

Ondersteunende methodieken
voor kwaliteitsbeoordeling van stedelijk geluid

Tools for Urban Sound Quality Assessment

Pieter Thomas

Promotoren: prof. dr. ir. D. Botteldooren, prof. dr. ir. T. Van Renterghem
Proefschrift ingediend tot het behalen van de graad van
Doctor in de Ingenieurswetenschappen: Elektrotechniek

Vakgroep Informatietechnologie
Voorzitter: prof. dr. ir. D. De Zutter
Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar 2014 - 2015



ISBN 978-90-8578-752-5
NUR 962
Wettelijk depot: D/2014/10.500/98



Universiteit Gent
Faculteit Ingenieurswetenschappen en Architectuur
Vakgroep Informatietechnologie

Promotoren:

Prof. Dr. Ir. Dick Botteldooren
Prof. Dr. Ir. Timothy Van Renterghem

Examencommissie:

Prof. Dr. Ir. Patrick De Baets (voorzitter)	Universiteit Gent
Prof. Dr. Ir. Hendrik Rogier (secretaris)	Universiteit Gent
Prof. Dr. Ir. Dick Botteldooren (promotor)	Universiteit Gent
Prof. Dr. Ir. Timothy Van Renterghem (promotor)	Universiteit Gent
Dr. Sc. techn. Kurt Heutschi	ETH Zürich - EMPA
Prof. Dr. Ir. Jean-Jacques Embrechts	Université de Liège
Prof. Dr. Ir. Dirk Stroobandt	Universiteit Gent
Prof. Dr. Ir. Jean-Pierre Martens	Universiteit Gent

Universiteit Gent
Faculteit Ingenieurswetenschappen en Architectuur
Vakgroep Informatietechnologie
Sint-Pietersnieuwstraat 41, B-9000 Gent, België

Tel.: +32-9-264.99.95
Fax.: +32-9-264.99.69



Proefschrift tot het behalen van de graad van
Doctor in de Ingenieurswetenschappen:
Elektrotechniek
Academiejaar 2014-2015

As you set out for Ithaka
hope the voyage is a long one,
full of adventure, full of discovery.
Laistrygonians and Cyclops,
angry Poseidon - don't be afraid of them:
you'll never find things like that on your way
as long as you keep your thoughts raised high,
as long as a rare excitement
stirs your spirit and your body.
Laistrygonians and Cyclops,
wild Poseidon - you won't encounter them
unless you bring them along inside your soul,
unless your soul sets them up in front of you.

Hope the voyage is a long one.
May there be many a summer morning when,
with what pleasure, what joy,
you come into harbors seen for the first time;
may you stop at Phoenician trading stations
to buy fine things,
mother of pearl and coral, amber and ebony,
sensual perfume of every kind -
as many sensual perfumes as you can;
and may you visit many Egyptian cities
to gather stores of knowledge from their scholars.

Keep Ithaka always in your mind.
Arriving there is what you are destined for.
But do not hurry the journey at all.
Better if it lasts for years,
so you are old by the time you reach the island,
wealthy with all you have gained on the way,
not expecting Ithaka to make you rich.

Ithaka gave you the marvelous journey.
Without her you would not have set out.
She has nothing left to give you now.

And if you find her poor, Ithaka won't have fooled you.
Wise as you will have become, so full of experience,
you will have understood by then what these Ithakas mean.

C.P. Cavafy

Translated by Edmund Keeley / Philip Sherrard.

Dankwoord

Het is nu ongeveer vijf jaar geleden dat ik aan dit onderzoeksavontuur begonnen ben. Mijn allereerste dag aan de onderzoeksgroep akoestiek begon met de woorden *je weet wel wat te doen hé*. Dat jaar had ik mijn masterthesis rond pleinakoestiek tijdens openlucht concerten afgewerkt, een onderwerp dat ik zelf had mogen voorstellen. Mijn werk in de eerste maanden richtte zich dan ook op het verder uitwerken van de resultaten.

De focus van mijn onderzoek is sindsdien lichtjes bijgeschaafd - veracademiseerd zeg maar - waarbij ik steeds de mogelijkheid gekregen heb mijn eigen *ding* te doen. De vele meetcampagnes (al dan niet bij vriestemperaturen), conferenties, projecten, het begeleiden van (thesis)studenten en de betrokkenheid bij de nieuwe geluidsnormen zorgden ervoor dat, in tegenstelling tot wat sommigen zich mis-schien inbeelden bij doctoraatsonderzoek, dit verre van een saaie bureaujob werd.

Het werk dat je nu voor je hebt beschrijft de resultaten van mijn tocht op de woe-lige golven van de akoestiek, die ik nooit tot een goed einde zou gebracht hebben zonder de hulp en steun van velen waar ik steeds op kon rekenen. Allereerst zou ik mijn promotor prof. Dick Botteldooren van harte willen bedanken voor de kan-sen die hij me gegeven heeft. Ondanks zijn drukke tijdsschema vond hij steeds de tijd om problemen te bespreken en mee te zoeken naar antwoorden. Zijn en-thousiasme werkt ongelooflijk aanstekelijk en motiveerde me enorm om door te zetten op moeilijke momenten. Evenzeer zou ik mijn co-promotor, prof. Timothy Van Renterghem, willen bedanken. Zonder zijn inzichten rond geluidspropagatie, numerieke simulaties en geluidsbermen, had ik dit werk nooit kunnen realiseren.

Verder wil ik ook heel graag mijn collega's bedanken die elke keer weer kleur ga-ven aan de dag. Many thanks for Weigang, for all his help and company during the E17 measurement campaign, his amazing stories and tips and tricks to cope with bureaucracy. Luc, voor al zijn hulp bij mijn praktische problemen en om mij wegwijst te maken in de wereld van Svantek, Swing en Pulse. Annelies, om de pijn van de statistiek wat te verzachten en te helpen om (al is het maar voor korte tijd) orde in de chaos van het labo te scheppen. I also want to thank Mima for the lively discussions and (Santa) Gemma for her company. And last, but not least, Arnaud, Bert, Damiano, Elisabeth, Kang, Karlo, Laurent, Lei, Maria, Michiel, Ramanan, Samuel, Sara, Vincent and Xiadong and all the EM-colleagues on the INTEC cor-ridor.

Bij onderzoek komen heel wat praktische zaken kijken. Voor het maken van meetopstellingen en prototypes kon ik steeds rekenen op de hulp van Luc en Peter. Isabelle maakte me wegwijs in de - soms absurde - administratieve rompslomp, terwijl ik op Kristien kon rekenen voor alle computerproblemen.

Een deel van deze thesis bestaat uit een uitgebreide meetcampagne aan een geluidsberm naast de E17. Graag zou ik de familie D'Hondt van harte willen bedanken om ons toe te laten die metingen uit te voeren, maandenlang apparatuur in hun tuin te verdragen en - wat zeer geapprecieerd werd - ons van drank en warmte te voorzien tijdens die lange en koude meetcampagne.

Verder wil ik alle mensen die deelgenomen hebben aan de luistertesten bedanken. Zonder jullie had ik nooit gegronde conclusies kunnen trekken over de prestatie van de ontwikkelde auralizatietechnieken. Hoe moeilijker jullie de test vonden, des te beter was mijn resultaat. En moeilijk bleek hij inderdaad te zijn...

Verder wil ik ook graag mijn vrienden bedanken. Zij zorgden voor afwisseling en hielpen om mijn gedachten eens *op iets anders* te zetten. Bedankt, Benjamin en Sofie, Ellen en Tom, Jan en Linde, Jørgen en Liesbeth, Lies en Stijn, Michiel, Olivier en Julie, Pieter en Sanne en de *Pironnekes*! Bedankt ook Karel, de man van het snelle(re) licht, voor de vele pitta's en de ontspannende momenten!

Een heel speciaal woord van dank gaat naar mijn ouders, mijn broer(ie) Michaël, mijn familie en lieve schoonfamilie; Lut en Jo, Renée, Nienke en Sven. Zij hebben me steeds alle kansen gegeven om me te ontplooien, en steunden me in elk project - hoe zot ook. Bedankt om er te zijn voor mij!

De afgelopen jaren zijn niet altijd gemakkelijk geweest. In enkele jaren tijd heb ik al mijn grootouders moeten missen. Ik ben er zeker van dat ze nu heel fier zouden geweest zijn. Meme Wivina, meme Irma, pepe Louis, jullie zullen altijd in mijn gedachten blijven!

Als laatste wil ik graag mijn lieve vriendin Ilke bedanken. Exact één dag na de start van mijn doctoraat begon ook ons verhaal. Daardoor heb je elk aspect van mijn onderzoek meegemaakt. Zowel de leuke momenten, in Lissabon en Lyon, maar ook de stressvolle momenten voor presentaties, meetcampagnes (inclusief aanvallen van monsterlijke kreeften en vallende glazen bollen) en de laatste maanden tijdens het afwerken van mijn onderzoek. Sinds die dag hebben we samen al veel meegemaakt: een huis gekocht, verbouwd en ondertussen wonen we al meer dan een jaar-en-half samen. Bedankt om me te steunen en om steeds in me te geloven - en vooral, bedankt voor al die mooie momenten samen!

*Pieter Thomas
Gent, november 2014*

Aut numquam tentes, aut perfice.

ARS AMATORIA - OVIDIUS

Table of Contents

Dankwoord	i
Nederlandse samenvatting	xvii
English summary	xxi
1 Introduction	1-1
1.1 Outline	1-3
1.2 Publications	1-4
1.2.1 Publications in international journals	1-4
1.2.2 Publications in international conferences	1-5
References	1-6
2 Recording and playback of a 3D outdoor sound environment	2-1
2.1 The human sound localization	2-2
2.2 Creating an immersive audio environment: an overview of existing techniques	2-4
2.2.1 Soundfield capturing	2-5
2.2.2 Soundfield reproduction	2-8
2.3 Design and characterization of a microphone array for binaural recording emulation	2-11
2.3.1 Characterization of the target HRTF directivity pattern	2-11
2.3.1.1 HRTF dataset and ear canal compensation	2-12
2.3.1.2 HRTF spherical harmonic representation	2-13
2.3.2 Microphone array design and characterization	2-20
2.3.2.1 Microphone array directivity pattern	2-21
2.3.2.2 Microphone directivity pattern evaluation	2-25
2.3.3 HRTF directivity pattern fitting	2-28
2.4 Subjective validation of the binaural reconstruction methodology	2-35
2.4.1 Acquisition of the audio samples	2-35
2.4.2 Listening test set-up and results	2-37
2.5 Conclusion	2-42
References	2-43

3	Auralizing the effect of noise mitigation measures	3-1
3.1	Auralization of a car pass-by behind a low finite-length vegetated noise barrier	3-2
3.1.1	Auralization of the effect of a small barrier on car passages	3-3
3.1.1.1	Source angle extraction	3-5
3.1.1.2	Simulation of barrier attenuation and diffraction	3-6
3.1.2	Validation of the auralization of the barrier effect	3-8
3.1.2.1	Listening test set-up	3-8
3.1.2.2	Results and discussion	3-10
3.1.3	Conclusion	3-12
3.2	Auralization of the effect of an L-shaped mound near a highway	3-12
3.2.1	Description of the site	3-13
3.2.2	Developing a point-to-point acoustical model	3-13
3.2.2.1	Determination of model parameters	3-19
3.2.3	Modeling the highway IL of the mound	3-23
3.2.3.1	Line source model	3-23
3.2.3.2	Vehicle source description	3-24
3.2.3.3	Calculation of $L_{eq,NM M}$ and IL	3-25
3.2.4	Model performance and validation	3-31
3.2.4.1	Measurement set-up and data clean-up	3-31
3.2.4.2	Extraction of $L_{eq,NM M}$	3-32
3.2.4.3	Estimation of the measured highway IL of the mound	3-34
3.2.4.4	Validation of the simulation models	3-36
3.2.5	Auralization of the effect of the mound	3-37
3.2.5.1	Auralization based on directivity pattern emulation	3-38
3.2.5.2	Auralization based on the MTB technique	3-42
3.2.6	Qualitative evaluation of the auralization of the mound	3-44
3.2.6.1	Audio sample acquisition and auralization	3-44
3.2.6.2	Listening test set-up and results	3-44
3.2.7	Conclusion	3-49
	References	3-50
4	Using room acoustical parameters for evaluating the quality of urban squares for open-air rock concerts	4-1
4.1	Introduction	4-2
4.2	Parameter definitions	4-4
4.2.1	Energy related parameters	4-4
4.2.2	Reverberation related parameters	4-5
4.2.3	Binaural related parameters	4-6
4.2.3.1	Objective measures for apparent source width ASW	4-6
4.2.3.2	Objective measures for listener envelopment LEV	4-6
4.3	Measurement methodology	4-7

4.3.1	Test signal and deconvolution technique	4-7
4.3.2	Measurement set-up	4-8
4.3.3	Impulse response post-processing	4-9
4.3.3.1	Octave band filtering	4-9
4.3.3.2	Impulse response truncation	4-10
4.4	Case-studies	4-11
4.4.1	Description of the squares	4-11
4.4.2	Correlation analysis	4-11
4.4.3	Influence of the delay-lines at the Lokerse Feesten	4-13
4.4.4	Comparison of four urban squares	4-17
4.4.4.1	Description by clarity C_{80}	4-17
4.4.4.2	Description by T_{30} and BR_{EDT}	4-19
4.4.4.3	Description by $[1 - IACC_{E3/L3}]$ and LEV_{calc}	4-19
4.4.4.4	Description by $\Delta L_{eq,A/C}$	4-20
4.5	Conclusion	4-22
	References	4-23
5	Reverberation-based urban street sound level prediction	5-1
5.1	Introduction	5-2
5.2	Measurement set-up	5-3
5.2.1	Impulse response extraction	5-3
5.2.2	Reflection ratio $RR(t_0)$	5-4
5.2.3	Validation in half-open space	5-6
5.3	Street reflection ratio measurements	5-7
5.3.1	Creating a model for $RR(12ms)$	5-8
5.3.1.1	Regression analysis based on street width	5-9
5.3.1.2	Multiple regression analysis	5-11
5.3.2	Model performance	5-14
5.4	Predicting sound pressure level based on the reflection ratio	5-15
5.4.1	General model for the SPL in a street canyon	5-15
5.4.2	Model validation in a real street canyon	5-17
5.4.3	Influence of street width and height on predicted SPL . . .	5-19
5.5	Conclusion	5-21
	References	5-22
6	Conclusions and perspectives	6-1
6.1	Perspectives	6-3

List of Acronyms

A

ANOVA	Analysis of Variance
ASW	Apparent Source Width

B

BC	Building Correction
BR	Bass Ratio
BQI	Binaural Quality Index

D

DFT	Discrete Fourier Transform
-----	----------------------------

E

EDT	Early Decay Time
END	Environmental Noise Directive

F

FDTD	Finite-Difference Time-Domain method
------	--------------------------------------

