\text{eq},3\text{D}}(kR)] \text{pinv}(\mathbf{Y}) \mathbf{p}_R \\ &= \mathbf{E} \mathbf{p}_R,\end{aligned}\tag{4.13}$$

where  $\mathbf{E}$  denotes the encoding matrix. The equalization diagonal matrix  $\text{diag}[\mathbf{H}_{\text{eq},3\text{D}}(kR)]$  forms its diagonal elements  $H_{\text{eq},m}(kR)$  by inverting, one-by-one, the terms of Eq. (4.12).

Inverting the mode-strength terms poses implementation problems due to the very low value of high-degree ( $m$ ) functions  $W_m(kR)$  at low frequencies (Rafaely, 2004). Increasing the radius  $R$  of the microphone shifts the resulting large gains to lower frequencies, at the expense of also shifting spatial aliasing problems to lower frequencies (Moreau *et al.*, 2006; Rafaely, 2005; Rafaely *et al.*, 2007; Abhayapala and Ward, 2002). Estimating the HOA components at low frequencies, without having to resort to a large radius  $R$  design, is commonly treated by applying Tikhonov regularization (Moreau *et al.*, 2006; Moreau, 2006; Poletti, 2005; Zotkin *et al.*, 2010) to the solution of Eq. (4.13). Effectively, the equalization terms are modified so that:

$$\tilde{H}_{\text{eq},m}(kR) = \frac{|W_m(kR)|^2}{|W_m(kR)|^2 + \lambda^2} \cdot \frac{1}{W_m(kR)},\tag{4.14}$$

where the parameter  $\lambda$  controls the amount of regularization, with higher values yielding a more regularised solution. The parameter  $\lambda$  tapers off the high magnitudes of the inverse of the mode strengths (radial equalization filters) at low frequencies (see Epain and Daniel, 2008, Section 3) so that the microphone self-noise is not excessively boosted. It should be noted here that this study applies a basic encoding procedure, as described for example by Moreau *et al.* (2006), and does not employ any further optimisation methods, such as the ‘Shape-matching’ method described in Section 2.8 or the alternative formulation presented by Jin *et al.* (2014), to control and optimise the high-frequency errors (due to spatial aliasing) of the encoding operation. For the example microphone array of the present work, such techniques provided a minor but significant improvement of the average error above 8 kHz. Hence, the reduction of the high frequency aliasing does not affect the estimation of the useful frequency range ( $< 5 - 6$  kHz) where the reproduction error (Section 4.3.3.1) is well minimal.

Combining Eqs. (4.7), (4.13) and (4.14), the loudspeaker signals can be derived as:

$$\tilde{\mathbf{s}} = \mathbf{D}\tilde{\mathbf{E}}\mathbf{p}_R = \tilde{\mathbf{T}}\mathbf{p}_R, \quad (4.15)$$

with the tilde ( $\tilde{\cdot}$ ) here denoting a regularised –concerning the microphone encoding– solution and  $\tilde{\mathbf{T}}$  being an  $[L \times Q]$  matrix combining the signals captured by  $Q$  microphones to create the signals driving  $L$  loudspeakers.

#### 4.2.5 Plane-wave assumption & near-field effects

In general, loudspeakers can be considered as plane-wave sources at realistic array radius distances (e.g.,  $r_l \sim 1-2$  m) when looking at a small portion of the wave-front in the vicinity of a human head centred in the array. In this study, loudspeakers and virtual sources are simulated using a set of HRTF data measured with a source at  $r_s = 1.2$  m (Oreinos and Buchholz, 2013c) from the centre of the dummy head. Hence, the plane-wave assumption is considered valid assuming such a source distance. However, regardless of the validity of that assumption, given that the distance of the (free-space) virtual sources equals the distance of the reproduction loudspeakers, the near-field compensation (see Favrot and Buchholz, 2010a, Eq. 1) reduces to a unity gain.

### 4.3 Methods

Fig. 4.1 summarises the simulation framework that this study employs to systematically analyse the limitations of HOA reproduction for HA testing applications. The framework consists of an acoustic scene simulation stage (left section), a sound-field coding stage (middle section) and a playback stage (right section). Reverberant scenes were simulated using the room acoustics software ODEON. The simulation resulted in the definition of a number of image sources, representing the first specular reflections of the room’s response, as well as an energy-based description of the diffuse (Late) room reverberation. This information was then mapped onto a grid of auxiliary plane-wave sources using the LoRA toolbox (Favrot and Buchholz, 2010b). The derived signals at the auxiliary sources formed the output of the acoustic scene simulation stage.

In the subsequent sound field coding stage, different methods were applied to map the auxiliary sound source signals to a (virtual) playback loudspeaker array. On the one hand, an idealised coding method (path 1) was realised to provide a suitable reference condition. On the other hand, a HOA ‘Decoding-only’ method (path 2) and an ‘Encoding and Decoding’ method (path 3) were realised to separately investigate how the HOA encoding and decoding operations contribute to the total reproduction error.

Finally, in the playback (synthesis) stage, binaural signals were derived by employing HRTFs as well as Behind-the-Ear-HA-related transfer functions, BTE-TFs, to simulate the acoustic path from each loudspeaker to the listener’s eardrums and HA microphone-pairs, respectively. For each HA microphone-pair the signals were then combined to form the output of a directional HA microphone. For reference purposes, the traditional free-field case was also considered inside the loudspeaker array by omitting the acoustic manikin. For this free-field case, the reconstructed pressure was examined at the fictitious ears of a ‘transparent head’, characterised by the spherical coordinates:  $(r_e, \theta_e, \delta_e) = (0.09 \text{ m}, \pm 90^\circ, 0^\circ)$ .

In the following sections, a detailed description of all the simulation stages shown in Fig. 4.1 is presented.

### 4.3.1 Acoustic scene simulation

In order to gain an understanding of the practical real-world limitations of HOA sound reproduction, this study considers simulated reverberant acoustic scenes. The advantage of simulated, as opposed to measured, acoustic scenes lies in having access to the detailed spatial, temporal and spectral information of sound waves as they are reflected, scattered and diffracted by objects and boundaries of the room.

#### 4.3.1.1 ODEON room model

A typical small office room ( $3.6 \times 4.4 \times 2.7 \text{ m}$ ) was modelled in ODEON, following the shape and acoustics of a meeting room at the National Acoustic Laboratories. This room was chosen so that a reasonably, yet not excessively, reverberant field was generated, providing a good balance between direct sound and reverberant energy. The sound sources that were used in the room acoustics simulation modelled the spectrum and directivity ( $\mathcal{Q} = 1.6$ ) of a human talker with a normal vocal effort, which in ODEON was realised by the directivity patten ‘Tlknorm\_NATURAL’. All sources were placed at a distance  $r_{\text{src}} = 1 \text{ m}$  from the receiver. The simulation resulted in an estimated reverberation time of  $T_{30} = 0.3 \text{ s}$  and a critical distance [Kuttruff, 2000, Eq. (5.40)] of  $r_c \approx 0.85 \text{ m}$ .

Fig. 4.2 illustrates the simulated acoustic scene. Source locations 1 and 2 ( $\theta_{\text{src}} = 0^\circ, -60^\circ$ ) were used for the error analysis of the HOA reconstruction of a front versus a lateral source (Sections 4.4.2.1 and 4.4.3.1). The location of source 2 was chosen to yield a substantial head shadowing effect (Oreinos and Buchholz, 2013c, Fig. 5), minimizing the direct energy reaching the left ear. Source locations 1 and 3 ( $\theta_{\text{src}} = 0^\circ, -150^\circ$ )

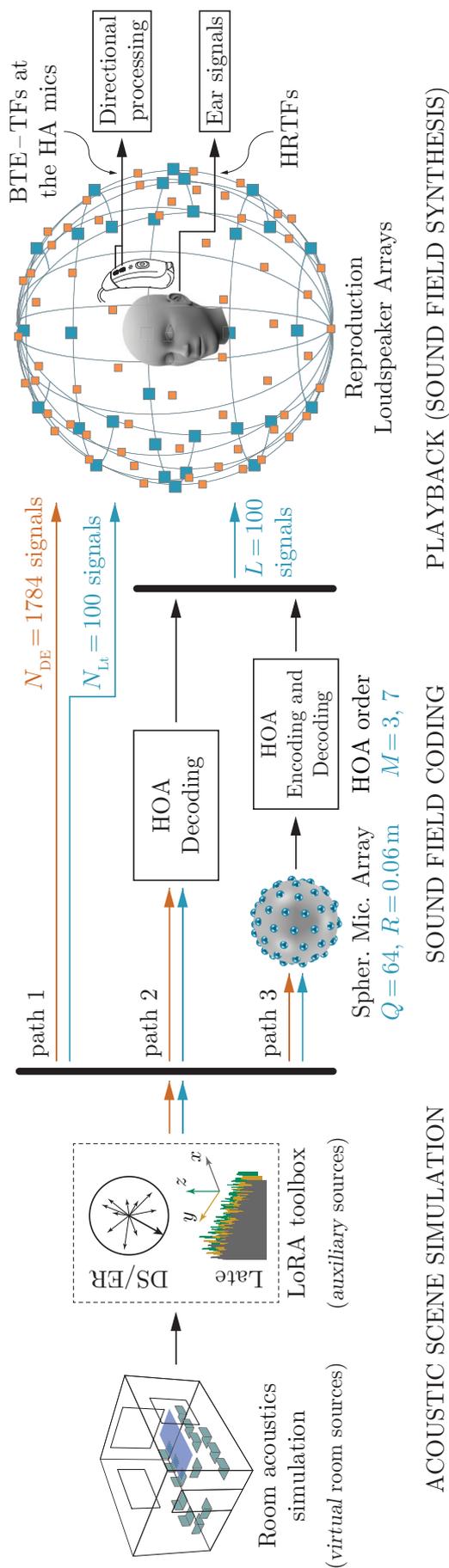


FIGURE 4.1: The complete simulation framework employed to validate reverberant HOA sound fields.

were used for the analysis of the HOA errors incurred on the response of the considered directional (cardioid) HA (Sections 4.4.2.2 and 4.4.3.2). The location of source 3 was chosen such that the HA provided a considerable directional benefit.

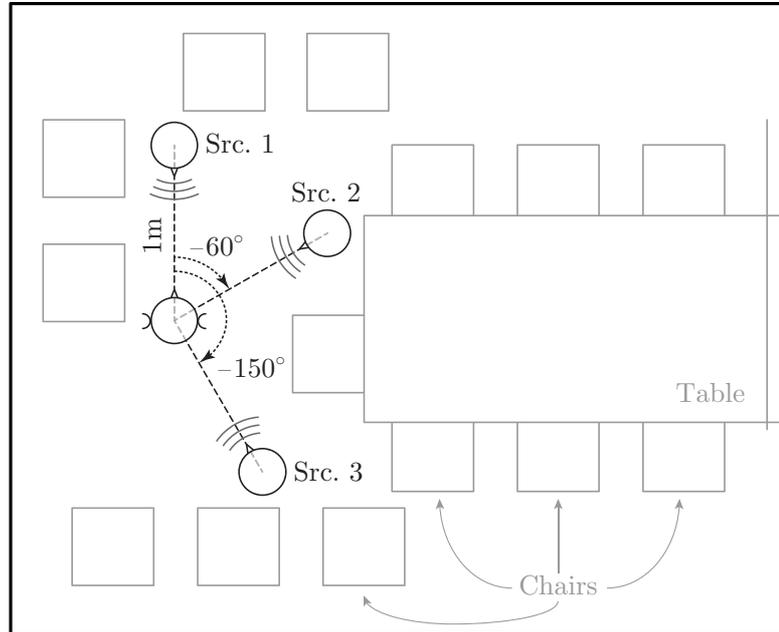


FIGURE 4.2: Top view of the modelled room, showing the receiver and 3 sources, 1 m away, at  $\theta_{\text{src}} = 0^\circ, -60^\circ$  and  $-150^\circ$  relative to the receiver (same height).

The acoustic simulation in ODEON resulted in a spatial, multi-dimensional description of the room impulse response (RIR) characterising the propagation of sound from all sources to the receiver inside the considered room. The direct signal (DS) and early reflections (ERs) were thereby treated as discrete image sources, with ODEON calculating the time and directions of arrival, as well as their spectrum in octave bands. The late diffuse reflections (Late) were not treated as discrete sources but were rather expressed as direction- and frequency-dependent energy decay curves, in the form of directional intensity vectors estimated at discrete time intervals (Christensen and Koutsouris, 2013).

A high transition order (Christensen and Koutsouris, 2013) value  $N_T = 4$  was employed in ODEON so that the main energy of the RIR was described deterministically, in terms of image sources, as opposed to being described stochastically, in terms of energy decay curves. The rationale was that reconstructing specular reflections would be more demanding, in terms of HOA accuracy, compared to reconstructing diffuse sound-field distributions.

In order to study the effect of source-receiver location inside the modelled room, four additional scenes were considered. These scenes only contained sources 1 and 2, albeit repositioned at four randomly-picked locations in the room with the constraint that the relative position to the receiver (also repositioned) was retained.

Moreover, an anechoic condition was simulated by keeping only the DS in the RIR of the initial scene (sources 1 and 2 considered in succession) RIR. On the one hand, this allowed studying the effect of adding reverberation on the HOA reconstruction error. On the other hand, it allowed direct comparison to literature results, which typically do not take into account room reverberation.

#### 4.3.1.2 Rendering to auxiliary sound sources

Following the ODEON room acoustic simulation, the LoRA toolbox (Favrot and Buchholz, 2009, 2010b) was employed to translate the RIR description, provided by ODEON, into the multichannel impulse responses  $h_u^{\text{DE}}[n]$ , with each channel  $u = 1 \dots N_{\text{DE}}$  referring to a single auxiliary sound source. The  $(\cdot)^{\text{DE}}$  superscript refers to the specular components of direct signal and early reflections. For that purpose, the DS and ERs were simply mapped to a discrete grid of auxiliary sources by shifting the direction of each individual image source to the direction of the closest available auxiliary source. A grid of  $N_{\text{DE}} = 1784$  auxiliary sources was chosen, corresponding to the positions of the extended HRTF data-set recorded by Oreinos and Buchholz (2013c), since these HRTFs were later used to derive ear and HA microphone signals (Section 4.3.3). Thanks to the very dense nature of the HRTF grid, the shifting process resulted in very small angular errors ( $\Delta\theta < 5^\circ$ ), which, in the following, were considered unimportant.

The procedure by which the diffuse (Late) part of the RIR description was mapped to a grid of auxiliary sound sources was somewhat more involved. ODEON describes the late reverberation via vector intensity decay curves in  $x$ -,  $y$ -,  $z$ - coordinates, as well as a total power decay curve. These decay curves were treated by LoRA as envelopes mapped to the  $W, X, Y, Z$  channels of a 1<sup>st</sup> order Ambisonics (B-format) coding scheme, following the rendering technique presented by Hollerweger (2006); Merimaa and Pulkki (2005). The B-format channels were decoded to  $N_{\text{Lt}} = 100$  auxiliary sources, initially quasi-regularly distributed on a sphere (Fliege, 2007) and then shifted so that each auxiliary source was aligned with its closest point from the HRTF setup introduced above. The  $N_{\text{Lt}} = 100$  B-format envelopes, were then multiplied (per time frame) with  $N_{\text{Lt}}$  sequences of uncorrelated Gaussian noise, yielding the impulse responses  $h_v^{\text{Lt}}[n]$ , with  $v = 1 \dots N_{\text{Lt}}$  and with the  $(\cdot)^{\text{Lt}}$  superscript referring to the diffuse (Late) part of the RIR.

The LoRA impulse responses  $h_u^{\text{DE}}[n]$  and  $h_v^{\text{Lt}}[n]$  were normalised such that their Discrete Fourier Transform (DFT) represented SPL power spectrum densities at the receiver position. A further normalisation was thereafter applied by dividing their DFT with the DFT of the DS impulse response alone ( $\mathcal{F}\{h_1^{\text{DE}}[n]\}$ ), as observed at the centre of the HATS with the HATS removed. This normalisation is in agreement with the

common definition of HRTFs (Blauert, 1997). Hence, the frequency response of the auxiliary sources, at the output of the ‘Acoustic Scene Simulation’ stage (Fig. 4.1), can be expressed as:

$$\begin{aligned} H_u^{\text{aux/DE}}(f) &= \frac{\mathcal{F}\{h_u^{\text{DE}}[n]\}}{\mathcal{F}\{h_1^{\text{DE}}[n]\}}, \quad u = 1 \dots N_{\text{DE}} \\ H_v^{\text{aux/Lt}}(f) &= \frac{\mathcal{F}\{h_v^{\text{Lt}}[n]\}}{\mathcal{F}\{h_1^{\text{DE}}[n]\}}, \quad v = 1 \dots N_{\text{Lt}}. \end{aligned} \quad (4.16)$$

### 4.3.2 Sound field coding

Sound field coding refers to the method employed to generate loudspeaker signals that best reconstruct the original sound field inside the loudspeaker array. In this study, the initial sound field consisted of a simulated reverberant environment, expressed as the combined output of all auxiliary (plane-wave) sound sources (Section 4.3.1.2). According to Fig. 4.1 (path 1), the reference sound field was identical to the initial sound field, and was realised by simply replacing the  $N_{\text{DE}}$  and  $N_{\text{Lt}}$  auxiliary sources by virtual loudspeakers at identical locations. The relevant loudspeaker signal responses can thus be expressed as:

$$\begin{aligned} S_u^{\text{Ref/DE}}(f) &= H_u^{\text{aux/DE}}(f), \quad u = 1 \dots N_{\text{DE}} \\ S_v^{\text{Ref/Lt}}(f) &= H_v^{\text{aux/Lt}}(f), \quad v = 1 \dots N_{\text{Lt}} \end{aligned} \quad (4.17)$$

In the HOA Decoding-only scheme (Fig. 4.1, path 2), a set of  $L$  loudspeaker filters was derived via Eqs. (4.7) and (4.8), coding the  $N_{\text{DE}}$  and  $N_{\text{Lt}}$  auxiliary source signals into  $L$  playback loudspeaker signals. Section 4.3.3.2 describes the geometry of the loudspeaker array.

In the HOA Encoding and Decoding scheme (Fig. 4.1, path 3), the sound field of the  $N_{\text{DE}}$  and  $N_{\text{Lt}}$  auxiliary sources was simulated by plane waves impinging from the directions of the individual auxiliary sources. This sound field was then captured by a virtual microphone array (see Section 4.3.2.1), resulting in  $Q$  microphone signals that were derived by calculating Eq. (4.9), using a high truncation order of  $M = 60$ , separately for each individual auxiliary source. Applying Eqs. (4.13) and (4.14), the microphone signals were encoded into HOA components  $\tilde{B}_{mn}^\sigma$ . Finally, the loudspeaker signals  $\tilde{s}_l(f)$  were derived using Eqs. (4.15) and (4.8).

Both HOA coding schemes resulted in loudspeaker signals that can be expressed as:

$$S_l^{\text{HOA}}(f) = \sum_{u=1}^{N_{\text{DE}}} \tilde{s}_{u,l}^{\text{DE}}(f) H_u^{\text{aux/DE}}(f) + \sum_{v=1}^{N_{\text{Lt}}} \tilde{s}_{v,l}^{\text{Lt}}(f) H_v^{\text{aux/Lt}}(f), \quad l = 1 \dots L, \quad (4.18)$$

where  $\tilde{s}_{u,l}^{\text{DE}}(f)$  and  $\tilde{s}_{v,l}^{\text{Lt}}(f)$  are the  $l$ -th loudspeaker HOA filters relevant to the decoding of the  $u$ -th DS/ER and  $v$ -th Late-part auxiliary source respectively.

#### 4.3.2.1 HOA microphone array considerations

The employed (simulated) HOA microphone array consisted of a rigid sphere of radius  $R = 0.06$  m, bearing  $Q = 64$  microphones, flush-mounted on the surface of the sphere. The individual microphones were simulated as pressure sensors with a flat frequency response, exhibiting the noise spectrum of standard Knowles HA microphones (Buchholz, 2013). They were placed at the nodes of a quasi-regular setup (Fliege, 2007), thus ensuring low SHF orthonormality errors. Choosing the radius  $R$  involved balancing contradictory design requirements. Increasing  $R$  shifts the excessive amplification introduced by the HOA encoding to lower frequencies (Moreau *et al.*, 2006, Fig. 11), thereby effectively increasing the usable bandwidth downwards. On the other hand, decreasing  $R$  reduces the solid angle between the microphones, which pushes the spatial aliasing frequency cut-off to higher frequencies (Li and North, 1997; Rafaely *et al.*, 2007), thereby increasing the usable bandwidth upwards. After a manual iterative optimization process that considered the total microphone array noise as well as the pressure reproduction errors (see Section 4.4.2.1) the chosen radius was found to be a good compromise. However, the described microphone array should be considered as an example configuration employed to highlight the principle factors that need controlling when designing a HOA microphone array. The reader is referred to Rafaely (2005) for an extensive discussion on other practical considerations of microphone arrays, such as gain and phase mismatch, and sensor positioning errors. For more best practices when employing real microphones, refer also to the encoding method of ‘Shape-Matching’ (Section 2.8) and to the work of Jin *et al.* (2014).

### 4.3.3 Playback and field synthesis

The final part of the simulation framework (Fig. 4.1) represents the playback of (virtual) loudspeakers, driven by the signals computed at the sound field coding stage (Section 4.3.2). The sound field synthesis was realised by employing a set of transfer functions  $H_{\text{TF}}(f, \theta, \delta)$  representing one of the following:

- (1) the phase delay of the plane-waves as they propagate from each loudspeaker to the ears of a transparent head [i.e.,  $H_{\text{TF}}(f, \theta, \delta) = e^{i2\pi f/cr_e \cos\gamma}$ , with  $r_e$  the transparent head radius and  $\gamma$  the angle between the loudspeaker  $(\theta, \delta)$  and the fictitious ear  $(\theta_e, \delta_e)$ ];

(2) the head-related, or BTE-HA-microphone-related, transfer functions (HRTF/BTE-TF, or simply just HRTF) from each loudspeaker to the ears or HA microphones of the HATS.

Considering the reference condition of the sound field coding stage (Fig. 4.1, path 1), the normalised transfer function describing the reverberant source-receiver path can be expressed as the superposition of all the relevant loudspeaker signals, given by Eq. (4.17), individually filtered by an appropriate  $H_{\text{TF}}(f, \theta, \delta)$  function:

$$H_{\text{Ref}}(f) = \sum_{u=1}^{N_{\text{DE}}} S_u^{\text{Ref/DE}}(f) H_{\text{TF}}(f, \theta_u, \delta_u) + \sum_{v=1}^{N_{\text{Lt}}} S_v^{\text{Ref/Lt}}(f) H_{\text{TF}}(f, \theta_v, \delta_v). \quad (4.19)$$

Note that for the anechoic-source case, where  $N_{\text{DE}} = 1$  and  $N_{\text{Lt}} = 0$ ,  $H_{\text{Ref}}(f)$  reduces, as expected, to the transfer function  $H_{\text{TF}}(f, \theta_1, \delta_1)$ , thanks to the normalisation of Eq. (4.16).

Considering paths 2 and 3 of the sound field coding stage (Fig. 4.1), the normalised transfer function describing the HOA-reconstructed reverberant source-receiver path can be expressed as the superposition of all  $L$  loudspeaker signals [given by Eq. (4.18)]:

$$H_{\text{HOA}}(f) = \sum_{l=1}^L S_l^{\text{HOA}}(f) H_{\text{TF}}(f, \theta_l, \delta_l) \quad (4.20)$$

#### 4.3.3.1 HOA reconstruction errors

Having calculated the reference (Fig. 4.1, path 1) and HOA-relevant (Fig. 4.1, paths 2 and 3) transfer functions that describe the sound propagation inside the considered acoustic scene, the HOA reconstruction errors were calculated (in dB) as:

$$\mathcal{E}_{\text{HOA}}(f) = 20 \log_{10} \left| \frac{H_{\text{HOA}}(f)}{H_{\text{Ref}}(f)} \right|, \quad (4.21)$$

power-smoothed using a running, 1/6-of-an-octave-wide Hanning window. Applying to heuristics, limiting the error to  $|\mathcal{E}_{\text{HOA}}(f)| \leq 2$  dB was expected to ensure a sufficiently accurate reproduction of the original sound field. The pressure-magnitude error  $\mathcal{E}_{\text{HOA}}(f)$  will be subsequently referred to as simply ‘pressure error’ for the sake of nomenclature compactness.

Alternative metrics that additionally consider the phase of the sound fields, e.g., the widely used normalised error (Ward and Abhayapala, 2001; Poletti, 2005; Moreau *et al.*, 2006):  $|[H_{\text{HOA}}(f) - H_{\text{Ref}}(f)]/H_{\text{Ref}}(f)|$ , were considered and resulted in the same trends

as the magnitude errors  $\mathcal{E}_{\text{HOA}}(f)$ . However, their values were difficult to interpret in terms of perceptual attributes (e.g., JND in spectral colorations) and, moreover, were too sensitive to relative errors where the pressure of the original sound field was small (i.e., for sources at the contralateral ear). Phase errors calculated as:  $\text{Arg}[H_{\text{HOA}}(f)/H_{\text{Ref}}(f)]$  were omitted since they simply confirmed the results illustrated by the magnitude errors  $\mathcal{E}_{\text{HOA}}(f)$ .

In any case, the performance analysis of a directional HA microphone operating inside a HOA field, might be the most sensitive metric for the validation of the phase reconstruction accuracy inside the regions of accurate HOA operation. This comes as a result of the differential operation required to produce the directional output, which exhibits high sensitivity to phase errors (Kuk *et al.*, 2000). Thus, the results shown in Sections 4.4.2.2 and 4.4.3.2 may be considered as an indirect measure of the phase reconstruction accuracy.

For a centred listening position, the consideration of the long-term magnitude and phase is sufficient for the case that a single anechoic sound source is reproduced. In the case that multiple sound sources or/and room reverberation are introduced, the different signal or reverberation components will overlap or rapidly fluctuate in time, frequency, and direction of arrival and thereby potentially provide time-frequency intervals with a highly increased or decreased (instantaneous) pressure error. The long-term pressure error of Eq. (4.21) will not be very sensitive to these short-term effects and will only provide a measure of an overall (frequency-dependent) deviation. However, since the signal analysis performed by the human auditory system as well as the processing of advanced HA enhancement methods will always involve some degree of temporal as well as spectral and spatial integration, it is expected that the long-term pressure error  $\mathcal{E}_{\text{HOA}}(f)$  represents at least a first best-case approximation of these integration effects on the overall accuracy of a given HOA system. This aspect will be revisited throughout Section 4.5.3.1.

#### 4.3.3.2 Loudspeaker array configuration

The geometry and number of loudspeakers of the HOA reproduction array were determined by considering the maximum HOA order of interest. Given the restriction  $L \geq (M + 1)^2$  (Ward and Abhayapala, 2001) and taking into consideration practical implementation issues (equipment cost and complexity), the maximum HOA order was set to  $M = 7$ . Although that would require only a minimum of  $L = 64$  loudspeakers, such an array (assuming a quasi-regular configuration) was found to yield a maximum orthonormality error (Moreau *et al.*, 2006) of  $\mathcal{E} = 0.578$  (specifically, for the

inner product of  $Y_{71}^{+1}(\theta, \delta)$  with itself), corresponding to a matrix condition number of  $\kappa(\mathbf{C}) = 6.26 \approx 16$  dB [for a discussion and insights on the condition number, refer to [Li et al. \(2004a\)](#); [Zotter \(2009\)](#)].

In order to reduce the orthonormality error to a more acceptable value of  $\mathcal{E} = 0.110$  – corresponding to a matrix condition number of  $\kappa(\mathbf{C}) = 1.11 \approx 1$  dB –, this study employed an  $L = 100$  loudspeaker array, following the same quasi-regular geometry ([Fliege, 2007](#)) used for the  $N_{\text{Lt}}$  auxiliary sources modelling the RIR Late part (Section 4.3.1). According to [Solvang \(2008\)](#), the larger than necessary number of loudspeakers impairs solely the region of the reconstructed sound field that is outside the sweet spot ( $kr > M$ ), which is not of interest in the present HOA analysis.

### 4.3.3.3 HOA system noise due to microphone array

Microphone array noise has traditionally been evaluated in literature using the White Noise Gain (WNG), a robustness metric originating from antenna theory ([Gilbert and Morgan, 1955](#); [Cox et al., 1987](#); [Veen and Buckley, 2010](#); [Meyer and Elko, 2004](#); [Rafaely, 2005](#); [Favrot and Marschall, 2012](#); [Jin et al., 2014](#)). The WNG is defined as the output power of a microphone array (beamformer) due to unit-variance white noise at its sensors ([Veen and Buckley, 2010](#)). However, in the case of a HOA encoding microphone, the microphone signals are combined to produce the driving loudspeaker signals, as seen in Eq. (4.15), rather than a single output.

Hence, in this study, the noise generated by the entire HOA system at the location of the listener inside the loudspeaker array was derived by assuming that the only input signal to the microphone array was the noise generated by the  $Q$  microphones of the array (i.e., microphone array operating in a source-free acoustic environment). Applying the encoding and decoding operations given in Eq. (4.15), the power spectral density (PSD) of the entire-system noise was derived from the theory of random signals [[Oppenheim et al., 1999](#), Eq. (2.195)] as:

$$N_{\text{HOA}}(f) = \sum_{q=1}^Q \left| \sum_{l=1}^L H_{\text{TF}}(f, \theta_l, \delta_l) \tilde{T}(f, l, q) \right|^2 B(f), \quad (4.22)$$

with  $\tilde{T}(f, l, q)$  being the  $l$ -th row,  $q$ -th column element of matrix  $\tilde{\mathbf{T}}$ , introduced in Eq. (4.15), representing the filter that maps signals from the  $q$ -th microphone to the  $l$ -th loudspeaker;  $H_{\text{TF}}(f, \theta_l, \delta_l)$  being the loudspeaker-to-ear transfer function (see introduction of Section 4.3.3); and  $B(f)$  being the equivalent input noise PSD of each microphone. The noise across single microphones was considered mutually uncorrelated.

It should also be noted at this stage the equivalent input noise PSD of microphones roughly follows a pink-noise or ‘1/ $f$ ’ law (constant power per fractional octave band) rather than a white-noise law.

#### 4.3.4 Directional hearing aid microphone and SNR benefit

Directional microphones produce an output that depends on both the direction and the frequency of the impinging wave. When a virtual source is reproduced in a HOA system, all loudspeakers emit sound waves of different amplitudes and phase. Thus, the output of a directional microphone (i.e., inherent directionality) might be distorted, compared to the ideal anechoic sound field, in regions where the accuracy of the HOA sound field reconstruction is limited, exhibiting erroneous secondary lobes. To evaluate the effect of HOA sound field reproduction on the benefit achieved with a directional HA microphone, the ‘effective directivity’ of a directional microphone was estimated. The term ‘effective’ is used here to denote the apparent directivity when such a microphone operates inside a HOA sound field. Such a subtle distinction is important in order to emphasise that the physical directivity of the microphone remains unaffected. A first order delay-and-subtract cardioid microphone output was derived by using measured front and rear BTE-HA-microphone transfer functions (Oreinos and Buchholz, 2013c) to map the HOA loudspeaker signals to HA microphone signals. The directivity parameter (Kates, 2008, p. 84) was adjusted to  $\alpha = 0.575$  to best approximate the directivity pattern (with the HAs placed on the HATS) of an ideal cardioid pattern. The associated rear microphone delay was given by:

$$\tau = \frac{d_{\text{mic}} \alpha}{c(1 - \alpha)}, \quad (4.23)$$

where  $d_{\text{mic}} = 0.012\text{m}$  was the inter-microphone distance for the employed BTE HAs. The resulting non-integer sample delay  $n_\tau = \tau f_s$  ( $f_s$ : sampling frequency) was implemented as a 64-sample-long sinc-type filter.

The directivity index (Beranek, 1986; Dittberner and Bentler, 2003) was employed to quantify the frequency-dependent distortions incurred by HOA on the effective directivity of the directional (cardioid) HA microphone. A quadrature approximation for the estimation of the DI was employed as below:

$$\text{DI}(f) \approx 10 \log_{10} \frac{4\pi |p_{\text{max}}(f)|^2}{\sum_{i=1}^I \alpha_i |p(f, \theta_i, \delta_i)|^2}, \quad (4.24)$$

where  $p(f, \theta_i, \delta_i)$  is the directional HA microphone output due to a unit-amplitude source located at  $(\theta_i, \delta_i)$ ,  $p_{\text{max}}(f)$  is the maximum directional output [occurring at the look direction  $(\theta_{\text{max}}, \delta_{\text{max}})$ ],  $\alpha_i$  are the quadrature weights (normalised so that  $\sum_{i=1}^I \alpha_i = 4\pi$ )

and  $I$  is the number of applied quadrature points. Here, a regular grid of  $I = 900$  quadrature points (Fliege, 2007) was employed, after shifting each point to align with the closest available HRTF measurement position (Oreinos and Buchholz, 2013c). The frequency-dependent look direction  $(\theta_{\max}, \delta_{\max})$  was estimated *only* for the reference case and the same value was used to derive the DI for the HOA case. Moreover, the search for  $(\theta_{\max}, \delta_{\max})$  was restricted to an area around the expected maximum direction defined by the spherical wedge (ungula):  $|\theta| < \pi/2$  and  $|\delta| < \pi/12$ . This restriction was introduced to prevent any erroneous direction estimations due to possible large side lobes. Finally, the resulting  $\text{DI}(f)$  function was spectrally smoothed using a rectangular 1/3-of-an-octave-wide power smoothing window.

Since the DI is derived based on the pressure output relevant to  $i = 1 \dots I$  anechoic sources, it is not suitable to quantify the effects of HOA on the HA microphone effective directivity when HOA is employed to reconstruct a reverberant scene. In this case, the signal-to-noise ratio (SNR) benefit was used instead and, for that purpose, an in-room scene (see Fig. 4.2) comprising Source 1 as the target (signal) and Source 3 as the distractor (noise), was simulated (see Section 4.3.1.1). After deriving the HA microphone transfer functions for both in-room sources, via Eqs. (4.19) and (4.20), their inverse DFTs were convolved with different 10 s-long anechoic recordings of male speech. The anechoic recordings were normalised to yield a long-term broadband SNR of 0 dB at the room receiver position, with the HATS removed.