G

GIS	Geographic Information System
GUI	Graphical User Interface

H

HATS	Head-and-Torso Simulator
HRIR	Head-Related Impulse Response
HRTF	Head-Related Transfer Function

I

IACC	Interaural Cross Correlation
IL	Insertion Loss
ILD	Interaural Level Difference
ITD	Interaural Time Difference
ITDG	Initial Time Delay Gap

K

KEMAR	Knowles Electronics Mannequin for Acoustic Research
KLE	Karhunen-Loeve Expansion

L

LAD	Leaf Area Density
LEV	Listener Envelopment
LF	Lateral (Energy) Fraction
LG	Lateral Sound Level

M

MAA	Minimal Audible Angle
MLS	Maximum-Length Sequence
MSE	Mean Square Error
MTB	Motion-Tracked Binaural sound

P

PCA	Principal Components Analysis
-----	-------------------------------

R

RR	Reflection Ratio
RT	Reverberation Time
rMSE	relative Mean Square Error

S

SDI	Surface Diffusion Index
SNR	Signal-to-Noise Ratio
SPL	Sound Pressure Level
SSH	Surface Spherical Harmonics
STI	Speech Transmission Index

T

TL	Transmission Loss
----	-------------------

V

VBAP	Vector Base Amplitude Panning
------	-------------------------------

W

WFS

Wave Field Synthesis

WHO

World Health Organization

WNG

White Noise Gain

List of Symbols

Mathematical symbols

j	imaginary unit
ϕ	azimuth, $\phi \in [0, 2\pi[$
θ	elevation, $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$
r	radial distance
\mathbf{I}_n	identity matrix of size n
$ \cdot $	magnitude
$\angle \cdot$	(unwrapped) phase
\cdot^T	matrix transpose
\cdot^*	complex conjugate
\cdot^H	Hermitian transpose

Acoustical symbols

α	absorption coefficient
Δ_{ref}	source-receiver distance in the RR measurement set-up (chapter 4 - $\Delta_{ref} = 2.48\text{m}$)
$\delta_{\mathbf{m}}$	discrete Dirac function, delayed with m samples
$\delta\phi$	angular segmentation size
λ	wavelength, $\lambda f = c$
μ	regularization parameter
$\rho_{mag}(f), \rho_{pha}(f)$	magnitude, unwrapped phase spatial correlation coefficient
ω	pulsation, $\omega = 2\pi f$
$A^{nm}(k)$	frequency dependent spherical harmonic (Ambisonic) coefficient of order n and degree m
$A_{ }^{nm}(f), A_{\angle}^{nm}(f)$	frequency dependent spherical harmonic coefficient of order n and degree m resulting from magnitude, unwrapped phase decomposition
a	spherical array radius
$\text{BR}, \text{BR}_{EDT}, \text{BR}_{T_{30}}$	bass ratio, defined as the ratio of the sum of the RT, EDT, T_{30} in the 125Hz and 250Hz octave band and the sum of the RT, EDT, T_{30} in the 500Hz and 1000Hz octave band
c	speed of sound (approx. 340m/s in air)

C_{80}	clarity, defined as the ratio between the energy contained in the early soundfield (energy measured in the first 80ms) and the energy in the reverberant soundfield
C_T, C_v	structure parameter of temperature, velocity fluctuation
D_{50}	Definition or Deutlichkeit, defined as the ratio between the energy contained in the early soundfield (energy measured in the first 50ms) and the total energy
E, E_{dir}, E_{rev}	sound energy, sound energy contained in the direct, reverberant part of the soundfield
EDT	Early Decay Time, defined as the time it takes for the level to decay 60dB, calculated from the level decay of the early soundfield
f	frequency
f_s	sampling frequency
f_X	cross-over frequency
G, G	normalized ground factor (chapter 3), sound strength (chapter 4), defined as the ratio of the SPL on a given location to the SPL, measured in anechoic environment at 10m from an omnidirectional source
H	average building height
$H(f)$	transfer function
$h(t)$	impulse response
h_S, h_R	source, receiver height
$(h_S)_{eq}, (h_R)_{eq}$	equivalent source, receiver height
$HRTF(f, \phi, \theta)$	complex-valued HRTF at frequency f and direction (ϕ, θ)
$HRTF_{ }(f, \phi, \theta),$	magnitude, unwrapped phase of the HRTF at frequency f
$HRTF_{\perp}(f, \phi, \theta)$	and direction (ϕ, θ)
IACC	interaural cross correlation, defined as the time (≤ 1 ms) where the cross correlation between the left and right ear signal reaches a maximum
$IL(\phi)$	angular insertion loss (IL)
k	wavenumber, $k = \frac{2\pi}{\lambda}$
L	propagation length through the vegetation
$L(t)$	level function, defined as the sound energy decay in function of time
l_w	leaf width
$L_{eq}, L_{eq,NM}, L_{eq,M}$	equivalent sound pressure level, equivalent sound pressure level before, after installation of the mound
$L_{eq,A}, L_{eq,C}$	A-, C-weighted equivalent sound pressure level
$\Delta L_{eq,oct}$	equivalent sound pressure level relative to a reference point in octave bands
$\Delta L_{eq,A}, \Delta L_{eq,C}$	A-, C-weighted equivalent sound pressure level relative to a reference point
$L_p, L_{p,NM}, L_{p,M}$	sound pressure level, sound pressure level before, after installation of the mound

$L_{early}(t_0)$,	sound pressure level of the early, reverberant part of the
$L_{reflections}(t_0)$	soundfield (with dividing time t_0)
$L_w, L_{w,P}, L_{w,R}$	source power level, source power level of propulsion, rolling noise
N, N_{max}	maximum resolved, resolvable order of the spherical harmonic coefficients
N_F	Fresnel number
$p(kr, \phi, \theta)$	frequency dependent pressure at location (r, ϕ, θ) and wavenumber k
r, r_{xz}	source-receiver distance, source-receiver distance projected on the xz -plane
r_S, r_R	distance between source, receiver and the closest barrier diffraction edge
r_W	distance between two diffraction edges
T	temperature
T_{30}	reverberation time, defined as the time it takes for the level to decay 60dB, calculated from the level decay of the reverberant soundfield
T_s	center time, defined as the gravity point of acoustical energy over time
t	time
t_0	dividing time between direct and reverberant soundfield (chapter 4 - $t_0 = 12\text{ms}$)
Q	average number of sources passing per hour
v	vehicle speed
v_i	quadrature weight for point i on the sphere
v_{wind}	wind speed
w	street width
$w(k, \phi, \theta)$	magnitude of a plane wave with wavenumber k , incident from (ϕ, θ)

Special functions

$h_n(\cdot)$	n -th order spherical Hankel function of the second kind
$j_n(\cdot)$	n -th order spherical Bessel function of the first kind
$P_n(\cdot)$	Legendre polynomial of order n
$P_n^m(\cdot)$	associated Legendre function order n and degree m
$Y_n^m(\phi, \theta)$	spherical harmonic of order n and degree m

Nederlandse samenvatting

–Summary in Dutch–

Sinds de tweede helft van de twintigste eeuw hebben talloze onderzoeken de invloed van omgevingslawaai op de gezondheid aangetoond. In een recent rapport schatte de wereldgezondheidsorganisatie (WGO) dat in West-Europa jaarlijks meer dan één miljoen gezonde levensjaren verloren gaan door blootstelling aan verkeerslawaai. Slaapstoornissen en ergernis wegen het zwaarst door, maar cardiovasculaire ziekten, cognitieve stoornissen en tinnitus zijn andere gekende gevolgen.

Met het doorvoeren van de END-richtlijn omtrent omgevingslawaai in 2002, erkende de Europese Unie de impact van omgevingslawaai, en verplichtte ze de aanmaak van strategische geluidskaarten om de blootstelling te beoordelen en het opstellen van actieplannen teneinde het omgevingslawaai te reduceren.

De meest efficiënte manier om het effect van verkeerslawaai te reduceren is het verminderen of helemaal verwijderen van de stoorbronnen. Aangezien een doordacht ontwerp van het stedelijke verkeersnetwerk echter niet altijd mogelijk is, zijn geluidsreducerende maatregelen zoals geluidsbarrières dikwijls nodig, terwijl stilte- en recuperatiegebieden voorzien kunnen worden om de inwoners te laten herstellen.

Traditioneel wordt het effect van geluidsreducerende maatregelen enkel uitgedrukt in geluidsniveaus, zonder rekening te houden met de perceptie van de luisteraar. Het doel van dit onderzoek is om stadsplanners met een meer uitgebreide set van ondersteunende methodieken te voorzien, teneinde een subtieler ontwerp mogelijk te maken. Zo worden een aantal auralizatietechnieken voorgesteld, die toelaten om een geluidsvoorbeeld van het effect van toekomstige geluidsreducerende maatregelen te creëren, en wordt het gebruik van zaalakoestische parameters geïntroduceerd, die de evaluatie van de stedelijke geluidsomgeving op basis van geluidsniveaus aanvullen.

In het eerste deel van dit werk worden twee verschillende technieken voor de auralizatie van geluidsreducerende maatregelen voorgesteld. Beide auralizatietechnieken zijn gebaseerd op a priori opnames van de geluidsomgeving, waarin het specifieke geluidsspectrum van het voertuig, de effecten van het wegooppervlak, de aantallen en types van voertuigen enz. reeds aanwezig zijn. Terzelfdertijd worden ook andere bronnen, typisch voor de beschouwde omgeving, opgenomen, wat resulteert in meer realistische auralizaties. Deze techniek verschilt danig van ab

initio auralizaties, waarbij het motor- en rolgeluid van elk voertuig individueel gemodelleerd wordt, terwijl andere bronnen niet in rekening worden gebracht.

Een eerste manier om een geluidsbarrière te auralizeren is gebaseerd op de tijdsafhankelijke verzwakking van een binauraal signaal, opgenomen met een hoofd-en-torso simulator. Hierbij wordt, na het opnemen van passages van individuele voertuigen, elk fragment verzwakt met het richtingsafhankelijke introductieverlies, corresponderend met de ogenblikkelijke positie van de bron.

Deze techniek werd toegepast in een eerste gevallenstudie, waarbij de passage van een voertuig achter een laag begroeid geluidsscherp met beperkte lengte werd geauraliseerd. Hierbij werden de introductieverliezen van het scherm gesimuleerd door gebruik te maken van zowel het ISO 9613-2 propagatiemodel als het Harmoise model en de FDTD simulatietechniek met verschillende gradaties van geometrisch detail. De richting van de bron werd bepaald door gebruik te maken van een biologisch geïnspireerde benadering. De evaluatie van deze auralizatietechniek op basis van luistertesten toont aan dat geauraliseerde fragmenten inderdaad het effect van een dergelijke lage barrière op de geluidsomgeving simuleren. Desalniettemin konden luisteraars, bij vergelijking met originele a posteriori opnames, de geauraliseerde fragmenten alsnog identificeren.

Een tweede auralizatietechniek is gebaseerd op de richtingsafhankelijke verzwakking van de geluidsomgeving. Voor deze techniek is een binaurale directiviteitspatroon emulatietechniek ontworpen, gebaseerd op een sferisch 32-kanaals microfoonrooster. Het microfoonrooster is ontworpen door bestaande technieken uit te breiden met een niet-uniforme microfoonverdeling, waarbij een compromis gezocht is tussen de ruimtelijke aliasing en de afmeting van het sferisch rooster. Dit wordt gerealiseerd door de microfoons in het horizontaal vlak te concentreren. De frequentie-afhankelijke gewichten van elk microfoonsignaal worden bepaald via een geregulariseerde kleinste-kwadratenbenadering, zodat het resulterende directiviteitspatroon van het microfoonrooster de linker en rechter hoofd-gerelateerde transferfuncties, verzwakt met het richtingsafhankelijke introductieverlies, benadert, om zo het effect van de geluidsreducerende maatregel te integreren. Voor hoge frequenties, waar het resultaat van de kleinste-kwadratenbenadering niet voldoende accuraat meer is, wordt een binauraal signaal gegenereerd door enkel gebruik te maken van het microfoonsignaal van de microfoon die het dichtste bij het linker- of rechteroor gelegen is, verzwakt met het introductieverlies berekend over de linker resp. rechter hemisfeer. Het voordeel van deze auralizatietechniek is dat op deze manier het effect van de geluidsreducerende maatregel op meerdere bronnen onderzocht kan worden, waarbij enkel de bronnen uit een bepaalde richting verzwakt worden. Tevens kan de beweging van het hoofd geïntegreerd worden door de gewichten van het microfoonrooster te herberekenen voor elke oriëntatie. Het potentieel van deze auralizatietechniek werd aangetoond in een tweede gevallenstudie, waarbij het effect van een L-vormige berm naast de E17 snelweg op de geluidsomgeving onderzocht werd. Voor de auralizatie werden drie verschillende modellen voor het richtingsafhankelijke introductieverlies ontwikkeld, gebaseerd op het ISO 9613-2 propagatiemodel, het Pierce diffractiemodel voor de berm en het Harmoise model. Door het gemodelleerde introductieverlies te vergelijken

met de werkelijk gemeten waarde, werd afgeleid dat het introductieverlies van de berm het best benaderd werd door het Harmonoise model, terwijl een grote afwijking gevonden werd wanneer het ISO 9613-2 model gebruikt werd. De prestatie van deze auralizatietechniek werd onderzocht in een tweede luistertest. In tegenstelling tot de eerste auralizatietechniek, hadden de luisteraars grote moeite om originele binaurale opnames te onderscheiden van geauralizeerde fragmenten.

In het tweede deel van dit werk wordt onderzocht of kwantitatieve zaalakoestische parameters gebruikt kunnen worden om de parameters gerelateerd met het geluidsniveau, die stadsplanners dikwijls gebruiken om de akoestiek op pleinen en straten te beschrijven, aan te vullen. Hiervoor werden parameters, gerelateerd met de akoestische energieverdeling, reverberatie en binaurale kwaliteit gemeten op vijf verschillende pleinen tijdens rockconcerten. Door gebruik te maken van correlatieanalyse is een statistisch onafhankelijke parameterset geconstrueerd, bestaande uit de helderheid C_{80} , reverberatietijd T_{30} , binaurale parameters $[1 - IACC_{E3/L3}]$ en het relatieve geluidsniveau $\Delta L_{eq,A/C}$, wat toelaat om een vergelijking tussen de verschillende pleinen te maken. Zo vond men dat de reverberatietijd T_{30} hoofdzakelijk afhangt van de vorm en afmetingen van het plein, terwijl de positionering van de bron slechts weinig invloed heeft.