Subsequently, frequency-dependent, long-term SNR values were calculated for: (1) an omnidirectional microphone output (i.e., only considering the front HA microphone) and (2) a cardioid HA microphone output. Those were respectively termed:  $\text{SNR}_o(f)$  and  $\text{SNR}_c(f)$ . The effective directional benefit was finally estimated, in dB, as:  $\Delta\text{SNR}(f) = \text{SNR}_c(f) - \text{SNR}_o(f)$ .

All SNR values were computed in frequency bands using a filter-bank of 3<sup>rd</sup> order, 1/3-of-an-octave-wide Butterworth bandpass filters (as specified in ANSI S1.11-2004 [R2009]), with centre frequencies spaced 1/12-of-an-octave apart, from 157 Hz to 8000 Hz. In addition to  $\text{SNR}_o(f)$  and  $\Delta\text{SNR}(f)$ , the intelligibility-weighted and frequency-integrated (broadband)  $\text{SNR}_o^I$  and  $\Delta\text{SNR}^I$  (Greenberg *et al.*, 1993) were computed, with the applied frequency weights derived from the speech intelligibility index theory (ANSI, 1997, Table 3).

## 4.4 Simulation results

### 4.4.1 Microphone array noise

The overall noise in a HOA sound reproduction system is mainly caused by the noise inherent in the microphones of the encoding array. The solid line in Fig. 4.3(a) shows the equivalent input noise (EIN) of an example single microphone integrated in one-ERB-wide [Equivalent Rectangular Bandwidth (Moore, 2012, p. 76)] bands, spaced 43 Hz apart. The employed integrated EIN spectrum, denoted by  $\bar{B}(f)$ , was derived by Buchholz (2013) for an example miniature microphone according to Beranek (1988, p. 640). Using microphones with a different EIN profile  $B'(f)$  would simply offset the total HOA noise, in dB, by  $B'(f) - B(f)$ .

The resulting HOA noise PSD [Fig. 4.3(a)] was calculated via Eq. (4.22), using  $\lambda = 0.01$  to derive the encoding-decoding matrix  $\tilde{\mathbf{T}}$ . The noise was estimated at the HOA reproduction array centre and at the left ear of a transparent head, considering a free-space condition inside the loudspeaker array. Both resulting noise PSDs were integrated in ERB-wide bands, spaced 43 Hz apart, as above for the single-microphone EIN case.

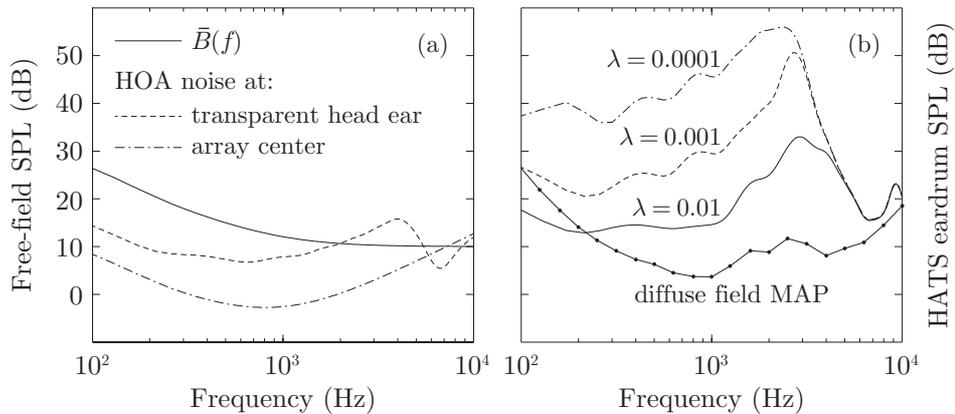


FIGURE 4.3: HOA Encoding and Decoding system ( $M = 7$ ) noise at the: (a) array centre and left-ear of a transparent head ( $\lambda = 0.01$ ); (b) left ear of the HATS. The total noise RMS power at the eardrum was computed as:  $N_{\text{rms}}^{\text{ear}} = 40, 55, 64$  dB for  $\lambda = 0.01, 0.001, 0.0001$ . For comparison purposes plot (a) shows the single-microphone  $\bar{B}(f)$  and plot (b) shows the minimum audible pressure (MAP) for a diffuse-field.

Figure 4.3(b) illustrates the effect on the system noise when placing an acoustic manikin (HATS) inside the HOA loudspeaker array. In this case, the noise PSD generated at the left HATS eardrum, integrated in ERB bands as before, was derived considering three regularization values ( $\lambda = 0.01, 0.001, 0.0001$ ) to illustrate how the regularization affects the HOA system noise. The broadband RMS power of the noise picked up by either HATS' eardrum microphone for the three above regularization values was estimated as:

$N_{\text{rms}}^{\text{ear}} = 40, 55, 64$  dB SPL. These eardrum RMS levels correspond to the following free-field RMS levels:  $N_{\text{rms}}^{\text{ff}} = 33, 42, 57$  dB SPL, considering a diffuse sound presentation.

For reference purposes, Fig. 4.3(b) shows the diffuse-field minimum audible pressure (MAP: Killion, 1978; Moore, 2012, p. 58). The MAP values were derived by mapping minimum audible field (MAF) values for diffuse-field listening, given in ISO-389-7:2005, to the HATS ears. This mapping was performed using the diffuse-field averaged HRTF, calculated from the employed HRTF data set as described by Jot *et al.* (1995) and Gardner (1998, p. 34). For all regularization values, the HOA system noise is higher than the MAP and as such, it should be considered audible.

In the remaining of the study, unless otherwise noted, the regularization parameter was fixed to a practically relevant value of  $\lambda = 0.01$ .

## 4.4.2 Anechoic virtual source reconstruction

### 4.4.2.1 HOA pressure errors

The upper panels of Fig. 4.4 show the reference pressure (Fig. 4.1, path 1)  $H_{\text{Ref}}(f)$  at the left ear of the HATS for anechoic sound sources placed at  $\theta_{\text{src}} = 0^\circ$  (left column) and  $\theta_{\text{src}} = -60^\circ$  (right column) on the horizontal plane. The lower panels of Fig. 4.4 show the corresponding pressure errors  $\mathcal{E}_{\text{HOA}}(f)$ , calculated via Eq. (4.21), for a 3<sup>rd</sup> and 7<sup>th</sup> order HOA sound field reconstruction. Decoding-only (Fig. 4.1, path 2) and Encoding and Decoding (Fig. 4.1, path 3) cases were considered separately to illustrate the effect of using a HOA microphone. The grey-shaded areas in Fig. 4.4 indicate an acceptable error range of  $\pm 2$  dB (see Section 4.3.3.1).

As known from literature, the HOA pressure errors are limited at low frequencies but increase significantly above a certain cut-off frequency. This cut-off frequency increases, in turn, with the HOA order. Additionally, at high frequencies [here above  $f_{\text{al}} \approx 7.4$  kHz as estimated using Moreau *et al.* 2006, Eq.(29)], spatial aliasing occurs, rendering the pressure errors uncontrollable.

While Fig. 4.4 presents HOA errors when using a basic decoding scheme, as given in Eq. (4.7), Fig. 4.5 shows the HOA pressure errors when applying the max- $r_{\text{E}}$  decoding, i.e., using the modified decoding matrix given in Eq. (4.8) at high frequencies. The max- $r_{\text{E}}$  decoding was found to be optimal when activated above  $f_{\text{x}} = M \cdot 800$  (Hz). Setting the crossover frequency such that  $f_{\text{x}} > f_{\text{HOA}}$  [Eq. (4.5)] ensured that the region of accurate HOA reconstruction was left intact. Comparing Figs. 4.4 and 4.5 illustrates that the max- $r_{\text{E}}$  decoding decreases the pressure errors at high frequencies and thereby increases the usable frequency range.

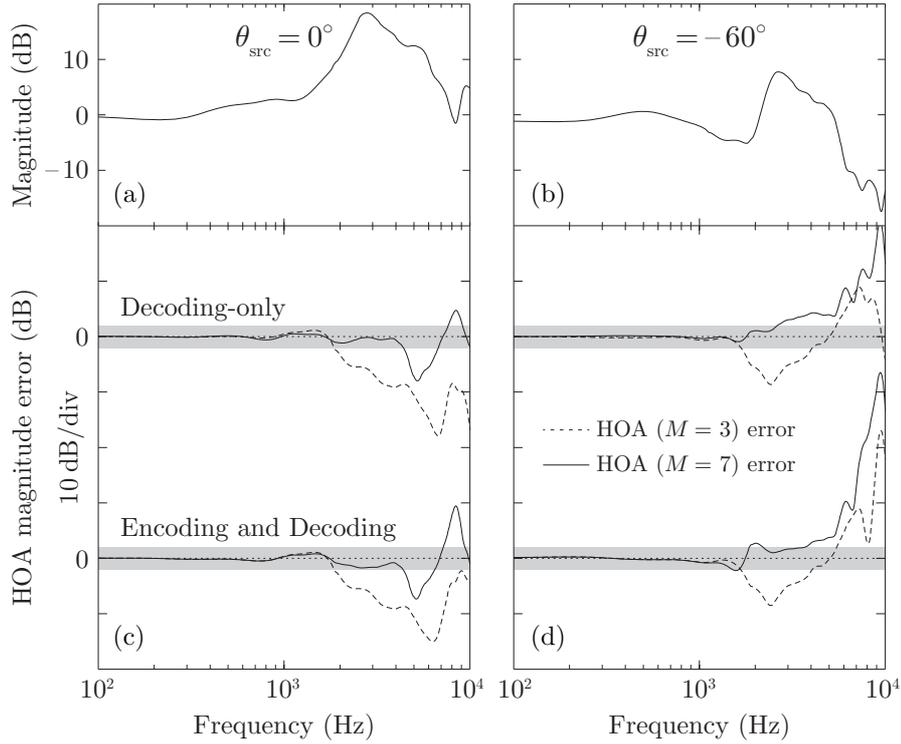


FIGURE 4.4: Pressure errors at HATS' left ear of an  $M = 3$  and  $M = 7$  HOA system reconstructing single anechoic sources placed on the horizontal plane, at  $\theta_{\text{src}} = 0^\circ$  (left column) and  $\theta_{\text{src}} = -60^\circ$  (right column). Plots (a) and (b) show the reference pressure spectra  $H_{\text{Ref}}(f)$ . Plots (c) and (d) show the corresponding HOA pressure-errors  $\mathcal{E}_{\text{HOA}}(f)$ . The shaded regions depict the  $\pm 2$  dB error boundaries.

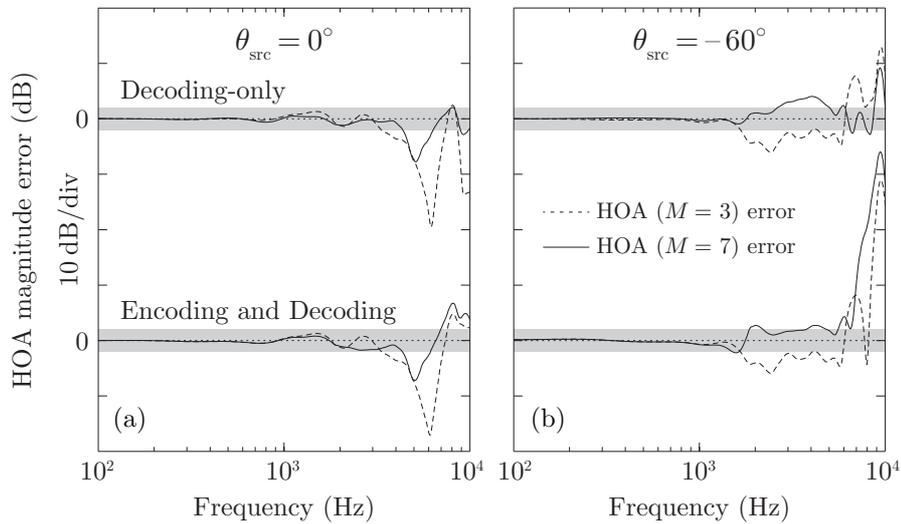


FIGURE 4.5: Same as Fig. 4.4 but using the  $\text{max-}r_{\text{E}}$  decoding activated at the HOA-order-dependent crossover frequency:  $f_{\text{x}} = M \cdot 800$  (Hz). The reference response  $H_{\text{Ref}}(f)$  is given in Figs. 4.4(a) and (b), respectively.

Whereas Figs. 4.4 and 4.5 illustrate the HOA pressure errors for two example virtual sources, Fig. 4.6 presents the pressure errors for evenly distributed (single) virtual sources covering the entire horizontal plane with an azimuth spacing of  $\Delta\theta = 5^\circ$ . Figure 4.6(a) presents the Decoding-only case (path 2) with no acoustic manikin inside the array (free-space). Figure 4.6(b) also presents the Decoding-only case but considers the HATS (left ear as denoted by the vertical dotted lines in the plots) inside the HOA loudspeaker array. Figure 4.6(c) presents the Encoding and Decoding case (path 3) while also including the HATS inside the loudspeaker array. In all graphs, the acceptable error threshold  $|\mathcal{E}_{\text{HOA}}(f)| \leq 2$  dB is denoted by the solid lines and the commonly employed HOA cut-off frequency  $f_{\text{HOA}}$  [Eq. (4.5)] is denoted by the horizontal dash-dotted lines. The sources that lie contralateral to the considered (left) ear yield the largest reconstruction errors (i.e., the smallest usable frequency range).

To further analyse the effect of HOA encoding on the pressure errors, Fig. 4.7 shows the 2 dB error threshold as a function of azimuth with the amount of regularization  $\lambda$  as parameter. Figure 4.7(a) shows the pressure errors for the left ear of a transparent head and Fig. 4.7(b) shows the pressure errors for the left ear of a HATS. The plots further highlight that the errors introduced by the microphone array encoding: (1) mainly occur when the HATS is considered inside the HOA loudspeaker array, (2) are pronounced for contralateral source directions, and (3) require unrealistically small regularization values  $\lambda$  if they are to be mitigated.

In order to illustrate the direction-dependent accuracy of the HOA reconstruction when a listener lies inside the reproduced sound field, pressure errors were calculated for 400 virtual sources quasi-regularly (Fliege, 2007) positioned on a sphere (after being shifted so that each node aligns with one of the HRTF measurement positions, as described in Section 4.3.1.2). Figure 4.8 shows the resulting errors, as surface plots on a sphere, for an example  $M = 7$  Encoding and Decoding system, calculated at the frequency of 5 kHz. As before, the errors were calculated for the left ear of a HATS placed inside the simulated sound-field. Similar to Fig. 4.6, the plots show that the reconstruction errors are heavily position-dependent, with the ipsilateral virtual sources yielding lower errors compared to the contralateral sources.

Aiming at better quantifying the effect of source direction on the pressure errors at the listener's ears, the described analysis was repeated (separately for each frequency point) using various HOA systems of orders  $M = 1, 2, \dots, 7$  and the percentage of the 400 reconstructed sources that resulted in  $|\mathcal{E}_{\text{HOA}}(f)| \leq 2$  dB was estimated. The left column in Fig. 4.9 shows the derived percentage of accurately reproduced sound sources as a function of frequency, with the HOA order  $M$  as parameter, for the cases of: a transparent head and HOA Decoding-only [panel (a)], a HATS and HOA Decoding-only [panel (b)],

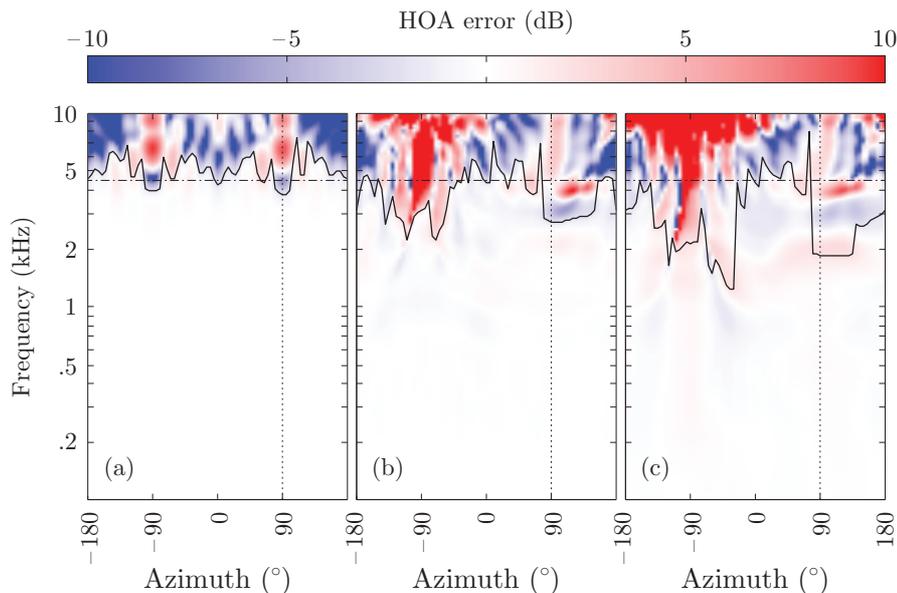


FIGURE 4.6: Pressure errors of an  $M = 7$  HOA system reconstructing anechoic sources lying on the horizontal plane. The errors were calculated at the left ear, denoted by the vertical dotted line at  $\theta = 90^\circ$ , for a: (a) transparent head, Decoding-only case; (b) HATS inside array, Decoding-only case; (c) HATS inside array, Encoding and Decoding case. The thick curves show the 2 dB pressure error cut-off frequency and the horizontal dash-dotted lines represent the rule-of-thumb  $f_{\text{HOA}}$  limit.

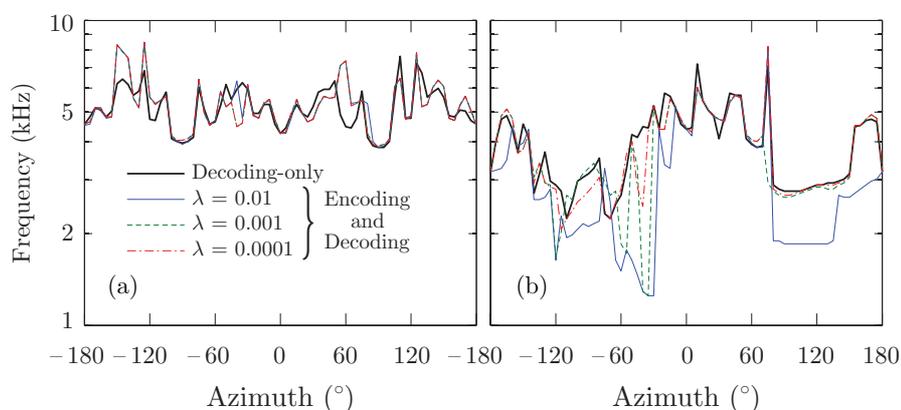


FIGURE 4.7: 2 dB pressure error cut-off frequency curves for an  $M = 7$  HOA system (path 3) plotted as a function of source azimuth (horizontal plane only) for different regularization values  $\lambda$ . The errors were calculated at the left ear of a: (a) transparent head and (b) HATS.

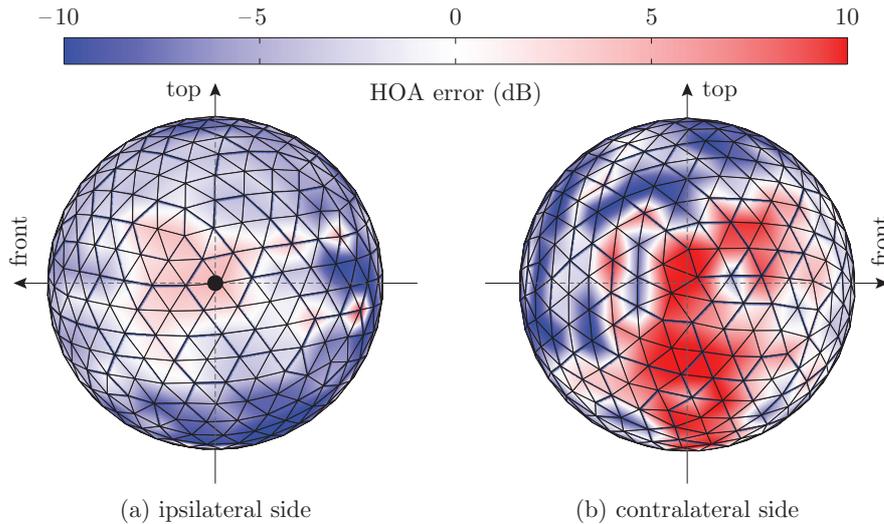


FIGURE 4.8: Pressure errors at 5 kHz for an  $M = 7$  order HOA Encoding and Decoding system measured at the left ear of a HATS relevant to 400 quasi-regularly distributed anechoic sources. Panel (a) shows the ipsilateral source directions and panel (b) the contralateral source directions. The black dot in panel (a) indicates the location of the left ear.

a HATS and HOA Encoding and Decoding [panel (c)]. For the transparent head, sound sources from all directions are successfully reproduced (within  $|\mathcal{E}_{\text{HOA}}(f)| \leq 2$  dB) up to an order-dependent cut-off frequency, above which, practically all sources exceed the error criterion thus yielding a very steep drop-off in the percentage of accurately reproduced sound sources [Fig. 4.9(a)]. This abrupt cut-off behaviour is not observed when a HATS is considered inside the HOA sound field [Fig. 4.9(b)–(c)]. Due to the source direction-dependent nature of the pressure errors shown in Figs. 4.6 and 4.8, sources from mainly the contralateral directions exceed the error criterion at lower frequencies than sources from mainly the ipsilateral directions. As a consequence, the order-dependent frequency roll-off of the percentage of accurately reproduced sound sources shown in Fig. 4.9(b) is much shallower than in Fig. 4.9(a). This effect is even more pronounced when the HOA encoding process is additionally taken into account [Fig. 4.9(c)]. Moreover, in the latter case the curves for orders  $M \geq 5$  roll off at almost the same frequency, highlighting that the regularization required in the encoding process limits the effective HOA order and thus the frequency bandwidth of the accurately reproduced HOA sound field.

The intersections of the three horizontal dashed-dotted lines shown in Fig. 4.9(a)–(c) with the above-described error curves are plotted in Fig. 4.9(d)–(f) as a function of the HOA order  $M$ . The resulting curves illustrate HOA cut-off frequencies below which 50%, 75%, or 95% of the 400 sources are reproduced within an error tolerance of  $|\mathcal{E}_{\text{HOA}}(f)| \leq 2$  dB. It can be observed that when a HATS is considered inside the HOA sound field [Fig. 4.9(e)], increasing the HOA order results in a frequency bandwidth benefit that diminishes when applying a stricter error criterion. As expected, this behaviour is

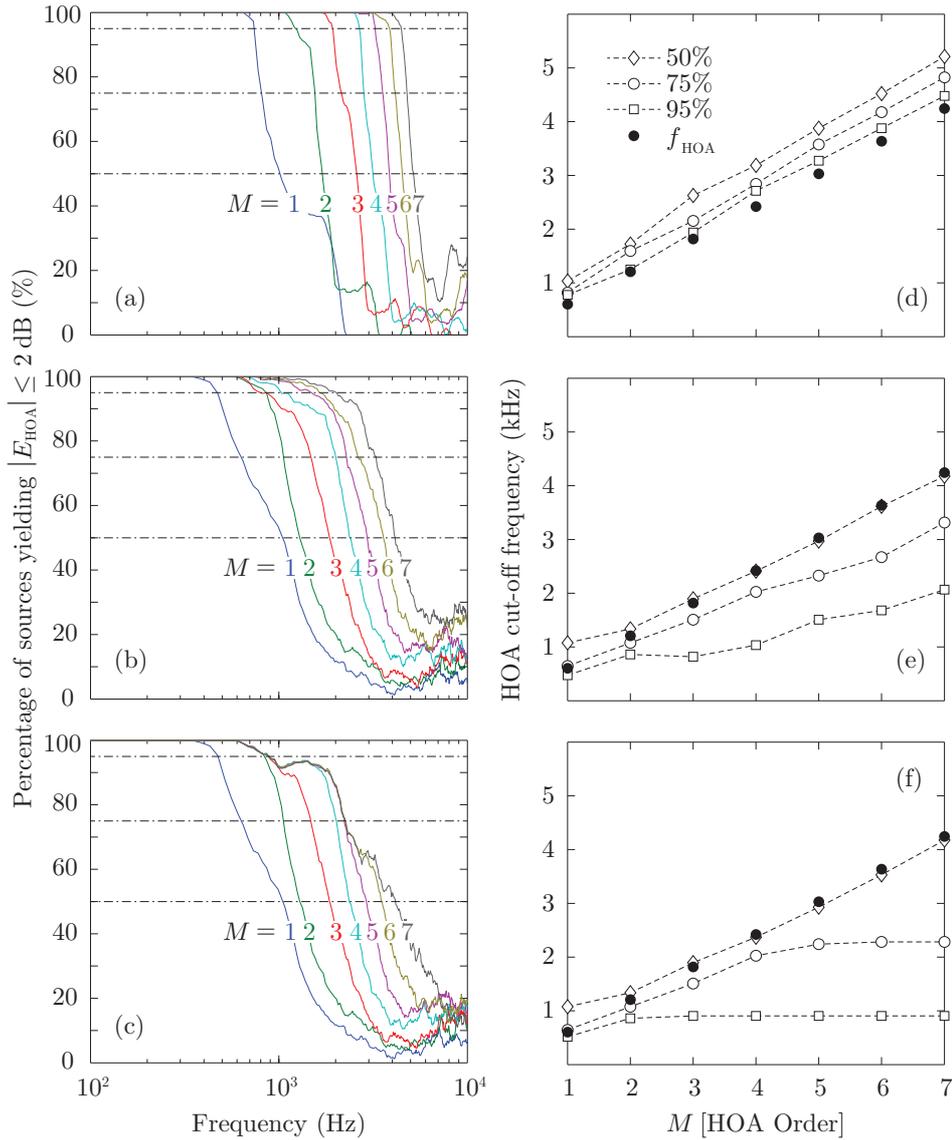


FIGURE 4.9: Frequency-dependent percentage of 400 anechoic sources that can be reconstructed with a pressure error of  $|\mathcal{E}_{\text{HOA}}(f)| \leq 2 \text{ dB}$ , using different order ( $M$ ) HOA systems. The errors were calculated at the left ear of: (a) a transparent head, Decoding-only case; (b) a HATS, Decoding-only case; (c) a HATS, Encoding and Decoding case. The horizontal dash-dotted lines show different error-tolerance criteria (50%, 75% or 95%) applied to calculate the HOA cut-off frequencies of the plots (d), (e) and (f) as a function of order  $M$ . The solid circles show the  $f_{\text{HOA}}$  rule-of-thumb value.

further pronounced when the HOA encoding process is additionally taken into account [Fig. 4.9(f)], showing that increasing the HOA order above  $M \geq 2$  has no practical effect on the 95% thresholds, while above  $M \geq 5$  no practical effect is either observed on the 75% thresholds. However, even when a HATS is considered inside the HOA sound field [Fig. 4.9(e)–(f)] the 50% threshold remains in good agreement with the rule-of-thumb  $f_{\text{HOA}}$  described in Section 4.2.2.

#### 4.4.2.2 Directional HA-microphone HOA errors

To evaluate the effect of HOA on the directional response of an example cardioid HA microphone worn on HATS, 3D ‘effective directivity’ patterns were derived by considering the HA microphone operating inside the sound field generated by the HOA loudspeaker array. These effective directivity patterns were calculated for the Reference case (Fig. 4.1, path 1) and the Encoding and Decoding case (Fig. 4.1, path 3), considering the HOA orders  $M = 3, 7$ . They were calculated as the directional HA microphone outputs relevant to all 1784 source directions provided by the employed HRTF database, then normalised by the microphone output maximum value.

Figure 4.10 shows the effective-directivity patterns in the horizontal plane, derived after averaging the (frequency-dependent) cardioid HA microphone output responses in 1/3-of-an-octave bands around the example centre frequencies  $f = 500$  Hz and  $f = 4000$  Hz. The minimum values were clipped at  $-30$  dB for improved readability. At  $f = 500$  Hz both HOA orders are sufficient to preserve the Reference (inherent) cardioid HA microphone directivity. At  $f = 4000$  Hz the 3<sup>rd</sup> order is well above its cut-off frequency, leading to heavily distorted effective-directivity patterns, while the 7<sup>th</sup> order just approaches that limit.

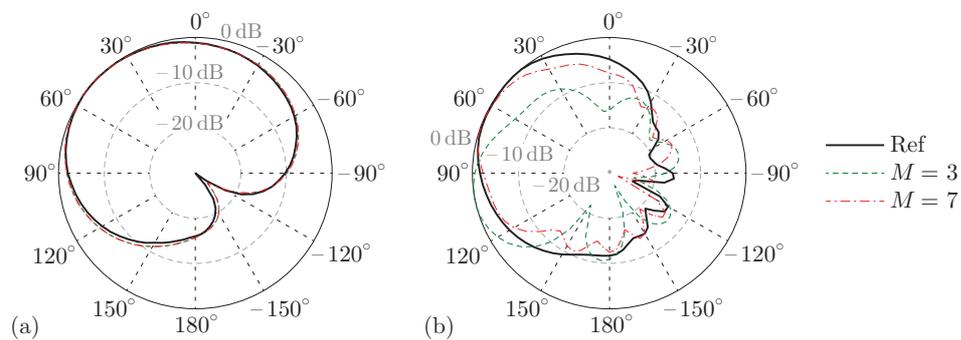


FIGURE 4.10: Effective-directivity plots of a cardioid HA microphone placed above HATS’ left ear and considered inside a Reference and a HOA Encoding and Decoding (orders  $M = 3, 7$ ) sound field at: (a)  $f = 500$  Hz and (b)  $f = 4000$  Hz.

Figure 4.11 shows the  $DI(f)$  relevant to the effective HA directivity patterns described above. It can be seen that HOA affects the  $DI(f)$  slightly at low frequencies (up to  $\approx 400$  Hz) but more significantly above a cut-off frequency of about 2 kHz for  $M = 3$  and about 5 kHz for  $M = 7$ . The high frequency errors are due to HOA introducing side-lobes and slightly rotating the maximum-directivity direction of the considered cardioid patterns as seen in Fig. 4.10(b).

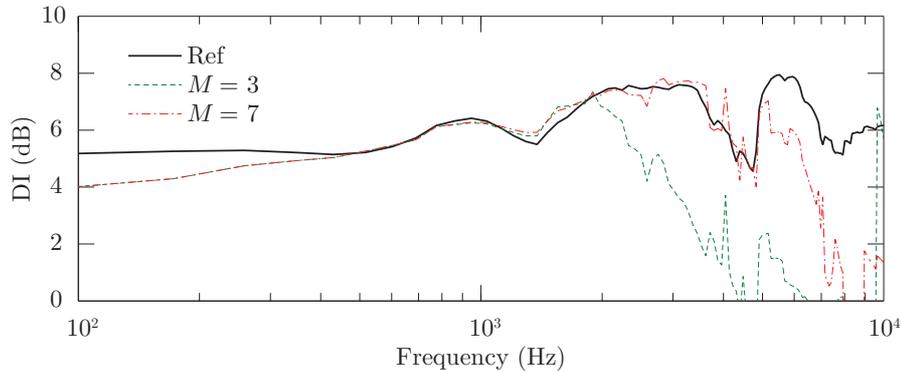


FIGURE 4.11: Directivity index (DI) relevant to the effective directivity of a cardioid HA microphone considered inside a Reference and a HOA Encoding and Decoding (orders  $M = 3, 7$ ) sound field. For reference, the DI of an ideal cardioid microphone in *free-space* (i.e., not placed on the head) is  $DI = 4.8$  dB (for frequencies where  $d_{\text{mic}} \ll c/f$ ).

### 4.4.3 Reverberant virtual source reconstruction

#### 4.4.3.1 HOA pressure errors

In the previous section the accuracy of the HOA sound field reproduction was evaluated for anechoic sound sources. Even though literature mainly considers this case, it does not necessarily reflect the behaviour inside realistic environments where reverberation is present. Hence, a similar HOA evaluation as described in Section 4.4.2 is herein presented, but this time with the sound sources considered inside a simulated room (see Section 4.3.1.1).

Applying the methods described in Section 4.3.3, the pressure at the left ear of the HATS, placed inside the HOA loudspeaker array, was calculated for the sound sources Src. 1 ( $\theta_{\text{src}} = 0^\circ$ ) and Src. 2 ( $\theta_{\text{src}} = -60^\circ$ ) inside the considered room of Fig. 4.2. The resulting pressure spectra  $H_{\text{Ref}}(f)$  for the reference case (Fig. 4.1, path 1) are shown in Figs. 4.12(a) and (b). Similar to Fig. 4.4, the corresponding errors  $\mathcal{E}_{\text{HOA}}(f)$  introduced by the HOA Decoding (Fig. 4.1, path 2) as well as the Encoding and Decoding (Fig. 4.1, path 3) are shown in Figs. 4.12(c) and (d) for an employed HOA order of  $M = 3$  and  $M = 7$ . By comparing to the anechoic case (Fig. 4.4), it is illustrated that the long-term pressure errors are significantly decreased when room reverberation is included.

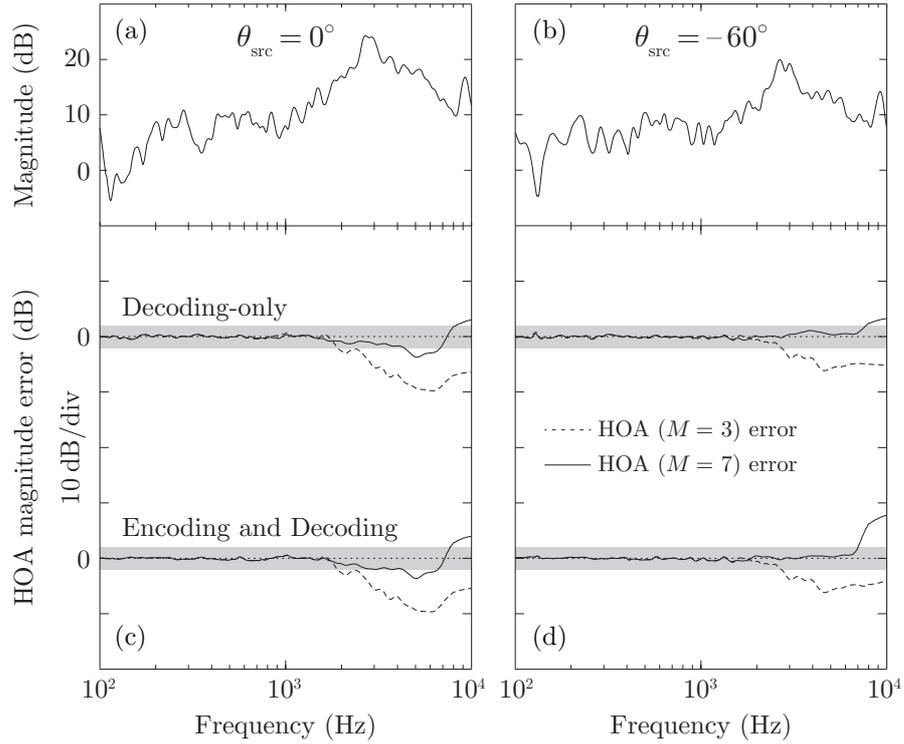


FIGURE 4.12: Pressure errors at HATS' left ear of an  $M = 3$  and  $M = 7$  HOA system reconstructing single reverberant-space sources placed on the horizontal plane, at  $\theta_{\text{src}} = 0^\circ$  (left column) and  $\theta_{\text{src}} = -60^\circ$  (right column) and at a distance of  $r_{\text{src}} = 1$  m referenced to the in-room receiver. Plots (a) and (b) show the reference pressure spectra  $H_{\text{Ref}}(f)$ . Plots (c) and (d) show the corresponding HOA pressure-errors  $\mathcal{E}_{\text{HOA}}(f)$ . The shaded regions depict the  $\pm 2$  dB error boundaries.