Geïnspireerd door de zaalakoestische parameters is een nieuwe parameter, de reflectieverhouding RR , gedefinieerd als de verhouding van het reverberante tot het directe geluidsveld, voorgesteld om de relatieve bijdrage van de reflecties in stedelijke straten te beschrijven. Het gebruik van deze parameter voor het aanmaken van geluidskaarten kan zowel de accuraatheid als efficiëntie verhogen, aangezien nu alle reflecties in rekening worden gebracht, terwijl de rekentijd sterk gereduceerd wordt, in vergelijking met de klassieke methodes voor het in kaart brengen van geluid. De reflectieverhouding is onderzocht in 99 straten tijdens een extensieve meetcampagne met een specifiek ontworpen en degelijk gekarakteriseerde meet setup, gemonteerd op het dak van een auto, om snelle en gemakkelijke metingen mogelijk te maken. Gebaseerd op deze data is een model ontworpen om de RR te voorspellen op basis van geometrische en architecturale karakteristieken van de straat. De RR is in hoge mate gecorreleerd met de breedte van de straat, terwijl een beperkt verband met de hoogte of de ruwheid van de gevels gevonden werd. Op basis van de RR en de energie van het rechtstreekse geluid is een model ontwikkeld om de verdeling van het geluidsniveau in een straat te voorspellen. Dit model werd gevalideerd door berekeningen van het geluidsniveau te vergelijken met metingen in een straat.

English summary

Since the second half of the twentieth century, numerous investigations have proven the link between exposure to environmental noise and adverse health effects. In a recent report, the WHO estimated that every year at least one million healthy life years were lost in Western Europe due to traffic noise exposure. Sleep disturbance and noise annoyance are the main burden, while cardiovascular diseases, cognitive impairment and tinnitus are other known effects.

With the Environmental Noise Directive (END), enacted in 2002, the EU recognized these impacts of environmental noise and ordained the creation of strategic noise maps to assess the exposure and action plans for the reduction of environmental noise.

The most efficient measure to reduce the effect of traffic noise is vehicle source reduction or removal. As a smart design of the urban traffic network is not always feasible, noise mitigation measures such as sound barriers are often needed, while silent areas and restoration zones can be provided to allow inhabitants to recover. Traditionally, the effect of such noise mitigation measures is expressed by noise levels only and the listener's perception is not accounted for. This work aims at providing urban planners with a more extensive set of tools to achieve a more subtle design. A set of auralization methodologies is provided that enables an auditive preview of future noise mitigation measures, while the use of a set of room acoustic parameters is introduced to complement the level-based evaluation of the urban sound environment.

In the first part of this work, two different auralization methodologies for urban noise mitigation measures are proposed. Both auralization methodologies are based on a priori recordings of the sound environment, already including the vehicle-specific source spectrum, effects of the local road surface, vehicle fleet composition, traffic flow etc., while other sources, typical for that site, are accounted for as well, resulting in more realistic auralizations. This methodology differs from *ab initio* auralizations, which start from modeling the individual vehicle engine and rolling noise sources and do not account for other sources present at the site.

A first technique to auralize a noise barrier is based on the time-dependent attenuation of a binaural signal, recorded with a head-and-torso simulator. Individual car passages were recorded and each fragment was attenuated with the angular insertion loss (IL), corresponding with the momentaneous position of the source.

This technique is applied in a real case-study, where a car passage behind a low

finite-length vegetated noise barrier is auralized. In this case, the IL of the barrier is simulated by means of the ISO 9613-2 propagation model, the Harmonoise point-to-point model and the full-wave FDTD technique, for different degrees of geometrical complexity. The direction of the source was extracted by using a biologically inspired approach. Listener-based evaluation of this methodology showed that auralized samples indeed incorporated the effect of such a low barrier on the sound environment, however, listeners were still able to identify auralized samples when opposed to original a posteriori recordings.

A second auralization methodology is based on a direction-dependent attenuation of the sound environment. In function of this auralization methodology a binaural pattern emulation technique has been developed based on a spherical 32-channel microphone array. The array has been designed by extending existing techniques to a non-uniform distribution of microphones that is a compromise between spatial sampling and spherical array size. This is achieved by concentrating the microphones towards the horizontal plane. By use of a regularized least-square fit, the frequency dependent weights of each microphone signal have been determined so that the resulting directivity pattern of the array approximates the left and right head-related transfer functions (HRTFs), attenuated with the angular dependent IL to incorporate the effect of the noise mitigation measure. For high frequencies, where the solution of the least-square fit is not accurate enough, the motion-tracked binaural sound (MTB) technique is implemented by selecting only the signal from the microphone closest to the left or right ear position, attenuated with the IL calculated over the left resp. right ear hemisphere. The advantage of this auralization methodology is that the effect of the noise mitigation measure on multiple sources can be investigated, only attenuating sources from a certain direction. Furthermore, movement of the head can be incorporated by recalculating the array weights for each orientation.

The potential of this auralization technique is shown in a second case-study, where the effect of an L-shaped mound near the E17 highway on the sound environment is investigated. Prior to auralization, three different models for the angular IL of the mound are created based on the ISO 9613-2 propagation model, Pierce barrier diffraction model and the Harmonoise point-to-point model. By comparing the predicted IL with the measured IL, it was concluded that the highway IL of the mound is best approximated by the Harmonoise model, while a large deviation is found when using the ISO 9613-2 model. The performance of the auralization technique is investigated in a second listening test. In contrast to the first auralization methodology, listeners had great difficulty in differentiating between original binaural a posteriori recordings and auralized samples.

In the second part of this work, it is investigated if quantitative room acoustic parameters can be used to complement existing sound pressure level parameters urban planners nowadays use to describe the acoustics at urban squares and streets. Therefore, parameters related to energy distribution, reverberation and binaural quality are measured on five different squares while used for rock concerts. A statistical independent set, containing the clarity C_{80} , reverberation T_{30} , binaural

parameters $[1 - \text{IACC}_{E3/L3}]$ and relative sound pressure level $\Delta L_{eq,A/C}$, is extracted by use of correlation analysis and a comparison between squares is made. It is found that T_{30} is mainly dependent on the shape and dimensions of the square, but only little on the used source positioning.

Inspired by the room acoustic measures, a new parameter, the reflection ratio RR, is defined as the ratio of the reverberant-to-direct sound energy to describe the relative contribution of reflections in urban street canyons. The use of this parameter in noise mapping can largely improve both their accuracy and efficiency, as it takes the contributions of all reflections into account, while the calculation time is strongly reduced relative to classical reflection-based noise mapping methods. In an extensive measurement campaign, this parameter is investigated in 99 urban street canyons with a dedicated and well-characterized set-up, mounted on top of a car to enable fast and easy measurements. Based on the multitude of data, a model is created to predict the RR based on geometrical and architectural characteristics of the street. It was concluded that the RR correlates mostly with street width, while it is dependent to a limited extent on either street height or facade roughness. Based on the RR and the knowledge of the direct sound energy, a model is developed to predict the sound pressure level distribution in a street canyon. The model was validated by comparing predicted levels with sound levels measured in a street canyon.

1

Introduction

The efficient organization of the urban structure has been a challenge ever since people started concentrating their living in cities. The ancient Greek architect Hippodamus of Miletus (498-408 BC) was the first known urban planner, with his design of the orthogonal city lay-out. However, urban planners already efficiently organized life in the city many centuries before his time.

In the ancient times urban city-planners did not bother about the minimization of the urban noise. As such, the city of Rome was known as one of the noisiest at that time. Around 50 BC, Rome counted about one million inhabitants, while thousands of merchants and tourists visited the city daily. In order to control the traffic, Julius Caesar ordained that no carts were allowed from sunset to two hours before sunrise. The effect of this daytime traffic ban on people's sleep is evident, and inspired the Roman poet Juvenalis while writing his Satires [1, 2]. At that time, the only way to escape the buzzing of the city was to move to the countryside or suburbs.

Since the Roman times a lot has changed, unfortunately not always for the best. Although over the ages more thought was given to efficient planning, the industrial revolution introduced completely different noises in the city, originating from steam engines, smithies, building activities etc., as no separation between residential and industrial areas was made [3]. Nowadays, the current urban noise issue is mainly caused by motorized road vehicles introduced by the end of the nineteenth century.

Only during the last 30 years, researchers identified the link between adverse health effects and environmental noise. In 1999, the WHO summarized the evidence of health impacts of noise exposure. It was found that environmental noise resulted in annoyance, sleep disturbance, cardiovascular problems, cognitive impairment of children and noise-induced hearing loss [4]. The WHO calculated that every year more than one million healthy life years are lost in Western Europe due to the negative effects of noise exposure [5]. Approximately one in three individuals experiences daytime annoyance and one in five suffers from noise-induced sleep disturbance.

Based on these findings, the European Council has taken initiatives to stimulate the reduction of the harmful effects from transport noise. This resulted in the enactment of the Environmental Noise Directive (END) 2002/49/EC for the reduction of environmental noise in 2002 [6], ordaining the calculation of noise maps to evaluate the noise exposure, and the design of action plans to limit the (traffic) noise. Ideally, the most efficient measure would be to remove or strongly reduce the vehicle noise sources, by guiding the traffic away from inhabited areas. However, as such smart design of the (existing) traffic network is not always feasible, noise mitigation measures such as barriers are often used to screen the people from traffic noise. Furthermore, quiet (green) areas and restoration zones can be included to allow the city-inhabitants to mentally recover.

The assessment of the effect of such noise measures on the sound environment is typically based on the (predicted) sound level only [7], and does not account for the perceptual evaluation.

This work aims at presenting a more comprehensive set of tools to aid city-planners in the design of the urban sound environment and the assessment of noise mitigation measures. Generally, well-established prediction models are used to evaluate the effect of mitigation measures in terms of sound pressure levels. Such quantitative characterization is however far from comprehensible for a broad public, and a more auditive preview is desired. Therefore, two new auralization methodologies are designed to allow a binaural impression of the effect of a given measure on the sound environment, and it will be investigated to what extent these methodologies can be used for the assessment of noise mitigation measures. The designed auralization methodologies are based on recordings prior to the installation of the noise mitigation measure. In contrast to existing *ab initio* auralization techniques, these methodologies already include effects of the local road surface, vehicle fleet composition and traffic flow, while other sources present at a given sound environment can be included as well, resulting in more realistic auralizations. For this purpose, a binaural pattern emulation technique has been developed based on a 32-channel spherical microphone array, allowing for spatial-selective source attenuation. Supplementary to the evaluation of the outdoor sound environment by means of

auditive previews, a set of room acoustic based parameters are proposed that can be used for a more extensive qualitative characterization of urban square and street acoustics, complementary to existing noise level parameters. As these room acoustic parameters can be linked to the quality perception of music, the use of these parameters to describe the acoustics on urban squares for open-air concerts is illustrated. For more general urban noise, street reverberance plays a significant role in the (perceived) level increase. Therefore, an extra parameter is proposed that estimates the contribution of the reverberant sound, solely based on the knowledge of the direct sound and architectural characteristics of the street. It is illustrated how this parameter can be used to aid in the fast and accurate calculation of noise maps.

This research contributed to the European HOSANNA project on the holistic and sustainable noise abatement by optimized combinations of natural and artificial means [8] and the IDEA project for the creation of (mobile) sensor networks for monitoring environmental noise [9]. Results regarding the characterization of square acoustics during rock concerts were used by the Flemish government in designing legislation concerning hearing loss prevention during music events.

1.1 Outline

This work is organized as follows. In chapter 2 a methodology is developed to create an immersive audio environment for the assessment of outdoor sound fields. Based on an extensive study of existing recording and reproduction techniques, a 32-channel spherical microphone array is proposed for outdoor sound recording combined with binaural reproduction over headphones. The binaural signal is recreated by means of the pattern emulation technique and the motion-tracked binaural sound (MTB) technique. The performance of the methodology for binaural reproduction of outdoor environments is evaluated by listening tests.

In chapter 3, two new methodologies are proposed to auralize the effect of noise mitigation measures on the sound environment based on a priori recordings. A first methodology is based on binaural HATS recordings, while the effect of a barrier is included by time-dependent source attenuation. A second auralization methodology uses the microphone array technique developed in chapter 2 for the recording and reproduction of the sound environment. The noise mitigation measure is accounted for by directionally-dependent attenuation of the sources. Both methodologies are used in two different case-studies, and a qualitative assessment is made by listening tests.

Chapter 4 introduces the use of (room) acoustic parameters, enabling a more qualitative analysis to complement existing noise level parameters for the evaluation of urban squares used for (rock) concerts. Based on measurements at five different squares, a suitable parameter set is selected and the behavior of the parameters as

a function of square geometry is analyzed.

In chapter 5 the reflection ratio RR is introduced to characterize the ratio of reverberant-to-direct energy and evaluate the level increase due to street reverberance. This parameter is measured in 99 streets with a dedicated and well-characterized measurement set-up. Based on these measurements, a model to predict the RR as a function of geometrical and architectural properties of the street is designed. Furthermore, it is shown to what extent this parameter can be used to calculate sound levels in a street canyon, based on knowledge of the direct sound energy only. In a last chapter, the main conclusions and perspectives have been bundled.

1.2 Publications

1.2.1 Publications in international journals

- P. Thomas, W. Wei, E. Steyaert, T. Van Renterghem, and D. Botteldooren. *Array-based auralization methodology for the assessment of noise mitigation measures*. Under review.
- M. Adnadevic, A. Bockstael, P. Thomas, D. Botteldooren, and H. Keppler. *Development, validation and application of a simulator for distortion product otoacoustic emissions*. Submitted to J. Acoust. Soc. Am..
- P. Thomas, T. Van Renterghem, E. De Boeck, L. Dragonetti, and D. Botteldooren. *Reverberation-based urban street sound level prediction*. J. Acoust. Soc. Am., 133(6):3929-3939, June 2013.
- T. Van Renterghem, P. Thomas, F. Dominguez, S. Dauwe, A. Touhafi, B. Dhoedt, and D. Botteldooren. *On the ability of consumer electronics microphones for environmental noise monitoring*. Journal of Environmental Monitoring, 13(3):544–552, December 2011.
- P. Thomas, T. Van Renterghem, and D. Botteldooren. *Using room acoustical parameters for evaluating the quality of urban squares for open-air rock concerts*. Applied Acoustics, 72(4):210–220, November 2011.
- A. Can, T. Van Renterghem, M. Rademaker, S. Dauwe, P. Thomas, B. De Baets, and D. Botteldooren. *Sampling approaches to predict urban street noise levels using fixed and temporary microphones*. Journal of Environmental Monitoring, 13(10):2710–2719, August 2011.

1.2.2 Publications in international conferences

- D. Botteldooren, T. Van Renterghem, D. Oldoni, S. Dauwe, L. Dekoninck, P. Thomas, W. Wei, M. Boes, B. De Coensel, B. De Baets, and B. Dhoedt. *The internet of sound observatories*. J. Acoust. Soc. Am. - Proceedings of Meetings on Acoustics, 133(5):3592, 2013.
- P. Thomas, M. Boes, T. Van Renterghem, D. Botteldooren, M. Hornikx, W. Desmet, and M. E. Nilsson. *Auralisation of a car pass-by behind a low finite-length vegetated noise barrier*. In Proceedings of the 9th European Conference on Noise Control, pages 932–937. European Acoustics Association (EAA), 2012.
- P. Thomas, E. De Boeck, L. Dragonetti, T. Van Renterghem, and D. Botteldooren. *Creating an urban street reverberation map*. In Proceedings of Forum Acusticum, pages 397–402. European Acoustics Association (EAA), 2011.
- P. Thomas, L. Dragonetti, T. Van Renterghem, and D. Botteldooren. *Detailed analysis of the sound field in a scale model of a street canyon*. In Proceedings of the EAA Euregio 2010 congress, pages 65:1–65:4, 2010.
- P. Thomas, T. Van Renterghem, L. Dekoninck, and D. Botteldooren. *Sound-field research of urban squares used for outdoor rock concerts*. In Proceedings of the 39th international congress and exposition on noise control engineering (Inter-Noise 2010) : paper 566, 2010.
- T. Van Renterghem, D. Botteldooren, P. Thomas, A. Can, S. Dauwe, B. Dhoedt, M. Rademaker, B. De Baets, A. Touhafi, F. Bominguez, J. Theunis, and R. O'Donoghue. *Towards an extensive noise and air quality measurement network*. In Proceedings of the 39th international congress and exposition on noise control engineering (Inter-Noise 2010) : paper 278, 2010.
- P. Thomas, T. Van Renterghem, L. Dekoninck, and D. Botteldooren. *Urban square sound fields during rock concerts*. In Proceedings of the Institute of Acoustics, volume 32, pages 292–297, 2010.
- T. Van Renterghem, P. Thomas, D. Botteldooren, S. Dauwe, B. Dhoedt, F. Dominguez, and A. Touhafi. *The use of cheap microphones in extensive outdoor noise monitoring networks*. In Proceedings of the Institute of Acoustics, volume 32, pages 374–377, 2010.
- P. Thomas, T. Van Renterghem, L. Dekoninck, and D. Botteldooren. *Acoustic evaluation of public squares used for outdoor concerts*. In Proceedings of the 8th European conference on Noise Control, 2009.

References

- [1] Juvenalis. *Satires III*, v.232-238. 60-140 AD.
- [2] K. Matthews. *The embattled driver in ancient Rome*. Expedition, pages 22–27, 1960.
- [3] A.E.G. Jacobs. *Het geluid van gisteren: waarom Amsterdam vroeger ook niet stil was*. PhD thesis, Maastricht University, 2014.
- [4] *Guidelines for community noise*. World Health Organization, Geneva, Switzerland, 1999.
- [5] *Burden of disease from environmental noise*. World Health Organization, Geneva, Switzerland, 2011.
- [6] *Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 relating to the assessment and management of environmental noise*. Official Journal of the European Communities, 189:12–25, 2002.
- [7] Gaetano Licitra, editor. *Noise mapping in the EU*. CRC Press, 2013.
- [8] *HOSANNA*. www.greener-cities.eu.
- [9] *IDEA project*. www.idea-project.be.

2

Recording and playback of a 3D outdoor sound environment

The use of spatial audio reproduction techniques is widely employed for the (subjective) analysis of concert halls, often in combination with a visual component [1, 2]. The fact that the listener can evaluate different auditory scenes under controllable circumstances, without the necessity of actually being at the location, is one of the biggest advantages of this technique.

Recently, the use of immersive techniques has been applied to the evaluation of outdoor environments and more particularly to the assessment of noise mitigation measures such as noise barriers [3, 4]. When combined with a proper auralization engine, a very powerful tool can be created, enabling city-planners to investigate the qualitative effect of their design prior to realization and to present the result in an audible way to the general public and decision makers [5].

The structure of this chapter is as follows. In the first section, an overview of the different human hearing mechanisms to localize sound sources is given. The second section provides a literature study of existing recording and reproduction techniques for immersive audio recreation. Special emphasis is put on the array recording techniques and binaural reproduction, as these techniques will be expanded in the next section, where the design and implementation of a spherical 32-channel microphone array to emulate the human HRTFs is discussed. In the last section an assessment of the proposed array recording technique is made. The different design choices are compared and evaluated by listening tests based on binaurally reproduced recordings of an outdoor environment.

2.1 The human sound localization

Prior to the design of an immersive audio environment, a proper insight in the different hearing mechanisms to localize sound sources is necessary. The ability of the human hearing to localize sound sources is remarkable, as only two signals, one from each ear, are used to enable proper source recognition, localization and extraction, even in noisy environments.

Since the duplex-theory of Lord Rayleigh [6], published in 1907, many different localization mechanisms or cues have been identified. In what follows, a summary of these cues is given, based on [7–11]. Although all these cues have to be in accordance to enable clear localization, some cues dominate for specific localization aspects such as azimuthal and elevational source detection and distance estimation.

For the detection of azimuthal angle, the human listener is largely dependent on interaural signal differences [12]. Lord Rayleigh identified the interaural time (ITD) and level (ILD) difference to be the main cues. He formulated a duplex-theory of sound localization for pure-tone stimuli, which has been extended ever since. At high frequencies, the wavelength of the sound is smaller than the dimensions of the head, and the contralateral ear will be shielded from the source, leading to an angle dependent level difference between the signal from both ears. At low frequencies, the wavelength is much larger than the dimensions of the head, and level differences are negligible. For these frequencies, the listener relies on time differences between sound arrival at both ears. The cross-over frequency is approx. 1.5kHz, although a transition zone is defined where both cues play a role.

The duplex-theory has been recently updated by Macpherson et al. [12] for broadband sources. Macpherson found that for low-pass signals the ITD was the most significant cue, while the influence of the ILD was negligible and vice-versa for high-pass signals. For wideband stimuli however, the importance of the ITD cue was at least equal to or dominated the ILD cue. Furthermore, it was seen that the ITD derived from onset difference and ongoing envelope could still play a role in high frequency stimuli localization.

The use of ITD and ILD cues alone, however, does not enable the listener to unambiguously determine the location of the sound source. Since a source in the frontal plane of the listener has comparable ITD and ILD values as with the mirrored source in the back plane, front-back (or back-front) confusion can occur. Surfaces of equal ITD and ILD values are called cones of confusion. Different cues can be used to resolve these ambiguities [13–16]. For sound sources with high frequency content, interaction and shielding by the pinnae of the ear can already deliver a sufficient cue. These effects typically play a role from 3kHz on. For lower frequency and broadband sources, dynamic cues are used. Such cues are delivered by (small) head movements, inducing changes in ITD and ILD, which are able to deliver a

sufficient localization cue. Already in 1939, Wallach suggested that dynamic cues dominate pinnae cues, which has been confirmed ever since.

Based upon these different cues, a listener is able to discriminate sources in the azimuthal plane with high accuracy. For azimuthal angles, the minimal audible angle (MAA), which is a measure of resolution of the auditory space that a listener can discriminate, can take values less than 1° for sources directly in front of the listener. However, for increasing azimuthal source positions, the MAA increases as well, to approx. 10° for sources close to the side [17–19].

The discrimination of source elevation is mainly based on monaural cues related to the pinnae and proves to be more difficult than azimuthal source angle detection. Interferences between direct sound and reflections on the cavum concha of the pinnae give rise to high frequency spectral peaks and notches (comb-filtering). Knowledge on source elevation can be extracted from the location of the interference notch in the spectrum, as the path length difference between reflections depends on source elevation. The use of monaural spectral cues is however based on a priori assumptions of the source spectrum, as the listener cannot know whether or not spectral notches were induced by the pinnae or were already present in the original source signal. Therefore, a certain smoothness of spectrum is assumed [8]. The use of the pinnae has little to no value for elevation discrimination at frequencies below 3kHz due to their size relative to the wavelength. Algazi found that for elevation detection of sources with low frequency spectral content, the listener mainly relies on cues caused by torso reflections and head diffraction [20], while Perrett indicated that dynamic cues could also improve elevation detection [21]. Based on the effectiveness of the different cues, the MAA for elevation discrimination was found to be 4° under optimal conditions, or two to four times larger than the MAA for azimuthal angle discrimination [18, 19]. Given the fact that the discrimination of source elevation is largely dependent on the shape of the pinnae, which differs significantly between listeners, a difference in elevation discrimination performance between listeners is to be expected. Furthermore, the way the brain interprets the monaural cues can be different. This proves to be problematic when generalized pinnae cues are used for the creation of an immersive audio environment, which can completely destroy the elevation discrimination ability of the listener and influence the perceived source elevation [10, 14].

In the foregoing, focus has been on the identification of the direction of a source, also called lateralization. For a complete localization of the source in the auditory space, mechanisms for distance perception should be included as well.

Different distance cues have already been identified [7, 10, 22, 23]. One of the main cues is based on the sound intensity of the source. If the listener is familiar with the original source level or the level can be referred to other source posi-

tions, an estimate of the distance can be made based on the geometrical divergence of the source. For sources such as speech a reasonable good distance estimation can be made. When the listener is unfamiliar with the source (level), the direct-to-reverberant energy ratio can provide an absolute cue together with an estimate of the acoustical properties of the room (room volume and reverberation time). Based on these parameters, Bronkhorst et al. developed a mathematical model to estimate perceived distance [24]. Another cue is given by the influence of different distance dependent mechanisms, such as air and reflective surface absorption, on the spectrum [23]. Again, certain familiarity with the original source spectrum ameliorates the effect of these spectral cues. Additionally, Brungart et al. illustrated that the low frequency ILD provides an extra cue for sources closer than 1m, given its short-range distance dependency [23, 25]. However, it was seen that performance degraded for source locations approaching the median plane, where interaural differences are zero, and for sources further away, when interaural differences become distance independent.

In contrast to the source lateralization ability, source distance estimation is subject to large errors. Performance strongly depends on the specific situation and usability of different cues, which varies with source spectral content and environment. Zahorik et al. suggested that the perceived distance can be estimated by a compressive power function of the physical distance [23]. Distances of far sources are generally underestimated, indicating the existence of an auditory horizon, while the distance of sources close to the listener tends to be overestimated. Furthermore, Zahorik reported large differences in responses between listeners, which increase with increasing distance.

Closely related with distance perception is the concept of sound source externalization. Here, the presence of distance cues aids to resolve in-head localization, which often occurs when listening over headphones. In [26], Gardner mentioned the most important acoustical cues to be the pinnae cues, reverberation cues and dynamic cues. Furthermore, Hartmann et al. denote the influence of interaural cues and indicate the importance of maintaining realistic monaural spectral cues [27]. Other research however showed that sources can still be externalized if cues other than individual spectral cues are still present.

2.2 Creating an immersive audio environment: an overview of existing techniques

The build-up of an immersive audio environment typically consists of a capture and reproduction phase. Optionally, a generation phase can be inserted, in which a virtual source is artificially generated and located in the auditory space. In this

work we will only focus on the recreation and adaptation of existing soundfields based on recordings. The main phase of this process is the capture phase, where the soundfield is recorded and decomposed. The sole constraint is that the auditory scene, which is reproduced later on from these recordings, is perceived by the listener the same way as if he was physically present during the recordings. This does not necessarily mean that the reproduced soundfield is physically exactly the same as the originally recorded soundfield. The listener should merely have the same perception [28].

2.2.1 Soundfield capturing

Different techniques exist to capture a given soundfield for immersive reproduction, each with different degrees of complexity. The most straightforward method is to record a soundfield with a Head-and-Torso Simulator (HATS). The HATS mimics the different auditory cues that enable a listener to identify, localize and separate different sound sources. The general principle is that it is sufficient to reproduce the pressure at the listener's eardrums in order to give the same perception [29]. Although its simplicity is attractive, there are some inherent drawbacks of this technique. Firstly, the HATS is constructed based on a generalized model for the human torso, head and ears. Individual physical properties of the listeners ears, head and torso are not taken into account, leading to increased front-back confusion, source internalization and decreased accuracy of elevational hearing. Furthermore, as the position of the HATS is fixed during recording, the orientation of the listener is restricted to that specific direction, disabling the use of motion cues [10, 14, 30].

By allowing rotation of the listener's head, front-back confusion can be resolved and externalization of the source can be improved [15]. This is supported by the Motion-Tracked Binaural sound (MTB) technique, developed by Algazi et al. [10, 31, 32]. Basically, the soundfield is sampled by multiple microphones, located on a sphere with dimensions comparable to a human head (a first order spherical-head model). Depending on the orientation of the head, derived from a head tracker, a binaural signal is extracted from the microphones closest to the listener's ears. At low frequencies, the signal can be constructed by interpolation between different microphone signals. At high frequencies, for which half of the wavelength is smaller than the inter-microphone distance, only the signal from the microphone closest to the ear is used to avoid spatial aliasing artifacts. The ability of the MTB technique to provide additional motion cues strongly reduces the need for listener customization as better source externalization and front-back localization can be achieved, even though the pinnae are omitted [13, 15]. However, for elevational source localization, the presence of the monaural spectral cues induced by the pinnae is still necessary. Additionally, Algazi proposed to introduce spectral

notches to compensate for this [31, 32].

Other techniques try to capture and decompose the soundfield directly, without accounting for the different auditory cues. These cues are generated later on, in the reproduction phase. A well used decomposition technique is the higher order Ambisonic recording technique. Ambisonic is based on a spatial harmonic decomposition of the soundfield. The basis functions in spherical coordinates can be written as a combination of spherical Bessel functions $j_n(kr)$, accounting for the radial distance, and the spherical harmonics $Y_n^m(\phi, \theta)$, a set of orthonormal functions grouping the dependency on azimuth and elevation:

$$p(kr, \phi, \theta) = \sum_{n=0}^{\infty} \sum_{m=-n}^n A^{nm}(k) j_n(kr) Y_n^m(\phi, \theta). \quad (2.1)$$

The coefficients $A^{nm}(k)$ represent the frequency dependent spherical harmonic (Ambisonic) coefficients, which can be calculated as follows:

$$A^{nm}(k) = \frac{1}{j_n(kr)} \int_0^{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} p(kr, \phi, \theta) Y_n^{m*}(\phi, \theta) \cos(\theta) d\theta d\phi. \quad (2.2)$$

The theory was first developed by Gerzon in 1973 [33, 34] and led to the design of a first order Ambisonic Soundfield microphone [35, 36]. To achieve higher spatial resolution, a higher order decomposition becomes necessary. For this purpose, spherical microphone arrays have been developed [37–39]. In theory, a continuous spherical microphone would be able to resolve all coefficients from Eq. 2.2. However, practical arrays can only sample the soundfield at discrete microphone locations, reducing the integrals in Eq. 2.2 to a discrete sum. Some limitations arise from this discretization [37, 38]. At first, spatial aliasing can occur when half of the wavelength is smaller than the inter-microphone distance. Furthermore, due to undersampling, the orthonormality relation breaks down for higher order spherical harmonics, making it impossible to unambiguously determine their coefficients. These problems can be solved by restricting the frequency range and order N of the coefficients (restricting the infinite sum in Eq. 2.1 to order N). Furthermore, the number of resolvable coefficients can be increased by using a larger number of microphones and optimizing their positions on the sphere, together with a suitable spherical quadrature [37, 38]. However, if the soundfield is not order limited, like e.g. a plane wave, or if the soundfield contains harmonics with order higher than the highest resolvable coefficient, higher order harmonics will be aliased onto the lower order coefficients, comparable to aliasing in the time domain [40].