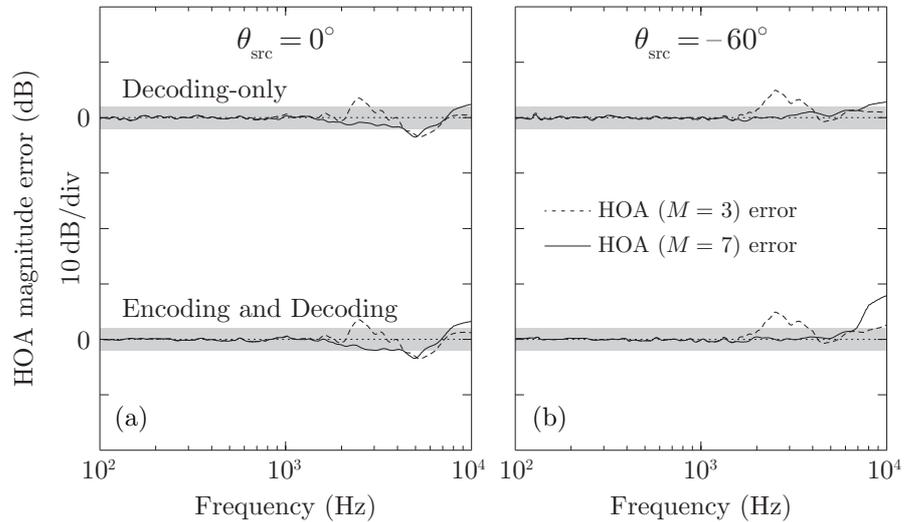


FIGURE 4.13: Same as Fig. 4.12 but using the  $\text{max-}r_{\text{E}}$  decoding activated at the HOA-order-dependent crossover frequency:  $f_{\text{x}} = M \cdot 800$  (Hz). The reference response  $H_{\text{Ref}}(f)$  is given in Figs. 4.12(a) and (b), respectively.

Whereas in Fig. 4.12 the basic HOA decoding scheme is applied, Fig. 4.13 shows the HOA pressure errors for the case that the  $\text{max-}r_E$  decoding is applied at high frequencies. Similar to the anechoic case (Fig. 4.5), the  $\text{max-}r_E$  decoding results in a significant reduction of the HOA errors at high frequencies and thereby increases the usable frequency range.

#### 4.4.3.2 Directional HA-microphone HOA errors

In Section 4.4.2.2 where anechoic sound sources were considered, the effect of HOA on the output of a directional HA microphone was evaluated by considering its effective directivity pattern as well as the relevant  $\text{DI}(f)$ . In the reverberant case, such measures cannot be applied and therefore the SNR and SNR benefit (Section 4.3.4) metrics were considered instead. Figures 4.14(a) and (b) present the frequency-dependent SNR in the reference condition (path 1) for an omnidirectional HA microphone  $\text{SNR}_o(f)$  at the left and right ear of the HATS (placed inside the reverberant scenario shown in Fig. 4.2 and using Src. 1 and Src. 3). Figures 4.14(c) and (d) show the corresponding SNR error incurred by a 3<sup>rd</sup> and 7<sup>th</sup> order HOA Encoding and Decoding (path 3) HOA system.

Moreover, Figs. 4.15(a) and (b) show the reference SNR benefit  $\Delta\text{SNR}(f)$  at the left and right ear of a HATS, achieved when the omnidirectional HA microphone is replaced by a cardioid HA microphone. Figures 4.15(c) and (d) show the corresponding error on the SNR benefit that is introduced by the HOA Encoding and Decoding process (path 3). Gray shaded areas indicate  $\pm 2$  dB error boundaries.

In contrast to the anechoic case shown in Figs. 4.10 and 4.11, where the HOA processing significantly affects the output of a directional HA microphone at high frequencies, in the reverberant case both the SNR measured with an omnidirectional HA microphone and the SNR benefit provided by a cardioid HA microphone are only slightly affected (mostly  $< 1$  dB errors and only few sub-octave wide notches of  $\sim 2$  dB) by the HOA processing (at least for  $M = 7$ ). Hence, the diffuse reflections seem to mitigate to some extent the sound field errors introduced by HOA at high frequencies.

So far, the SNR has been reported in a frequency-dependent manner. Applying an intelligibility-motivated weighting, as described in Section 4.3.4, the broadband SNR at the output of an omnidirectional HA microphone ( $\text{SNR}_o^I$ ) and the SNR benefit achieved by a cardioid HA microphone ( $\Delta\text{SNR}^I$ ), as well as the corresponding errors introduced by the HOA Encoding and Decoding (path 3) for  $M = 3$  and  $M = 7$  are summarised in Table 4.1. The results highlight that in a reverberant environment even a low order ( $M = 3$ ) HOA system could probably be successfully used to evaluate the effect of a directional HA microphone.

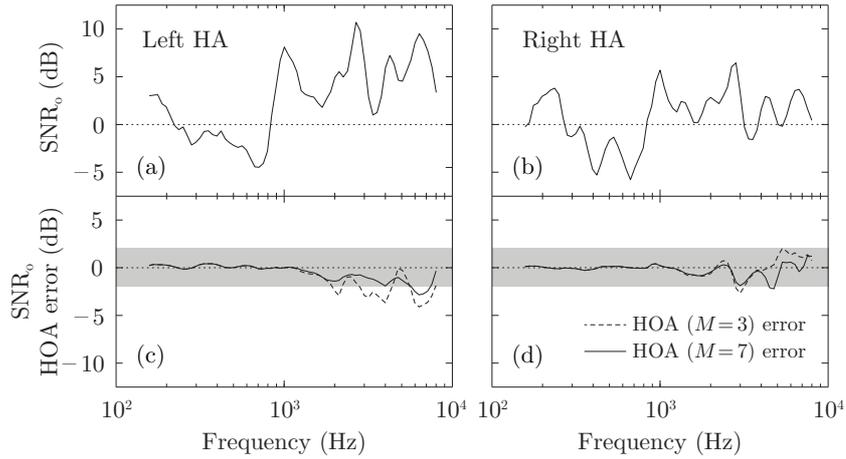


FIGURE 4.14:  $\text{SNR}_o(f)$  for an omnidirectional microphone at the left [(a) and (c)] and right [(b) and (d)] HAs for the reverberant scenario shown in Fig. 4.2, with Src. 1 as target and Src. 3 as distractor. The SNR for the Reference system (path 1) is shown in plots (a) and (b), while the SNR errors due to HOA system ( $M = 3, 7$ ; path 3) are shown in plots (c) and (d). The grey-shaded area indicate the  $\pm 2$  dB error boundaries.

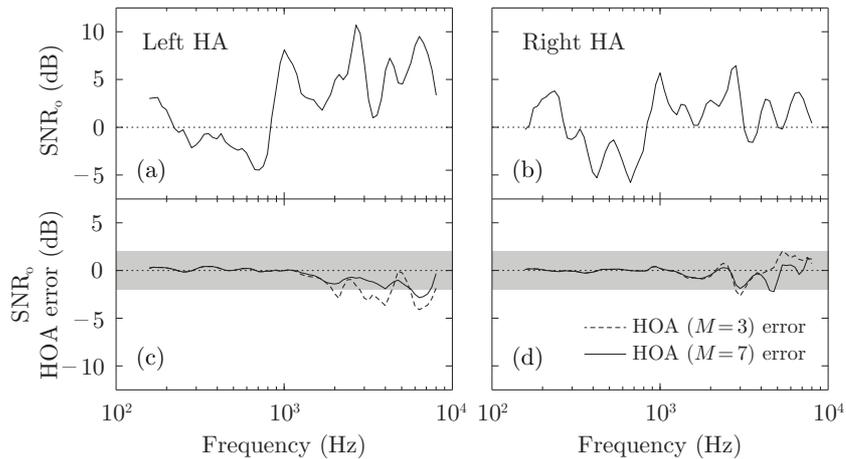


FIGURE 4.15: SNR-benefit,  $\Delta\text{SNR}(f)$ , provided by a cardioid HA microphone and the corresponding HOA error, derived for the same acoustic scene and HOA system as used in Fig. 4.14. The grey-shaded area indicate the  $\pm 2$  dB error boundaries.

TABLE 4.1: Intelligibility-weighted omnidirectional HA microphone SNR ( $\text{SNR}_o^I$ ) and SNR-benefit ( $\Delta\text{SNR}^I$ ) due to a cardioid HA microphone, in a Reference and HOA-reconstructed (path 3) reverberant environment.

side	$\text{SNR}_o^I$ (dB)			$\Delta\text{SNR}^I$ (dB)		
	Ref	HOA 3 error	HOA 7 error	Ref	HOA 3 error	HOA 7 error
Left	3.0	-1.1	-0.7	2.5	-0.3	-0.3
Right	0.7	-0.1	-0.3	5.2	-1.0	0.0

## 4.5 Discussion

### 4.5.1 HOA system noise

The analysis of Section 4.4.1 illustrates that the HOA system noise strongly depends on the estimation position inside the loudspeaker array. For a reasonable regularization value of  $\lambda = 0.01$ , Fig. 4.3(a) shows that, when no listener is present, the integrated PSD of the total HOA system noise at the centre of the loudspeaker array (dash-dotted line) is well below the EIN of the single sensors of the microphone array (solid line). This is due to different noise components being correlated and out-of-phase across loudspeakers and, thus, cancelling out in the centre of the array. Considering an off-centre position, i.e., the ear position of a transparent head, reduces this cancellation effect and significantly increases the corresponding PSD (dashed line). Inspecting Eq. (4.15) can give a further understanding of the HOA system noise. The  $[L \times Q]$  matrix  $\tilde{\mathbf{T}}$  linearly combines the uncorrelated noisy outputs of  $Q = 64$  sensors on the microphone array, to create  $L = 100$  spatially-separated signals driving the loudspeaker array. Hence, the loudspeakers end up emitting partially correlated noise signals.

Considering the integrated noise at the ears of a listener (HATS) inside the loudspeaker array, results in an even larger total HOA noise PSD, as shown in the three top curves of Fig. 4.3(b). Besides the reduced cancellation of correlated noise components, this increase can be explained by acoustic diffraction and resonance effects of the body, head, and pinnae. Interestingly, for the chosen microphone and loudspeaker array topologies, the PSD of the HOA system noise has a maximum at  $\approx 3$  kHz. This is the combined effect of the ear-canal resonance (Blauert, 1997) and the specifics of the HOA microphone equalization filters  $\tilde{H}_{\text{eq},m}(kR)$  [see Eq. (4.14)]. Unfortunately, this peaky behaviour occurs at a region where the normal auditory system is most sensitive. The comparison to the MAP in Fig. 4.3(b) also shows that for all regularization levels considered in this study, the total HOA system noise is clearly audible to a normal-hearing listener.

However, the audibility of the HOA system noise will also depend on the level and complexity (or type) of the recorded acoustic scene. Due to auditory masking effects the HOA system noise may be inaudible, or at least may not significantly disturb the listening experience. Hence, future research should look into optimizing the regularization in the HOA encoding process by considering the individual sensor EIN and the HOA microphone and loudspeaker array configuration in conjunction with the specifics of the scene to be recorded. In that direction, acoustic scene-dependent regularization, or even time-dependent regularization controlled by a short-time signal power analysis could be considered. Moreover, future research needs to separately examine the effect of the HOA

system noise on various signal enhancement algorithms implemented on hearing devices which operate inside the HOA sound field.

## 4.5.2 HOA errors for anechoic source reconstruction

### 4.5.2.1 HOA pressure errors

Section 4.4.2.1 evaluated the HOA pressure errors in a loudspeaker array with and without a listener being present. The case omitting the listener was mainly introduced as a point of reference. The good agreement between the results of this study and the results presented in the relevant literature (Ward and Abhayapala, 2001; Daniel, 2001; Poletti, 2005; Moreau *et al.*, 2006; Epain *et al.*, 2010) confirmed the general applicability of the described HOA simulation and evaluation methods. For example, the previously reported cut-off frequency  $f_{\text{HOA}}$  of Eq. (4.5), above which the HOA pressure errors drastically increase, was herein confirmed [see Fig. 4.9(a) and (d)]. Similarly, this study confirmed the decrease of HOA pressure errors at high frequencies due to the max- $r_E$  decoding, as previously reported (Daniel *et al.*, 1998; Daniel, 2001; Murillo *et al.*, 2014).

The systematic evaluation of the case when a listener (HATS) is present inside the loudspeaker array provides one of the main contributions of this study. In this case, the aforementioned cut-off frequency  $f_{\text{HOA}}$  does not apply any more and needs to be revised. As illustrated in Figs. 4.6 and 4.7, the HOA pressure errors at the HATS in-ear microphone are highly source-direction dependent. Applying a threshold criterion such that the HOA pressure errors are restricted to  $|\mathcal{E}_{\text{HOA}}(f)| \leq 2$  dB, this cut-off frequency is almost unaffected for frontal sources, but highly reduced for sources that are contralateral to the considered ear. This holds true regardless of whether the processing chain includes the HOA encoding microphone or not. The reduced accuracy in contralateral source reproduction can be explained by the higher spatial-order nature of the head and torso, compared to a source-free region. This higher spatial-order necessitates, in turn, higher sound-field orders for accurate reproduction.

This observation is important when taking into account binaural hearing in complex acoustic environments. The importance can be best illustrated by considering the simple scenario of a listener attending to a frontal target talker, in the presence of a lateral distractor. Due to head shadow effects, the ear that is contralateral to the distractor provides a better SNR. The same consideration applies in more complex scenes, where auditory better-ear glimpsing mechanisms take advantage of short-term SNR differences between the two ears (Glyde *et al.*, 2013). Hence, any HOA reconstruction error that affects the contralateral ear, potentially affects the ear with the better (short-term) SNR and thus, might reduce the ability of a listener to understand speech at low SNR values.

The described reduction of the usable HOA bandwidth when a listener is considered inside the loudspeaker array is further aggravated when a HOA microphone array is considered in the sound-field reconstruction chain. This is particularly true for the reproduction of sources that lie contralateral to the considered ear. Whereas for a regularization value of  $\lambda = 0.01$  the microphone array has no noticeable effect on the HOA pressure errors for frontal sources, a much smaller value of  $\lambda = 0.0001$  is required to eliminate errors induced by the microphone array for sources lying around  $\theta_{\text{src}} = -60^\circ$  (right side). Since lower regularization activates higher-order HOA components at lower frequencies, it can be deduced that those higher-order HOA components are required to correctly reproduce contralateral sources, even at relatively low frequencies. This aspect is also the reason why in Figs. 4.9(b) and 4.9(c) even a 7<sup>th</sup> order HOA system provides only a very narrow frequency range where 100% of the 400 considered sources are correctly reproduced. On the contrary, in the case when no listener is considered inside the loudspeaker array, the HOA pressure error is much less affected by the microphone array and a regularization value of  $\lambda = 0.01$  seems to be sufficient for all source directions.

#### 4.5.2.2 Directional HA-microphone HOA errors

Figures 4.10(a) and 4.11 show that the effective directivity of a directional HA is not affected below the HOA cut-off frequency described in Section 4.4.2.1. However, above this cut-off frequency, HOA rotates the direction of the main lobe maximum and introduces erroneous side lobes, as illustrated in Fig. 4.10(b). Both phenomena affect the DI (Fig. 4.11) relevant to the considered HA microphone effective directionality.

Moreover, Fig. 4.11 illustrates that the HOA sound field consistently underestimates the DI at low frequencies. This error does not appear in the HOA Decoding-only case (results not shown here), and, hence, can be purely attributed to HOA encoding over-regularization occurring at low frequencies. That over-regularization does not cause significant pressure errors at the ear or single HA microphone (Section 4.4.2). However, it affects the output of a directional HA microphone where the small pressure errors of the single microphones are amplified by up to 30 dB due to the inherent differentiation operation (Kates, 2008, Figs. 4-7).

### 4.5.3 HOA errors for reverberant source reconstruction

#### 4.5.3.1 HOA pressure errors

Comparing Fig. 4.12 with Fig. 4.4 illustrates that the reproduction of a reverberant sound source gives rise to lower HOA pressure errors than the reproduction of an anechoic sound

source and thereby decreases the required order  $M$  of the HOA reproduction system. That advantage is pronounced at the contralateral ear for lateral sources, that is, the condition where, in the anechoic case, the HOA sound field reproduction appeared most problematic (Section 4.5.2.1). This is due to the head shadow effect that reduces the directional energy of the direct sound at the contralateral ear but not the diffuse energy of the reverberation and thus, reduces the direct-to-reverberant energy ratio.

Hence, it can be concluded that even though the reproduction of the spectral, temporal, and spatial details of the individual reverberant components (or reflections) is limited in the same way as discussed for an anechoic sound source, the individual errors seem, in average, to partially cancel out each other. This consideration may be compared to moving the source-receiver pair (see Fig. 4.2) to different locations inside the virtual room. Even small changes in location drastically change the spatial, temporal, and spatial details in the room's response, but won't significantly change the overall behaviour. Listeners inside the room are typically not even aware of these acoustic changes. Nevertheless, a further analysis using listening tests that involve directional HA microphones is required.

In order to ensure that the above observations were not merely an artefact of the chosen scenario, the source-receiver configuration shown in Fig. 4.2 was re-simulated in different locations inside the virtual room. Although the detailed behaviour of the HOA pressure errors differed from location to location, the above findings were confirmed. Moreover, this analysis revealed that the HOA pressure errors were well-behaved up to a frequency of at least 10 kHz, even for the 3<sup>rd</sup> order system.

It should be noted that although the amount of reverberation provided by the considered acoustic scenario was rather low, it provided a significant amount of reduction of the HOA pressure errors. The simulated acoustic space was a meeting room with a reverberation time of  $T_{30} = 0.3$  s and a critical distance of  $r_c \approx 0.85$  m. Thus, the distance between the sources and the receiver (1 m) was only slightly larger than the critical distance. To further investigate the effect of reverberation on the accuracy of HOA sound field reproduction, a larger and more reverberant room was simulated and a similar analysis as described above was carried out. The acoustic space was a medium-sized lecture room ( $6.7 \times 9.5 \times 3.0$  m) with a reverberation time of  $T_{30} = 0.6$  s and a critical distance of  $r_c \approx 1.4$  m. Although details are not reported here, for a source-receiver configuration as shown in Fig. 4.2, the reduction in HOA pressure errors provided by this larger room was slightly stronger than for the meeting room discussed before. Moreover, the case of an increased source-receiver distance of 2 m was considered for a frontal sound source. As expected, the increased amount of reverberation resulted in an even reduced HOA pressure error.

The spatial smoothing induced by the reverberation effectively increases the usable frequency range of a HOA system. The latter can be even further increased by replacing the basic HOA decoding with the  $\text{max-}r_E$  decoding at high frequencies (Fig. 4.13). In this case a 7<sup>th</sup> order HOA system that employs the example microphone and loudspeaker array described in Sections 4.3.2.1 and 4.3.3.2 can achieve a usable frequency range of more than 7 kHz, which is significantly higher than the  $\sim 4$  kHz cut-off frequency found in the anechoic case (Fig. 4.5).

#### 4.5.3.2 Directional HA-microphone HOA errors

Figures 4.14 and 4.15 illustrate the effect of HOA sound field reproduction on the SNR at the output of a single (frontal) HA microphone as well as the SNR benefit at the output of a directional HA microphone for the reverberant scenario shown in Fig. 4.2. The incurred errors appear well controlled for most of the considered frequency range of up to 10 kHz. This is even true for low frequencies, where the (anechoic) directivity pattern was distorted by the regularization applied in the HOA encoding process (Section 4.5.2.2).

Even though an asymmetric scene is considered here with a frontal target and a lateral distractor source, the HOA-incurred errors of both the SNR and SNR benefit are very similar for the left and right HA. Since the SNR is derived by dividing the energy of the reverberant target and distractor sources, the HOA errors of the two sources are combined in a complicated manner. Hence, no clear conclusions can be drawn on whether the HOA errors are lower at the ear ipsilateral or contralateral to the distractor source.

Finally, the intelligibility-weighted SNR and SNR benefit for the discussed reverberant scenario were consistently underestimated, albeit by no more than 1.1 dB even for the 3<sup>rd</sup> order HOA system. It remains unclear, however, if errors of that magnitude are perceptually important. Hence, listening tests using either a real HOA system or simulated HOA sound fields presented via headphones need to be conducted to quantify such errors in terms of their effect on speech understanding. This is particularly important, given that the intelligibility-weighted SNR benefit does not take into account the temporal and spatial variations of the reverberant masker.

## 4.6 Conclusions

As widely discussed in the relevant literature, HOA can only successfully reproduce a given sound field inside a limited area (the sweet spot) of the playback loudspeaker array

and for a limited frequency range. In the free-field case, the size of this sweet spot as well as the upper cut-off frequency are mainly determined by the employed HOA order. However, this study additionally showed that when a listener is considered inside the loudspeaker array, which implies that the sweet spot is at least the size of the head, the usable frequency range is also dependent on the direction of the reconstructed sources. This is particularly prominent for anechoic sources lying contralateral to the considered ear, where the usable frequency range is considerably reduced due to head shadow effects.

It was also shown that the accuracy of the HOA sound reproduction is considerably improved when reverberant sound fields are considered. Diffuse reverberation effectively reduces the required spatial resolution of the sound reproduction system and thus, for a fixed HOA order, increases the usable frequency range. These considerations even hold true when considering the SNR and SNR benefit at the output of a directional HA microphone placed on a listener's head. At least this is the case when long-term (or temporally-integrated) performance measures are applied, as it is usually done in HA research and development. Temporal details such as transients or echo-free onsets are still limited by HOA in a similar manner as anechoic signals. Thus, it needs to be further evaluated with appropriate listening tests involving HAs if/when these temporal details are objectively or perceptually relevant when listening in reconstructed realistic environments.

Following the detailed analysis provided in this study, a 7<sup>th</sup> order HOA system appears to provide a usable frequency range of about 6 – 7 kHz to a listener when realistic (reverberant) environments are considered. Since the considered frequency range for most HA applications is about 5 – 6 kHz, it is suggested that a 7<sup>th</sup> order HOA system should be suitable for testing HAs. However, this HOA order requires at least 64 loudspeakers and as many microphones, which should optimally be positioned in a quasi-regular manner on the surface of a sphere. Such large arrays are expensive and difficult to realise and thus, are often not available.

A promising approach to limit the complexity of the required loudspeaker array could involve the variant of ‘mixed-order Ambisonics’ (MOA: Favrot *et al.*, 2011; Marschall *et al.*, 2012; Favrot and Marschall, 2012). MOA provides an increased spatial resolution in the horizontal plane by combining a given 3D HOA system with a higher-order 2D system. This is achieved by employing loudspeaker and microphone arrays with an increased density of sensors on the equator (i.e., in the horizontal plane). The approach takes advantage of the fact that most real-life sound sources are located in the horizontal plane, while the remaining off-horizontal plane components are primarily related to the early reflections and the (diffuse) reverberation. The previous rationale is moreover

supported by the fact that the spatial resolution of the auditory system is much higher in the horizontal plane than in the vertical plane.

Yet another strategy that could be employed to increase the effective spatial resolution of a HOA system involves the theory of compressed sensing ([Wabnitz \*et al.\*, 2011](#); [Noohi \*et al.\*, 2013](#)). However, this approach relies on the spectral, temporal, and spatial ‘sparsity’ of the acoustic scene, and may not be easily applied to very complex and reverberant scenes.

It should be mentioned that even though a number of practical limitations were taken into account in the described HOA system analysis (e.g., the realistic noise inherent in the microphone array), a number of other practical limitations were not considered. For instance, the employed microphone and loudspeaker arrays involved a considerable number of sensors/sources that were quasi-regularly positioned on the surface of a sphere. Reducing the number of the microphones and the loudspeakers can have a significant impact on the overall system performance. Moreover, mismatches in the sensitivity/phase and positioning errors on the microphone array side, as well as a non-reflection-free environment on the loudspeakers playback-room side will have adverse effects on the system’s accuracy and thus, need to be carefully compensated.

A final remark that should be made is that the present study quantifies the HOA errors as deviations from the ideal real-life (albeit simulated) sound field. That may be an overly sensitive or even misleading performance measure. Since listening tests that are currently performed in the laboratory employ rather simple and artificial scenes, the increase in realism, complexity and flexibility that can be achieved even with a low-order HOA system may well outweigh the introduced sound field errors. This consideration again highlights the need for further behavioural studies that evaluate the subjective importance of HOA errors in an applied context.

## Acknowledgements

The authors acknowledge the financial support of the HEARing CRC, established and supported under the Cooperative Research Centres Program – an initiative of the Australian Government.

# Chapter 5

## Effect of higher-order Ambisonics on evaluating beamformer benefit in realistic acoustic environments\*

### Abstract

Multi-channel loudspeaker systems have been proposed to assess the real-life benefit of devices such as hearing aids, cochlear implants, or mobile phones. This paper investigates to what extent sound fields recreated by Higher-Order Ambisonics (HOA) can be used to evaluate the performance of spatially selective multi-microphone processing schemes (beamformers) inside complex acoustic environments. Two example schemes are considered: an adaptive directional microphone (ADM) and a contralateral suppression bilateral beamformer (BBF), both implemented in the context of a hearing aid device. The acoustic scenarios consist of a single speech target ( $0^\circ$ ) competing against three speech distractors ( $\pm 90^\circ$  and  $180^\circ$ ) set either in an anechoic or in a reverberant simulated classroom ( $T_{30} = 0.6$  s). The HOA effect on the directional algorithm performance is quantified through: (a) the adaptive, frequency-dependent, algorithm gains, (b) the SNR improvement calculated in one-third octave bands, and (c) the processed target frequency response.

The HOA reconstruction errors influence the beamformers in mainly two ways; first, by altering the spatial characteristics of the sound field, which in turn modifies the adaptation of the algorithms, and second, by affecting the spectral content of the sources. The results suggest that although HOA (here 7<sup>th</sup> order) does not degrade the broadband, long-term, intelligibility-weighted SNR improvement of the two beamformers, it imposes a low-pass effect on the processed target. This renders the HOA coding problematic above the system's cut-off frequency.

---

\*Chapter published as:

Oreinos, C., Buchholz, J. M. and Mejia, J. (2013), "Effect of Higher-Order Ambisonics on Evaluating Beamformer Benefit in Realistic Acoustic Environments", in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA* (New Paltz).

## 5.1 Introduction

As microphone array processing in hearing aids (HAs) becomes more sophisticated, a growing need arises to correctly assess the end-user's benefit in real-life situations. Current listening tests aimed at evaluating the directional benefit of HAs are mainly conducted in highly controlled but unrealistic environments. A conventional testing setup, for example, uses two to eight loudspeakers in order to represent the target and distractors, placed in an acoustically treated laboratory room. However, such evaluations poorly predict the real-life benefit (Cord *et al.*, 2004). On the other hand, field studies are highly uncontrollable, unrepeatable and impractical for large-scale testing of algorithms, especially at their early stage of development.

Virtual sound environments (VSE) created by an array of loudspeakers could potentially be used to allow ecologically valid testing of signal processing schemes. Higher-Order Ambisonics (HOA) (Ward and Abhayapala, 2001; Poletti, 2005; Moreau *et al.*, 2006; Daniel, 2001) is one of the most popular methods for creating VSEs. The physical and perceptual limitations of HOA are well known but its use to test HA multi-microphone signal enhancement schemes (i.e., beamformers) has not been studied.

This study evaluates the extent to which HOA-reconstructed sound fields can be used to successfully evaluate beamformers, herein implemented in the context of a HA device. The whole sound path is considered, including scene generation, reproduction of virtual sources, capturing by HA microphones and beamformer processing. The simulation framework utilises different distortion metrics to compare the beamformer benefit in a simulated reference and a HOA-decoded sound reproduction system.

## 5.2 Methods

The first stage of the simulation framework (Fig. 5.1) models the acoustic scene inside which the beamformers operate. Here, it consisted of a target directly in front of a dummy listener and three distractors at  $\pm 90^\circ$  and  $180^\circ$ , all on the horizontal plane. In the anechoic scenario (path 1), the four talkers were modelled as free-space plane wave sources. In the reverberant scenario (path 2), a classroom (reverberation time:  $T_{30} = 0.6$  s; dimensions:  $6.7\text{ m} \times 9.5\text{ m} \times 3\text{ m}$ ) was modelled using ODEON. The target was placed at 1 m and the distractors at 2 m distance from the listener (room's critical distance was 1.4 m). All distractors were normalised to the same RMS level. After the ODEON simulation, the LoRA (Loudspeaker-based Room Auralisation) framework (Favrot and Buchholz, 2010b) was used to map the direct sound (DS) and specular

early reflections (ER) of all talkers to a dense, 1784-node spherical array. The late reverberation was recreated using a virtual 196-node quasi-regular (Fliege, 2007) virtual loudspeaker array according to the method described in (Favrot and Buchholz, 2010b).

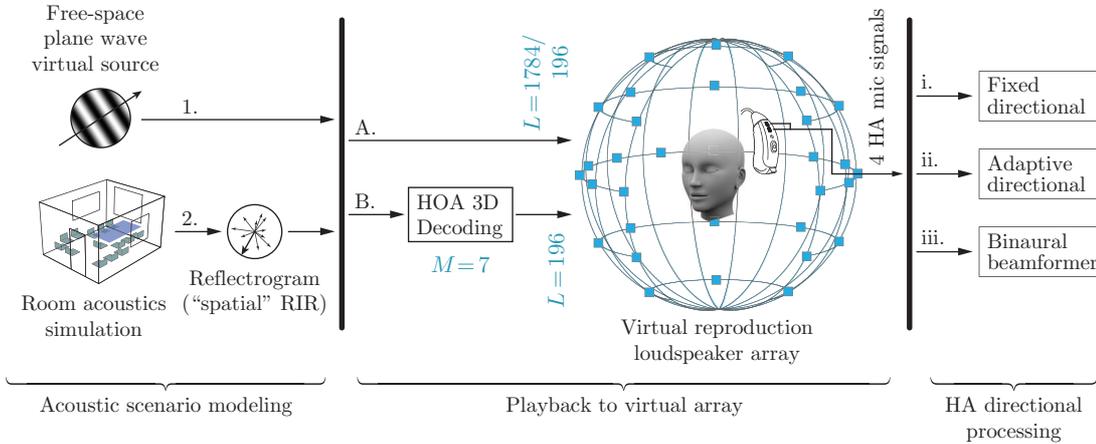


FIGURE 5.1: The simulation framework including the: (1) anechoic, and (2) in-room acoustic scenario modelling, the playback to a virtual (simulated) spherical loudspeaker array using the: (A) Reference system –single loudspeaker per modelled wave–, and (B) 7<sup>th</sup>-order HOA system for the reconstruction of the HA microphone signals which are then fed to the three examined hearing aid directional algorithms (i – iii).

Two methods were then used to reproduce the virtual scene created at the first stage of the framework. In the reference ‘REF’ method (path A), the LoRA DS/ER output signals drove directly the 1784-node virtual loudspeaker array, and the LoRA late reverberation drove the 196-node virtual loudspeaker array. In the ‘HOA’ method (path B), an  $M = 7$  order 3D HOA system was employed to decode all DS/ER virtual sources to the 196-node quasi-regular virtual-loudspeaker array. The late reverberation was reproduced as in the REF method.

Since this study aims at looking into the effect of HOA, rather than the absolute performance of beamformers in real life, the simulated classroom scenario reproduced by the REF system can be considered as a good approximation to a real room.

### 5.2.1 Higher-order Ambisonics (HOA)

HOA is based on the decomposition of sound fields to spherical harmonics (Williams, 1999). Effectively, the sound field of a given virtual source placed at  $(\theta_s, \delta_s, r_s)$  is reproduced by feeding weighted versions of the source signal to a spherical array of  $L$  loudspeakers. Applying the plane-wave assumption, the optimum –in the least square sense– loudspeaker weights are given (Moreau *et al.*, 2006; Daniel, 2001) by solving the

equation below:

$$\sum_{l=1}^L s_l \sum_{m=0}^M \sum_{n=0}^m \sum_{\sigma=\pm 1} Y_{mn}^{\sigma}(\theta_l, \delta_l) = \sum_{m=0}^M \sum_{n=0}^m \sum_{\sigma=\pm 1} Y_{mn}^{\sigma}(\theta_s, \delta_s), \quad (5.1)$$

with  $Y_{mn}^{\sigma}(\theta, \delta)$  being the *real-valued* spherical harmonics defined in (Daniel, 2001) and  $M$  the HOA order. Expressing Eq. (5.1) in matrix form and solving the linear equations (Moreau *et al.*, 2006; Daniel, 2001) results in the *basic* HOA decoding scheme:

$$\begin{bmatrix} s_1 \\ \vdots \\ s_L \end{bmatrix} = \text{pinv} \left( \begin{bmatrix} Y_{00}^1(\theta_1, \delta_1) & \dots & Y_{00}^1(\theta_L, \delta_L) \\ \vdots & Y_{mn}^{\sigma}(\theta_l, \delta_l) & \vdots \\ Y_{MM}^{-1}(\theta_1, \delta_1) & \dots & Y_{MM}^{-1}(\theta_L, \delta_L) \end{bmatrix} \right) \begin{bmatrix} Y_{00}^1(\theta_s, \delta_s) \\ \vdots \\ Y_{MM}^{-1}(\theta_s, \delta_s) \end{bmatrix} \quad (5.2)$$

where  $\text{pinv}(\cdot)$  represents the Moore-Penrose pseudoinverse. For the applied quasi-regular loudspeaker setup, since  $(M+1)^2 < L$ , Eq. (5.2) represents an under-determined system that was found to be well-behaved (i.e., not ill-posed). The  $\text{pinv}$  operation was performed in MATLAB using a singular-value decomposition and only keeping singular values above a machine-specific threshold. No further regularization methods (as in Poletti, 2005) were necessary.

The reconstruction error of HOA increases with the frequency and distance from the centre of the array and decreases with the order  $M$  (Ward and Abhayapala, 2001). Additionally, the presence of a listener inside the array further increases the errors (Oreinos and Buchholz, 2013a). The order  $M = 7$  was chosen as an upper practical limit, requiring  $(M+1)^2 = 64$  loudspeakers in 3D- and  $(2M+1) = 15$  loudspeakers in 2D-HOA. However, a denser loudspeaker array setup was employed here so that the HOA reconstruction error is minimised in the  $kr < M$  zone (Solvang, 2008), albeit at the expense of possible spectral colorations above the  $kr = M$  threshold.

Above the cut-off frequency  $f_c$  of HOA, the  $\text{max-}r_E$  decoding (Daniel, 2001; Daniel *et al.*, 1998) was used. Here,  $f_c = M \cdot 600$  was applied as derived in (Oreinos and Buchholz, 2013a) for the case of a listener inside the reconstructed sound field. For a sweet-spot radius of  $r \approx 0.09$  m this agrees well to the rule of thumb derived in (Moreau *et al.*, 2006) for the free-field case, i.e.  $f_c = cM/2\pi r$ . Combining the basic and  $\text{max-}r_E$  gains at  $f_c$  (5<sup>th</sup> order Butterworth crossover filters) resulted in the frequency-dependent loudspeaker gains, i.e., filters,  $s_l(f)$ .

After the final mapping of the 10s-duration anechoic/in-room target and distractor signals to the virtual loudspeaker array via either of the REF/HOA methods, the signals in all four HA microphones were calculated by convolving the virtual loudspeaker signals with a set of impulse responses from all 1784 nodes to the microphones of HAs worn by

a HATS (previously measured by the authors; Oreinos and Buchholz, 2013c) and then adding all loudspeaker signals separately for each individual HA microphone. The target level was adjusted to give an initial intelligibility-weighted SNR (Greenberg *et al.*, 1993) (averaged across HA microphones) of  $-5$  dB.

### 5.2.2 Hearing aid beamforming

The HA microphone signals were fed at the last stage of the framework to three spatially-selective algorithms (beamformers), operating at sampling frequency  $f_s = 24$  kHz. The first one (path i) was a fixed (non-adaptive) cardioid microphone. It was only used to derive a baseline SNR improvement value (Table 5.1) and otherwise was not further analysed.

The second considered algorithm was a multi-channel adaptive directional microphone (Elko and Pong, 1995; Luo *et al.*, 2002), referred to as ADM (path ii). Its output is given by:

$$y_i(n) = c_{i_{\text{front}}}(n) - b(i, k) c_{i_{\text{back}}}(n), \quad (5.3)$$

where  $y_i(n)$  is the ADM output,  $c_{i_{\text{front}}}(n)$  and  $c_{i_{\text{back}}}(n)$  are the outputs of a front- and rear-facing cardioid microphone respectively and  $b(i, k)$  is the mixing gain. All of those quantities are relevant to the  $i$ -th filter-bank channel and  $k$ -th frame of the block processing (frame size: 24 samples). A 49-channel uniform DFT filter-bank was used for the multichannel analysis and synthesis.

The optimum mixing gains  $\underline{b}(i, k)$  that minimise the ADM output energy, under the constraint of an undistorted target (imposed by restricting  $\underline{b}(i, k)$  to  $[0, 1]$ ), were calculated as the ratio of the estimates of the cardioids' cross-power over the power of the rear-facing cardioid as follows:

$$\underline{b}(i, k) = \frac{R_{C_f C_b}(i, k)}{R_{C_b C_b}(i, k)} = \frac{C_{\text{front}}(f_i) C_{\text{back}}^*(f_i)}{|C_{\text{back}}(f_i)|^2}, \quad (5.4)$$

where  $C_{\text{front}}(f_i)$  and  $C_{\text{back}}(f_i)$  are the DFT outputs of the  $i$ -th channel at the  $k$ -th frame. The asterisk  $(\cdot)^*$  denotes the complex conjugate signal. In the current realization, the cross-power and power estimates  $R_{C_f C_b}(i, k)$ ,  $R_{C_b C_b}(i, k)$  were updated at every frame  $k$  using a running estimate (see Luo *et al.*, 2002) to smooth the temporal variations and thus minimise unwanted artefacts.

The third considered algorithm (path iii) was a recently proposed scheme referred to as the bilateral beamformer (BBF) (Mejia *et al.*, 2007; Mejia and Dillon, 2010). It operates by combining the signals (front-facing cardioid outputs in this realization) from both

sides of the head and then estimating their similarity based on the running average of a correlation-based metric. The calculated frame- and frequency-dependent gains for the left and right HAs,  $W_L(i, k)$  and  $W_R(i, k)$ , aim at preserving the on-axis target with minimum distortions, while suppressing all laterally-impinging sounds. In the present implementation, the same 49-channel uniform DFT filter-bank was used as in the ADM case.

Both ADM and BBF algorithms were implemented so that they can also operate in ‘freeze’ mode; the total target plus distractors signal were fed to the adaptive schemes and the gains  $\underline{b}(i, k)$  and  $W_{L/R}(i, k)$  were stored per frame and frequency channel. Then the target and distractors were used in succession as inputs to the algorithms in freeze mode utilizing the stored gain values. In that way the processed target and distractor signals were available separately at the output of the simulation framework.

### 5.3 Performance Analysis

The effect of the HOA sound field reproduction on the ADM and BBF beamformer processing and overall performance was evaluated using different metrics. The differences in these metrics derived in the HOA sound field and in the reference (REF) sound field quantify the errors introduced by the HOA coding.

Fig. 5.2 shows example results for a number of short-term metrics derived for the case of the ADM beamformer when applied to the simulated classroom scenario. Fig. 5.2(a) shows the ADM mixing gains  $\underline{b}(i, k)$  for the REF and Fig. 5.2(b) for the HOA system in a dB scale, clipped below  $-15$  dB. As described in Section 5.2.2 the gains are provided in 49 frequency channels with a temporal resolution of 24 samples (i.e., 10 ms). Up to at least 4 kHz the gains look very similar between the two playback systems while above that frequency, HOA seems to cause a general increase of the mixing gains providing a stronger weighting on the forward facing cardioid microphone of the ADM.

The error introduced by the HOA coding on the absolute value of the short-term (segmental; Hansen and Pellom, 1998) SNR improvement [Fig. 5.2(c)] was calculated using a sliding (50% overlap) 30 ms Hann window and 18 one-third octave frequency bands centred from 156 Hz to 8 kHz (ANSI S3.5-1969). The segmental SNR improvement is correctly estimated for the ADM beamformer up to about 3 kHz.

Fig. 5.2(d) shows the error introduced to the target short-term spectrum (30 ms Hann window, 50% overlap, 1024-point DFT) at the output of the ADM beamformer, which is largely unaffected (error less than 2 dB) by the HOA system up to about 4 – 5 kHz. Above this frequency, the processed target appears low-pass filtered as commonly seen in

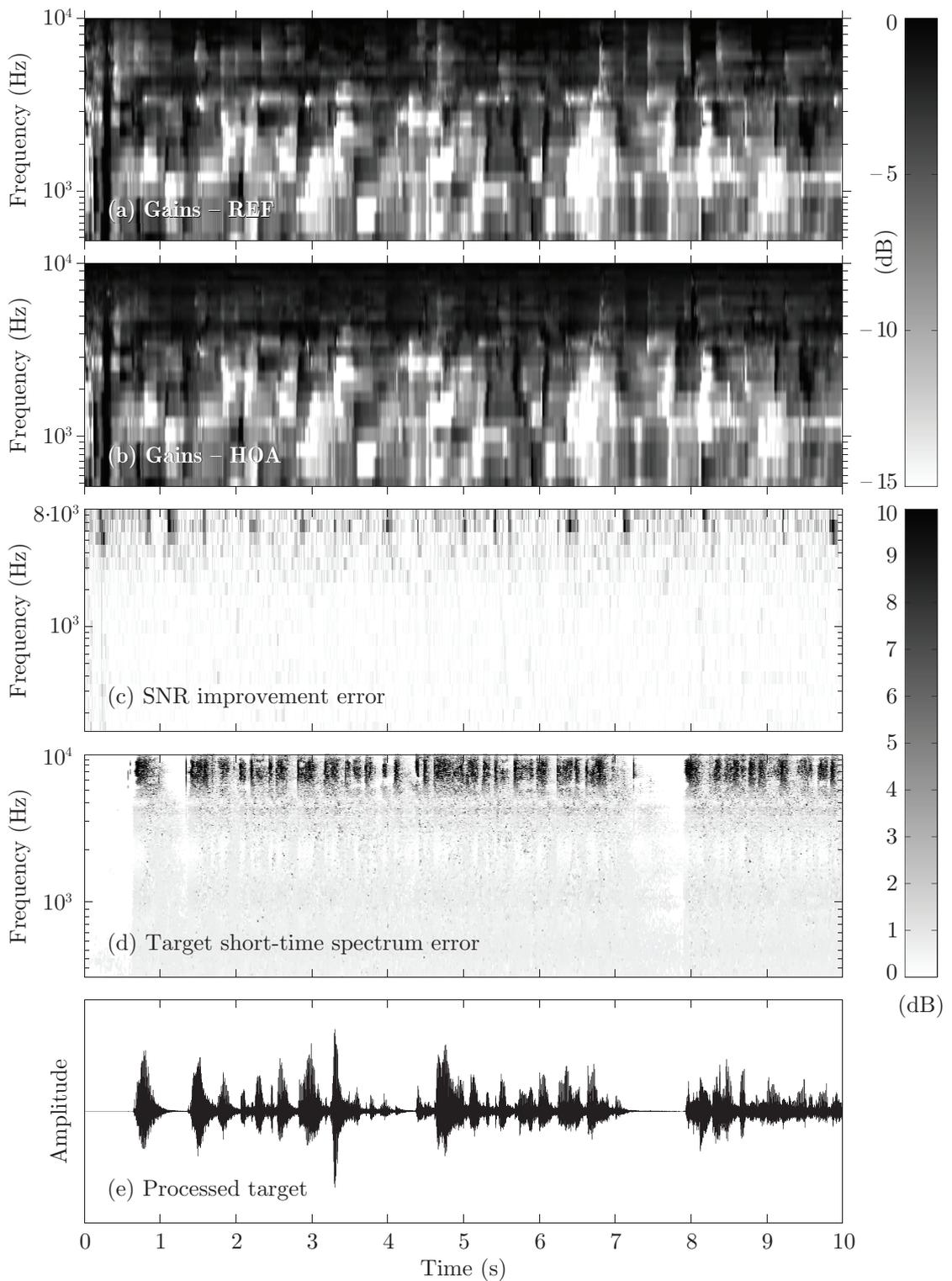


FIGURE 5.2: Short-time analysis of the ADM beamformer operating in the simulated classroom scenario: adaptive gains when using (a) REF, and (b) HOA; (c) absolute SNR improvement error, and (d) absolute output target spectrum error (errors due to HOA coding); (e) processed target.

HOA systems. This is further confirmed by the long-time average spectra of the target [Fig. 5.3(a)], calculated using the Welch estimation method (1024-point STFT, Hann window, 50% overlap). The high-frequency spectrum of the distractors on the contrary is less affected [Fig. 5.3(b)].

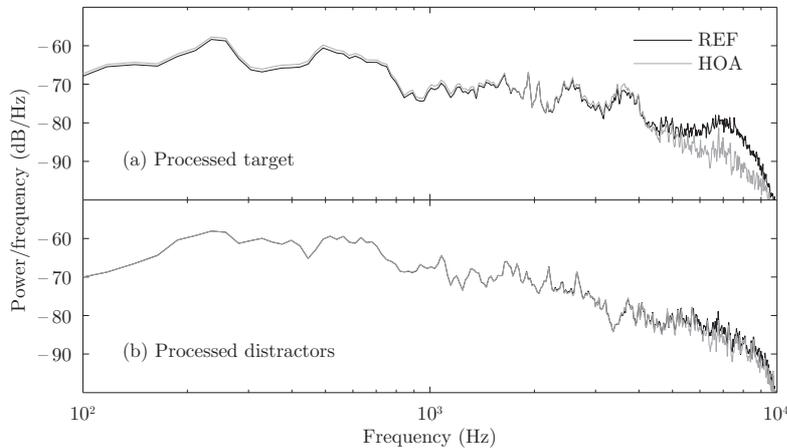


FIGURE 5.3: Long-term spectra of ADM output target and distractors in the simulated classroom scenario, using a REF and a HOA system.

Fig. 5.4 shows the median values and interquartile ranges (IQR) over time of the gain errors introduced by the HOA coding [i.e., difference between Fig. 5.2(a)/(b)] for both beamformers and both acoustic scenarios. For both beamformers, the gain errors are lower and more controlled in the reverberant scenario. Moreover, at high frequencies the BBF gains exhibit increasing errors while the ADM gain errors taper off.

The narrowband, long-time averaged ADM and BBF SNR improvements shown in Fig. 5.5 exhibit similar trends. They were calculated using the full length of the signals, analysed in 18 one-third octave bands from 156 Hz to 8 kHz (ANSI S3.5-1969).

## 5.4 Discussion

The short-time analysis shown in Fig. 5.2 illustrates that the HOA sound field is equivalent to the reference field in terms of the ADM operation and benefit, up to a frequency roughly corresponding to  $f_c = M \cdot 600 = 4.2$  kHz (Oreinos and Buchholz, 2013a). Above  $f_c$ , the ADM time- and frequency-dependent gains are affected, the short-time SNR is overestimated and the processed target (but not the distractors) is effectively low-pass filtered (Fig. 5.3). Although not shown here, the same behaviour can be observed for the BBF algorithm.

The average gain error plots of Fig. 5.4 suggest that the HOA error starts from lower frequencies, especially when the anechoic scenario is considered. Moving to the reverberant

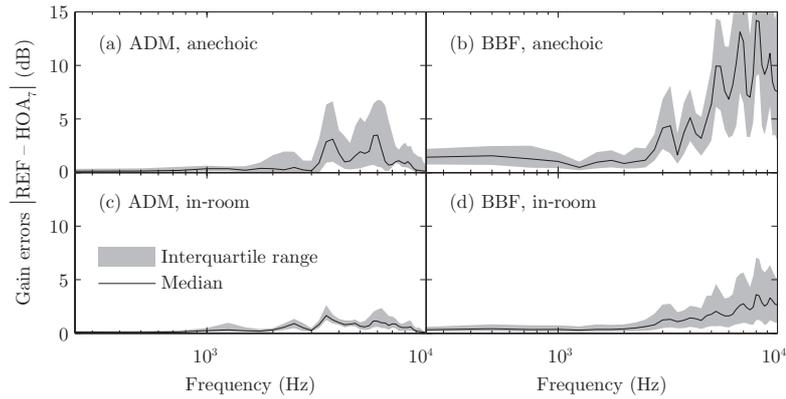


FIGURE 5.4: Median and IQR of ADM and BBF gain errors caused by HOA.

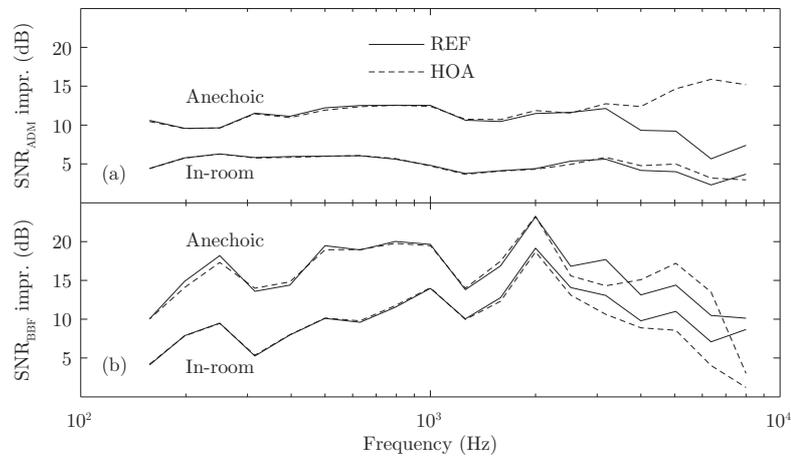


FIGURE 5.5: SNR improvement provided by (a) the ADM, and (b) the BBF algorithms in an anechoic and in-room scenario, using the REF (solid line) and HOA (dotted line) system.

scenario, the gain errors for both ADM/BBF are greatly reduced [Fig. 5.4(c)/(d)]. The ADM gain errors taper off to 0 dB in both acoustic scenarios due to the clipping of the  $\underline{b}(i, k)$  gains in the range  $[0, 1]$ , needed to constrain the nulls to the rear hemisphere (Section 5.2.2). As Fig. 5.2(a) shows, the gains are already saturating at high frequencies in the REF system. HOA only seems to push the gains further into saturation, which results into a low gain error. Although it appears tempting to compare plots (a) & (c) with (b) & (d), such a comparison between ADM and BBF gain errors lacks any physical relevance since the ADM produces mixing gains  $\underline{b}(i, k)$  while the BBF produces gains  $W_{L/R}(i, k)$  that directly filter the input signals.

Fig. 5.5 suggests that the long-term narrowband SNR improvement of ADM is almost perfectly replicated in a HOA reverberant system [bottom curves of Fig. 5.5(a)]. The BBF algorithm appears more prone to erroneous SNR improvement estimations above 2 kHz even in the reverberant scenario. This sensitivity of the BBF to the use of HOA may be attributed to its operation principle; it uses correlation-based metrics along with

TABLE 5.1: Intelligibility-weighted SNR improvement (dB).

	Fixed Directional		Adaptive Directional		Bilateral Beamformer	
	anechoic	in-room	anechoic	in-room	anechoic	in-room
REF	5.6	3.1	11.1	5.1	16.7	11.7
HOA	5.6	3.0	12.2	5.2	16.5	11.1

a series of complex rules to define the adaptive gains. Thus, even small field errors can cause its operation to drift away from the optimum range. HOA, for example, spreads the energy of virtual sources above the cut-off frequency  $f_c$  around the virtual source location. Sound energy appearing from other than the target's position could make the BBF algorithm treat those portions of the useful signal as unwanted noise. On the contrary, the ADM operation is hard constrained by design, limiting nulls to the rear hemisphere. Hence, HOA reconstruction errors cannot cause the algorithm to suppress the target.

Despite the gain and spectrum errors that the HOA coding causes, it does not affect the broadband intelligibility-weighted SNR improvement (Greenberg *et al.*, 1993) of the ADM and BBF –in both acoustic scenarios– as shown in Table 5.1. This occurs because the high frequency SNR values are weighted less than those at low and middle frequencies, reducing the errors' contribution. It could be then hypothesised that Figs. 5.1- 5.5 exaggerate the practical limitation of HOA coding for evaluating beamforming algorithms.

## Acknowledgements

The authors acknowledge the financial support of the HEARing CRC, established and supported under the Cooperative Research Centres Program – an initiative of the Australian Government. In addition, we would like to thank Cong-Van Nguyen for the filter-bank code as well as for the many fruitful discussions and general support on signal processing implementation issues.

## Chapter 6

# An objective and subjective validation of applied loudspeaker-based virtual sound environments used for directional hearing aid testing\*

### Abstract

Assessing the real-life benefit of hearing aids (HAs)<sup>†</sup> inside realistic virtual sound environment (VSEs) has long been desired. However, no verified methods exist in literature. This study first created a ‘cocktail party’ scene inside a real-room which was then modelled using acoustic simulation software. The acoustic model was reproduced inside a 41-loudspeaker array using either a direct mapping technique or a mixed-order Ambisonics method, giving rise to the ‘Model’ and ‘MOA’ VSEs, respectively. The accuracy of the VSEs was assessed by comparing the performance of two directional HA algorithms inside all three acoustic environments. For that purpose, a number of objective and subjective performance measures were applied. The objective analysis of both the ‘Model’ and ‘MOA’ VSEs revealed only minor long-term errors in pressure and SNR, but considerable variations in their short-time behaviour. The subjective analysis demonstrated that the general behaviour and trends in the benefit measured in speech intelligibility as well as acceptable noise levels were preserved in both VSEs. However, the estimated directional HA benefits were slightly reduced in the ‘Model’ VSE, and further reduced in the ‘MOA’ VSE. Hence, it can be concluded that the considered VSEs can be used for testing directional HAs, but the provided sensitivity is reduced when compared to a real environment. This can result in an underestimation of the provided directional benefit.

---

\*Manuscript to be shortly submitted for publication.

<sup>†</sup>For a detailed list of the employed abbreviations refer to page [xvii](#)

## 6.1 Introduction

Researchers as well as the hearing devices industry have been recently demonstrating a growing interest in using ecologically-valid, but at the same time, highly controlled listening tests to estimate the benefit of signal enhancement strategies. This interest has been notably motivated by the observation (Ricketts, 2000; Revit *et al.*, 2007; Compton-Conley *et al.*, 2004) that large discrepancies occur between the laboratory-derived improvement provided by hearing aids (HAs) and their real-life benefit. Binaural technology has been employed (Van den Bogaert *et al.*, 2008; Rychtáriková *et al.*, 2009; Völk and Fastl, 2010; Mueller *et al.*, 2012; Fels, 2013) as a means to introduce some realism into laboratory-based listening tests. However, presenting the auditory scene via headphones: (1) ideally requires individual head-related transfer functions (HRTFs) to the microphones of HAs worn on the subject's head; (2) precludes head movements (unless a head-tracker is used); (3) disregards the acoustic path contribution (leakage) of the HA ear-mould; and (4) precludes the use of off-the-shelf or the subject's individual HAs.

One possible solution to overcome the drawbacks associated with the headphone presentation of stimuli is to use loudspeaker-based virtual sound environments (VSEs) for administering realistic listening tests (Revit *et al.*, 2002; Minnaar *et al.*, 2010). In this way, complex acoustic scenes can be reconstructed either by employing a (room) acoustics simulation software (Favrot and Buchholz, 2010b; Seeber *et al.*, 2010), or by recording real-life environments using a microphone array (Park and Rafaely, 2005; Moreau *et al.*, 2006). The coded acoustic scene information can then be reproduced by loudspeaker arrays, with the appropriate loudspeaker signals derived either by methods such as the vector-based amplitude panning method (VBAP: Pulkki, 2001), or via sound-field reconstruction methods such as wave-field synthesis (WFS: Berkhout *et al.*, 1993) or higher-order Ambisonics (HOA: Daniel, 2001; Ward and Abhayapala, 2001; Poletti, 2005).

For the purpose of conducting listening test with HAs, sound-field synthesis methods (WFS, HOA) appear to be better candidates than VBAP, since the latter method only focuses on recreating perceptual attributes as opposed to accurately recreating the actual sound-field. Hence, VBAP is expected to be insufficient for listening test purposes since the HA processing, especially when considering multi-microphone enhancement schemes, significantly differs from the auditory system processing. Between the sound-field reconstruction methods of WFS and HOA, the latter has the advantage (Daniel *et al.*, 2003; Spors and Ahrens, 2008) that: (1) it aims to optimise the sound-field at the centre of the loudspeaker array, resulting in a centred sweet-spot, which suits the

listener-centred nature of many listening tests; (2) it is a flexible and scalable format, i.e., the sound-field encoding can be separated from the decoding, which, in turn, can be conducted on a number of different loudspeaker arrays; and (3) it directly offers a method of using a spherical microphone array to decompose the sound-field, i.e., to estimate the HOA components (Moreau *et al.*, 2006; Oreinos and Buchholz, 2015a).

At the same time, however, a number of limitations characterise the HOA method: (1) it has a limited spectral and spatial accuracy, determined by the size and geometry of the microphone and loudspeaker arrays; (2) it exhibits a sweet spot whose size diminishes with frequency; (3) it is affected by amplified microphone noise induced during the encoding operation. While the theoretical limitations of HOA have been well studied (Ward and Abhayapala, 2001; Poletti, 2005), only limited knowledge exists (Oreinos and Buchholz, 2013b, 2015a) on its applicability to reconstruct realistic reverberant sound-fields intended for listening tests on HA-fitted subjects.

The alternative method of generating predominantly static acoustic scenes via acoustic modelling techniques avoids the sound-field analysis with a microphone array. Hence, the limitations introduced by the microphone array encoding are removed. The reverberant sound-field can then for example be described by a set of auxiliary sources (Oreinos and Buchholz, 2015a) that encode the specular reflections and the diffuse (late) reverberation (Favrot and Buchholz, 2010b). The encoded information can then be reproduced by a loudspeaker array via any of the above discussed methods (VBAP, WFS or HOA). Alternatively, the above auxiliary sources can be assigned, one-by-one, to the single closest loudspeaker available at the reproduction array (Favrot and Buchholz, 2010b). This direct, ‘single-loudspeaker’ source-mapping method eliminates the spectral and spatial limitations of the previous (VBAP, WFS and HOA) methods which use more than one loudspeakers to place virtual sources even in positions where a physical loudspeaker does not exist. This is, however, accomplished at the expense of changing the spatial properties of the model-generated scene. VSEs based on the single-loudspeaker reproduction of a room acoustics model have been previously perceptually validated by Favrot and Buchholz (2010b). Nevertheless, their suitability for listening tests involving multi-microphone signal enhancement HAs still remains unknown.

The present study investigates the applicability of two example methods for realising VSEs used in evaluating the performance (e.g., provided speech intelligibility benefit) of directional HAs tested inside them. The investigation follows both an objective and a subjective path, where the traditional terms ‘objective’ and ‘subjective’ refer to ‘employing technical estimators of perceptual attributes’ and ‘directly quantifying what is actually perceived’ respectively. For the sake of compactness, however, the terminology ‘objective’ and ‘subjective’ will be employed in the remaining of the work.

The tested VSEs were both generated from a room acoustics model description which was reproduced by loudspeakers either using a direct coding method as described by Favrot and Buchholz (2010b) or via using a mixed-order Ambisonics (MOA), i.e., a variation of HOA (Favrot *et al.*, 2011). The VSEs were compared to a real-life reverberant scene in terms of the directional HA benefit estimated either by using a dummy head or test subjects. To that end, a ‘cocktail party’ scene was first created inside a real-room (Section 6.2.1.1). That acoustic environment was then modelled, resulting in a set of auxiliary sources (Section 6.2.1.2). Those auxiliary sources were subsequently reproduced by the 41-channel loudspeaker array available at the National Acoustic Laboratories (NAL) via: (1) a direct, single-loudspeaker mapping, and (2) a MOA encoding and decoding operation (Section 6.2.1.3). The MOA encoding was realised via a simulated 62-channel microphone array, currently being built at NAL. Listening test with HA-fitted hearing-impaired subjects were conducted in all 3 (real-room plus two VSEs) acoustic environments (Sections 6.2.4, 6.3.2 and 6.4.2) using two different directional HA algorithms. Additionally, signals at the microphones of HAs worn on an acoustic manikin were derived in all 3 environments to objectively evaluate the processed HA (output) signals (Sections 6.2.3, 6.3.1 and 6.4.1). The study concludes with a discussion (Section 6.4) and a brief summary of the most important findings (Section 6.5).

## 6.2 Methods

### 6.2.1 Acoustic scene generation

Figure 6.1 presents the acoustic environments inside which listening tests were performed to evaluate the performance of two directional enhancement algorithms, as implemented on hearing aids. One of the acoustic environments was a real room (path ‘Room’) and the remaining two (paths ‘Model’ and ‘MOA’) were virtual sound environments (VSEs) reproduced inside the loudspeaker array built at NAL. The ‘Model\*’ path was a special version of the ‘Model’ environment, solely employed to derive the ‘MOA’ VSE, and otherwise did not represent a realisable VSE suitable for listening tests.

#### 6.2.1.1 Real-room environment

The real-room environment (path ‘Room’ in Fig. 6.1) consisted of a meeting-room at the National Acoustic Laboratories (NAL) of average size with a volume of  $V \approx 172 \text{ m}^3$  and a reverberation time of  $T_{20}$  as shown in Fig. 6.4a. The floor plan is shown in Fig. 6.2. The critical distance was estimated [Kuttruff 2000, Eq. (5.40)] at  $r_c \approx 1.5 \text{ m}$

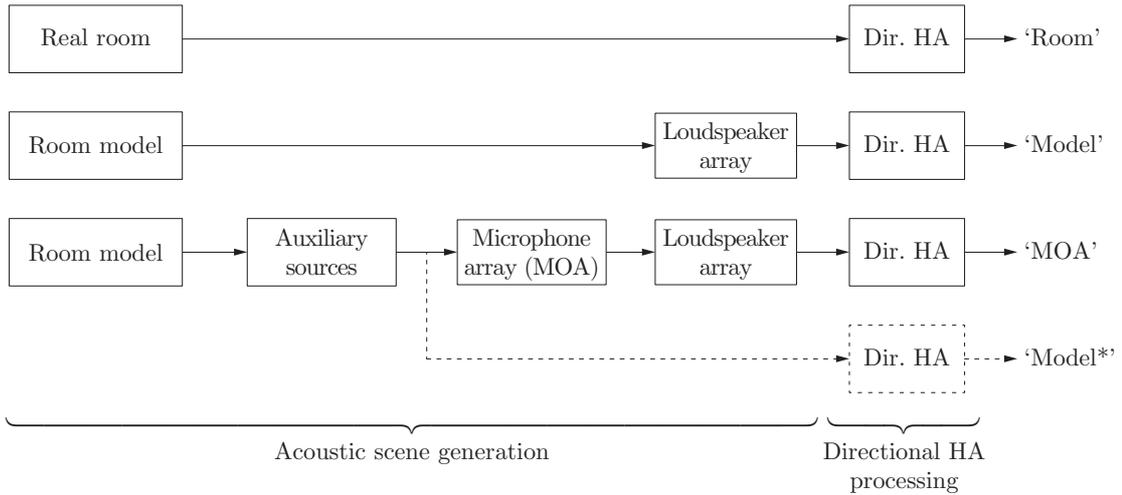


FIGURE 6.1: Illustration of the acoustic environments that were evaluated, using different subjective and objective performance measures, at the output of directional hearing aids.

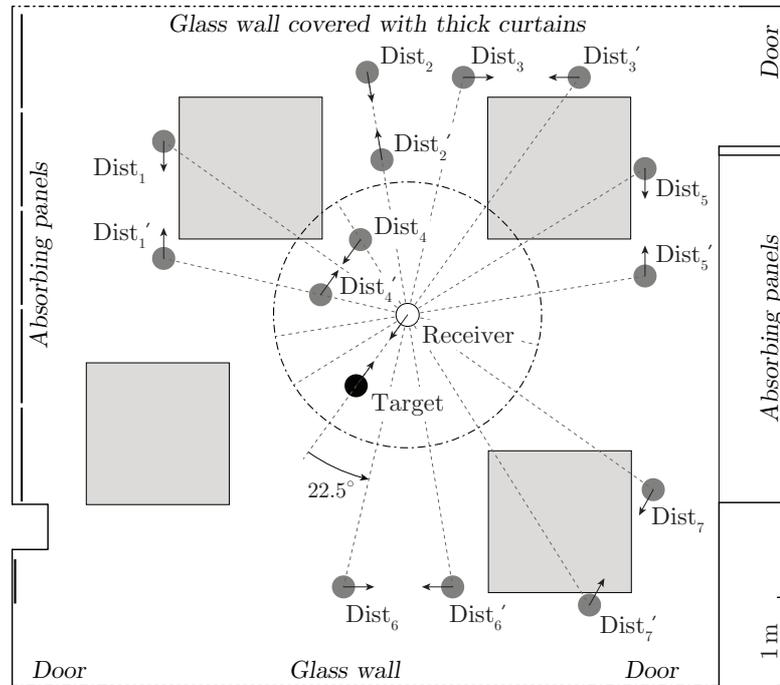


FIGURE 6.2: Room floor plan showing the position and orientation of all sources.

and is indicated in Fig. 6.2 by the dash-dotted circle around the receiver. Seven pairs of distractor sources and one target source were positioned in the room as seen in Fig. 6.2. All of them were reproduced by Genelec 8020C active (self-amplified) loudspeakers, positioned with their tweeters at a height of  $h = 1.15$  m from the floor. The loudspeakers were driven by two ADI-8 DS digital-to-analogue (DA) converters and an RME Fireface UFX interface, connected to a desktop PC. Prior to positioning them in the room, the gains of all loudspeakers were adjusted inside a sound-treated test booth to exhibit the same sensitivity.

The loudspeakers reproducing distractor sources were facing one another, in pairs, imitating the natural orientation of people engaged in a dialogue. The chosen source configuration provided a diffuse and non-symmetric scene, featuring a range of nearby and more distant distractors. It was designed to approximate a ‘cocktail party’ scene in a realistic way, while not purposely leading to an over-estimation of the considered directional HA algorithm benefit.

A mixture of seven male and female speech dialogues was assigned one-to-one to the seven distractor source pairs. The speech material was previously recorded in an anechoic chamber at NAL and was subsequently calibrated to the same long-term average power (with the speech pauses removed) before being fed to the loudspeakers. An overall gain was applied to all stimuli such that the total RMS level at the receiver point (no listener present) equalled 69 dB SPL, or equivalently 64 dBA (A-weighted level). That value was derived from the ODEON room model by assigning a male talker with normal vocal effort (ODEON pattern: T1knorm\_NATURAL.S08) as the source type for all distractors and then deriving the total power at the receiver point. Three experienced listeners confirmed that the presentation level matched their expectations based on the number of talkers and size of room. A similar calibration procedure was applied to the target speech material (see Section 6.2.4.3) so that it could be presented at specified SNR values.

The subjects that participated in the psychoacoustic experiment (see Section 6.2.4) were seated with their head centred at the receiver position and facing the target-loudspeaker. The chair was adjusted such that their ear canals were at a height of  $h = 1.15$  m from the floor.

### 6.2.1.2 Direct room acoustics model-based sound rendering

The first virtual sound environment (path ‘Model’ in Fig. 6.1) consisted of the direct auralisation of an acoustic model of the real-room environment described in Section 6.2.1.1. The VSE was generated inside the NAL 3D loudspeaker array (Fig. 6.3a). The array consisted of 41 Tannoy V8 concentric loudspeakers positioned on the surface of a virtual sphere of radius  $R_L = 1.85$  m, built inside an anechoic chamber of size  $6.7 \times 9.4 \times 6.3$  m. The loudspeakers were driven via a desktop PC hosting an RME HDSPe MADI interface with two M-32 DA converters, followed by 11 Yamaha XM4080 quad-channel power amplifiers.