Furthermore, it should be pointed out that the use of an open sphere configuration in Eq. 2.2 gives rise to numerical instabilities due to zeros in the spherical Bessel function. This can be accounted for by using a rigid sphere configuration, or even

cardioid microphones [38, 39]. However, at low frequencies (low kr) higher order coefficients are still difficult to resolve, as the inverse of the (rigid sphere) radial function takes small - but non-zero - values for increasing order n .

In [38], Rafaely estimates the influence of errors due to measurement noise, position errors and microphone response. At high frequencies, the ability of the array to decompose the soundfield is limited by spatial aliasing (from $kr > N$ on), while at low frequencies, the measurement noise and position errors impose a limit.

A special form of soundfield decomposition, based on the spherical harmonic formalism, is the plane wave decomposition [41, 42]. Here, the soundfield is written as an infinite sum of plane waves with different arrival directions (ϕ_l, θ_l) , each with a directional amplitude $w(k, \phi_l, \theta_l)$. Based on the property that a plane wave with unit amplitude, incident from direction (ϕ_l, θ_l) , decomposes into harmonic coefficients $Y_n^{m*}(\phi_l, \theta_l)$ and the orthonormality relation of the spherical harmonics, the directional amplitudes, also called plane wave components, can be written as

$$w(k, \phi_l, \theta_l) = \sum_{n=0}^{\infty} \sum_{m=-n}^n A^{nm}(k) Y_n^m(\phi_l, \theta_l), \quad (2.3)$$

with $A^{nm}(k)$ the spherical harmonic decomposition of the soundfield, given by Eq. 2.2. This indicates that if the spherical harmonic coefficients for a given soundfield are known, the directional amplitudes of the plane waves can easily be extracted. In [43], Duraiswami showed that this decomposition can be extended for the near-field case as well.

Another way to decompose the soundfield into its plane wave components, is to make use of a beamformer. The design of such a beamformer is based on the spherical harmonic decomposition of a spatial Dirac function, pointed towards a given look-angle (ϕ_l, θ_l) , and the assumption that the soundfield is constructed out of plane waves [44–47]. The plane wave decomposition of the soundfield can be found by steering the beamformer to all possible directions and recording the output as the plane wave components. The performance of the practical design of the beamformer is bound to limitations inherent to spatial sampling and thus limits the degree of spherical harmonic coefficients that can be used to approximate the spatial Dirac function. In practice, a restriction to lower orders broadens the approximation of the spatial Dirac function, and plane waves with other directions than the look-angle will influence the decomposition.

Another technique to capture the soundfield is to use microphone arrays to generate virtual microphones with dedicated directivity patterns [48–50]. The virtual microphones are constructed in such a way that each of their output signals can be directly used by the loudspeakers in the reproduction system, without depending on an intermediate decomposition in spherical harmonic coefficients or plane

wave components. The desired virtual microphones are created by making a linear combination of the measured directivity patterns of the different microphones in the array. By using measured directivity patterns, errors related to deviations from the theoretical microphone directivity patterns (caused by the presence of cable outlets or array fixation systems), microphone position errors, calibration errors and deviation in the frequency response can implicitly be taken into account.

In this work, an adapted version of the virtual microphone technique will be preferred, due to the ability to directly emulate the directivity pattern of an individual human listener, based on the head-related transfer functions (HRTFs) [50, 51]. In this way, individual physiological characteristics of the listener can be taken into account, while head rotation can be implemented by updating the microphone weights when the listener rotates his head.

Furthermore, in view of the auralization of noise mitigation measures, this technique allows to directly incorporate the effect of the measure by including spatial selectivity of the target directivity pattern.

2.2.2 Soundfield reproduction

Together with the soundfield recording phase, the sound reproduction phase plays the key role in the construction of an immersive audio environment. Many techniques exist to offer a submerged experience to the listener, each with different advantages and performances. Although reproduction techniques are often coupled with a specific recording technique, some soundfield representations allow an interchange between reproduction techniques.

Binaural audio reproduction is the most cost-efficient reproduction technique, as the soundfield is directly reconstructed at the listener's eardrums [10]. The degree of immersivity depends on how well the different (listener-specific) auditory cues are accounted for in the left and right ear signal. Therefore, the different auditory cues should already be present in the left and right ear signal. It follows that HATS recordings or recordings based on the MTB technique can be applied directly. However, personal ear geometry is ignored. Other representations of the soundfield can be used as well, under the condition that auditory cues are inserted by using (listener-specific) HRTFs. For example, the Ambisonic representation can be converted by calculating the signal for a virtual set of speakers and filtering each loudspeaker signal with the related HRTF, while for a plane wave representation one only needs to filter each incident plane wave with the corresponding HRTF [29, 52]. In [53], Avni created a binaural signal by using an Ambisonic decomposition of the soundfield and HRTFs. This study showed that the resulting binaural representation was perceived as more externalized, thinner and more dis-

tant, while the timbre is more balanced when higher order coefficients are used. Lower order decomposition introduces audible artifacts at higher frequencies due to spatial aliasing.

For the reproduction at the eardrums, headphones are often used, allowing a very portable and easy-to-use immersive system. The use of a mere headphone, however, only allows a static configuration while in reality the auditory scene will move with the listener when he changes orientation. A head tracker can account for this by registering current head orientation and adapting the auditory cues accordingly. This will also improve externalization and diminish front-back confusion [15, 54].

A more advanced technique for binaural reproduction is the crosstalk cancellation technique [26]. The crosstalk cancellation system typically consists of two closely spaced loudspeakers in front of the listener. Filtering is applied to each loudspeaker in order to prevent signals reaching the contralateral ear (crosstalk), so that the left resp. right eardrum only receives sound from the left resp. right loudspeaker. By comparison with binaural reproduction over headphones, Gardner found that the crosstalk cancellation technique improves externalization, while front-back confusion was reduced, even for static recordings [26]. When deployed in an anechoic environment, Damaske found a nearly ideal localization in the horizontal plane [55]. The inclusion of a head tracker to allow listener rotation and an extra pair of speakers behind the listener improves performance even more. Although this technique allows head motion (translation), the sweet spot (artifact-free zone) is rather limited and can only accommodate for one listener [54].

In order to overcome the problems related with binaural reproduction, more complex reproduction mechanisms have been developed, without attaching headphones or head trackers to the listener's head and relying on the presence of auditory cues in the recordings [28].

Panning techniques have been in use for many years. By level or time-delay differences, a virtual source can be located in between loudspeakers. Stereophony is a basic example of this technique. In [56, 57], Pulkki developed the vector-based amplitude panning (VBAP) technique, an extension of pairwise amplitude panning to tripletwise panning in 3D by using a spherical loudspeaker configuration around the listener. VBAP is a local panning approach, as maximum three loudspeakers are active for the reproduction of a virtual source. A bigger sweet spot than with the crosstalk cancellation algorithm can be achieved. Outside the sweet spot the localization of sources collapses to the nearest active loudspeaker. The use of more loudspeakers tends to broaden the sweet spot and increases localization ability.

The Ambisonic reproduction technique is based on the spherical harmonic representation of the soundfield. Typically, Ambisonic can be categorized as a global

panning technique, as all loudspeakers are active, even for the reproduction of a source at a single position [28]. In its basic form, the speaker signals can be calculated by Eq. 2.1. Other decoders have been suggested as well, based on velocity and energy localization vector theories, optimizing the energy concentration towards the source direction rather than spreading energy over all loudspeakers [58, 59]. Such decoders are extremely useful when only lower order coefficients are available.

Optimally, loudspeakers should be regularly placed around the listener, and their lay-out and numbers are strongly dependent on the order of the spherical harmonic coefficients one wishes to reproduce. The use of higher order coefficients increases angular discrimination ability and reduces front-back confusion [60, 61]. The sweet spot decreases with increasing frequency by spatial sampling artifacts due to the discretization of the loudspeaker array [28]. However, the use of too many loudspeakers can lead to coloration near the sweet spot due to comb-filtering effects caused by the simultaneous loudspeaker activity [62].

The wave field synthesis technique (WFS) is another technique to enlarge the sweet spot to a wider area. WFS is based on the Kirchoff-Helmholtz integral theorem, which states that the pressure inside a closed volume can be reconstructed if the sound pressure and the normal component of the particle velocity on the boundary of that volume are known [63]. A discrete implementation of this theory typically requires a vast amount of loudspeakers around the listener as the distance between loudspeakers should be smaller than half a wavelength to avoid spatial aliasing. Below the aliasing frequency, the whole listening area is free of artifacts. At higher frequencies, spatial artifacts can be observed over the entire listening area and no artifact-free sweet spot can be defined [64]. Different studies show that localization is not impaired for loudspeaker spacings lower than 22cm and the localization is independent of the listener position [28, 65]. Coloration, however, is already noticed for lower loudspeaker distances.

In view of the development of a flexible auralization tool to aid urban planners with their design and enable demonstrations for the larger public, reproduction over headphones is selected. In contrast to other techniques, headphone-based reproduction puts only limited constraints on the hardware (low-cost), while being very flexible and portable. In this work a static reproduction will be used. To allow head movement a head tracker could be included. Attempts have been made [54], but at this moment, real-time update of the given auralization methodology was not possible with the given calculation resources.

2.3 Design and characterization of a microphone array for binaural recording emulation

In this section, a system for the recording of outdoor sound environments is designed to enable the assessment of the effect of noise mitigation measures by urban city planners and general public. Based on the previous review, two different recording techniques are selected. In a first method, a binaural signal of the sound environment is directly generated from HATS recordings. No further processing of the signal is needed, however, only limited information on the direction of the sources, necessary for the auralization of the effect of noise mitigation measures later on, is provided (see section 3.1). A second technique, which will further be described here, is based on a state-of-the-art dedicated microphone array to provide much more detailed information on the spatiality of the sound environment. This technique will prove to be specifically useful for a dedicated auralization methodology, based on the generation of a binaural signal with the incorporation of spatial selectivity (see section 3.2).

The microphone array proposed here is optimized for the simulation of a virtual microphone with directional characteristics that approximate the directivity of the human hearing, encompassing different auditory cues. Therefore, a set of weights is calculated for each microphone of the array. The set of optimal weights is frequency dependent, but will be restricted towards the high frequency range as directional characteristics, described by the HRTFs, become highly sensitive to spectral fluctuations due to interferences and individual differences of the pinnae. For frequencies higher than a given cross-over frequency f_X , only the signal from the microphone closest to the virtual ear location will be selected, similar as in the MTB technique. However, the MTB technique is not used for the full frequency range, as in that case the ability to deliver detailed information on the soundfield would be lost.

2.3.1 Characterization of the target HRTF directivity pattern

For pure binaural signal generation, the target directivity pattern of the microphone array is completely determined by the right and left ear head-related transfer function (HRTF) for each source direction. The angle dependent HRTF is defined as the transfer function describing the path from source to eardrum, incorporating effects of the head, torso, pinnae and outer ear canal. Interaural cues are inserted by comparing left and right ear HRTF for a given source location.

In general, the HRTFs are listener dependent, given individual physiological differences in head size and pinnae shape. As a consequence, the target HRTF should be adapted for every listener for optimal reconstruction. This increases complexity

Table 2.1: Different loudspeaker orientations for HRTF sampling. The elevation is defined as the angle between source and equatorial plane [66].

Elevation	Number of measurements	Azimuth increment
-40°	56	6.43°
-30°	60	6.00°
-20°	72	5.00°
-10°	72	5.00°
0°	72	5.00°
10°	72	5.00°
20°	72	5.00°
30°	60	6.00°
40°	56	6.43°
50°	45	8.00°
60°	36	10.00°
70°	24	15.00°
80°	12	30.00°
90°	1	—

of the system, as HRTFs should be measured for each listener beforehand, which is a lengthy procedure. Therefore, it is decided to choose a set of non-individualized HRTFs, as taken from a dummy head (HATS). Wenzel et al. showed that in this case listeners performed equally well regarding azimuthal angle discrimination, while an increase of front-back/up-down confusions and elevational errors has been reported [14]. It is suggested that these errors originate from the absence of individualized pinnae cues. However, it has been shown previously that such errors can be minimized by enabling dynamic cues.

2.3.1.1 HRTF dataset and ear canal compensation

The set of target HRTFs is taken from measurements by Gardner at MIT [66]. Gardner measured the response of a HATS (KEMAR model DB-4004) with small pinnae (model DB-061) and occluded ear simulators (model DB-100) with ear canal extensions (model DB-050) for 710 different orientations (Table 2.1). The source was positioned at 1.4m distance. Impulse responses were truncated to 512 samples at $f_s = 44.1\text{kHz}$, which is found to be more than sufficient for auditory scene recreation [67].

The spectral distortion introduced by the speaker and measurement equipment is compensated for by an inverse filter, calculated from the impulse response measured at the center of the HATS, without HATS present, which was included in the Gardner dataset.

Furthermore, a compensation filter for the ear canal and headphones (Sennheiser HD-280 pro) is designed to account for the spectral distortion introduced by the headphones and the fact that the listener uses his own ear canals. As the response of the Sennheiser HD-280 pro headphones was not included in the dataset, the inverse filter is extracted from measurements of the impulse response between the headphones and inner ear simulator (B&K Type 4158C) of the HATS (B&K Type 4128C) available at Ghent University.

For the calculation of the inverse filters, a time-domain least-squares solution with regularization was used to minimize performance error and effort penalty of the inverted response [68, 69]. The exact solution for the inverse filter is given by

$$\mathbf{h}_{\text{inv}} = [\mathbf{C}^T \mathbf{C} + \mu \mathbf{I}]^{-1} \mathbf{C}^T \delta_{\mathbf{m}}, \quad (2.4)$$

with \mathbf{C} the convolution matrix in Toeplitz form, constructed from the measured impulse response, $\delta_{\mathbf{m}}$ the discrete Dirac function, delayed with m samples, and μ the regularization parameter.

The regularization parameter μ allows to control the energy of the inverse filter, to avoid dips and peaks being overcompensated for at ill-conditioned frequencies. Additionally, the length of the inverse filter can be controlled, so that circular convolution effects can be minimized.

Fig. 2.1(a) shows the magnitude of the compensation filter for the influence of the measurement set-up. In Fig. 2.1(b) the magnitude and efficiency of the headphones and ear canal compensation filter is given. The effect of ear canal resonances is clearly illustrated by the peak at 2-3kHz and dip at 7kHz. To limit overcompensation at frequencies higher than 15kHz, while assuring proper canal resonance compensation, the original impulse response has been filtered with a low-pass filter with cut-off frequency at 16kHz prior to inversion.

After the design, the filters can easily be applied to Gardner's HRTF dataset by spectral multiplication. Here, the HRTFs were zero-padded to 1024 samples to avoid circular convolution effects.