In order to calculate appropriate signals to drive the loudspeakers, the real-room environment described in Section 6.2.1.1 was modelled using the room acoustics simulation software ODEON. Initial estimations for the acoustic properties (diffusion/scattering)

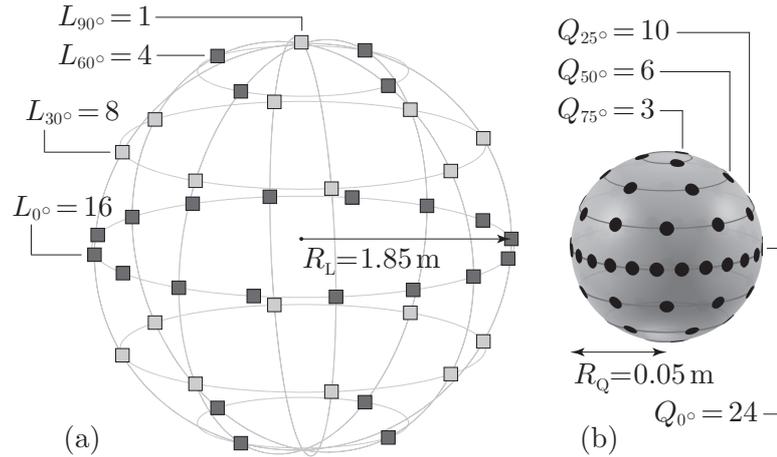


FIGURE 6.3: The employed: (a)  $L = 41$ -loudspeaker, and (b)  $Q = 62$ -microphone arrays (not plotted to scale). Both array configurations exhibit horizontal-plane symmetry, apart from a missing loudspeaker at nadir ( $\delta_l = -90^\circ$ ).

of all materials were considered. The exact 3D directivity of the Genelec 8020C loudspeaker, as measured and provided by the manufacturer, was assigned to all sources of the room model. Following the acoustics simulation, descriptions of the room impulse response (RIR) were derived for all sources shown in Fig. 6.2. These RIR descriptions were afterwards transformed into 41 loudspeaker signals using the LoRA toolbox (Favrot and Buchholz, 2010b). Thereby, the direct sound (DS) and early (specular) reflections (ER) were each realised by a single loudspeaker while the late (Lt) diffuse reverberation was reproduced as described by Oreinos and Buchholz (2015a).

Since the specular room reflections (DS/ER) were reproduced by single loudspeakers, all room sources in the real-room environment were positioned, without loss of generality, at angles where an horizontal-plane loudspeaker (i.e.,  $L_{0^\circ}$  ring in Fig. 6.3a) existed. Those angles are shown in Fig. 6.2 with the dashed lines radially extending outward from the receiver. This placement excluded systematic errors that would have otherwise been incurred in the VSE due to misalignment between the available loudspeakers and each source's DS (carrying the largest part of the energy in most cases). The ERs, whose directions were defined by the room geometry, were assigned to their nearest loudspeaker in the playback array.

In order to fine-tune the model, RIRs for each source of the considered scene were computed as the summed contribution of all loudspeaker signals. Then, in an iterative manner, the absorption and scattering coefficients of the room-model materials were adjusted inside ODEON so that the reverberation time  $T_{20}$  and early decay time  $T_{\text{EDT}}$  of the reproduced RIRs closely matched the values derived by the real-room RIR measurements. Fig. 6.4 illustrates the good agreement between the RIRs reproduced in the ODEON-LoRA sound environment and the real-room considering: (a) the median

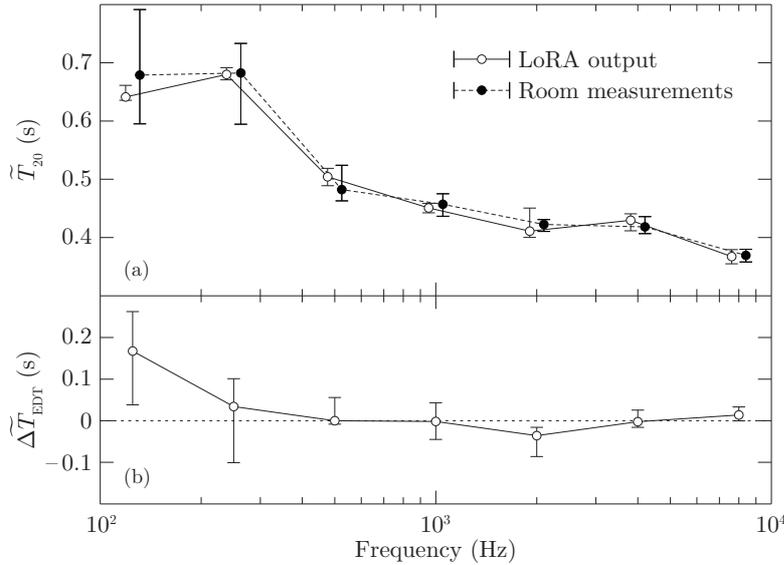


FIGURE 6.4: Room-model fine-tuning. Median and IQR, across all 15 positions, of the: (a) reverberation time  $T_{20}$  as estimated in the real room and in the room model; (b) difference of the early decay time  $\Delta T_{EDT}$  between the real room and room model estimations.

and interquartile range (IQR) of the reverberation time  $T_{20}$ , and (b) the median and IQR of the early decay time errors  $\Delta T_{EDT} = T_{EDT}^{\text{Model}} - T_{EDT}^{\text{Room}}$ . Since the  $T_{20}$  is largely independent of room position, the values were averaged across all 15 positions shown in Fig. 6.2. This was not the case for  $T_{EDT}$ , which varies strongly across room position. Instead, the individual  $T_{EDT}$  errors were averaged across all 15 room positions. No further adjustments were made to the room model at any later stage.

Following the process described above, 15 multichannel impulse responses (mIRs) were created, one for each source, to auralise the considered room scene. The same anechoic speech material as presented in the real-room was convolved, source-by-source, with the derived mIRs. Before feeding the convolved signals to the array, all 41 loudspeakers were equalised, using minimum-phase FIR filters, and calibrated so that at the centre of the array they had a flat (within  $\pm 2$  dB) frequency response in the range of 100 – 16000 Hz. Following, an overall gain was applied to all loudspeaker channels such that the generated long-term level of the combined distractor sources equalled 69 dB SPL at the array centre (no listener present). A similar signal generation and calibration procedure was applied to the target speech material (see Section 6.2.4.3).

The subjects were seated with their head centred at the loudspeaker array and facing the frontal ( $\theta_l = 0^\circ$ ) loudspeaker. The chair was adjusted such that their ear canals were at the same height as the equator of the loudspeaker array.

### 6.2.1.3 Mixed-Order Ambisonics sound-field reconstruction

The second virtual sound environment (path ‘MOA’ in Fig. 6.1) was also generated inside the NAL loudspeaker array. It was created using the same room acoustics model as described in Section 6.2.1.2, but this time applying the mixed-order Ambisonics (MOA) method to create the sound field inside the loudspeaker array. MOA is a variation of the higher-order Ambisonics method that enhances the sound-field reproduction (increased sweet-spot region or higher frequency cut-off) of virtual sources lying on the horizontal plane (Favrot *et al.*, 2011; Käsbaach *et al.*, 2011; Favrot and Marschall, 2012; Marschall *et al.*, 2012). MOA applies different orders,  $M_{3D}$  and  $M_{2D}$ , to reproduce the periphonic (3D) and horizontal-plane only (2D) sound-field components. It includes all spherical harmonic function (SHFs) up to the degree  $m = M_{3D}$  plus all sectorial SHFs (i.e.,  $m = n$ ) with degrees  $M_{3D} < m \leq M_{2D}$ . The system orders were set to  $M_{3D} = 4$ ,  $M_{2D} = 7$ , which were the highest orders that could be sufficiently reproduced using NAL’s 41-channel loudspeaker array [Ward and Abhayapala, 2001, Eqs. (22) and (30)]. For the interested reader, alternative SHF selection schemes, similar to the herein considered MOA scheme, can be found in the work of Travis (2009).

The MOA signals were derived from the acoustic room model (same as in Section 6.2.1.2) using a simulated microphone array with a radius of  $R_Q = 0.05$  m and  $Q = 62$  microphones (see Fig. 6.3b for details). The simulations were made for a microphone array that is currently developed at NAL. This microphone configuration was used as an example array to illustrate the artefacts that a real MOA recording system would introduce, including spatial aliasing and amplified microphone self-noise. In the process of simulating the output signals picked up by that array, the effect of real-life positioning errors as well as mismatches of the amplitude and phase of the microphones were taken into consideration (details can be found in Appendix D).

To provide appropriate input signals to the 62 microphones, signals were first derived for a very dense array of auxiliary sources (or ‘virtual loudspeakers’) following the procedures described in Section 6.2.1.2. Hence,  $N_{DE} = 1784$  sources were used to simulate the DS and ER components and  $N_{Lt} = 100$  sources were used to simulate the late reverberation components. Afterwards, the pressures at all 62 microphones were derived individually for each of the auxiliary (plane-wave) sources using a hard sphere model [Daniel, 2001, Eq. (A.50)]. The derived microphone signals were then encoded into MOA sound-field components by applying the ‘shape-matching’ operation (Oreinos and Buchholz, 2014), optimised using the 41-loudspeakers array, and calculated with a regularisation factor of  $\lambda_{SM} = 0.4$ . The sound-field was decoded into 41 loudspeaker signals using a basic decoding scheme.

At the final stage, the 41-channel mIRs derived from the MOA coding stage, one for each of the room sources, were convolved, source-by-source, with the same anechoic speech material as presented in Section 6.2.1.1. The resulting signals were then mixed with the simulated microphone array noise (see Appendix D) and fed to the NAL loudspeaker array. The loudspeakers were equalised as described in Section 6.2.1.2, but the calibration procedure was slightly modified. Here, the omnidirectional microphone that was used for calibration in the centre of the loudspeaker array was replaced by a Brüel & Kjær type 4128C head and torso simulator (HATS). A calibration gain was applied on all loudspeaker channels such that the combined distractor sources generated the same long-term level at the in-ear microphones of the HATS as that generated in the (calibrated) VSE described in Section 6.2.1.2. This modified calibration procedure was applied to account for the level variations that are observed in MOA (and HOA) sound-fields when moving away from the array centre (Solvang, 2008). A similar calibration procedure was applied to the target speech material.

Finally, the subjects that participated in the psychoacoustic experiment (Section 6.2.4) were seated inside the loudspeaker array in the same way as described in Section 6.2.1.2.

## 6.2.2 Directional hearing aid processing

Two directional processing methods were implemented to validate the different virtual sound environments described in Section 6.2.1: (1) a first-order delay-and-subtract cardioid (CRD) directional microphone (Kates, 2008), and (2) the NAL proprietary adaptive bilateral beamformer (BBF) processing (Mejia *et al.*, 2007; Mejia and Dillon, 2010). The bilateral correlation-based BBF scheme combines signals from HAs at the left and right ear aiming at preserving on-axis sounds with minimum distortions, while suppressing laterally-impinging sounds. Both algorithms were programmed in MATLAB as well as C, operating at the sampling frequency of  $f_s = 24$  kHz. The MATLAB version was used for off-line processing, which was applied in the objective analysis described in Section 6.2.3. The C version was used for real-time processing as required for the listening tests (see Section 6.2.4) and was run on a standard laptop PC. The laptop PC was connected to a pair of behind-the-ear HA satellites via a custom-built pre-amplifier and an RME Hamerfall DSP Multiface II AD/DA interface. The HA satellites, together with a dedicated pre-amplifier, were built and provided by Phonak AG.

The CRD processing was optimised to minimise the output power for a rear source at  $(180^\circ, 0^\circ)$  while maintaining a flat-spectrum output for a frontal source at  $(0^\circ, 0^\circ)$ , with the HAs placed on a HATS. The CRD outputs at the left and right HA formed the input to the BBF processing, hence no more adjustment was necessary for the BBF stage.

To allow an objective analysis of the detailed spectro-temporal behaviour of the adaptive BBF algorithm in the different VSEs, the off-line implementation in MATLAB featured both an ‘adaptive mode’ and a ‘freeze-mode’ (Oreinos and Buchholz, 2013b). First, the combined target and distractor signals were processed at a given SNR using the ‘adaptive mode’. Thereby, all frame- and frequency-dependent gains were stored. Afterwards, the identical target and distractor signals were fed separately to the BBF processing operating in ‘freeze mode’, i.e., using the stored (pre-adapted) gains.

### 6.2.3 Objective evaluation

In order to objectively validate the effect of the considered VSEs on the outputs of the different HA signal processing schemes described in Section 6.2.2, stimuli were created as described in Section 6.2.1 and then converted to signals at the HA microphones. For the real-room case (‘Room’ path in Fig. 6.1), a HATS, fitted with a pair of behind-the-ear HA satellites, was placed inside the real-room environment at the receiver (listener) location shown in Fig. 6.2. IRs were measured from each loudspeaker inside the room to all four HA microphones and convolved with the anechoic speech recordings that were used in the real-room environment to drive all the loudspeakers (see Section 6.2.1.1). For the virtual sound environments (paths ‘Model’ and ‘MOA’ in Fig. 6.1), the HATS was placed in the centre of the NAL loudspeaker array and IRs were measured from each of the 41 loudspeakers to each of the four HA microphones. The resulting IRs were convolved with the loudspeaker signals described in Sections 6.2.1.2 and 6.2.1.3 and the final filtered loudspeaker signals were superimposed to form four HA microphone signals. To establish a suitable reference condition for the MOA system, the same processing was applied as for the VSEs, but instead of the 41 playback loudspeakers, 1784+100 auxiliary sources were considered, as described in Section 6.2.1.3, giving rise to the ‘Model\*’ path in Fig. 6.1. The IRs associated with the auxiliary sources were taken from Oreinos and Buchholz (2013c), who used the same HATS and HA satellites as applied in this study.

The objective analysis considered an example SNR of  $-10$  dB as derived at the centre of the head with no head present. That value was equal to the average SNR at which the listeners that participated in the speech intelligibility experiment (described in Section 6.2.4) were tested.

#### 6.2.3.1 Pressure reconstruction errors

To evaluate the accuracy of the different virtual sound environments described in Section 6.2.1, the output signals of the CRD and BBF processing were first calculated.

Separate output signals were calculated with the target only and with all of the distractors at the input of the HAs operating in ‘freeze’ mode (see Section 6.2.2). Due to the different delays introduced by the different HA signal processing schemes, all input and output signals of both HA processing schemes had to be time-aligned. The required delays were estimated using a cross-correlation analysis. The RMS levels of all output signals were further normalised to the same value. Next, the power spectral density  $S[n, f]$  at the output of the different HA processing schemes was calculated for the different acoustic scenes using a short-time Fourier transform (STFT) and applying a sliding Hamming window of length  $\tau_{\text{ST}} = 0.02$  s with 50% overlapping and an FFT length of  $N_{\text{FFT}} = 1024$  samples. Short-time spectral errors (eventually representing pressure errors) were then calculated in dB as:

$$\begin{aligned}\mathcal{E}_S^{\text{Model}}[n, f] &= S^{\text{Model}}[n, f] - S^{\text{Room}}[n, f] \\ \mathcal{E}_S^{\text{MOA}}[n, f] &= S^{\text{MOA}}[n, f] - S^{\text{Model}^*}[n, f],\end{aligned}\tag{6.1}$$

where  $S^{\{\text{Env}\}}[n, f]$  represents the short-time power spectral density of the CRD/BBF output signals expressed in dB,  $n$  represents the time frame,  $f$  the discrete frequency and the superscript  $(\cdot)^{\{\text{Env}\}}$  refers to the different acoustic environments shown in Fig. 6.1. The error introduced by the ‘Model’ VSE is given by  $\mathcal{E}_S^{\text{Model}}$  and the error introduced by the ‘MOA’ VSE is given by  $\mathcal{E}_S^{\text{MOA}}$ .

Based on the derived short-time pressure errors given in Eq. 6.1, integrated (or long-term) pressure errors were additionally calculated as a function of frequency. The first and third quartiles ( $Q1$ ,  $Q3$ ) as well as the median values were calculated across time after pauses in the signals were removed. The derived quartiles were then spectrally smoothed using a sliding  $1/3$ -of-an-octave Hanning window.

### 6.2.3.2 SNR benefit errors

Similar to the pressure errors described in Section 6.2.3.1, short-time errors in the SNR benefit ( $\Delta\text{SNR}$ ) provided by the directional HA algorithms (CRD/BBF) were also derived for the different acoustic environments. The  $\Delta\text{SNR}$  was thereby given as the difference in SNR measured at the frontal omnidirectional HA microphone and the SNR measured at the corresponding output of either the CRD or BBF processing. The short-time SNR was derived in  $\tau_{\text{ST}} = 0.02$  s—long time segments using a 3<sup>rd</sup>—order Butterworth filter bank with 18  $1/3$ -of-an-octave wide bands spanning a frequency range from 156 Hz to 8 kHz. The error in the resulting short-time benefit  $\Delta\text{SNR}$  was then calculated in dB

as:

$$\begin{aligned}\mathcal{E}_{\Delta\text{SNR}}^{\text{Model}}[n, f] &= \Delta\text{SNR}^{\text{Model}}[n, f] - \Delta\text{SNR}^{\text{Room}}[n, f] \\ \mathcal{E}_{\Delta\text{SNR}}^{\text{MOA}}[n, f] &= \Delta\text{SNR}^{\text{MOA}}[n, f] - \Delta\text{SNR}^{\text{Model}^*}[n, f],\end{aligned}\tag{6.2}$$

where  $\Delta\text{SNR}^{\{\text{Env}\}}[n, f]$  refers to the short-term SNR benefit, expressed in dB, for the different acoustic environments shown in Fig. 6.1.

Time-integrated SNR benefit errors ( $\mathcal{E}_{\Delta\text{SNR}}$ ) were derived in a similar way as described for the pressure errors (Section 6.2.3.1). However, in this case, preceding the calculation of the frequency-dependent quartiles, the pauses were only removed from the target signal and the  $\Delta\text{SNR}[n, f]$  values were clipped to the range  $[-5 \text{ dB}, 20 \text{ dB}]$ . Hence, a metric very similar to the segmental SNR (Tribolet *et al.*, 1978; Hansen and Pellom, 1998) was applied, except that in this study the mean value calculation was replaced by the median.

## 6.2.4 Subjective evaluation

The listening experiments consisted of two parts; an automated speech-in-noise test and an acceptable noise level (ANL) test (Nábělek *et al.*, 1991; Walravens *et al.*, 2014). Both methods followed closely the procedures described by Best *et al.* (2015), and are briefly summarised in the subsequent sections. The overall aim was to measure the benefit that the adaptive BBF beamformer provided to hearing impaired (HI) subjects in the different acoustic environments (see Fig. 6.1) on top of the benefit already provided by the CRD processing. Thereby, the underlying hypothesis was that if this additional benefit was equal for all the different acoustic environments, then the model-based VSE (Section 6.2.1.2) as well as the MOA-based VSE (Section 6.2.1.3) are equally suitable for testing directional HAs as the real-world environment (Section 6.2.1.1). The experiment was conducted in two appointments: during the first, testing was carried out inside the real-room (path ‘Room’ in Fig. 6.1) and during the second, testing was carried out inside the NAL 3D loudspeaker array (paths ‘Model’ and ‘VSE’ in Fig. 6.1).

### 6.2.4.1 Test subjects

Eighteen ( $N = 18$ ) adults with a moderate, sloping sensorineural hearing loss, aged between 66 and 78 years old (mean = 73.8 yo) participated in the study. All subjects had a symmetric hearing loss with an across-ear difference of less than 10 dB (at frequencies between 500 – 3000 Hz). Median audiometric thresholds are shown in Fig. 6.5 together with their interquartile range. All participants had English as a first language.

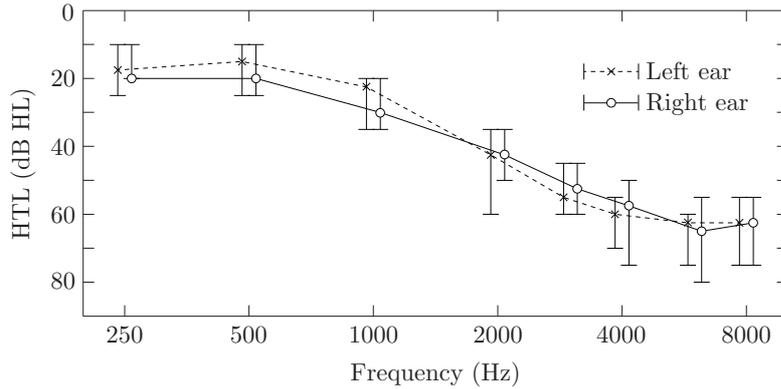


FIGURE 6.5: Median and interquartile range of hearing thresholds (in dB HL), for the left and right ear, of all hearing impaired test participants.

#### 6.2.4.2 Hearing aid fitting

Following the audiometric threshold measurement, the subjects were fitted with the calibrated pair of Phonak behind-the-ear HA satellites (Section 6.2.2). The hearing aids were fitted to the subjects' ears using highly occluding insert foam tips. Real-ear insertion gains were then measured using a Siemens Unity 2 fitting system, with the HAs operating in CRD mode and the subjects seated at the receiver position inside the real-room. The gains were adjusted in one-third-of-an-octave wide bands to match the linear NAL-RP (Byrne *et al.*, 1990) prescription targets relevant to their audiometric thresholds. The adjusted gains were also used in the BBF mode without repeating the fitting procedure and were then reused during the second appointment (testing inside the NAL 3D loudspeaker array).

#### 6.2.4.3 Speech intelligibility

For the automated speech-in-noise test, a large corpus of material based on the Bamford-Kowal-Bench (BKB) sentences (Bench *et al.*, 1979) was used. The total 464 sentences were spoken by an Australian talker and recorded in an anechoic environment. The sentences were presented, one-by-one, to the test subjects who were instructed to repeat as much as they could hear or infer from context. The test was administered via a MATLAB application (featuring a graphic user interface, GUI) built in-house at NAL. An audiologist was registering the number of correct phonemes on the GUI, based on the verbal responses of the test subjects. Each test subject was presented with two 32-sentence blocks for each of the two HA algorithms (CRD, BBF). The four, in total, sentence blocks were presented to subjects in a randomised order. Final scores of percent correct phoneme recognition were calculated by averaging the scores derived from the pair of sentence blocks relevant to each HA algorithm. The intelligibility benefit of the

BBF compared to the CRD processing was thereby calculated as the percentage increase of correct phonemes recognised by the test subjects.

The speech-in-noise test was conducted at a fixed SNR value, as described by [Best \*et al.\* \(2015\)](#). That value was calculated by performing an adaptive speech-in-noise test ([Keidser \*et al.\*, 2013b](#)), with the HAs operating in CRD mode, aiming at 50% speech intelligibility. The adaptive test was conducted inside all three acoustic environments. After deriving the speech reception threshold in noise (SRT<sub>n</sub>), the SNR at which the experiment was conducted was initially adjusted to SRT<sub>n</sub> – 2 dB. If, in rare cases, the test subject performed particularly poorly during the experiment (e.g., < 20–30% of correct phoneme recognition), the SNR was readjusted upwards (iteratively, if needed) in steps of 1 dB. That procedure ensured testing at the middle-low region of the psychometric function, thus providing enough headroom to measure the improvement of speech intelligibility when switching from CRD to BBF mode.

#### **6.2.4.4 Acceptable noise level**

To characterise the performance of the HA algorithms at higher, and thus, more realistic SNR values than considered in the speech-in-noise test (Section 6.2.4.3), the ANL was estimated in the second part of each of the two testing appointments. The applied procedure was adopted from [Best \*et al.\* \(2015\)](#) and consisted of the following two steps. Initially, the subjects were asked to set the level of a target signal (presented in quiet) to their most comfortable level (MCL). Then, with the target level fixed at MCL, they were asked to increase the background noise level (BNL) to the maximum value that allowed them to follow the target for a long period without becoming tense or tired. The ANL was thereby computed as:  $ANL = MCL - BNL$  (in dB). The BBF benefit on the ANL was finally calculated as the decrease of the ANL provided by the BBF, compared to the CRD processing. A positive ANL benefit value was used to denote a *lower* ANL, or else, a higher tolerance against background noise. Due to technical complications, only 16 out of the total 18 subjects completed the ANL test. Moreover, while the ANL test was performed twice for each subject inside the real-room and then the results were averaged, it was only performed once for every VSE condition due to time limitations.

## 6.3 Results

### 6.3.1 Objective analysis

#### 6.3.1.1 Pressure reconstruction errors

Figure 6.6 shows the short-time pressure errors of the processed target (top row) and the combined distractors (bottom row) at the output of the CRD processing at the left HA. The plotted errors  $\mathcal{E}_S^{\text{Model}}[n, f]$  are due to the ‘Model’ VSE [Eq. 6.1] and exhibit a quasi-random behaviour between successive time blocks, with maximum error values in the  $[-20 \text{ dB}, +20 \text{ dB}]$  range. The time-averaging of those errors, however, yields a median value that is mostly confined in the  $\pm 2 \text{ dB}$  range and an IQR extending roughly 4 dB about the median, as the panels (b) and (d) illustrate. Comparing the time-averaged spectral errors shown in panels (b) and (d) suggests that the ‘Model’ VSE incurs smaller errors on the target compared to the combined distractors.

The short-time pressure errors at the output of the BBF processing at the left HA, operating inside the ‘Model’ VSE as considered above, exhibited similar trends as those illustrated in Fig. 6.6. For that reason, only the time-averaged errors are presented in Fig. 6.7. Comparing plots (a) and (b) of that figure to plots (b) and (d) of Fig. 6.6 suggests that the BBF processing is more sensitive on inaccuracies of the VSE, thus yielding spectral errors with a slightly higher IQR.

In a similar fashion, Figs. 6.8 and 6.9 show the short-time and time-averaged pressure errors induced by the ‘MOA’ VSE (Section 6.2.1.3) on the CRD and BBF outputs at the left HA. The errors here were referenced to the outputs of the HAs operating inside the ‘Model’ VSE employing a dense auxiliary-source setup (Section 6.2.1.3; output ‘Model\*’ in Fig. 6.1) and not inside the ‘Room’ environment. In this case, the short-time pressure errors still show the quasi-random character that was seen in the ‘Model’ VSE (Fig. 6.6), but now at a much smaller scale. Hence, average errors relevant to both directional algorithms appear minimal up to  $\sim 2 \text{ kHz}$  for the target and up to  $\sim 3 \text{ kHz}$  for the combined distractors. Comparing panels (a) and (b) to panel (c) and (d) in Fig. 6.8 suggests that the median pressure errors of the combined distractors (at the CRD output) are smaller than those of the targets, although the IQR of the former is wider than that of the latter.

Pressure errors at the BBF output (Fig. 6.9) exhibit a similar median as illustrated for the CRD output, but an IQR roughly twice as large. This observation is in agreement with the increased short-time pressure errors at the BBF output, as compared to the CRD output, illustrated by Fig. 6.7.

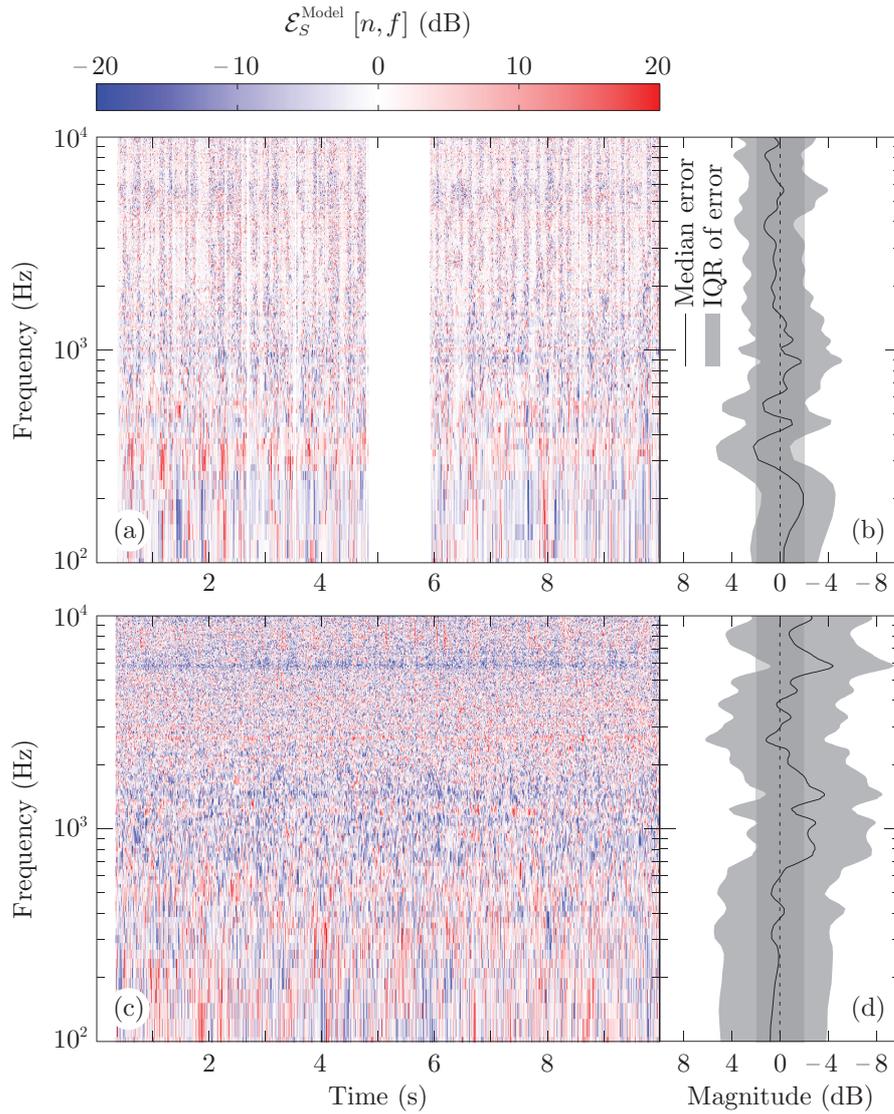


FIGURE 6.6: Short-time pressure errors  $\mathcal{E}_S^{\text{Model}}[n, f]$  [(a), (c)] along with their median and IQR values [(b), (d)] at the output of the CRD processing at the left HA, incurred by the ‘Model’ VSE. The top row [(a), (b)] shows the errors relevant to the target and the bottom row [(c), (d)] shows the errors relevant to all (summed) distractors. Blank areas in plots (a) and (c) denote pauses in the processed signal. The darker vertical zones in plots (b) and (d) show the  $\pm 2$  dB error region.

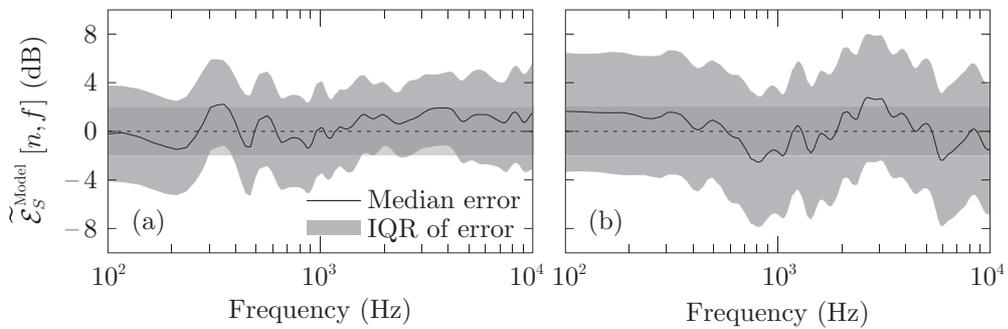


FIGURE 6.7: Median and IQR of pressure errors  $\mathcal{E}_S^{\text{Model}}$  incurred by the ‘Model’ VSE and measured at the output of the BBF processing (left HA) for the: (a) target, and (b) all (summed) distractors. The darker horizontal zones show the  $\pm 2$  dB error region.

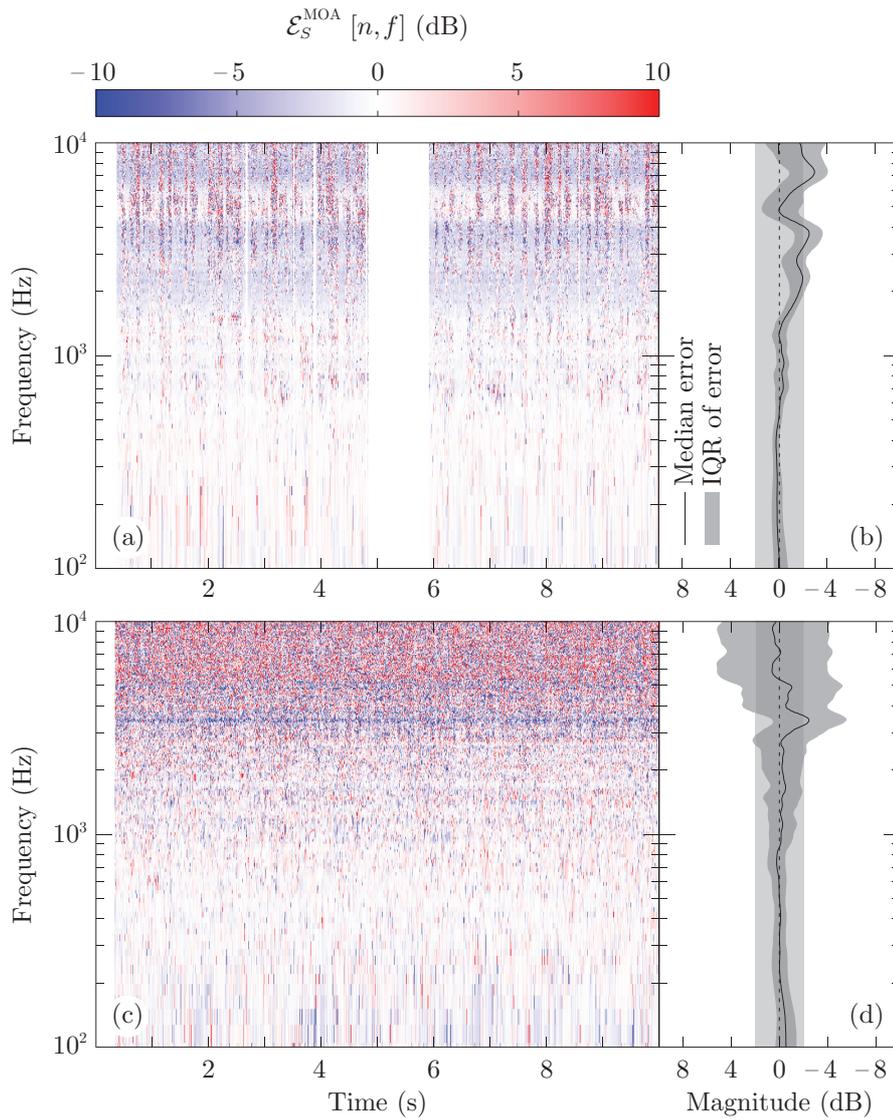


FIGURE 6.8: Same as Fig. 6.6 albeit showing the median and IQR of the pressure errors  $\mathcal{E}_S^{\text{MOA}}$  incurred by the ‘MOA’ VSE at the CRD output. Note the different colourbar scales.

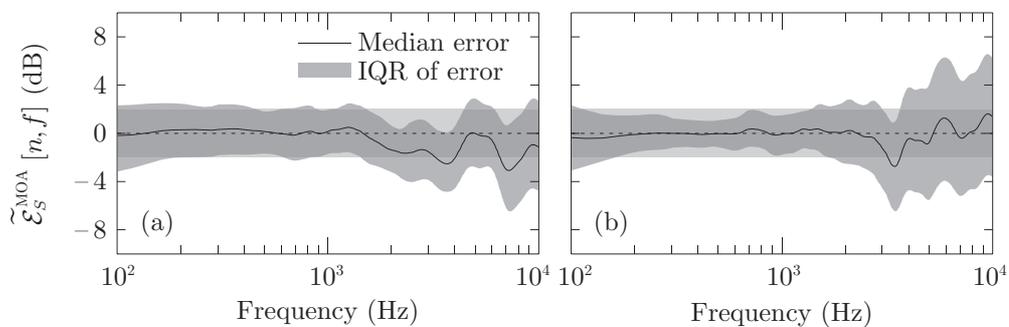


FIGURE 6.9: Same as Fig. 6.7 albeit showing the median and IQR of the pressure errors  $\mathcal{E}_S^{\text{MOA}}$  incurred by the ‘MOA’ VSE at the BBF output.

### 6.3.1.2 SNR benefit errors

Figure 6.10 shows the reference median SNR benefit ( $\widetilde{\Delta\text{SNR}}^{\text{Room}}$ ), as estimated inside the ‘Room’. The SNR benefit was estimated for both directional algorithms compared to an omnidirectional microphone (the front microphone of the HA satellites). Figure panels (a) and (b) illustrate an advantage of about 2–3 dB at middle frequencies offered by the BBF algorithm, compared to the CRD, to both ears. The plots also suggest a higher overall benefit of both directional strategies for the right ear. The additional short-time SNR benefit offered by BBF, compared to CRD, agrees with the significant improvement of intelligibility scores reported by [Best \*et al.\* \(2015\)](#).

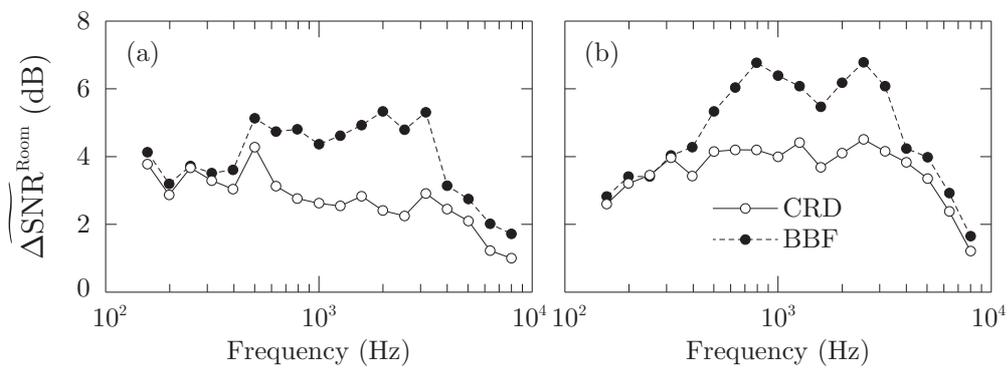


FIGURE 6.10: Median of the  $\Delta\text{SNR}^{\text{Room}}[n, f]$  relevant to both CRD and BBF algorithms, compared to an omnidirectional microphone, at the output of the: (a) left, and (b) right HA satellites. The values were estimated inside the ‘Room’ environment only.

Figures 6.11 and 6.12 show the median and IQR of the SNR benefit errors  $\mathcal{E}_{\Delta\text{SNR}}^{\text{Model}}[n, f]$  and  $\mathcal{E}_{\Delta\text{SNR}}^{\text{MOA}}[n, f]$  respectively, for both CRD/BBF algorithms. Similar to the pressure errors described in Section 6.2.3.1, the incurred short-time SNR benefit errors appear higher in the ‘Model’ than in the ‘MOA’ VSE. For the ‘Model’ VSE, the median errors are mostly limited to the  $\pm 2$  dB range with the IQRs extending as much as  $\pm 5$  dB about the median. For the ‘MOA’ VSE, the median errors are limited to the  $\pm 1$  dB range while the IQRs depend heavily on the directional algorithm, with the CRD yielding very low values (not exceeding the  $\pm 0.5$  dB range) and the BBF yielding considerably higher values (not exceeding the  $\pm 2.5$  dB range).

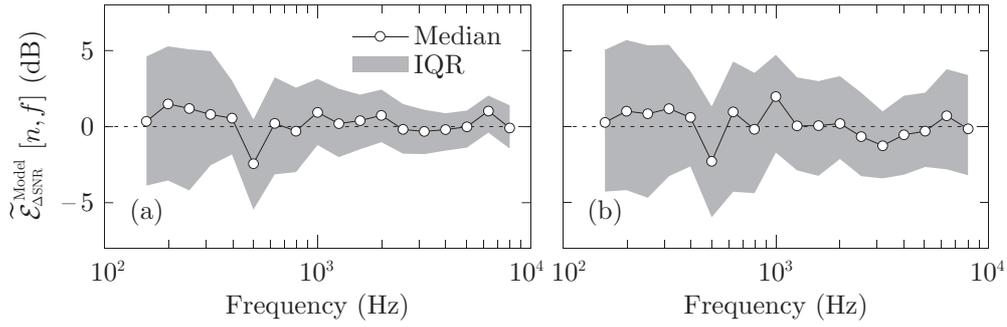


FIGURE 6.11: Median and IQR of the SNR error  $\mathcal{E}_{\Delta\text{SNR}}^{\text{Model}}[n, f]$  incurred by the ‘Model’ VSE and measured at the output of the: (a) CRD, and (b) BBF processing.

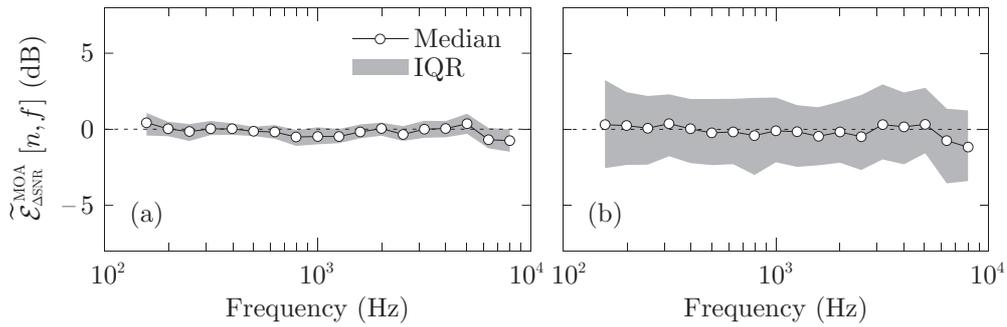


FIGURE 6.12: Same as Fig. 6.11 albeit showing the median and IQR of  $\mathcal{E}_{\Delta\text{SNR}}^{\text{MOA}}[n, f]$  incurred by the ‘MOA’ VSE.

## 6.3.2 Subjective measures

### 6.3.2.1 Cardioid microphone performance

The performance of the CRD processing can be derived from the SRTn values that were adaptively measured inside the 3 considered acoustic environments (Section 6.2.4.3). Figure 6.13 shows the derived SRTn values for all  $N = 18$  subjects, sorted by increasing four-frequency average hearing loss – 4FAHL (hearing threshold average at 500, 1000, 2000, 4000 Hz). The SRTn and 4FAHL values exhibit a strong correlation inside all 3 environments [Room:  $r = 0.65$ ,  $p = 0.004$ ; Model:  $r = 0.77$ ,  $p < 0.001$ ; MOA:  $r = 0.63$ ,  $p = 0.005$ ] and show a very similar behaviour to the data measured by Glyde *et al.* (2012, Fig. 4).

Figures 6.14(a) and (b) further illustrate that the SRTn estimations inside the ‘Model’ and ‘MOA’ VSEs closely follow the estimations inside the relevant reference environments (i.e., ‘Room’ and ‘Model’ respectively). The slopes of the linear regression lines are close to the ideal value of unity in both the Room/Model and Model/MOA plots, qualitatively suggesting that the estimated CRD microphone performance is on average well predicted inside the considered VSE.

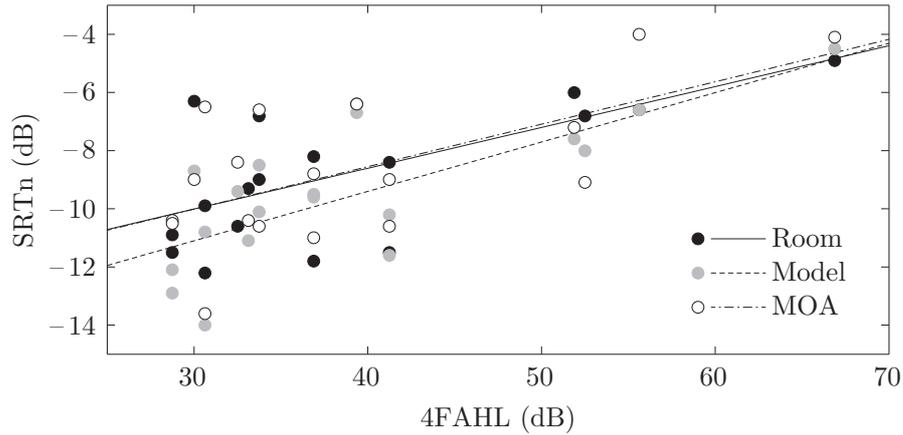


FIGURE 6.13: Speech reception thresholds in noise (SRTn), as derived in all acoustic environments, plotted as a function of 4FAHL. Linear regression lines are also shown [Room:  $r = 0.65$ ,  $p = 0.004$ ; Model:  $r = 0.77$ ,  $p < 0.001$ ; MOA:  $r = 0.63$ ,  $p = 0.005$ ].

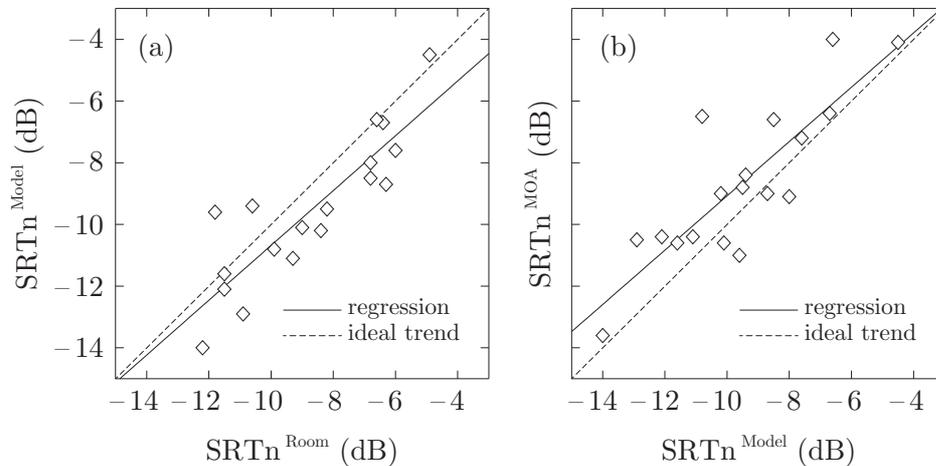


FIGURE 6.14: Plot (a) shows the SRTn values derived in the ‘Model’ VSE as a function of the values derived inside the reference ‘Room’ environment. In a similar way, plot (b) shows the SRTn values, derived in the ‘MOA’ VSE, as a function of the values derived in the ‘Model’ VSE. Regression lines [Room/Model:  $r = 0.87$ ,  $p < 0.001$ ; Model/MOA:  $r = 0.84$ ,  $p < 0.001$ ] are plotted and compared to the ideal, error-free (slope = 1; dashed line) scenario.

The mean SRTn error is  $-0.8$  dB when comparing the ‘Model’ and ‘Room’ environment and  $0.9$  dB when comparing the ‘Model’ and ‘MOA’ VSEs. Such differences appear well within one standard deviation of the expected test-retest accuracy and thus, may be considered of low practical relevance. According to [Keidser \*et al.\* \(2013b\)](#), the standard deviation of the test-retest accuracy for the SRTn, measured with the herein applied BKB sentences, albeit presented in diffuse babble noise, is about  $1.1$  dB. For the current experiment, the test-retest accuracy is expected to be even larger due to the stronger fluctuations inside the applied multi-talker masker. Comparing the SRTn values between the three acoustics environments, a paired t-test provided a significant difference for both the ‘Model-Room’ [ $t(17) = 2.89$ ,  $p = 0.010$ ] and the ‘MOA-Model’ comparison [ $t(17) = 2.68$ ,  $p = 0.016$ ].

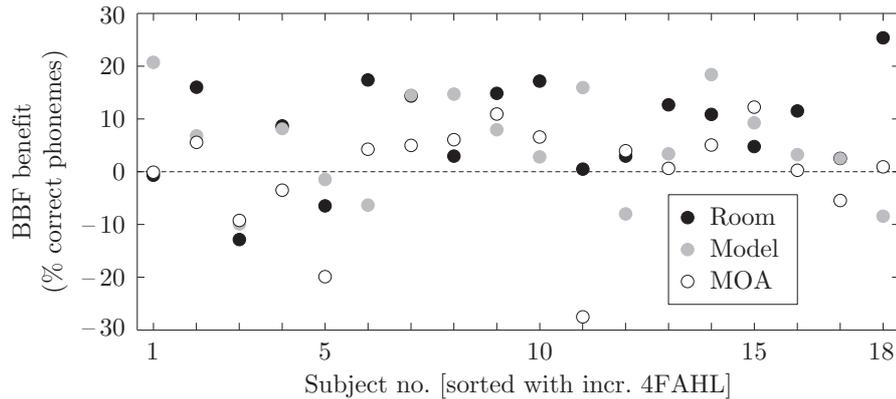


FIGURE 6.15: Intelligibility benefit offered by the BBF compared to the CRD processing, as derived in all 3 acoustic environments for all 18 subjects. The benefit is quantified as additional percent-correct phoneme recognition.

Concerning the use of the t-test, it should be noted that the data was tested for normality using the one-sample Kolmogorov-Smirnov test but normality could not be shown, which most likely was due to the small sample size of  $N = 18$  hearing impaired subjects. However, plotting the histograms of our data did not reveal any obvious deviations from a normal distribution. Moreover, the robustness of the t-test to deviations from the normality assumption makes that choice further justified. Finally, employing t-tests facilitated the direct comparison of the present results with those of [Best \*et al.\* \(2015\)](#).

### 6.3.2.2 Beamformer benefit on speech intelligibility

The average intelligibility benefit offered by the BBF, compared to the CRD microphone, is presented in Fig. 6.15 as the additional percentage of phonemes that each subject successfully identified inside all acoustic environments. Similar to [Best \*et al.\* \(2015\)](#), the measured benefit varies strongly between subjects, ranging from about  $-30\%$  to almost  $+30\%$ . However, a tendency towards positive percentage values can be observed for all three acoustic environments, suggesting that the BBF, in general, provided a benefit in addition to the CBD processing.

Figure 6.16(a) presents the mean plus/minus one standard deviation for the BBF benefit on the percentage-scores of correct phoneme recognition as derived inside all three acoustic environments. Inside the ‘Room’ environment a small mean benefit of 8% is provided by the BBF processing, which according to a paired t-test is highly significant [ $t(17) = 3.53, p = 0.003$ ]. The BBF benefit is still significant in the ‘Model’ VSE [ $t(17) = 2.35, p = 0.031$ ], but the mean value is slightly reduced to 5%. In the case of the ‘MOA’ VSE no significant benefit could be observed [ $t(17) = 0.10, p = 0.921$ ] on the BBF benefit, which averaged to an almost zero value. However, considering the scatter-plot dots of Fig. 6.16, it can be seen that the majority of the subjects do

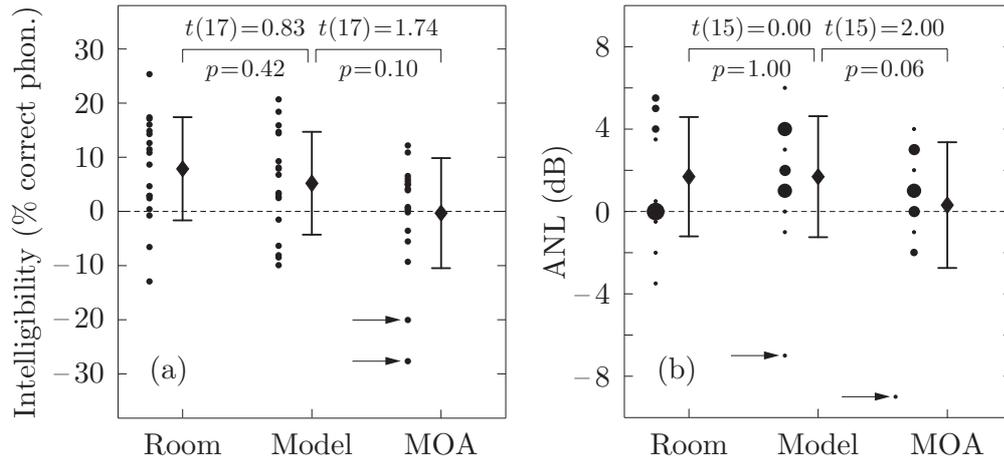


FIGURE 6.16: Mean and scatter plots illustrating the benefit of the BBF, compared to the CRD, on: (a) intelligibility (additional % correct phonemes), and (b) ANL (in dB). The whiskers show  $\pm 1$  standard deviation about the mean, which is represented by the diamond. The scattered dots show all benefit observations, while the outlying values are indicated with the right-pointing arrows. The size of the dots in plot (b) increases with the number of underlying observations.

actually provide a small benefit, but the mean benefit is highly influenced by the very large negative benefits measured for subjects no. 5 and no. 11 (indicated using arrows). Excluding these two subjects from the data analysis would result in an almost significant BBF benefit [ $t(15) = 1.91, p = 0.075$ ] of almost 3% for the ‘MOA’ VSE, while practically having no effect on the benefits measured in the other two environments. Although the responses of these two subjects across the different conditions as well as in-between the two 32-sentence block testing (Section 6.2.4.3) show some suspicious inconsistencies, no hard argument can be provided that would justify their exclusion from the statistical analysis.

A paired t-test revealed neither a significant difference in BBF benefit between the ‘Room’ and ‘Model’ environments [ $t(17) = 0.83, p = 0.42$ ] nor between the ‘Model’ and ‘MOA’ environments [ $t(17) = 1.74, p = 0.10$ ], as seen in Fig. 6.16. But even if significant differences would have been found (potentially by increasing the sample size), the practical importance of changes in phoneme recognition rate by about 3–5% should be regarded as debatable.

Similar to Best *et al.* (2015), we additionally analysed a number of other potential dependencies, including the BBF intelligibility benefit as a function of the SNR as well as the 4FAHL. However, in agreement with the results of Best *et al.* (2015), neither showed any dependency for any of the three acoustic environments.

### 6.3.2.3 Beamformer benefit on acceptable noise level

The effect of the testing acoustic environment on the BBF acceptable-noise-level (ANL) benefit was found to be very similar to the behaviour observed in the speech-in-noise test (Section 6.3.2.2). The BBF benefit on the ANL is illustrated in Fig. 6.16(b), which indicate the mean plus/minus one standard deviation. Individual benefits are shown by dots, with the size of the dots illustrating the number of underlying observations, i.e., number of subjects exhibiting equal benefit values. A small BBF benefit was found inside the ‘Room’ environment (mean:  $1.7 \pm 2.9$  dB) as well as in the ‘Model’ VSE (mean:  $1.7 \pm 2.9$  dB). A paired t-test revealed that both benefits were significant [Room:  $t(15) = 2.33, p = 0.034$ ; Model:  $t(15) = 2.30, p = 0.036$ ] and not significantly different from each other [ $t(15) = 0.00, p = 1.00$ ].

On the contrary, the BBF benefit on the ANL (mean:  $0.3 \pm 2.2$  dB) was not found to be significant in the ‘MOA’ VSE [ $t(15) = 0.41, p = 0.688$ ] and was almost significantly reduced in comparison to the benefit measured in the ‘Model’ VSE [ $t(15) = 2.00, p = 0.06$ ]. However, if the single outlier (subject no. 5) is excluded from the analysis of all environments, then statistical significance remains in the ‘Room’ and ‘Model’ environments [ $t(14) = 2.99, p = 0.010$  and  $t(14) = 4.70, p < 0.001$  respectively], while the benefit in the ‘MOA’ VSE almost reaches significance [ $t(14) = 1.97, p = 0.068$ ].

## 6.4 Discussion

### 6.4.1 Objective analysis

#### 6.4.1.1 Pressure reconstruction errors

Section 6.3.1.1 evaluated the pressure errors at the left HA output, as incurred by the ‘Model’ and ‘MOA’ VSEs, for both CRD and BBF processing schemes. The errors were computed with reference to the ‘Room’ and ‘Model\*’ environments respectively. Both the target and the combined distractors were analysed.

The ‘Model’ VSE incurred notable short-time pressure errors of up to  $\pm 20$  dB at the output of the CDR processing [Figs. 6.6(a) and 6.6(c)]. The quasi-random deviations can be explained by: (1) the stochastic nature of generating the spatial room impulse response in the LoRA toolbox (Favrot and Buchholz, 2010b; Oreinos and Buchholz, 2015a), and (2) the unspecified phase of the specular reflections in the ODEON simulation software (Christensen and Koutsouris, 2013). Considering these limitations, the ‘Model’ VSE can only be expected to correctly reconstruct the pressure in terms of

long-term statistics, but not in terms of the detailed temporal structure. This may be compared to measuring a signal at two different room locations, where the detailed reflections may notably differ, although the underlying statistics and long-term behaviour exhibit a similar pattern.

In Section 6.2.1.2 (Fig. 6.4) it has already been confirmed that the long-term statistics of the considered ‘Model’ RIRs, as estimated by a number of room acoustic parameters, closely agree to the ‘Room’ environment parameters. Figures 6.6(b) and 6.6(d) furthermore show that the median values of the pressure errors for both the target and distractor signals are confined within a region of about  $\pm 2$  dB. Comparing the IQRs of Fig. 6.6(b) and 6.6(d) illustrates that the error range for the target signal is smaller than for the combined distractors. This can be explained by the different diffuseness of the two signals; the nearby target energy is dominated by the direct signal while in the case of the combined distractors most of the energy reaches the HA microphones via reflections. Hence, while the target direct signal can be described deterministically in the ODEON/LoRA framework, the reflections of the combined distractors incur: (1) random errors due to the stochastic nature of the reconstruction method (see above), and (2) errors due to the limited prediction accuracy of the acoustic properties of the materials used in the room acoustics model.

Similar observations are derived from the time-averaged plots of Fig. 6.7 for the processed outputs of the BBF algorithm. In this case, the errors appear slightly increased by roughly  $\pm 1$  dB compared to the CRD outputs. This can be attributed to the BBF processing being more sensitive to the short-time interaural cross-coherence, which is influenced by the detailed behaviour of the reverberation in a short-time level.

Examining the short-time pressure errors for the ‘MOA’ VSE (Fig. 6.8), reveals that the errors incurred by ‘MOA’ are smaller than those incurred by the ‘Model’ VSE. Although the errors show a quasi-random behaviour across successive time blocks, the deviations are much smaller than those observed for the ‘Model’ VSE (Fig. 6.6; note the different colourbar scales). The sound-field coding strategy applied in the ‘MOA’ VSE directly reproduces the phase and time information of the original sound field (i.e., the ‘Model\*’ VSE in this case), and any incurred pressure errors are solely due to the limited accuracy of the applied coding strategy and not due to any stochastic processes as used for the ‘Model’ VSE creation (Oreinos and Buchholz, 2015a). In this case, the median error related to the processed target is larger than the median error related to the combined distractors. However, the opposite trend is observed for the IQR of the errors. These findings can be interpreted using the results presented in Oreinos and Buchholz (2015a) about the Ambisonics errors when reconstructing reverberant sound-fields. It was thereby shown that the reverberation effectively provides a spatial smoothing to

the sound field and thereby decreases the mean pressure errors incurred by Ambisonics. Even though the coding of each individual reflection component will be limited by Ambisonics like any other sound-field component (Ward and Abhayapala, 2001; Moreau *et al.*, 2006), when averaging across all reflection components these individual errors appear to partially cancel each other. Applying this observation to the ‘MOA’ VSE explains why the more diffuse distractors exhibit a lower time-averaged error than the more direct-signal-dominated target. It also explains the increased error IQR observed in the combined distractors case [Fig. 6.8(c)] through an increased number of Ambisonics reconstruction errors associated with individual components.

A very similar behaviour of the pressure error as discussed for the CRD processing is observed for the BBF processing in the MOA VSE (Fig. 6.9). In this case the errors are slightly more pronounced than for the CRD case, suggesting that MOA coding impairs sensitive spatial interaural-coherence cues which are utilised by the BBF processing.

#### 6.4.1.2 SNR benefit errors

Figure 6.10 illustrates that the time-averaged benefit  $\Delta\text{SNR}[n, f]$  provided by the BBF processing in the ‘Room’, is about 2 – 3 dB higher, within a frequency range of 400 – 4000 Hz, than the benefit provided by the CRD processing. That theoretical  $\Delta\text{SNR}$  estimation agrees with the observed significant intelligibility advantage offered by the BBF scheme to the hearing impaired listeners that participated in the conducted experiment (see Section 6.3.2.2).

Comparing Figs. 6.10(a) and 6.10(b) shows that both the CRD and BBF processing schemes provide a higher SNR benefit at the right ear of the HATS than at the left ear. This is due to the nearby distractors shown in Fig. 6.2 that are on the right side of the acoustic manikin. Due to their proximity, these sources provide a significant portion of the total distractor energy and are dominated by their direct sound component. Hence, the CRD and BBF processing attenuate these nearby sources more efficiently than the more distant, and thus diffuse, sources lying ipsilateral to the left side.

The short-time analysis of the SNR benefit errors  $\mathcal{E}_{\Delta\text{SNR}}[n, f]$ , incurred by the ‘Model’ and ‘MOA’ VSEs, produced similar quasi-random errors (not presented due to space limitations) as was shown for the short-time pressure errors (e.g., Figs. 6.6 and 6.8). The time-averaged errors  $\mathcal{E}_{\Delta\text{SNR}}$  were also found to agree with the relevant conclusions drawn in Section 6.4.1.1. Hence, Fig. 6.11 shows a small median error  $\mathcal{E}_{\Delta\text{SNR}}^{\text{Model}}$  but a considerable IQR of almost  $\pm 5$  dB. The IQR region is also more pronounced at the BBF output, especially at high frequencies, given the more aggressive processing that is more sensitive to short-time errors in the interaural coherence.

Figures 6.11(a) and 6.11(b) illustrate a systematic error incurred by the ‘Model’ VSE at 500 Hz which cannot be easily explained. This error was not apparent in the pressure error plots of Figs. 6.6 and 6.7 or even in the room acoustic parameters ( $T_{20}$ ,  $T_{EDT}$ ) plots of Fig. 6.4 and may be due to some destructive phase-interaction of the different sound-field components.

As was already observed for the pressure errors in the ‘MOA’ VSE (Section 6.4.1.1), the averaged SNR benefit errors  $\mathcal{E}_{\Delta\text{SNR}}^{\text{MOA}}$  presented in Fig. 6.12, exhibit a much more controlled behaviour, than in the ‘Model’ VSE, which is due to the direct reconstruction of the reference ‘Model\*’ sound-field that does not include any stochastic processes. Specifically for the CRD output at the left HA, the benefit  $\Delta\text{SNR}$  can be reproduced with minimal deviations up to even 8 kHz. The small error values can again be attributed to the spatial smoothing effect that reverberation has on the Ambisonics errors (Oreinos and Buchholz, 2015a). The median SNR benefit errors at the output of the CRD processing at the right ear (plot not shown) are very similar to the left ear, but the IQR is slightly increased by no more than 1 dB above approximately 5 kHz. This minor increase can be explained by the less diffuse nature of the field at the right HA (see Fig. 6.2).

Figure 6.12(b) illustrates that the averaged errors  $\mathcal{E}_{\Delta\text{SNR}}^{\text{MOA}}$  observed for the BBF processing inside the ‘MOA’ VSE exhibit a larger, than observed for the CRD processing, IQR of about  $\pm 2$  dB. Again, this is due to the BBF processing being very sensitive to the interaural coherence (Mejia *et al.*, 2007; Mejia and Dillon, 2010), which is disturbed by the short-time pressure errors incurred by the ‘MOA’ VSE.

The effect of the short-time pressure errors incurred by the VSEs on the interaural coherence have been analysed in this study only indirectly, via the measured decrease in performance of the BBF processing. However, the interaural coherence is a sound-field property that is important from a perception point of view Blauert (1997); Boehnke *et al.* (2002) as well as for other technical applications such as coherence-based noise-reduction and de-reverberation methods (Westermann *et al.*, 2013). Hence, the effect of VSEs on the interaural coherence should be systematically investigated, but this lies out of the current study scope.

## 6.4.2 Subjective evaluation

### 6.4.2.1 Cardioid microphone performance

The adaptive SRTn data measured with 18 HI subjects fitted with directional (CRD) HAs show very similar behaviour across the three acoustic environments (Section 6.3.2.1).

Hence, the SRTn data are highly correlated between the ‘Room’ and the ‘Model’ environment [Fig. 6.14(a);  $r = 0.87$ ,  $p < 0.001$ ] as well as between the ‘Model’ and ‘MOA’ VSEs [Fig. 6.14(b);  $r = 0.84$ ,  $p < 0.001$ ]. Moreover, all the environments show an almost identical significant dependency on the 4FAHL (Fig. 6.13). Errors of less than 1 dB were observed when considering mean SRTn differences between environments. These errors across environments were moreover significant for both the ‘Room-Model’ and the ‘Model-MOA’ comparison.

However, such small absolute errors may be considered of minor practical relevance, considering that the mean errors are significantly smaller than: (1) the mean benefit of 3 – 4 dB expected by cardioid HA microphones (Keidser *et al.*, 2013a), and (2) the test-retest accuracy of the applied test procedures (Keidser *et al.*, 2013b). The previous authors have shown that the standard deviation of the test-retest accuracy is about 1.1 dB for speech reception thresholds measured with the herein applied BKB sentences, albeit presented in diffuse babble noise. In the present study, an even reduced test-retest accuracy (i.e., higher standard deviation) is expected due to the increase in the temporal fluctuations and in the spatial and semantic complexity provided by the multi-talker distractor environment (see Fig. 6.2). Further differences in SRTn values between environments may have also been introduced by the HA fitting with the highly-occluding foam tips (Section 6.2.4.2). Moreover, due to practical constraints, the data for the ‘Room’ environment were always measured during the first testing appointment and the data for the ‘Model’ and ‘MOA’ VSEs were measured during the second appointment. Hence, slight differences in the fitting of the HAs could have resulted in small systematic differences between the ‘Room’ and the ‘Model’/‘MOA’ environments.

#### 6.4.2.2 Beamformer benefit on speech intelligibility

The mean benefit in phoneme-scores of 8% that was observed for the BBF processing in the ‘Room’ environment (Section 6.3.2.2), when compared to the CRD processing, is very similar to the mean benefit of 7% observed by Best *et al.* (2015) for the same HA processing inside a comparable seven-talker background environment. This BBF benefit was clearly significant in the present study as well as in Best *et al.* (2015). Although the exact psychometric function is not known, a 8% benefit measured at approximately 2 dB below the 50% correct point (Section 6.2.4.3) may suggest an effective (mean) SNR benefit of roughly 1 dB. This value is smaller than the 2 – 3 dB benefit objectively measured on a manikin (Section 6.3.1.2). This reduction in the BBF benefit could be explained by: (1) small head movements (or misalignments) during the listener testing, which could have reduced the effectiveness of the directivity provided by the BBF processing, and

(2) distortions, incurred by the BBF processing, that reduce the intelligibility of the negative-SNR target.

It should be emphasised at this stage that the relatively small BBF intelligibility benefit, as estimated even in the ‘Room’, was due to the complex nature of the distractors scene, as well as the low SNR values at which the algorithms was operating. However, in environments with more temporal fluctuations, i.e., smaller number of less diffuse distractors, higher benefits can be achieved (Best *et al.*, 2015).

Moving to the ‘Model’ VSE, a slight decrease in the BBF benefit to 5% was observed (Fig. 6.16), which was still significant. A paired t-test comparison revealed no significant difference between the benefit observed in the ‘Room’ and ‘Model’ environments. Additionally, the BBF benefit measured inside the ‘Model’ VSE was found to be statistically significant. This fact suggests that the realised ‘Model’ environment can be similarly employed as the real ‘Room’ environment to test highly directional bilateral signal processing schemes, such as the present BBF algorithm.

The conclusions for the effect of the ‘MOA’ environment on the BBF benefit are not as straightforward. The measured benefit showed similar variations across subjects as for the other two environments, but the mean benefit was zero and a paired comparison with the ‘Model’ VSE revealed almost significant differences. However, excluding the two outliers from the statistical analysis, as discussed in Section 6.3.2.2, would result in an almost significant BBF benefit of roughly 3%. This suggests that the ‘MOA’ VSE does reconstruct most of the cues employed by the BBF processing, but due to the already small benefit of 5%, as observed in the reference ‘Model’ environment, the slightly deteriorated cues render that small benefit just insignificant. Hence, it can be concluded that the ‘MOA’ (as well as the ‘Model’) VSE slightly reduces the sensitivity of the considered subjective benefit measure, which can become a problem if very small benefits have to be resolved. However, since the main tendencies of the BBF processing seem to be retained in the ‘MOA’ VSE, the sensitivity might have been restored by considering an increased number of test subjects and, thus, reducing the influence of the outlying observations.