2.3.1.2 HRTF spherical harmonic representation

The set of equalized HRTFs is only available for a number of discrete directions. Therefore, an interpolation strategy needs to be developed to estimate the HRTFs at intermediate orientations. Furthermore, no HRTF data is known for elevations lower than -40° as at such low elevations, the loudspeaker would practically coincide with the torso of the HATS. This implies that the HRTF data needs to be extrapolated to avoid pattern blow-up when calculating optimal microphone array weights later on.

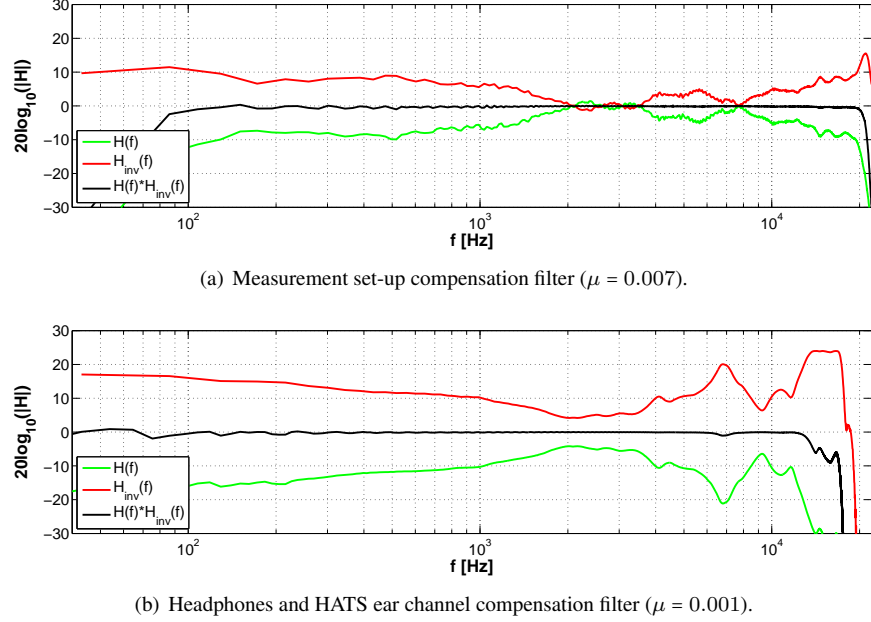


Figure 2.1: Magnitude of the original (blue) and inverse compensation filters (red). In black is the magnitude response of the combination of original and inverse filter responses.

Different strategies have been developed to perform HRTF interpolation [70, 71]. Local interpolation methods such as bilinear, bicubic and morphing techniques make use of neighboring HRTFs [72, 73]. These techniques, however, cannot account for data extrapolation towards the polar gap. Global interpolation methods rely on the decomposition of the complete set of HRTFs. Techniques such as the principal component analysis (PCA) [74], the Karhunen-Loeve expansion (KLE) [75] and the surface spherical harmonic decomposition (SSH) [76] allow the representation of the HRTFs into a set of basis functions.

In this work, a spherical harmonic decomposition of the equalized KEMAR HRTF data set is chosen, allowing a representation in function of a set of continuous predefined orthonormal basis functions, the spherical harmonics. Evans et al. [76] showed that a spherical harmonic decomposition of HRTFs in frequency domain, by separately analyzing magnitude and unwrapped phase, achieves better accuracy than decomposing the head-related impulse responses (HRIR) in time domain. Furthermore, he illustrated that higher degree spherical harmonics, corresponding with a higher amount of spatial variation, contribute progressively less to the reconstruction.

Generally, the N^{th} -order spherical harmonic synthesis of a square integrable function, either HRTF magnitude or unwrapped phase, can be written as

$$\widehat{\text{HRTF}}_{\parallel, \angle}(f, \phi, \theta) = \sum_{n=0}^N \sum_{m=-n}^n A_{\parallel, \angle}^{nm}(f) Y_n^m(\phi, \theta), \quad (2.5)$$

with $A_{\parallel, \angle}^{nm}(f)$ the frequency dependent spherical harmonic coefficients of order n and degree m for either magnitude (subscript \parallel) or unwrapped phase (subscript \angle). The spherical harmonics, $Y_n^m(\phi, \theta)$, are the angular dependent solution of the Helmholtz equation:

$$Y_n^m(\phi, \theta) = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m(\cos \theta) e^{jm\phi}, \quad (2.6)$$

with $P_n^m(\cdot)$ the associated Legendre functions and $|m| \leq n$. The spherical harmonics form a complete orthonormal basis over the sphere \mathbb{S}_2 and thus satisfy

$$\int_{\mathbb{S}_2} Y_n^m(\phi, \theta) Y_{n'}^{m'*}(\phi, \theta) ds = \delta_{n-n'} \delta_{m-m'}. \quad (2.7)$$

The decomposition in spherical harmonics is similar to the Ambisonic representation (Eq. 2.1) apart from the radial distance term, which is omitted as the far-field HRTFs are distance independent and measured at fixed range $r = 1.4\text{m}$.

In an analogue way, the frequency dependent spherical harmonic coefficients $A_{\parallel, \angle}^{nm}(f)$ can be extracted by direct integration over the unit sphere \mathbb{S}_2 :

$$A_{\parallel, \angle}^{nm}(f) = \int_0^{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \text{HRTF}_{\parallel, \angle}(f, \phi, \theta) Y_n^{m*}(\phi, \theta) \cos(\theta) d\theta d\phi. \quad (2.8)$$

However, as data is only available at L discrete orientations (ϕ_i, θ_i) , the integral in Eq. 2.8 is replaced with a discrete summation over all directions, limiting the number of resolvable coefficients (and thus the maximum order N_{max}). Furthermore, the discrete summation causes a breakdown of the orthonormality relation of the spherical harmonics (Eq. 2.7), which implies that energy from orders other than (n, m) will leak to $A_{\parallel, \angle}^{nm}(f)$.

To minimize the orthonormality error, different sampling schemes have been suggested that minimize the necessary amount of measurement directions L , while maximizing the resolvable order N_{max} [38, 45, 77]. In addition, a set of quadrature weights v_i , optimized for the L available directions (ϕ_i, θ_i) of the sampling grid, can be calculated so that

$$\int_{\mathbb{S}_2} \text{HRTF}_{\parallel, \angle}(f, \phi, \theta) Y_n^{m*}(\phi, \theta) ds \approx \sum_{i=1}^L v_i \text{HRTF}_{\parallel, \angle}(f, \phi_i, \theta_i) Y_n^{m*}(\phi_i, \theta_i). \quad (2.9)$$

These quadrature weights are readily available for well-known sampling grids such as uniform Fliege grids, Gauss-Legendre grids etc. [38]. For non-standard grids, quadrature weights can be estimated by the area of the Voronoi surfaces on \mathbb{S}_2 , corresponding with each measurement point [71, 78].

The exact value of N_{max} is function of the chosen sampling grid and quadrature weights, but is generally bounded by $L = (N_{max} + 1)^2$.

The direct integration technique works well for regular sampling schemes. However, for the irregular 710-point sampling scheme with polar gap another technique will be necessary. Zotkin et al. [79] proposed to calculate the spherical harmonic coefficients directly as a regularized least-square solution of a discretized version of Eq. 2.5. An extended version of this solution is given in Eq. 2.10, in matrix form [71]:

$$\mathbf{A}_{\parallel, \angle}(f) = [\mathbf{Y}^H \mathbf{V} \mathbf{Y} + \mu \mathbf{D}]^{-1} \mathbf{Y}^H \mathbf{V} \mathbf{HRTF}_{\parallel, \angle}(f), \quad (2.10)$$

with $\mathbf{A}_{\parallel, \angle}(f)$ the $(N+1)^2 \times 1$ spherical harmonic coefficient matrix for magnitude resp. unwrapped phase, $\mathbf{HRTF}_{\parallel, \angle}(f)$ the $L \times 1$ magnitude resp. unwrapped phase of the HRTF dataset and \mathbf{Y} the $L \times (N+1)^2$ matrix with spherical harmonic basis functions:

$$\mathbf{A}_{\parallel, \angle}(f) = \begin{pmatrix} A_{\parallel, \angle}^{00}(f) \\ A_{\parallel, \angle}^{1-1}(f) \\ A_{\parallel, \angle}^{10}(f) \\ A_{\parallel, \angle}^{11}(f) \\ \vdots \\ A_{\parallel, \angle}^{NN}(f) \end{pmatrix}, \quad \mathbf{HRTF}_{\parallel, \angle}(f) = \begin{pmatrix} \text{HRTF}_{\parallel, \angle}(f, \phi_1, \theta_1) \\ \text{HRTF}_{\parallel, \angle}(f, \phi_2, \theta_2) \\ \vdots \\ \text{HRTF}_{\parallel, \angle}(f, \phi_i, \theta_i) \\ \vdots \\ \text{HRTF}_{\parallel, \angle}(f, \phi_L, \theta_L) \end{pmatrix}$$

and

$$\mathbf{Y} = \begin{pmatrix} Y_0^0(\phi_1, \theta_1) & Y_1^{-1}(\phi_1, \theta_1) & Y_1^0(\phi_1, \theta_1) & \cdots & Y_N^N(\phi_1, \theta_1) \\ Y_0^0(\phi_2, \theta_2) & Y_1^{-1}(\phi_2, \theta_2) & Y_1^0(\phi_2, \theta_2) & \cdots & Y_N^N(\phi_2, \theta_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_0^0(\phi_L, \theta_L) & Y_1^{-1}(\phi_L, \theta_L) & Y_1^0(\phi_L, \theta_L) & \cdots & Y_N^N(\phi_L, \theta_L) \end{pmatrix}.$$

\mathbf{V} is an $L \times L$ diagonal matrix, containing the weighting coefficients derived from the Voronoi surfaces. Inclusion of \mathbf{V} is optional, but its inclusion decreases the orthonormality error. To further account for the open irregular grid, Tikhonov regularization is applied. Here, μ is the regularization parameter and \mathbf{D} is a $(N+1)^2 \times (N+1)^2$ diagonal matrix, with diagonal elements $1 + n(n+1)$ to decrease the influence of higher order harmonics.

The use of the least-squares methodology can cope with the polar gap when regularization is applied. However, the presence of higher order coefficients, resulting from the regularized fit on the known data, gives rise to meaningless (large) pattern fluctuations in the polar gap. In order to have more control on the data and to guarantee that the behavior of the directivity pattern is smooth in the polar gap, the spherical harmonic coefficients are calculated from a combination of the known data at available directions and the extrapolated data at the polar gap, resulting from a lower order ($N = 3$) approximation. Additionally, the magnitude value at the south pole of the polar gap was forced to take the value of the overall minimum of HRTF magnitudes.

The performance of the spherical harmonic representation of the HRTFs can be evaluated in terms of the relative mean square error (rMSE), indicating the quality of reconstruction. The rMSE can be defined for the complex-valued HRTF (Eq. 2.11) and individually reconstructed magnitude and unwrapped phase (Eq. 2.12 and Eq. 2.13):

$$\text{rMSE}(f) = 10 \log_{10} \frac{\sum_{i=1}^L v_i |\widehat{\text{HRTF}}(f, \phi_i, \theta_i) - \text{HRTF}(f, \phi_i, \theta_i)|^2}{\sum_{i=1}^L v_i |\text{HRTF}(f, \phi_i, \theta_i)|^2}, \quad (2.11)$$

$$\text{rMSE}_{||}(f) = 10 \log_{10} \frac{\sum_{i=1}^L v_i ||\widehat{\text{HRTF}}(f, \phi_i, \theta_i)| - \text{HRTF}_{||}(f, \phi_i, \theta_i)|^2}{\sum_{i=1}^L v_i \text{HRTF}_{||}(f, \phi_i, \theta_i)^2}, \quad (2.12)$$

$$\text{rMSE}_{\angle}(f) = 10 \log_{10} \frac{\sum_{i=1}^L v_i |\angle \widehat{\text{HRTF}}(f, \phi_i, \theta_i) - \text{HRTF}_{\angle}(f, \phi_i, \theta_i)|^2}{\sum_{i=1}^L v_i |\text{mod}(\text{HRTF}_{\angle}(f, \phi_i, \theta_i) + \pi) - \pi|^2}. \quad (2.13)$$

Fig. 2.2 shows the rMSE in function of frequency for a spherical harmonic decomposition, with maximum order of $N = 18$. At frequencies lower than 2kHz, the relative error of both magnitude and unwrapped phase reconstruction remains below -30dB . At higher frequencies, the error regarding unwrapped phase reconstruction increases much faster compared to the magnitude error, due to discontinuities in the original unwrapped phase pattern arising from pinnae reflections, and negatively influences overall rMSE performance. This is accounted for when fitting the microphone directivity pattern later on, as at frequencies higher than a given cross-over frequency f_X only the microphone signal closest to the locations of the ears will be used, approximating a first-order spherical-head model, and thus omitting individual pinnae effects.

These results clearly illustrate the advantage of separately decomposing magnitude, related to the ILD, and unwrapped phase, related to the ITD, as it allows an increased focus on the modeling of each individual interaural cue, and thus enables to account for their specific range of dominance. Good localization is to be expected at frequencies below 2kHz, given the excellent reconstruction accuracy of both magnitude and phase. However, as the ILD cue is dominant at higher

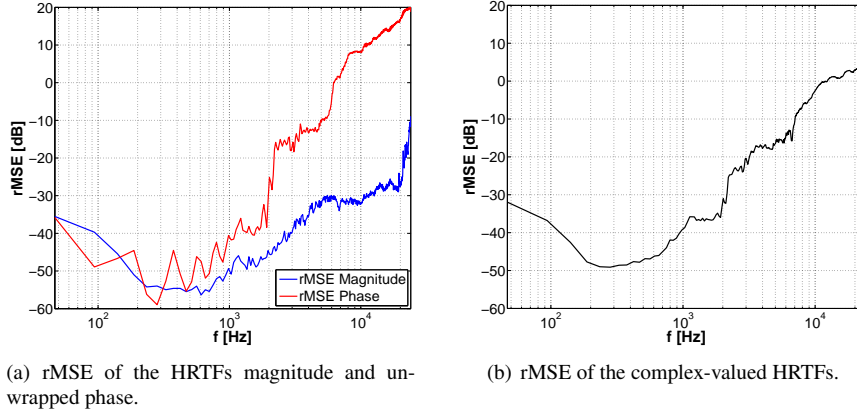


Figure 2.2: *rMSE of the recomposed HRTFs with $N = 18$.*

frequencies, the separate magnitude and phase decomposition allows to further extend the frequency range of localization ability by only using reconstructed magnitude, while substituting the deteriorated reconstruction of the high frequency phase with a linearized version [67].

In Fig. 2.3 the overall rMSE for the left ear directivity pattern is given for different directions. The values in this figure result from an interpolation between places where the HRTF is available and the rMSE can actually be calculated (encircled values). Here, an adapted version of Eq. 2.11 is used, where the sum over all directions is replaced by a sum over all frequencies until $f_{max} = 10\text{kHz}$. Good performance is seen for ipsilateral directions (in the neighborhood of $\phi = 90^\circ$). The rMSE progressively degrades towards contralateral directions ($\phi = 270^\circ$), where at some points the rMSE becomes larger than 0dB.