#### 6.4.2.3 Beamformer benefit on acceptable noise level

The BBF benefit on the ANL, as measured inside the ‘Room’, turned out to be 1.7 dB, which lies below the mean value of 3.6 dB reported by Best *et al.* (2015). This discrepancy could be due to: (1) the differences in the multi-talker distractor environments used in the two studies, or (2) to the milder hearing losses involved in the subjects of this study.

That ANL BBF benefit was, however, shown to be statistically significant inside the ‘Room’ environment.

When considering the ‘Model’ and ‘MOA’ VSEs, very similar conclusions can be drawn as for the BBF benefit on the intelligibility scores (see Section 6.4.2.2). Hence, the ‘Model’ VSE resulted in an indistinguishable ANL benefit estimation, which was statistically significant. The ‘MOA’ VSE resulted in an insignificant almost zero-value ANL benefit (0.3 dB). Removing the single outlier from the statistical analysis, as discussed in Section 6.3.2.3, would result in an almost significant BBF benefit in ANL of 0.9 dB. Hence, as already discussed in Section 6.4.2.2 for the intelligibility measurements, the described VSEs seem to retain most of the cues that are important for the BBF (as well as CDR) processing, but their small sound-field deteriorations do slightly reduce the different subjective outcome measures.

## 6.5 Conclusion

The present study investigated the applicability of two different methods for creating loudspeaker-based virtual sound environments (VSE) to evaluate different directional HA signal processing methods. Different objective and subjective performance measures were applied, and results were compared between the VSEs and the corresponding real environment. The real environment consisted of a target talker and seven pairs of distracting talkers, all realised by loudspeakers, spatially distributed inside a large-size meeting room. The first (‘Model’) VSE applied room acoustic modelling techniques to create a computer-based model of the real environment, which was then recreated inside a 3D loudspeaker array. The second (‘MOA’) VSE recreated the modelled environment using a simulation of an entire mixed-order Ambisonics (MOA) reproduction chain, including the limitations introduced by a realistic microphone array. Two different directional HA processing methods were investigated: (1) a first-order directional (cardioid) microphone (CRD) implemented separately on two behind-the-ear HAs placed on the left and right ear of a listener, and (2) an adaptive bilateral beamformer (BBF) that provided an increased directionality by combining the output of the two directional (CRD) HAs.

The objective analysis considered the short-time and long-term pressure spectrum and SNR benefit at the output of the two directional HA signal processing methods. It was shown that both the long-term pressure and SNR benefit were very well reproduced by both VSEs, with the error being slightly increased by the more aggressive processing of the BBF method. In contrast, the detailed behaviour of the short-time pressure varied significantly between environments, in particular for the ‘Model’ VSE. This was mainly

due to the stochastic processes that were applied to recreate the room reverberation, an effect that may be compared to the detailed changes observed when a receiver position is moved inside a real room.

The subjective evaluation employed 18 hearing impaired subjects to measure speech intelligibility as well as acceptable noise levels (ANL) using the CRD/BBF directional HAs, inside all three acoustic environments. It was shown that the behaviour and tendencies of the data were similar across environments, suggesting that the main benefits provided by the different HA processing methods could be preserved in the two VSEs. However, the overall benefit provided by the BBF processing was slightly reduced in the ‘Model’ VSE and even further reduced in the ‘MOA’ VSE. Due to the small mean benefits already observed in the real room (8% in phoneme scores and 3.9 dB in ANL), this resulted in an insignificant benefit inside the ‘MOA’ VSE. These observations indicate that the herein considered VSEs can be applied to test directional HA processing schemes, recognising however that the estimated benefits might be slightly reduced in comparison to the real environment.

Although the present results are encouraging, there is still room to improve the accuracy of the VSEs and particularly the ‘MOA’ one. For example, by conducting informal listening tests inside the loudspeaker array, we found that the diffuse multi-talker background (distractors) as reconstructed by ‘MOA’ was almost identical to the relevant ‘Model’ reconstruction. In contrast, the MOA processing clearly increased the apparent source width (Nowak *et al.*, 2013) of the direct sound-dominated target source, when compared to the ‘Model’ reproduction. The increased apparent source width may be related to the observed reduction in the benefits provided by the directional HA processing methods. This situation could have been remedied, potentially leading to much more accurate estimation results, if some technique (Meyer and Elko, 2008; Wabnitz *et al.*, 2011; Bernschütz, 2012; Khaykin and Rafaely, 2012; Noohi *et al.*, 2013) had been used to estimate the direction and time of arrival of the direct signal and a number of early reflections of the target. If such an estimate was available, the relevant specular reverberation components could have been directly mapped to the closest loudspeaker of the array, thus improving the reproduction of the target. Finally, it should also be emphasised that in future research the microphone array, which was only simulated here, should be replaced by a real array. That would provide direct encoding of the real-room acoustic environments, thus circumventing the need to rely on room acoustics modelling.

## **Acknowledgements**

This work was financially support by the HEARing CRC, established and supported under the Cooperative Research Centres Program – an initiative of the Australian Government. The authors would also like to thank Margot McLelland for the data collection, Cong-Van Nguyen for signal processing and implementation support, Jorge Mejia for providing us with the BBF algorithm and Mark Seeto for providing statistical support.

# Chapter 7

## Concluding discussion

### 7.1 Summary of findings

The aim of this thesis was to validate the use of virtual sound environments, primarily based on Ambisonics, for listening tests including hearing aids. A prerequisite to perform such a validation in an objective way consisted of measuring a dense, full-3D set of HRTFs to the in-ear and behind-the-ear hearing aid microphones of a head and torso simulator. The measuring procedure and its verification were presented in Chapter 3. A low-frequency extension method that combined a spherical head model with the measured responses was then proposed and validated. It was emphasised that the key advantage of that frequency extension method lies in its ability to maintain the sensitive amplitude and phase differences between the responses of the HA microphone pairs. Such an amplitude and phase accuracy is, in turn, essential for the implementation of optimum delay-and-subtract directional microphones (Kuk *et al.*, 2000; Kates, 2008).

The established HRTF data set was extensively used in later sections (Chapters 4 – 6) of this thesis in order to derive results and conclusions about the applicability of higher-order Ambisonics for listening test applications. Specifically, the HRTFs were used to simulate any given virtual source or virtual loudspeaker, i.e., transform source signals to in-ear and HA microphone signals. The complete HRTF data set was further made publicly available ([http://www.nal.gov.au/download/HATS\\_BTE\\_hrirDatabase.zip](http://www.nal.gov.au/download/HATS_BTE_hrirDatabase.zip)), with the expectation that it will assist the research community to create binaural representations of virtual auditory spaces that can be used for the research and development of HA algorithms. It is also hoped that it will be used by researchers pursuing similar work as herein presented, i.e., exploring the possibilities and limitations of virtual sound environments for administering realistic and ecologically valid listening tests (with or without HAs).

Proceeding to the main topic of this thesis, Chapter 2 summarised the main concepts and mathematical tools describing the higher-order Ambisonics coding. The notation of real-valued spherical harmonic functions was used, as applied by a number of researchers (Daniel, 2001; Moreau, 2006; Favrot *et al.*, 2011). The equivalence of the real-valued SHFs to the more commonly used complex-valued SHFs was demonstrated in

Appendix A. Most importantly, however, Chapter 2 presented and validated the ‘shape-matching’ encoding method. Although this method for encoding sound fields using practical microphone arrays had been previously presented (Moreau *et al.*, 2006), specific results proving its robustness and advantages did not exist in literature. The herein presented simulations illustrated that the ‘shape-matching’ encoding can effectively take into consideration the gain, phase and placement mismatches of the array microphones, while at the same time controlling the detrimental high-frequency spatial aliasing effects in a better way compared to the traditional ‘Ambisonics’ encoding (Daniel, 2001; Moreau *et al.*, 2006). In other words, it was demonstrated that the ‘shape-matching’ method inherently calibrates the microphone array, which would have otherwise been a tedious task.

Chapter 4 presented a unified VSE-creation framework, which was employed to yield some of the most fundamental conclusions of this work concerning HOA. To begin with, it was shown that the presence of a listener inside the reconstructed HOA sound-field introduced direction-dependent artefacts, i.e., virtual sources lying contralateral to the considered ear yielded higher pressure errors compared to ipsilateral virtual sources. That effect was even more pronounced when a spherical microphone array was employed for the sound-field encoding (as opposed to ideally HOA-coding virtual sources of known position).

Initial results pertinent to directional HAs were then derived by analysing the performance of a first-order cardioid microphone, implemented on head-worn HAs. The simulation results illustrated that the HOA sound-field did not degrade the HA directionality in the operation frequency range of HOA (i.e., up to roughly  $f_c = M \cdot 600$  (Hz),  $M$ : HOA order). Moreover, it was shown that when considering a HOA microphone for the sound-field encoding operation, the necessary regularisation can degrade the low-frequency performance of head-worn HA directional microphones (e.g., reduction of the directivity index DI by approximately 1 dB up to roughly 500 Hz for a regularisation of  $\lambda = 0.01$  and a spherical microphone array radius of  $R = 0.06$  m). It should be noted that such low-frequency errors were not observed on the pressure errors of the HA omnidirectional microphones which suggests that the errors are due to the differential operation of the directional microphones. This differential operation, in turn, increases the requirements for higher-order accuracy in the reconstructed sound field.

Most importantly, though, Chapter 4 considered the effects of reverberation on the HOA reconstruction errors. The simulation results illustrated that reverberant (i.e., in-room) virtual sources produced lower HOA pressure error, compared to anechoic virtual sources, when considering the responses in a time-integrated (i.e., long-term) way. The lower (in absolute values) pressure errors associated with reverberant sources

were explained through the spatially diffuse character of the reverberant sound scene that effectively provides a spatial smoothing to the field and thereby decreases the required accuracy (order  $M$ ) of the HOA reproduction system. However, it was emphasised that if the HOA pressure errors of reverberant source were to be calculated by considering a short-time analysis, then the spectral, temporal and spatial details of the individual reverberant components (or reflections) would have been limited in the same way as for anechoic sound sources.

Chapter 4 also presented a method to fine-tune the HOA microphone regularisation by considering the actual noise generated at the ears and HA microphone of a listener, seated at the centre of the reproduction loudspeaker array. That method was further pursued in a co-authored paper (Appendix B) that looked into optimising the mixed-order Ambisonics (MOA: a variant of higher-order Ambisonics) microphone regularisation. It was thereby illustrated for the example microphone array configuration ( $R = 0.05$  m,  $Q = 62$  sensors) that above about 500 Hz, setting the regularisation such that the generated MOA system noise lies below the normal-hearing audibility threshold would compromise the sound-field pressure accuracy.

The previous fundamental conclusions were further put to test in Chapter 5 in the context of two adaptive directional HA microphone schemes operating inside a HOA sound-field. Here, the HOA accuracy was implicitly analysed in a short-time sense since the HA algorithms were time-adaptive and thus relied on accurate timing cues. Despite these increased accuracy requirements, introducing the natural in-room reverberation of virtual sources was shown to reduce the (internal; not observed) adaptive coefficient errors as well as the (observed) SNR benefit and pressure errors. However, it was also shown that the HOA accuracy improvement, due to taking the reverberation into account, was lower when considering the adaptive bilateral beamformer output, compared to when considering the output of an adaptive first-order directional microphone. The sensitivity of the more aggressive bilateral beamformer scheme on the reconstructed sound-field errors was attributed to its use of binaural cross-correlation cues which were impaired by the HOA reconstruction.

A listening experiment was then conducted (Chapter 6) aiming to validate the previous theoretical results and conclusions. For that purpose, a ‘cocktail party’ scene was created inside a real-room and then that environment was modelled using acoustic simulation software. After fine-tuning the model, so that a number of acoustic parameters matched with real-room measurements, the different reverberation components (direct signal, specular early reflections and diffuse late reverberation) were decoded to the NAL 41-loudspeaker array. Two different coding methods were considered to derive loudspeaker signals: (1) a direct single-loudspeaker mapping of components to

the closest loudspeaker of the array, and (2) a mixed-order Ambisonics encoding/decoding scheme. The MOA coding was implemented as a 4/7 order system (i.e.,  $M_{3D}=4$ ,  $M_{2D}=7$ ) that employed a simulated 62-microphone array, which at that time was being built at NAL. Having constructed a real and two virtual sound environments, the performance of a first-order cardioid (CRD) HA microphone and an adaptive bilateral beamformer (BBF) were analysed, both objectively (through short-time pressure and SNR benefit errors) and subjectively (through listening-in-noise and acceptable noise level tests administered on hearing-impaired subjects).

The objective analysis illustrated that the HA output short-time pressure errors generated by the room acoustic model, compared to the real-room scene, exhibited a quasi-random nature across the time frames ( $\tau_{ST} = 0.02$  s). These errors were attributed to the stochastic processes employed in the synthesis of the model room impulse response descriptions as well as the decoding operation of the previous RIR descriptions to loudspeaker signals. The time-average of these pressure errors appeared well-controlled suggesting that the temporal fluctuations may be similar to the natural changes in the room's response when measured at different room locations.

The short-time pressure errors due to the 'MOA' VSE, as compared to the model-based description based on which the Ambisonics coding was derived, were found to be lower than the errors due to the 'Model' VSE. Hence, despite the limited MOA order, the median pressure errors of the target and of all (summed) distractors were mostly limited in the  $\pm 2$  dB region for frequencies up to 10 kHz. This held true at the output of both HA directional schemes.

The short-time errors of the directional HA benefit in SNR exhibited very similar trends with the above-described pressure errors for both the 'Model' and 'MOA' VSEs. The errors associated with the 'MOA' VSE appeared almost surprisingly low for the CRD operation. However, an increased spread of errors was demonstrated for the BBF processing.

The subjective speech intelligibility test conducted inside the real-room environment and inside the two VSEs suggested that the CRD microphone performance was well predicted in all acoustic environments. Moreover, the results relevant to the BBF operation illustrated that the statistically significant intelligibility benefit of 8 percentage points (in average) was underestimated, but still significant, in the 'Model' VSE. However, the 'MOA' errors further reduced the benefit yielded by the modelled scene so that no statistical significance was eventually observed between the BBF and CRD performance.

Finally, the estimation of the acceptable noise level inside all environments reinforced the above conclusion, i.e., that the 'Model' VSE still provides statistical significance, for

the BBF benefit over the CRD, but at a lower benefit, and that the ‘MOA’ VSE fails to produce any significant benefit in the considered experiment setup, and incorrectly suggests that the BBF does not offer any advantage compared to the CRD processing.

## 7.2 Limitations of this work

Much as this work offers valuable conclusions about the applicability and validity of VSEs for the evaluation of hearing aids, it is also characterised by a number of limitations. To start with, the considered simulations and the conducted experiment involved only directional HA schemes (i.e., a static cardioid microphone, an adaptive first-order directional microphone and an adaptive bilateral beamformer). The rationale for considering these spatially-selective processing schemes is that they were expected to be particularly sensitive inside a reconstructed sound-field since their output is strongly affected by the exact spatial, temporal and spectral information of the signals picked up by the HA microphones. However, the performance of other important HA processing schemes, such as dynamic range compression, noise reduction, frequency compression, etc. inside HOA-based VSEs was not investigated. Such an analysis could have offered a better understanding of the complex interactions between the VSEs and the HA performance.

Moving to the objective part of the analysis presented in this thesis, some binaural quantities, and most notably the interaural cross-coherence, could have been additionally examined. Such an analysis is expected to have given better insights on the localisation and spaciousness distortions incurred by the VSEs. For example, a similar study looking only into the cross-correlation metric, was presented by [Avni and Rafaely \(2009\)](#).

Perhaps the most crucial limitation of the experimental part of this work has been the inability to use the 62-microphone array that was still under development at the National Acoustic Laboratories at the time when the experiment was scheduled to take place. If the array had been available at the time of the experiment, actual room recordings would have been used in order to perform a fair validation of Ambisonics. In that case the ‘MOA’ VSE would have been directly comparable to the reference real-room environment, thus providing more meaningful comparisons and results.

Given the range of values in the experimental results (e.g., the separate %-correct phoneme scores when using the CRD or BBF, from which the presented BBF intelligibility benefits were calculated), a larger number of hearing impaired participants would have been highly desirable. Moreover, listeners with more severe hearing loss

profiles could have added valuable information, since they are expected to gain a higher benefit from the BBF directional scheme.

On the other end of the spectrum, listening tests with normal hearers could have been beneficial in order to identify the key perceptual drawbacks of the VSEs. For example, informal listening tests conducted by two experienced listeners (including the author of this thesis) pointed to the localisation blur of the MOA-coded target as probably being the primary issue for the observed directional HA benefit errors. Identifying such artefacts can appear useful in suggesting improvements to the methods employed to create VSEs.

### 7.3 Recommendations and prospects for future work

This thesis has presented objective and subjective results that support using virtual sound environments (main focus has been given on Ambisonics-based ones) for listening tests involving hearing aids. We should emphasise that at the present stage the subjective results (Chapter 6) demonstrated that the limited accuracy of the applied Ambisonics sound-field reconstruction (i.e., when considering an applicable  $M = 7$  order HOA system, which necessitates no more than 64 loudspeakers and microphones) reduces the sensitivity of a listening test, and thereby, may underestimate the benefits provided by a directional signal enhancement method. However, the limitations of the study of Chapter 6 suggest that some improvement can be expected, should those limitations be carefully dealt with. Considering the observations and conclusions derived from the studies included in this thesis, we can make a number of recommendations, which should be interpreted more as a rough guide rather than as a universally applicable rule.

Hence we can support that a 7<sup>th</sup> order Ambisonics system appears as a viable compromise between sound-field reconstruction accuracy (e.g., providing a usable frequency range of about 6 – 7 kHz when reverberant scenes are considered, which should suffice given that the considered frequency range for most hearing applications is about 5 – 6 kHz) and feasibility of implementation (64 microphones and loudspeakers are required in minimum for the encoding and decoding of the sound field).

Such a 7<sup>th</sup> order HOA system ideally requires quasi-regularly positioned microphones and loudspeakers on the sphere in order for the discretisation errors to remain as low as possible. This restriction can be alleviated by using a mixed-order Ambisonics system. A MOA system provides the biggest advantage when the sound sources of interest lie on the horizontal plane, since a higher system order can then be used for the horizontal components, combined with a lower order for the elevated components. This scheme

then necessitates a ring of microphones and loudspeakers on the horizontal plane, which moreover, is easier to build. For example a 4/7 MOA system was considered in this work, yielding excellent objective analysis results while at the same time providing encouraging subjective results when used to create a VSE for advanced HA algorithm testing. Even higher horizontal accuracy MOA systems, e.g. operating at the orders of 4/11, lie within the range of a feasible design (e.g., adopting the configuration of the 62-microphone array recently built at NAL). Certainly, exploring the performance and limitations of MOA systems remains a field for future work.

A further recommendation, although still being an open research question, is related to the concept of applying a hybrid VSE, as briefly discussed at the last sections of Chapter 6. Following the author's strong belief that this practice can enhance the accuracy of Ambisonics-based VSEs, it is suggested that a direction-of-arrival method is employed in order to estimate the direct signal and first early reflections of the target signal (assuming it is required to reconstruct a 'cocktail party' environment for listening-in-noise experiments). These virtual sources should then be directly decoded to the closest loudspeaker of the reproduction array, letting the Ambisonics coding handle the diffuse components of the target's RIR. Prospective future work might investigate how to derive and validate the signal processing framework to implement this concept.

Finally, validating Ambisonics-based VSEs through the performance of other HA processing schemes (dynamic range compression, noise reduction, etc.) directly emerges as one of the next logical steps to pursue. Such research will provide even more insights to the field of applying virtual acoustic environments for testing hearing devices.



# Appendix A

## Employed HOA equation formulation

The definitions and notations differ between classic Fourier Acoustics (Williams, 1999) and the Higher-Order Ambisonics formulation followed by Daniel (2001). This work adheres to the latter (HOA) formulation. The steps that follow illustrate how the Fourier Acoustics notation is transformed to the HOA notation. The interested reader is also referred to Poletti (2009).

First, the role of  $m$  and  $n$  is interchanged ( $m \longleftrightarrow n$ ) so that  $m$  represents the degree and  $n$  the order in the HOA notation. Second, the HOA notation employs  $\theta$  instead of  $\phi$  to denote azimuth. Additionally, the elevation angle  $\delta$  is used in place of the inclination angle  $\theta$ , such that:

$$\theta = \frac{\pi}{2} - \delta, \quad (\text{A.1})$$

which results in the below transformation for the argument of the associated Legendre polynomials ( $P_{mn}$ ):

$$\cos \theta = \sin \delta. \quad (\text{A.2})$$

Applying the above changes, Eq. (6.20) of Williams (1999), is transformed to:

$$Y_m^n(\theta, \phi) \longleftrightarrow \tilde{Y}_{mn}(\theta, \delta) = \sqrt{\frac{(2m+1)(m-n)!}{4\pi(m+n)!}} P_{mn}(\sin \delta) e^{in\theta}, \quad (\text{A.3})$$

inclination
azimuth
azimuth
elevation

where the tilde ( $\tilde{\cdot}$ ) above  $Y_{mn}$  denotes a complex formulation of the spherical harmonics.

*Real-valued* spherical harmonic functions (SHF) can be derived from Eq. (A.3) by combining complex conjugate functions corresponding to opposite values of order  $n$ , as derived for example by Blanco *et al.* (1997) excluding nonetheless the Condon-Shortley phase factor  $(-1)^n$ :

$$Y_{mn}^{\Re}(\theta, \delta) = \begin{cases} \frac{1}{\sqrt{2}} \left( \tilde{Y}_{mn} + (-1)^n \tilde{Y}_{m\bar{n}} \right) = \frac{1}{\sqrt{2}} \left( \tilde{Y}_{mn} + \tilde{Y}_{mn}^* \right) & \text{if } n > 0 \\ \tilde{Y}_{m0} & \text{if } n = 0, \\ \frac{1}{i\sqrt{2}} \left( \tilde{Y}_{m\bar{n}} - (-1)^n \tilde{Y}_{mn} \right) = \frac{1}{i\sqrt{2}} \left( \tilde{Y}_{m|n|} - \tilde{Y}_{m|n|}^* \right) & \text{if } n < 0 \end{cases}, \quad (\text{A.4})$$

where  $(\cdot)^{\Re}$  denotes real-valued SHFs *normalised as in the Fourier Acoustics formulation*, the asterisk  $(\cdot)^*$  denotes complex conjugation and  $\bar{n} = -n$ . The  $(\theta, \delta)$  dependence is dropped in the right hand side of the equations for brevity. The above equation uses the following property of the SHFs, similar to [Williams, 1999](#), Eq. (6.44):

$$\tilde{Y}_{m\bar{n}}(\theta, \delta) = (-1)^n \tilde{Y}_{mn}^*(\theta, \delta). \quad (\text{A.5})$$

Substituting Eq. (A.3) to Eq. (A.4) yields:

$$Y_{mn}^{\Re}(\theta, \delta) = N_{m|n|} P_{m|n|}(\sin \delta) \cdot \begin{cases} \cos n\theta & \text{if } n \geq 0 \\ \sin |n|\theta & \text{if } n < 0 \end{cases}, \quad (\text{A.6})$$

with the orthonormalisation factor  $N_{mn}$  being equal to:

$$N_{mn} = \sqrt{\frac{(2m+1)}{4\pi} (2 - \delta_{0,n}) \frac{(m-|n|)!}{(m+|n|)!}}. \quad (\text{A.7})$$

The HOA notation defines the real-valued SHFs as:

$$Y_{mn}^{\sigma}(\theta, \delta) = \sqrt{(2m+1)(2 - \delta_{0,n}) \frac{(m-n)!}{(m+n)!}} P_{mn}(\sin \delta) \cdot \begin{cases} \cos n\theta & \text{if } \sigma = +1 \\ \sin n\theta & \text{if } \sigma = -1 \text{ (ignored if } n = 0) \end{cases} \quad \text{with } 0 \leq n \leq m, \quad (\text{A.8})$$

Comparing Eq. (A.6) and Eq. (A.7) to Eq. (A.8) illustrates that the HOA formulation of real-valued SHFs differs to the classic Fourier Acoustics formulation only by a factor of  $\sqrt{1/4\pi}$ , i.e., the real SHFs are orthogonal but *not* orthonormal. The HOA SHFs are normalised for unit-power instead. The extra factor  $1/4\pi$  remaining from the inner product of the SHFs is absorbed by the HOA components  $B_{mn}^{\sigma}$ .

Finally, the HOA formulation includes a parameter  $i^m$  in the interior problem solution as seen below:

$$p(kr, \theta, \delta) = \sum_{m=0}^{\infty} i^m j_m(kr) \sum_{n=0}^m \sum_{\sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \delta), \quad (\text{A.9})$$

which is not included in the Fourier Acoustics formulation [[Williams, 1999](#), Eq. (6.140)]:

$$p(r, \theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^n A_{mn} j_n(kr) Y_n^m(\theta, \phi). \quad (\text{A.10})$$

This  $i^m$  factor is again absorbed by the HOA components  $B_{mn}^\sigma$ . Including this factor and adopting the unit-power SHFs normalisation of Eq. (A.7), simplifies the HOA description of the plane-wave field expansion, which is expressed in Fourier Acoustics notation [Williams, 1999, Eq. (6.175)] as:

$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi \sum_{n=0}^{\infty} i^n j_n(kr) \sum_{n=-m}^m Y_n^m(\theta, \phi) Y_n^m(\theta_k, \phi_k)^*, \quad (\text{A.11})$$

with  $\mathbf{k} : (k = \|\mathbf{k}\|, \theta_k, \phi_k)$  being the direction of plane-wave propagation.

Applying all the above transformations yields the following plane-wave expansion according to the HOA formulation:

$$e^{i\mathbf{k}\cdot\mathbf{r}} = \sum_{m=0}^{\infty} i^m j_m(kr) \sum_{n=0}^m \sum_{\sigma=\pm 1} Y_{mn}^\sigma(\theta, \delta) Y_{mn}^\sigma(\theta_k, \delta_k), \quad (\text{A.12})$$

with  $\mathbf{k} : (k, \theta_k, \delta_k)$  here being the direction of plane-wave incidence.

Comparing Eq. (A.12) to Eq. (A.9) gives the HOA components expressing the field of a unit-amplitude plane wave arriving from position  $(\theta_k, \delta_k)$  simply as the SHFs sampled at the direction of the wave incidence:

$$B_{mn}^\sigma \Big|_{\text{pl.wave from } (\theta_k, \delta_k)} = Y_{mn}^\sigma(\theta_k, \delta_k). \quad (\text{A.13})$$

Table A.1 summarises the differences between the Fourier Acoustics and HOA equation formulation.

TABLE A.1: Comparison of 'Fourier Acoustics' (Williams, 1999) to 'higher-order Ambisonics' (Daniel, 2001) notation.

'Fourier Acoustics' notation	Employed 'higher-order Ambisonics' notation
Complex-valued spherical harmonics	Real-valued spherical harmonics
$Y_n^m(\theta, \phi) = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m(\cos\theta) e^{im\phi}$	$Y_{mn}^\sigma(\theta, \delta) = \sqrt{(2m+1)(2-\delta_{0,n})} \frac{(m-n)!}{(m+n)!} P_{mn}(\sin\delta) \cdot \begin{cases} \cos n\theta, & \sigma = +1 \\ \sin n\theta, & \sigma = -1 \end{cases}$ $\delta_{0,n} = 1$ only if $n = 0$
Legendre polynomials	$P_{m0}(x) \equiv P_m(x) = \frac{1}{2^m m!} \frac{d^m}{dx^m} (x^2 - 1)^m$
Associated Legendre polynomials	$n > 0 : P_{mn}(x) = (1-x^2)^{n/2} \frac{d^n}{dx^n} P_{m0}(x)$
$m > 0 : P_n^m(x) = (-1)^m (1-x^2)^{m/2} \frac{d^m}{dx^m} P_n(x)$	$n < 0 : \text{Not needed since } 0 \leq n \leq m \text{ in the HOA formulation}$
Interior problem solution	$p(kr, \theta, \delta) = \sum_{m=0}^{\infty} \sum_{n=m}^{\infty} A_{mn} j_n(kr) Y_n^m(\theta, \phi)$
Time dependence	$e^{+i\omega t}$
Radial functions for the exterior problem	$h_n^{(1)}(kr) : \text{incoming wave}$ $h_n^{(2)}(kr) : \text{outgoing wave}$
Spherical coordinate system conventions	$[x, y, z] = [r \cos \delta \cos \theta, r \cos \delta \sin \theta, r \sin \delta]$

# Appendix B

## Co-authored work

The below peer-reviewed, conference-proceedings paper (only abstract included here) employs the same framework for computing the Ambisonics system noise, as presented in Chapter 3. The author of this thesis mainly contributed to that paper by: (1) providing the MATLAB code generating the Ambisonics system noise at the ears of an acoustic manikin, and (2) providing insights into the pressure-reconstruction simulation procedures.

Weller T., Buchholz, J. M., and Oreinos C. (2014), “Frequency dependent regularization of a mixed-order Ambisonics encoding system using psychoacoustically motivated metrics,” in *AES 55<sup>th</sup> Conference on Spatial Audio* (Helsinki).

### Abstract

In mixed-order Ambisonics (MOA) encoding the right choice of the regularization parameter  $\lambda$  is crucial to the reproduction quality of the system. The choice of  $\lambda$  is a trade-off between low microphone noise amplification and high directivity. Commonly used performance metrics are not suited to determine an optimum  $\lambda(f)$ -curve due to their limited psychoacoustical relevance. Therefore, it is proposed here to use two perceptually motivated metrics that consider the reproduced signals at the listener’s ear to quantify the reproduction error and the audibility of the total microphone noise. Regularisation was then optimised towards previously defined requirements on both of these metrics separately. The results showed that at frequencies below 500 Hz both requirements can be met. Above this frequency one or both of the requirements have to be relaxed in order to find a suitable  $\lambda(f)$ .



# Appendix C

## Ethics application approval

Below is the approval granted by the Australian Hearing Human Research Ethics Committee following my application relevant to the experiment described in Chapter 6.



Australian Hearing Hub, Level 5  
16 University Avenue, Macquarie University,  
NSW 2109, Australia  
T +61 2 9412 6872  
F +61 2 9412 6769  
www.nal.gov.au



<b>Australian Hearing Human Research Ethics Committee</b> <i>APPROVAL FOR RESEARCH INVOLVING HUMAN SUBJECTS</i>	
<b>APPROVAL NUMBER: AHHREC2014-3</b>	
<b>Project Number</b>	CRC R4.3.3b sub-project
<b>Project Title</b>	Real-world laboratory environments: Validation of realistic loudspeaker-based environments through speech-in-noise tests using directional hearing aids
<b>Classification</b>	<i>Class 1: Project with negligible risk</i>
<b>Principal Investigators authorized to conduct research</b>	Chris Oreinos, Jörg Buchholz
<b>Date Approved/Ratified</b>	24/3/2014
<b>Approval Method</b>	Approved by the Research Director plus one other uninvolved senior NAL scientist as a Class 1 project with negligible risk.
<p>This approval is based on the information contained in the ethics application that was presented to the Research Director on 10/3/2014 and is conditional upon your continuing compliance with the National Statement on Ethical Conduct in Human Research (2007) available at: <a href="http://www.nhmrc.gov.au/files/nhmrc/publications/attachments/e72_national_statement_130624.pdf">http://www.nhmrc.gov.au/files/nhmrc/publications/attachments/e72_national_statement_130624.pdf</a>.</p> <p>A duplicate set of the documents is enclosed for your record.</p> <p>Annual reporting to the Committee on progress of the project is required including a final report when the work is completed or discontinued for any reason. Reminders will be sent when progress reports are due.</p> <p>The Committee expects to be notified of any changes to the approved protocol or other issues that may have an impact on the ethics of the project either by means of the annual progress reports (checklists) or as an application for variation. Adverse or unforeseen events that affect the continued ethical acceptability of the project should be reported to the Chairman immediately.</p> <p>All future correspondence relating to the ethical aspects of this project must quote the above Approval Number.</p>	
	
<p>Dr Tim Gainsford Operations &amp; Finance Manager, NAL and AHHREC Secretary</p>	



## Appendix D

### Practical considerations for the microphone array simulation applied in Chapter 6

A number of technical limitations that become relevant when considering a real microphone array have been considered in the applied microphone array simulations used in Section 6.2.1.3. To simulate positioning errors, the array microphones were shifted one-by-one from their ideal position  $(\theta_q, \delta_q)$  to a new position  $(\theta'_q, \delta'_q) = (\theta_q + \theta_e, \delta_q + \delta_e)$ , with the azimuth and elevation offsets independently drawn from the normal distribution:  $\theta_e, \delta_e \sim \mathcal{N}(\mu = 0, \sigma^2 = 25/9)$ , expressed in degrees ( $^\circ$ ). The standard deviation of the previous distribution,  $\sigma = 5/3^\circ$ , implies that 99.7% of the offset values lie in the interval  $[-5^\circ, 5^\circ]$ . Similarly, to simulate both phase and sensitivity mismatches between microphones, the ideal unity-gain microphone sensors were multiplied by complex gains  $g_q e^{i\phi_q}$ , with  $g_q \sim \mathcal{N}(\mu = 0, \sigma^2 = 25/9)$  [in dB] and  $\phi_q \sim \mathcal{N}(\mu = 0, \sigma^2 = 25/9)$  [in degrees ( $^\circ$ )].

Moreover, the effect of the inherent microphone noise was also considered. For that purpose, 62 uncorrelated, 10 s-long noise signals with a power spectral density equal to the equivalent input noise of the sensors [as previously measured by Buchholz (2013)] were treated as the only microphone array input and then decoded to loudspeaker signals. An overall gain was applied on those loudspeaker signals, such that the long-term level at the ears of a centred HATS (middle of interaural axis placed at the centre of the loudspeaker array) was equal to the expected (derived from simulation) noise-level.



# Bibliography

- Abhayapala, T. D. and Ward, D. B. (2002), “Theory and Design of High Order Sound Field Microphones Using Spherical Microphone Array,” in *Proc. of the IEEE Int. Conf. Acoustics, Speech, Signal Processing, ICASSP* (Florida).
- Abramowitz, M. and Stegun, I. A. (1972), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover), Tenth GPO printing ed.
- Acoustic Research Institute (n.d.), “ARI HRTF Database,” URL <http://www.kfs.oeaw.ac.at/index.php?view=article&id=608&lang=en>, last accessed 17<sup>th</sup> February 2015.
- Adelman, R., Gumerov, N. A., and Duraiswami, R. (2014), “Semi-Analytical Computation of Acoustic Scattering by Spheroids and Disks,” *J. Acoust. Soc. Am.* **136**, EL405–EL410.
- Ahrens, J. and Spors, S. (2009), “Alterations of the Temporal Spectrum in High-Resolution Sound Field Reproduction of Different Spatial Bandwidths,” in *Proc. of the AES 126<sup>th</sup> Convention* (Munich).
- Ahrens, J., Wierstorf, H., and Spors, S. (2010), “Comparison of Higher Order Ambisonics and Wave Field Synthesis with Respect to Spatial Discretization Artifacts in Time Domain,” in *Proc. of the 40<sup>th</sup> Int. Conf. of the AES – Spatial Audio: Sense the Sound of Space* (Tokyo).
- Algazi, V. R., Duda, R. O., and Thompson, D. M. (2002), “The Use of Head-and-Torso Models for Improved Spatial Sound Synthesis,” in *Proc. of the AES 113<sup>rd</sup> Convention* (Los Angeles).
- Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (2001), “The CIPIC HRTF Database,” in *Proc. of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA* (New Paltz).
- ANSI (1997), “Methods for Calculation of the Speech Intelligibility Index,” Tech. Rep. ANSI S3.5-1997, American National Standards Institute.
- Avni, A. and Rafaely, B. (2009), “Interaural Cross Correlation and Spatial Correlation in a Sound Field Represented by Spherical Harmonics,” in *Proc. of the Ambisonics Symposium* (Graz).
- Bamford, J. S. and Vanderkooy, J. (1995), “Ambisonic Sound for Us,” in *Proc. of the AES 99<sup>th</sup> Convention* (New York).

- Baumgartner, R., Pomberger, H., and Frank, M. (2011), “Practical Implementation of Radial Filters for Ambisonic Recordings,” in *Proc. of the International Conference on Spatial Audio* (Detmold).
- Bench, J., Kowal, Å., and Bamford, J. (1979), “The BKB (Bamford-Kowal-Bench) Sentence Lists for Partially-Hearing Children,” *Br. J. Audiol.* **13**, 108–112.
- Beranek, L. L. (1986), *Acoustics* (American Institute of Physics).
- Beranek, L. L. (1988), *Acoustical Measurements* (American Institute of Physics).
- Berkhout, A. J., de Vries, D., and Vogel, P. (1993), “Acoustic Control by Wave Field Synthesis,” *J. Audio Eng. Soc.* **93**, 2764–2778.
- Bernschütz, B. (2012), “Bandwidth Extension for Microphone Arrays,” in *Proc. of the AES 133<sup>rd</sup> Convention* (San Francisco).
- Best, V., Mejia, J., Freeston, K., Dillon, H., Derleth, P., Koenig, G., and Hofbauer, M. (2015), “An Evaluation of the Performance of Two Binaural Beamformers in Spatially Diverse and Dynamic Multitalker Environments,” Under revision at the *Int. J. Audiol.*
- Blanco, M. A., Flórez, M., and Bermejo, M. (1997), “Evaluation of the Rotation Matrices in the Basis of Real Spherical Harmonics,” *J. Mol. Struct.-Theochem* **419**, 19–27.
- Blauert, J. (1997), *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press), Revised ed.
- Boehnke, S. E., Hall, S. E., and Marquardt, T. (2002), “Detection of Static and Dynamic Changes in Interaural Correlation,” *J. Acoust. Soc. Am.* **112**, 1617–1626.
- Bovbjerg, B. P., Christensen, F., Minnaar, P., and Chen, X. (2000), “Measuring the Head-Related Transfer Functions of an Artificial Head with a High Directional Resolution,” in *Proc. of the AES 109<sup>th</sup> Convention* (Los Angeles).
- Boyd, S. and Vandenberghe, L. (2004), *Convex Optimization* (Cambridge University Press).
- Brungart, D. S. and Rabinowitz, W. M. (1999), “Auditory Localization of Nearby Sources. Head-Related Transfer Functions,” *J. Acoust. Soc. Am.* **106**, 1465–1479.
- Buchholz, J. M. (2013), “A Real-Time Hearing-Aid Research Platform (HARP): Realization, Calibration, and Evaluation,” *Acta Acust. United Ac.* **99**, 477–492.
- Burkardt, J. (2010), “Quadrature Rules for the Sphere,” URL [http://people.sc.fsu.edu/~jburkardt/datasets/sphere\\_lebedev\\_rule/sphere\\_lebedev\\_rule.html](http://people.sc.fsu.edu/~jburkardt/datasets/sphere_lebedev_rule/sphere_lebedev_rule.html), last accessed 27<sup>th</sup> May 2014.

- Busson, S., Nicol, R., and Katz, B. F. G. (2005), "Subjective Investigations of the Inter-aural Time Difference in the Horizontal Plane," in *Proc. of the AES 118<sup>th</sup> Convention* (Barcelona).
- Byrne, D., Parkinson, A., and Newall, P. (1990), "Hearing Aid Gain and Frequency Response Requirements for the Severely/Profoundly Hearing Impaired," *Ear and Hearing* **11**(1), 40–49.
- Carpentier, T., Bahu, H., Noisternig, M., and Warusfel, O. (2014), "Measurement of a Head-Related Transfer Function Database with High Spatial Resolution," in *Proc. of the Forum Acusticum, European Acoustics Association* (Kraków).
- Christensen, C. L. and Koutsouris, G. (2013), *ODEON Room Acoustics Software, Version 12*, URL <http://www.odeon.dk/pdf/ODEONManual12.pdf>.
- Compton-Conley, C. L., Neuman, A. C., Killion, M. C., and Levitt, H. (2004), "Performance of Directional Microphones for Hearing Aids: Real-World versus Simulation," *J. Am. Acad. Audiol.*, 440–455.
- Cord, M. T., Surr, R. K., Walden, B. E., and Dyrland, O. (2004), "Relationship Between Laboratory Measures of Directional Advantage and Everyday Success with Directional Microphone Hearing Aids," *J. Am. Acad. Audiol.* **15**, 353–364.
- Cord, M. T., Surr, R. K., Walden, B. E., and Olson, L. (2002), "Performance of Directional Microphone Hearing Aids in Everyday Life," *J. Am. Acad. Audiol.* **13**, 295–307.
- Cox, H., Zeskind, R. M., and Owen, M. M. (1987), "Robust Adaptive Beamforming," *IEEE Trans. Acoust. Speech Sig. Processing* **ASSP-35**, 1365–1376.
- Daniel, J. (2001), "Représentation de Champs Acoustiques, Application à la Transmission et à la Reproduction de Scènes Sonores Complexes dans un Context Multimédia (Representation of acoustic fields, application in the transmission and reproduction of complex sound scenes in a multimedia context)," Ph.D. thesis, Université Paris 6.
- Daniel, J. and Moreau, S. (2004), "Further Study of Sound Field Coding with Higher Order Ambisonics," in *Proc. of the AES 116<sup>th</sup> Convention* (Berlin).
- Daniel, J., Nicol, R., and Moreau, S. (2003), "Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging," in *Proc. of the AES 114<sup>th</sup> Convention* (Amsterdam).
- Daniel, J., Rault, J.-B., and Polack, J.-D. (1998), "Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions," in *Proc. of the AES 105<sup>th</sup> Convention* (San Francisco).

- Desloge, J. G., Rabinowitz, W. M., and Zurek, P. M. (1997), “Microphone-Array Hearing Aids with Binaural Output—Part I: Fixed-Processing Systems,” *IEEE Trans. Speech and Audio Proc.* **5**, 529–542.
- Dittberner, A. and Bentler, R. (2003), “Interpreting the Directivity Index (DI),” *The Hearing Review* **10**, 16–21.
- Duda, R. O. and Martens, W. L. (1998), “Range Dependence of the Response of a Spherical Head Model,” *J. Acoust. Soc. Am.* **104**, 3048–3058.
- Duraiswami, R., Li, Z., Zotkin, D. N., Grassi, E., and Gumerov, N. A. (2005), “Plane-Wave Decomposition Analysis for Spherical Microphone Arrays,” in *Proc. of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA* (New Paltz).
- Duraiswami, R., Zotkin, D. N., and Gumerov, N. A. (2004), “Interpolation and Range Extrapolation of HRTFs,” in *Proc. of the IEEE Int. Conf. Acoustics, Speech, Signal Processing, ICASSP* (Québec).
- Elko, G. W. and Pong, A.-T. N. (1995), “A Simple Adaptive First-Order Differential Microphone,” in *Proc. of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA* (New Paltz).
- Epain, N. and Daniel, J. (2008), “Improving Spherical Microphone Arrays,” in *Proc. of the AES 124<sup>th</sup> Convention* (Amsterdam).
- Epain, N., Guillon, P., Kan, A., Kosobrodov, R., Sun, D., Jin, C., and van Schaik, A. (2010), “Objective Evaluation of a Three-Dimensional Sound Field Reproduction System,” in *Proc. of the 20<sup>th</sup> International Conference on Acoustics, ICA* (Sydney).
- Epain, N. and Jin, C. (2014), Personal Communication, e-mail exchanged on Apr. 2, 2014.
- Epain, N., Jin, C. T., and Zotter, F. (2014), “Ambisonic Decoding with Constant Angular Spread,” *Acta Acust. United Ac.* **100**, 928–936.
- Farina, A., Capra, A., Chiesi, L., and Scopece, L. (2010), “A Spherical Microphone Array for Synthesizing Virtual Directive Microphones in Live Broadcasting and in Post Production,” in *Proc. of the 40<sup>th</sup> Int. Conf. of the AES – Spatial Audio: Sense the Sound of Space* (Tokyo).
- Favrot, S. and Buchholz, J. M. (2009), “Validation of a Loudspeaker-Based Room Auralization System Using Speech Intelligibility Measures,” in *Proc. of the AES 126<sup>th</sup> Convention* (Munich).

- Favrot, S. and Buchholz, J. M. (2010a), “Impact of Regularization of Near Field Coding Filters for 2D and 3D Higher-Order Ambisonics on Auditory Distance Cues,” in *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics* (Paris).
- Favrot, S. and Buchholz, J. M. (2010b), “LoRA: A Loudspeaker-Based Room Auralization System,” *Acta Acust. united Ac.* **96**, 364–375.
- Favrot, S. and Marschall, M. (2012), “Metrics for Performance Assessment of Mixed-Order Ambisonics Spherical Microphone Arrays,” in *Proc. of the AES 25<sup>th</sup> UK Conference – Spatial Audio in today’s 3D world* (York).
- Favrot, S., Marschall, M., Käsbach, J., Buchholz, J. M., and Weller, T. (2011), “Mixed-Order Ambisonics Recording and Playback for Improving Horizontal Directionality,” in *Proc. of the AES 131<sup>st</sup> Convention* (New York).
- Fazi, F. and Nelson, P. (2007), “A Theoretical Study of Sound Field Reconstruction Techniques,” in *Proc. of the 19<sup>th</sup> International Congress on Acoustics, ICA* (Madrid).
- Fels, J. (2013), “Trends in Binaural Technology,” in *Proc. of the Annual Conference on Acoustics, AIA-DAGA* (Merano).
- Fliege, J. (2007), “Integration Nodes for the Sphere,” URL <http://www.personal.soton.ac.uk/jf1w07/nodes/nodes.html>, last accessed 10<sup>th</sup> August 2013.
- Fliege, J. and Maier, U. (1999), “The Distribution of Points on the Sphere and Corresponding Cubature Formulae,” *IMA J. Numer. Anal.* **19**, 317–334.
- Frank, M. (2009), “Phantom Sources using Multiple Loudspeakers in the Horizontal Plane,” Ph.D. thesis, University of Music and Performing Arts Graz, Austria.
- Gardner, W. G. (1998), *3-D Audio Using Loudspeakers* (Kluwer Academic Publishers).
- Gardner, W. G. and Martin, K. D. (1995), “HRTF Measurements of a KEMAR,” *J. Acoust. Soc. Am.* **97**, 3907–3908.
- Gauthier, P. A., Camier, C., Pasco, Y., Berry, A., Chambatte, É., Lapointe, R., and Delalay, M. A. (2011), “Beamforming Regularization Matrix and Inverse Problems Applied to Sound Field Measurement and Extrapolation Using Microphone Array,” *J. Sound and Vib.* **330**, 5852–5877.
- Gerzon, M. A. (1973), “Periphony: With-Height Sound Reproduction,” *J. Audio Eng. Soc.* **21**, 2–10.
- Gerzon, M. A. (1992), “General Metatheory of Auditory Localisation,” in *Proc. of the AES 92<sup>nd</sup> Convention* (Vienna).

- Gilbert, E. N. and Morgan, S. P. (1955), “Optimum Design of Directive Antenna Arrays Subject to Random Variables,” *Bell Syst. Tech. J.* **34**, 637–663.
- Glyde, H., Buchholz, J., Best, V., Hickson, L., and Cameron, S. (2013), “The Effect of Better-Ear Glimpsing on Spatial Release from Masking,” *J. Acoust. Soc. Am.* **134**, 2937–2945.
- Glyde, H., Cameron, S., Dillon, H., Hickson, L., and Seeto, M. (2012), “The Effects of Hearing Impairment and Aging on Spatial Processing,” *Ear and Hearing* **34**(1), 15–28.
- Golub, G. H. and Loan, C. F. V. (1996), *Matrix Computations* (John Hopkins University Press).
- Granados, A., Jacobsen, F., and Fernandez-Grande, E. (2013), “Regularised Reconstruction of Sound Fields with a Spherical Microphone Array,” in *Proc. of the 21<sup>st</sup> International Conference on Acoustics, ICA* (Montréal).
- Grassi, E., Tulsi, J., and Shamma, S. (2003), “Measurement of Head-Related Transfer Functions Based on the Empirical Transfer Function Estimate,” in *Proc. of the Int. Conf. on Auditory Display, ICAD* (Boston).
- Greenberg, J. E., Peterson, P. M., and Zurek, P. M. (1993), “Intelligibility-Weighted Measures of Speech-to-Interference Ratio and Speech System Performance,” *J. Acoust. Soc. Am.* **94**, 3009–3010.
- Gregorčič, G. (2001), “The singular value decomposition and the pseudoinverse,” Tech. rep., University College Cork, Ireland, URL [http://www.cs.bgu.ac.il/~na131/wiki.files/SVD\\_application\\_paper.pdf](http://www.cs.bgu.ac.il/~na131/wiki.files/SVD_application_paper.pdf), last accessed 24<sup>th</sup> July 2014.
- Hansen, J. H. L. and Pellom, B. L. (1998), “An Effective Quality Evaluation Protocol for Speech Enhancement Algorithms,” in *Proc. of the International Conference on Spoken Language Processing, ICSLP* (Sydney).
- Heller, A. J. and Benjamin, E. M. (2014), “The Ambisonic Decoder Toolbox: Extensions for Partial-Coverage Loudspeaker Arrays,” in *Proc. of the Linux Audio Conference* (Karlsruhe).
- Hollerweger, F. (2006), “Periphonic Sound Spatialization in Multi-User Virtual Environments,” Master’s thesis, University of California at Santa Barbara.
- Horn, R. A. and Johnson, C. R. (1994), *Topics in Matrix Analysis* (Cambridge University Press).
- Jin, C., Epain, N., and Parthy, A. (2014), “Design, Optimization and Evaluation of a Dual-Radius Spherical Microphone Array,” *IEEE/ACM Trans. Audio, Speech and Lang. Proc.* **22**, 193–204.

- Jot, J.-M., Larcher, V., and Warusfel, O. (1995), “Digital Signal Processing Issues in the Context of Binaural and Transaural Stereophony,” in *Proc. of the AES 98<sup>th</sup> Convention* (Paris).
- Käsbaach, J., Favrot, S., and Buchholz, J. (2011), “Evaluation of a Mixed-Order Planar and Periphonic Ambisonics Playback Implementation,” in *Proc. of the Forum Acusticum, European Acoustics Association* (Aalborg, Denmark), pp. 2191–2196.
- Kates, J. M. (2008), *Digital Hearing Aids* (Plural Publishing).
- Kayser, H., Ewert, S. D., Anemüller, J., Rohdenburg, T., Hohmann, V., and Kollmeier, B. (2009), “DATABASE of Multichannel in-Ear and Behind-the-Ear Head-Related and Binaural Room Impulse Responses,” *EURASIP, J. on Advances Signal Processing*.
- Keidser, G., Dillon, H., Convery, E., and Mejia, J. (2013a), “Factors Influencing Individual Variation in Perceptual Directional Microphone Benefit,” *J. Am. Acad. Audiol.* **24**, 955–968.
- Keidser, G., Dillon, H., Mejia, J., and Nguyen, C.-V. (2013b), “An Algorithm that Administers Speech-in-Noise Testing to a Specified Reliability at Selectable Points on the Psychometric Function,” *Int. J. Audiol.* **52**, 795–800.
- Kerber, S. and Seeber, B. U. (2013), “Localization in Reverberation with Cochlear Implants,” *J. Assoc. Res. Otolaryngol.* **14**, 379–392.
- Khaykin, D. and Rafaely, B. (2012), “Acoustic Analysis by Spherical Microphone Array Processing of Room Impulse Responses,” *J. Acoust. Soc. Am.* **132**, 261–270.
- Killion, M. C. (1978), “Revised Estimate of Minimum Audible Pressure: Where is the ‘Missing 6dB’?” *J. Acoust. Soc. Am.* **63**, 1501–1508.
- Kistler, D. J. and Wightman, F. L. (1992), “A model of Head-Related Transfer Functions Based on Principal Components Analysis and Minimum-Phase Reconstruction,” *J. Acoust. Soc. Am.* **91**, 1637–1647.
- Kuk, F., Bækgaard, L., and Ludvigsen, C. (2000), “Design Considerations in Directional Microphones,” *The Hearing Review* **7**, 68–73.
- Kuttruff, H. (2000), *Room Acoustics* (Spon Press), Fourth ed.
- Li, T.-H. and North, G. R. (1997), “Aliasing Effects and Sampling Theorems of Spherical Random Fields when Sampled on a Finite Grid,” *Ann. Inst. Statist. Math.* **49**, 341–354.

- Li, Z. and Duraiswami, R. (2007), “Flexible and Optimal Design of Spherical Microphone Arrays for Beamforming,” *IEEE Trans. Audio, Speech and Lang. Proc.* **15**, 702–714.
- Li, Z., Duraiswami, R., Grassi, E., and Davis, L. S. (2004a), “Flexible Layout and Optimal Cancellation of the Orthonormality Error for Spherical Microphone Arrays,” in *Proc. of the IEEE Int. Conf. Acoustics, Speech, Signal Processing, ICASSP* (Québec).
- Li, Z., Duraiswami, R., and Gumerov, N. A. (2004b), “Capture and Recreation of Higher Order 3D Sound Fields via Reciprocity,” in *Proc. of the Int. Conf. on Auditory Display, ICAD* (Sydney).
- Luo, F. L., Yang, J., Pavlovic, C., and Nehorai, A. (2002), “Adaptive Null-Forming Scheme in Hearing Aids,” *IEEE Trans. Signal Processing* **50**, 1583–1590.
- Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., Dillier, N., Houben, R., Dreschler, W. A., Froehlich, M., Puder, H., Grimm, G., Hohmann, V., Leijon, A., Lombard, A., Mauler, D., and Spriet, A. (2010), “Multicenter Evaluation of Signal Enhancement Algorithms for Hearing Aids,” *J. Acoust. Soc. Am.* **127**, 1491–1505.
- Maj, J.-B., Royackers, L., Wouters, J., and Moonen, M. (2006), “Comparison of Adaptive Noise Reduction Algorithms in Dual Microphone Hearing Aids,” *Speech Communication* **48**, 957–970.
- Marschall, M., Favrot, S., and Buchholz, J. M. (2012), “Robustness of a Mixed-Order Ambisonics Microphone Array for Sound Field Reproduction,” in *Proc. of the AES 132<sup>nd</sup> Convention* (Budapest).
- Mathworks (2013), *MATLAB R2013b Mathematics*, URL [http://www.mathworks.com/help/releases/R2013b/pdf\\_doc/matlab/math.pdf](http://www.mathworks.com/help/releases/R2013b/pdf_doc/matlab/math.pdf), last accessed 25<sup>th</sup> January 2014.
- Mejia, J. P., Carlile, S., and Dillon, H. A. (2007), “A Method and System for Enhancing the Intelligibility of Sounds,” Tech. Rep. WO 2007/137364, World Intellectual Property Organization, WIPO.
- Mejia, J. P. and Dillon, H. A. (2010), “A System and Method for Producing a Directional Output Signal,” Tech. Rep. WO 2010/051606, World Intellectual Property Organization, WIPO.
- Merimaa, J. and Pulkki, V. (2005), “Spatial Impulse Response Rendering I: Analysis and Synthesis,” *J. Audio Eng. Soc.* **53**, 1115–1127.
- Meyer, J. (2001), “Beamforming for a Circular Microphone Array Mounted on Spherically Shaped Objects,” *J. Acoust. Soc. Am.* **109**, 185–193.

- Meyer, J. and Elko, G. (2002), “A Highly Scalable Spherical Microphone Array Based on an Orthonormal Decomposition of the Soundfield,” in *Proc. of the IEEE Int. Conf. Acoustics, Speech, Signal Processing, ICASSP* (Florida).
- Meyer, J. and Elko, G. (2008), “Handling Spatial Aliasing in Spherical Array Applications,” in *Proc. of the Hands-Free Speech Communication and Microphone Arrays, HSCMA* (Trento).
- Meyer, J. and Elko, G. W. (2004), “Spherical Microphone Arrays for 3D Sound Recording,” in *Audio signal processing for next generation multimedia communication systems*, edited by Y. A. Huang and J. Benesty (Kluwer Academic Publishers), pp. 77–86.
- Minnaar, P., Albeck, S. F., Simonsen, C. S., Søndersted, B., Oakley, S. A. D., and Bennedbæk, J. (2013), “Reproducing Real-Life Listening Situations in the Laboratory for Testing Hearing Aids,” in *Proc. of the AES 135<sup>th</sup> Convention* (New York).
- Minnaar, P., Favrot, S., and Buchholz, J. M. (2010), “Improving Hearing Aids Through Listening Tests in a Virtual Sound Environment,” *The Hearing Journal* **63**, 40–44.
- Moore, B. C. J. (2012), *An Introduction to the Psychology of Hearing* (Emerald Group Publishing), Sixth ed.
- Moreau, S. (2006), “Étude et Realisation d’Outils Avancés d’Encodage Spatial pour la Technique de Spatialisation Sonore Higher Order Ambisonics: Microphone 3D et Contrôle de Distance (Study and realization of advanced spatial encoding tools for the spatial-sound technique of higher order ambisonics: 3D microphone and distance control),” Ph.D. thesis, Université du Maine.
- Moreau, S., Daniel, J., and Bertet, S. (2006), “3D Sound Field Recording with Higher Order Ambisonics - Objective Measurements and Validation of Spherical Microphone,” in *Proc. of the AES 120<sup>th</sup> Convention* (Paris).
- Mueller, M. F., Kegel, A., Schimmel, S. M., Dillier, N., and Hofbauer, M. (2012), “Localization of Virtual Sound Sources with Bilateral Hearing Aids in Realistic Acoustical Scenes,” *J. Acoust. Soc. Am.* **131**, 4732–4742.
- Murillo, D. M., Fazi, F. M., and Shin, M. (2014), “Evaluation of Ambisonics Decoding Methods with Experimental Measurements,” in *Proc. of the EAA Joint Symposium on Auralization and Ambisonics* (Berlin).
- Nábělek, A. K., Tucker, F. M., and Letowski, T. R. (1991), “Toleration of Background Noises: Relationship With Patterns of Hearing Aid Use by Elderly Persons,” *J. Speech Lang. Hear. Res.* **34**, 679–685.

- Nakado, T., Nishino, T., and Takeda, K. (2008), “Head-Related Transfer Function Measurement in Sagittal and Frontal Coordinates,” *Acoust. Sci. & Tech.* **29**, 335–337.
- Nam, J., Abel, J. S., and Smith, J. O. (2008), “A Method for Estimating Interaural Time Difference for Binaural Synthesis,” in *Proc. of the AES 125<sup>th</sup> Convention* (San Francisco).
- Neukom, M. (2007), “Ambisonic Panning,” in *Proc. of the AES 123<sup>rd</sup> Convention* (New York).
- Noohi, T., Epain, N., and Jin, C. T. (2013), “Direction of Arrival Estimation for Spherical Microphone Arrays by Combination of Independent Component Analysis and Sparse Recovery,” in *Proc. of the IEEE Int. Conf. Acoustics, Speech, Signal Processing, ICASSP* (Vancouver).
- Nowak, J., Liebetrau, J., and Sporer, T. (2013), “On the Perception of Apparent Source Width and Listener Envelopment in Wave Field Synthesis,” in *Proc. of the 5<sup>th</sup> Int. Workshop on Quality of Multimedia Experience, QoMEX* (Klagenfurt am Wörthersee).
- Oppenheim, A. V., Schaffer, R. W., and Buck, J. R. (1999), *Discrete-Time Signal Processing* (Prentice Hall International), Second ed.
- Oreinos, C. and Buchholz, J. (2013a), “Objective Analysis of HOA Sound-Field Reproduction for Hearing Aid Applications,” in *Proc. of the 21<sup>st</sup> International Conference on Acoustics, ICA* (Montréal).
- Oreinos, C. and Buchholz, J. (2014), “Validation of Realistic Acoustic Environments for Listening Tests Using Directional Hearing Aids,” in *Proc. of the International Workshop on Acoustic Signal Enhancement, IWAENC* (Antibes).
- Oreinos, C. and Buchholz, J. M. (2013b), “Effect of Higher-Order Ambisonics on Evaluating Beamformer Benefit in Realistic Acoustic Environments,” in *Proc. of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA* (New Paltz), [Manuscript presented in Chapter 5].
- Oreinos, C. and Buchholz, J. M. (2013c), “Measurement of a Full 3D set of HRTFs for in-Ear and Hearing Aid Microphones on a Head and Torso Simulator (HATS),” *Acta Acust. United Ac.* **99**, 836–844, [Manuscript presented in Chapter 3].
- Oreinos, C. and Buchholz, J. M. (2015a), “Objective Analysis of Ambisonics for Hearing Aid Applications: Effect of Listener’s Head, Room Reverberation, and Directional Microphones,” Accepted for publication at the J. Acoust. Soc. Am. [Initial manuscript presented in Chapter 4].

- Oreinos, C. and Buchholz, J. M. (2015b), “An Objective and Subjective Validation of Applied Loudspeaker-Based Virtual Sound Environments Used for Directional Hearing Aid Testing,” Manuscript prepared for submission to journal [Presented in Chapter 6].
- Park, M. and Rafaely, B. (2005), “Sound-Field Analysis by Plane-Wave Decomposition Using Spherical Microphone Array,” *J. Acoust. Soc. Am.* **118**(5), 3094–3103.
- Poletti, M. (2009), “Unified Description of Ambisonics Using Real and Complex Spherical Harmonics,” in *Proc. of the Ambisonics Symposium* (Graz).
- Poletti, M., Fazi, F. M., and Nelson, P. A. (2010), “Sound-Field Reproduction Systems Using Fixed-Directivity Loudspeakers,” *J. Acoust. Soc. Am.* **127**, 3590–3601.
- Poletti, M. A. (2005), “Three-Dimensional Surround Sound Systems Based on Spherical Harmonics,” *J. Audio Eng. Soc.* **53**, 1004–1025.
- Pulkki, V. (2001), “Spatial Sound Generation and Perception by Amplitude Panning Techniques,” Ph.D. thesis, Helsinki University of Technology.
- Rafaely, B. (2004), “Plane-Wave Decomposition of the Sound Field on a Sphere by Spherical Convolution,” *J. Acoust. Soc. Am.* **116**, 2149–2157.
- Rafaely, B. (2005), “Analysis and Design of Spherical Microphone Arrays,” *IEEE Trans. Speech Audio Processing* **13**, 135–143.
- Rafaely, B., Weiss, B., and Bachmat, E. (2007), “Spatial Aliasing in Spherical Microphone Arrays,” *IEEE Trans. Signal Processing* **55**, 1003–1010.
- Revit, L. J., Killion, M. C., and Compton-Conley, C. L. (2007), “Developing and Testing a Laboratory Sound System That Yields Accurate Real-World Results,” *The Hearing Review* **14.11**, 54–61.
- Revit, L. J., Schulein, R. B., and Julstrom, S. D. (2002), “Toward Accurate Assessment of Real-World Hearing Aid Benefit,” *The Hearing Review* **9.8**, 34–38 & 51.
- Ricketts, T. (2000), “Impact of Noise Source Configuration on Directional Hearing Aid Benefit and Performance,” *Ear and Hearing* **21**(3), 194–205.
- Rychtáriková, M., Van den Bogaert, T., Vermeir, G., and Wouters, J. (2009), “Binaural Sound Source Localization in Real and Virtual Rooms,” *J. Audio Eng. Soc.* **57**, 205–220.
- Seeber, B. U., Kerber, S., and Hafter, E. R. (2010), “A System to Simulate and Reproduce Audio-Visual Environments for Spatial Hearing Research,” *Hearing Research* **260**, 1–10.

- Sneeuw, N. (1994), “Global Spherical Harmonic Analysis by Least-Squares and Numerical Quadrature Methods in Historical Perspective,” *Geophys. J. Int.* **118**, 707–716.
- Solvang, A. (2008), “Spectral Impairment of Two-Dimensional Higher Order Ambisonics,” *J. Audio Eng. Soc.* **56**, 267–279.
- Spors, S. and Ahrens, J. (2008), “A Comparison of Wave Field Synthesis and Higher-Order Ambisonics with Respect to Physical Properties and Spatial Sampling,” in *Proc. of the AES 125<sup>th</sup> Convention* (San Francisco).
- Travis, C. (2009), “A New Mixed-Order Scheme for Ambisonic Signals,” in *Proc. of the Ambisonics Symposium* (Graz).
- Trevino, J., Okamoto, T., Iwaya, Y., and Suzuki, Y. (2010), “High Order Ambisonic Decoding Method for Irregular Loudspeaker Arrays,” in *Proc. of the 20<sup>th</sup> International Conference on Acoustics, ICA* (Sydney).
- Trevino, J., Okamoto, T., Iwaya, Y., and Suzuki, Y. (2011), “Evaluation of a New Ambisonic Decoder for Irregular Loudspeaker Arrays Using Interaural Cues,” in *Proc. of the Ambisonics Symposium* (Lexington).
- Tribolet, J., Noll, P., McDermott, B., and Crochiere, R. (1978), “A Study of Complexity and Quality of Speech Waveform Coders,” in *Proc. of the IEEE Int. Conf. Acoustics, Speech, Signal Processing, ICASSP* (Tulsa).
- Van den Bogaert, T., Doclo, S., Wouters, J., and Moonen, M. (2008), “The Effect of Multimicrophone Noise Reduction Systems on Sound Source Localization by Users of Binaural Hearing Aids,” *J. Acoust. Soc. Am.* **124**, 484–497.
- Van den Bogaert, T., Doclo, S., Wouters, J., and Moonen, M. (2009), “Speech Enhancement with Multichannel Wiener Filter Techniques in Multimicrophone Binaural Hearing Aids,” *J. Acoust. Soc. Am.* **125**, 360–371.
- Veen, B. V. and Buckley, K. M. (2010), “Beamforming Techniques for Spatial Filtering,” in *The DSP Handbook; Wireless, Networking, Radar, Sensor Array Processing, and Nonlinear Signal Processing*, edited by V. Madisetti (CRC Press), pp. 61.1–61.22.
- Völk, F. and Fastl, H. (2010), “Advantages of Binaural Room Synthesis for Research and Fitting of Hearing Aids, Cochlear Implants, Electro-Acoustical Stimulation, and Combined Systems,” in *Proc. of the 20<sup>th</sup> International Conference on Acoustics, ICA* (Sydney).
- Wabnitz, A., Epain, N., van Schaik, A., and Jin, C. (2011), “Time Domain Reconstruction of Spatial Sound Fields Using Compressed Sensing,” in *Proc. of the IEEE Int. Conf. Acoustics, Speech, Signal Processing, ICASSP* (Prague).

- Walravens, E., Keidser, G., Hartley, D., and Hickson, L. (2014), “An Australian Version of the Acceptable Noise Level Test and its Predictive Value for Successful Hearing Aid Use in an Older Population,” *Int. J. Audiol.* **52**, 1–8.
- Ward, D. B. and Abhayapala, T. D. (2001), “Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers,” *IEEE Trans. Speech Audio Process.* **9**, 697–707.
- Weller, T., Buchholz, J. M., and Oreinos, C. (2014), “Frequency Dependent Regularization of a Mixed-Order Ambisonics Encoding System Using Psychoacoustically Motivated Metrics,” in *Proc. of the 55<sup>th</sup> Int. Conf. of the AES* (Helsinki), [Abstract presented in Appendix B].
- Wenzel, E. M., Arruda, M., Kistler, D. J., , and Wightman, F. L. (1993), “Localization Using Nonindividualized Head-Related Transfer Functions,” *J. Acoust. Soc. Am.* **94**, 111–123.
- Westermann, A., Buchholz, J. M., and Dau, T. (2013), “Binaural dereverberation based on interaural coherence histograms,” *J. Acoust. Soc. Am.* **133**, 2767–2777.
- Williams, E. G. (1999), *Fourier Acoustics* (Academic Press).
- Zaunschirm, M. and Zotter, F. (2014), “Measurement-Based Modal Beamforming Using Planar Circular Microphone Arrays,” in *Proc. of the EAA Joint Symposium on Auralization and Ambisonics* (Berlin).
- Zotkin, D. N., Duraiswami, R., and Gumerov, N. A. (2010), “Plane-Wave Decomposition of Acoustical Scenes via Spherical and Cylindrical Microphone Arrays,” *IEEE Trans. Audio, Speech and Lang. Proc.* **18**, 2–16.
- Zotter, F. (2009), “Analysis and Synthesis of Sound-Radiation with Spherical Arrays,” Ph.D. thesis, University of Music and Performing Arts Graz, Austria.
- Zotter, F., Pomberger, H., and Frank, M. (2009), “An Alternative Ambisonics Formulation: Modal Source Strength Matching and the Effect of Spatial Aliasing,” in *Proc. of the AES 126<sup>th</sup> Convention* (Munich).