Fig. 2.4 displays a comparison between original and reconstructed HRIR, with frequencies until $f_{max} = 10\text{kHz}$ for the ipsilateral and contralateral side. Fair agreement is seen between original and reconstructed HRIRs, even for the contralateral side with worse overall rMSE values. It is seen that time-delay differences between both HRIRs are close to zero, while magnitude differences reach relatively low values.

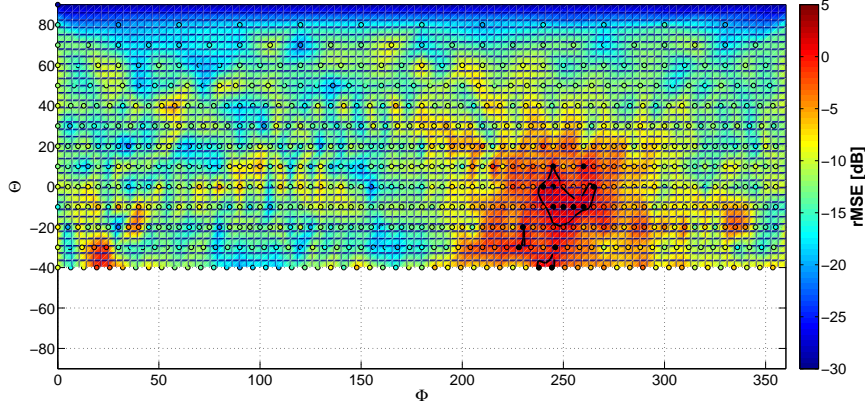
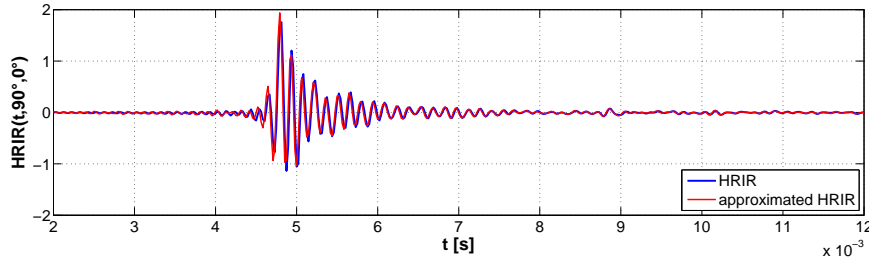
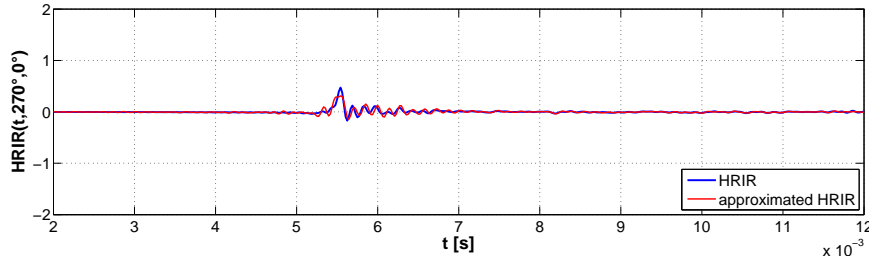


Figure 2.3: Interpolated spatial distribution of the overall rMSE, calculated over frequencies until $f_{max} = 10\text{kHz}$. Encircled values represent the actual rMSE values, calculated at the available measurement directions. The black contour and circles indicate an rMSE > 0dB.



(a) ($\phi = 90^\circ, \theta = 0^\circ$), rMSE = -5.6dB



(b) ($\phi = 270^\circ, \theta = 0^\circ$), rMSE = -3.8dB

Figure 2.4: Original and approximated HRIR, recomposed with frequency components until $f_{max} = 10\text{kHz}$.

2.3.2 Microphone array design and characterization

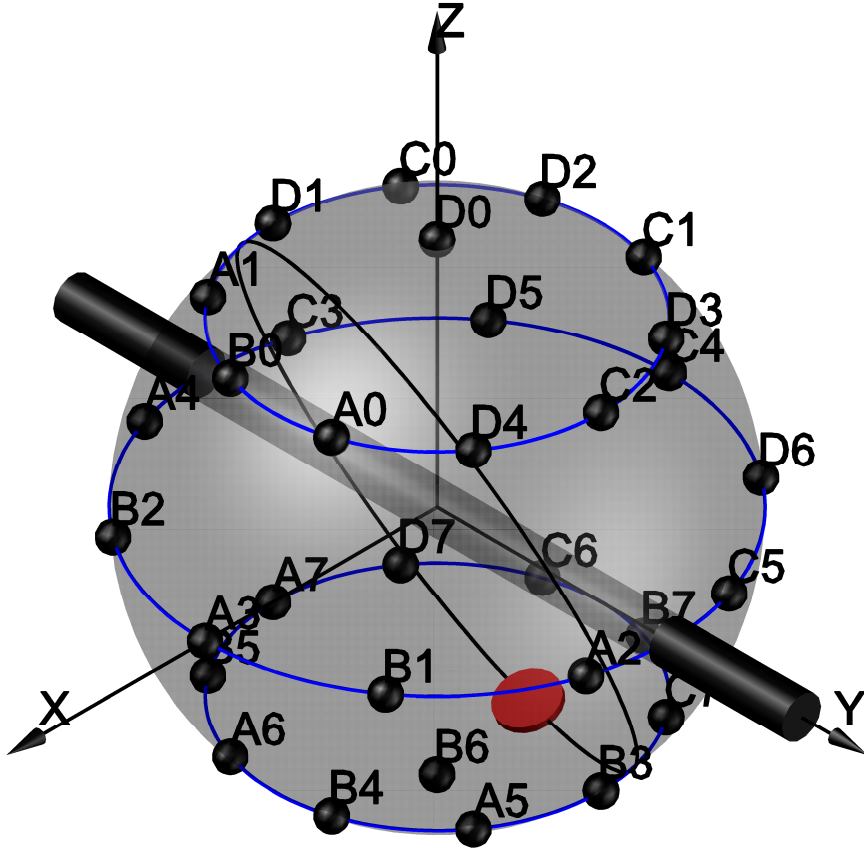


Figure 2.5: Spherical microphone array geometry. The positions of the 32 microphones are indicated. The bar along the y-axis is part of the fixation mechanism. The red circle indicates the outlet of microphone cables.

The microphone array is specifically designed to enable HRTF pattern fitting, while allowing the MTB technique above a certain cross-over frequency f_X as a back-up when pattern fitting fails. The design consists of a hard plastic sphere with 7cm radius and 32 omnidirectional microphones distributed along the surface. A hard sphere is chosen as a substitute for the human head, which allows to recreate the interaural differences. The chosen radius is somewhat smaller than the average radius of 8.7cm that Algazi et al. found (from measured radii ranging from 7.9cm to 9.5cm), based on the spherical-head model [80]. In this way, the frequency at which spatial aliasing occurs can be increased, as the inter-microphone distance

decreases with the radius, while the MTB technique is still possible at higher frequencies.

The 32 microphones are distributed on the sphere in such a way that focus is put on azimuthal rather than elevational pattern reconstruction. Humans are more sensitive for azimuthal detection, while elevational accuracy is lower and harder to obtain given the necessity to reconstruct the high frequency effects of the individual pinnae. Therefore, a more sparse microphone distribution is chosen along the elevational direction: apart from the two poles, the remaining 30 microphones are uniformly distributed along three circles at elevations of -45° , 0° and 45° . In Fig. 2.5 the microphone array is depicted, with the different microphone positions indicated.

The microphones used in the array are manufactured by Knowles Acoustics (type FG-23329), and are typically used for hearing aids. These are low cost microphones (approx. 30€), carefully selected after long-term outdoor comparison with several other low cost microphone capsules [81]. The noise floor and spectral characteristics of each microphone have been evaluated. The microphones have been calibrated prior mounting into the array, although unnecessary as magnitude and phase differences are accounted for when using measured directivity patterns in the pattern fitting technique. For the data-acquisition (DAQ), a National Instruments PXIe-1082 chassis with three NI-4498 DAQ cards, counting 16 channels each, is used. The DAQ hardware allows simultaneous sampling of all microphones, so no extra phase mismatch will be introduced.

2.3.2.1 Microphone array directivity pattern

If microphone capsules and mounting would be perfect, the theoretical directivity pattern of a microphone-on-a-sphere could be directly used in the HRTF pattern fitting procedure. However, due to deviations in microphone placement, level and phase differences between microphones, and influences of windshield, cable outlet and fixation bars, the real directivity pattern of each microphone will deviate from the theoretical one. Therefore, a characterization of the real directivity pattern of each microphone on the sphere should be made for usage in the fitting procedure.

To enable fast and precise measurements, a dedicated measurement framework has been developed (Fig. 2.6). Two rotation stages are used to rotate the microphone array around the Z- and Y-axis (Newport - model URM80CC), controlled by a Newport Motion Controller MM2000. Due to reciprocity, rotating the microphone array around its axes is acoustically equivalent to moving the sound source around the array. However, care must be taken, as due to the rotation of the array, the position of the array relative to the position of cables (cable outlet) and fixation bars may cause a deviation from the true directivity pattern. Due to practical reasons however, it was impossible to move the speaker around the array, so the set-up

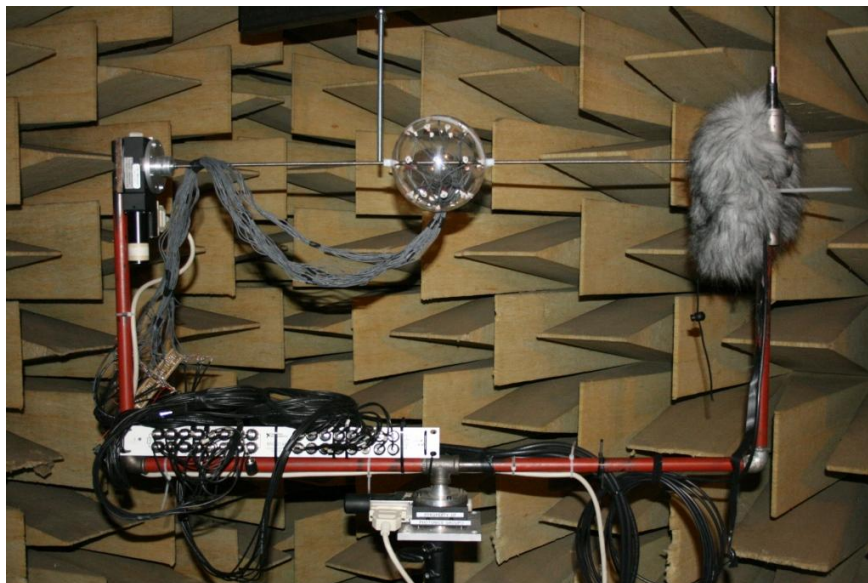


Figure 2.6: Mounting of the microphone array on the rotating measurement framework in the anechoic chamber. One Newport rotation stage is mounted at the bottom (rotation around the Z-axis) while the second one is mounted on the left side of the fixation axis (rotation around the Y-axis). Microphones are connected to a breakout box, mounted on the lower left, which connects signals to the DAQ hardware outside the anechoic chamber. The array windshield has been removed for the sake of the picture, but is still visible on the right end of the array fixation bar.

with rotating microphone array was preferred. An advantage of this set-up is the possibility to install an omnidirectional reference microphone (B&K Type 4189 with SV12L preamp) on a position symmetrical to the microphone array and loudspeaker (Adam S1X), cf. Fig. 2.7.

With the aid of the measurement framework, the array was rotated around the Y-axis from -90° to 90° , in steps of 10° , while rotating around the Z-axis in steps of 10° from 0° (microphone A3 pointed towards the loudspeaker) to 360° , clockwise defined. The transfer function, relative to free field, has been estimated for each microphone on the sphere for all 703 directions (including duplicates) by emitting a 20s-long pink noise and applying the Welch averaged periodogram method with a 1024-point Hamming window and 50% overlap, with the signal recorded at the reference microphone as input and the array microphone signal as output. In this way level and phase fluctuations of the emitting loudspeaker have been compensated for.

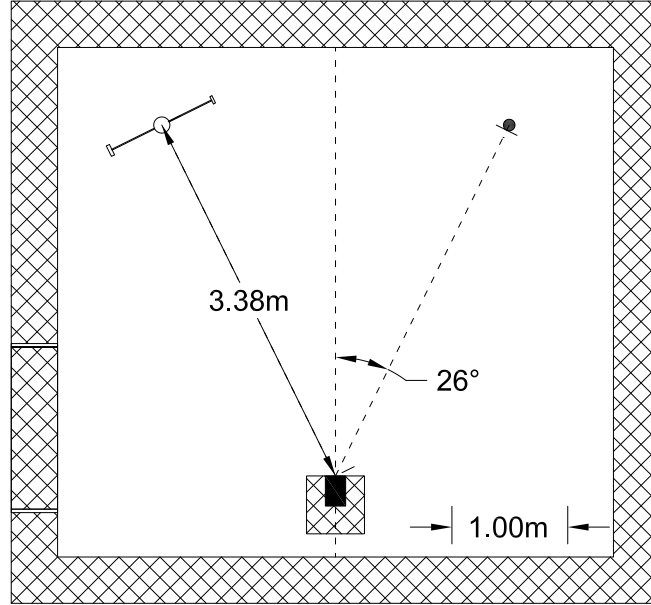


Figure 2.7: Measurement set-up in the anechoic chamber. In the upper left corner is the rotating measurement framework with microphone array, oriented towards the loudspeaker. The omnidirectional reference microphone is placed in the upper right corner. The loudspeaker is placed on the symmetry axis.

Due to the reciprocal measurement arrangement, the rotation angles around the axes indicate the orientation of the microphone array relative to the source, rather than the position of the source relative to the array (as should be the case when displaying the directivity pattern in function of azimuth and elevation). The directivity pattern of the microphone on the sphere needs to be known in the coordinate system fixed to the microphone array, which is the same as the HRTF directivity coordinate system (cf. Fig. 2.5).

Transformation Algorithm 1 gives the relation between both coordinate systems, with Φ resp. Θ the rotations around the Z- resp. Y-axis and ϕ' and θ' the azimuth and elevation in the spherical coordinate system with polar axis according to the Y-axis. This transformation can be understood by interpreting a rotation of the array around the Y-axis as an azimuthal movement of the source relative to the stationary array and a rotation around the Z-axis as an elevational source movement, with the Y-axis as polar axis. Further rotations have been applied to mutate the derived spherical coordinate system to the coordinate system attached to the array, with polar axis aligned with the Z-axis and X-axis through microphone A3.

```

if  $0^\circ \leq \Phi < 90^\circ$  then
     $\phi' = \Theta + 90^\circ$ 
     $\theta' = \Phi$ 
end if
if  $90^\circ \leq \Phi < 180^\circ$  then
     $\phi' = \Theta + 270^\circ$ 
     $\theta' = 180^\circ - \Phi$ 
end if
if  $180^\circ \leq \Phi < 270^\circ$  then
     $\phi' = \Theta + 270^\circ$ 
     $\theta' = 180^\circ - \Phi$ 
end if
if  $270^\circ \leq \Phi \leq 360^\circ$  then
     $\phi' = \Theta + 90^\circ$ 
     $\theta' = \Phi - 360^\circ$ 
end if

```

Algorithm 1: Transformation algorithm to express the orientation of the speaker relative to the stationary microphone array (in spherical coordinates (ϕ', θ') , with polar axis according to the Y-axis) in function of the rotation of the array around Z-and Y-axis, given by (Φ, Θ) .

After removing duplicate directions, the directivity pattern of each microphone on the sphere contains measurements at 614 directions. In order to enable a continuous representation and to smooth out pattern irregularities, a spherical harmonic decomposition has been conducted on the magnitude and spectrally unwrapped phase, in a similar way as on the target HRTF directivity pattern. A lower order decomposition allows for a significant degree of pattern smoothing, but increases the rMSE. Here, the same order $N = 18$ as for the HRTF decomposition has been used, with a Tikhonov regularization that favors lower orders [71, 82].

Fig. 2.8 shows the average rMSE for magnitude, unwrapped phase and overall performance of the 32 microphone directivity pattern reconstructions. It can be seen that the magnitude rMSE is bounded between -35dB and -20dB at frequencies higher than 100Hz, until the upper frequency range. The unwrapped phase rMSE remains below max. -12dB, with a minimum of -35dB, in the frequency range between 100Hz and 2kHz. From 2kHz on however, phase rMSE increases drastically due to approximation errors, which arise from spatial discontinuities in the high frequency phase directivity pattern that could not be resolved by spectral unwrapping. The overall rMSE of the complex-valued microphone patterns corresponds with the magnitude and unwrapped phase rMSE: between 100Hz and 2kHz, the rMSE remains below approx. -20dB.

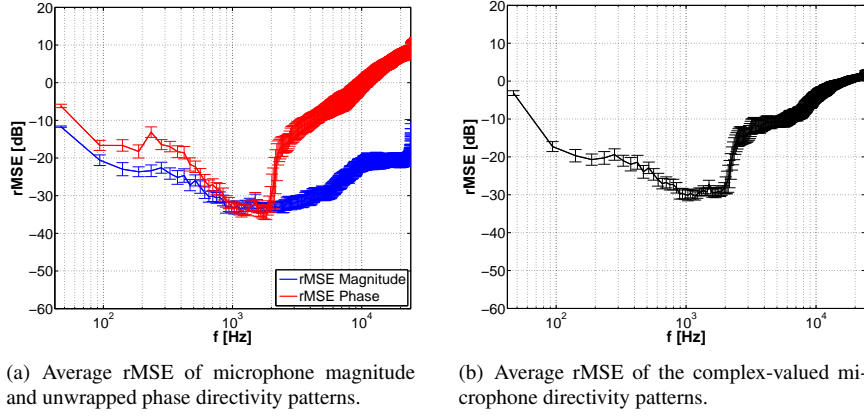


Figure 2.8: rMSE, averaged over all 32 microphone directivity patterns, recomposed with $N = 18$. The error bars indicate the standard deviation.

The rMSE values shown here are worse than the rMSE values from the HRTF approximation (Fig. 2.2). However, as it was the intention of the spherical harmonic approximation to smooth out irregularities of the measured microphone directivity patterns, a larger error between the smoothed approximation and original irregular directivity pattern is to be expected.

2.3.2.2 Microphone directivity pattern evaluation

In order to make a proper evaluation of the measured directivity pattern of a microphone on the array, the directivity pattern can be compared to the corresponding theoretical pattern. In [83], an expression for the total pressure near the surface of a sphere is given, resulting from scattering of a plane wave, incident from the negative Z-axis (equivalent with the microphone orientation):

$$p_{tot}(r, \theta', k) = p_i(r, \theta', k) + p_s(r, \theta', k), \quad (2.14)$$

with $p_i(r, \theta', k)$ and $p_s(r, \theta', k)$ the incident plane wave resp. scattered wave:

$$p_i(r, \theta', k) = \sum_{m \geq 0} (2m+1)(-j)^m P_m(\cos \theta') j_m(kr) \quad (2.15)$$

$$p_s(r, \theta', k) = \sum_{m \geq 0} D_m(ka) (2m+1)(-j)^m P_m(\cos \theta') h_m(kr), \quad (2.16)$$

where $D_m(ka) = -\frac{(m/ka)j_m(ka) - j_{m+1}(ka)}{(m/ka)h_m(ka) - h_{m+1}(ka)}$. $P_m(\cdot)$ is the Legendre polynomial, $j_m(\cdot)$ the spherical Bessel function of the first kind and $h_m(\cdot)$ the spherical Hankel function of the second kind. a is the radius of the sphere and $k = 2\pi/\lambda$ the wave

number. In this equation, the plane wave is incident from the negative Z-axis, and θ' is the elevation angle, defined between the positive Z-axis and the field points P . The equations are independent of azimuth angle ϕ due to rotational symmetry. In order to simulate the complex directivity pattern of a certain microphone on the sphere in standard spherical coordinates, with θ defined as the angle between the equatorial plane and the field points $P = (P_x, P_y, P_z)$ on the sphere, θ' is substituted with

$$\theta' = \pi - \arccos\left(\frac{P_x M_x + P_y M_y + P_z M_z}{a^2}\right), \quad (2.17)$$

with $M = (M_x, M_y, M_z)$ the location of the specified microphone on the sphere.

To evaluate the similarity between the theoretical directivity pattern and the measured and smoothed directivity pattern, the spatial correlation coefficient can be used, generally defined as

$$\rho(f) = \frac{\sum_{i=1}^L v_i X(\phi_i, \theta_i, f) Y(\phi_i, \theta_i, f)}{\sqrt{\sum_{i=1}^L v_i X(\phi_i, \theta_i, f)^2 \sum_{i=1}^L v_i Y(\phi_i, \theta_i, f)^2}}, \quad (2.18)$$

with v_i the area of the Voronoi surfaces and $X(\phi_i, \theta_i, f)$ and $Y(\phi_i, \theta_i, f)$ the directivity patterns to be compared. For evaluation of magnitude similarity, $\rho_{mag}(f)$ is calculated by substituting the measured and theoretical magnitude patterns in $X(\phi_i, \theta_i, f)$ resp. $Y(\phi_i, \theta_i, f)$, after normalizing with their magnitude rms-value. The correlation between measured and theoretical unwrapped phase, $\rho_{pha}(f)$, is calculated in a similar way. However, instead of using the rms-scaled values, $X(\phi_i, \theta_i, f)$ and $Y(\phi_i, \theta_i, f)$ is substituted with the unwrapped phase minus the average phase of the measured resp. theoretical pattern. In this way, initial phase delay differences, which do not contribute to the spatial variability of the unwrapped phase directivity pattern, are accounted for.

Fig. 2.9 displays $\rho_{mag}(f)$ for all 32 microphones. Between 100Hz and 1.25kHz the correlation between measured and theoretical magnitude directivity is higher than 0.97. The low correlation value at the lowest frequencies originates from the low SNR of the speaker at these frequencies. At frequencies higher than 1.25kHz, the magnitude correlation coefficient gradually decreases. However, $\rho_{mag}(f)$ remains higher than 0.9 until 10kHz, except for microphone A7, B3 and B6, which suffer the largest pattern deviations. As these microphones are oriented towards the lower bar of the fixation mechanism, deviations are probably caused by reflections on this bar. The decreasing similarity at higher frequencies is to be explained by the influence of the measurement set-up as well, as reflections on fixation bars, cables etc. influence the measured directivity pattern.

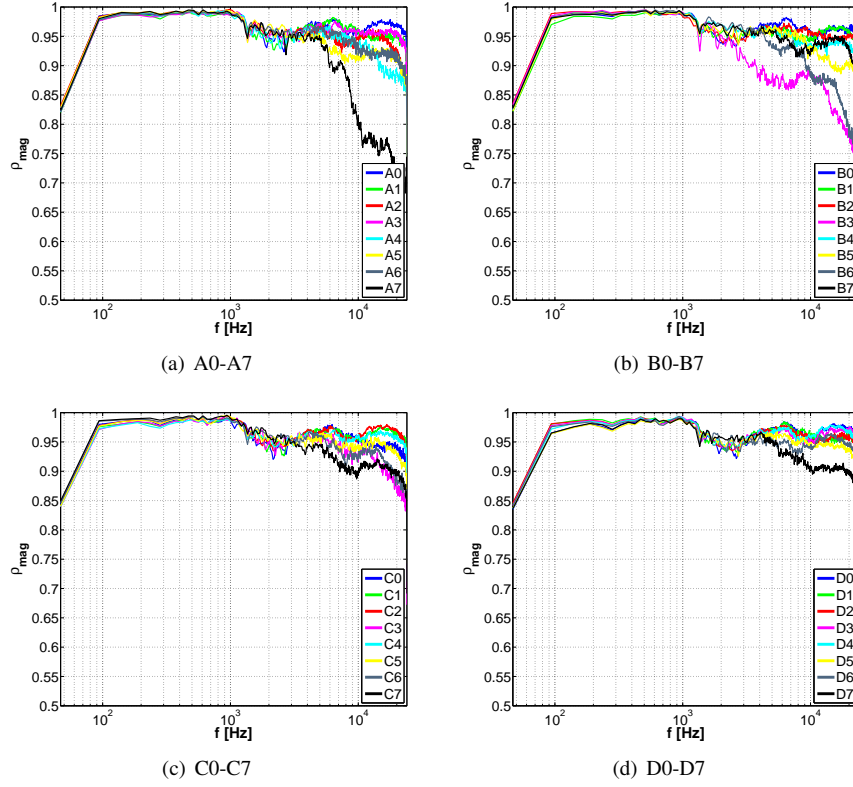


Figure 2.9: Spatial magnitude correlation coefficient $\rho_{mag}(f)$ between the 32 measured and theoretical microphone-on-sphere directivity patterns. The 32 microphones are clustered in four groups of 8 microphones, from A0 until D7.

The evaluation of phase similarity shows somewhat worse results, Fig. 2.10. At frequencies below 300Hz, $\rho_{pha}(f)$ takes values much lower than 0.9. At these frequencies, the difference between maximum and minimum absolute phase value is limited to approx. 0.4rad. Therefore, phase errors, resulting from measurement inaccuracies and microphone positioning, become relatively more significant, largely degrading $\rho_{pha}(f)$. At frequencies between 400Hz and 2kHz, the absolute phase value increases relative to the phase errors, and $\rho_{pha}(f)$ takes values between 0.9 and 0.99. From 2kHz on, similarity between the theoretical and measured phase decreases steadily due to influences of the measurement set-up (fixation bars, cables etc.).

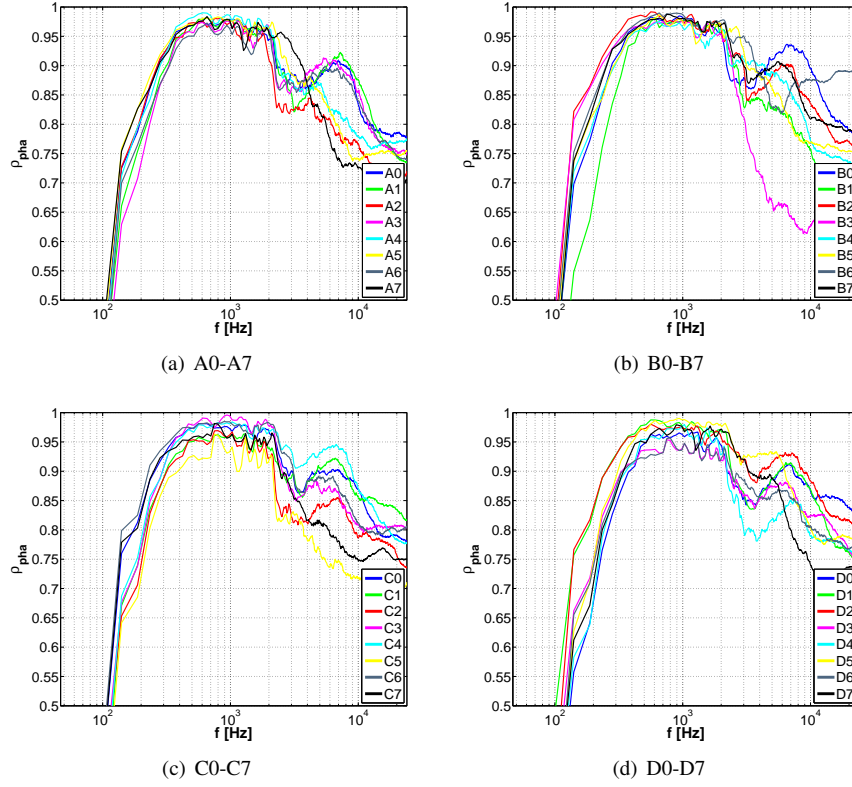


Figure 2.10: Spatial phase correlation coefficient $\rho_{pha}(f)$ between the 32 measured and theoretical microphone-on-sphere directivity patterns. The 32 microphones are clustered in four groups of 8 microphones, from A0 until D7.

2.3.3 HRTF directivity pattern fitting

In previous sections, the magnitude and unwrapped phase directivity patterns of all 32 microphones on the sphere and the target directivity pattern, extracted from the HRTFs, have been characterized in terms of spherical harmonic coefficients. This section describes how to find the appropriate filter coefficients $W_i(f)$ for each microphone signal so that the resulting array directivity pattern approaches the HRTF directivity pattern:

$$\text{HRTF}(f, \phi, \theta) = \sum_{i=1}^{32} W_i(f) \text{MIC}_i(f, \phi, \theta), \quad (2.19)$$

with $\text{HRTF}(f, \phi, \theta)$ and $\text{MIC}_i(f, \phi, \theta)$ the complex-valued directivity pattern of the HRTF, resp. i -th microphone on the sphere. Different techniques have been

developed to find the frequency dependent weights [48, 51]. In this work, we will estimate the microphone weights $W_i(f)$ from a regularized least-square solution of a discretized version of Eq. 2.19:

$$\mathbf{W}(f) = [\mathbf{MIC}^H(f)\mathbf{MIC}(f) + \mu(f)\mathbf{I}_{32}]^{-1} \mathbf{MIC}^H \mathbf{HRTF}(f), \quad (2.20)$$

with

$$\mathbf{HRTF}(f) = \begin{pmatrix} \text{HRTF}(f, \phi_1, \theta_1) \\ \text{HRTF}(f, \phi_2, \theta_2) \\ \vdots \\ \text{HRTF}(f, \phi_M, \theta_M) \end{pmatrix}, \quad \mathbf{W}(f) = \begin{pmatrix} W_1(f) \\ W_2(f) \\ \vdots \\ W_{32}(f) \end{pmatrix}$$

and

$$\mathbf{MIC}(f) = \begin{pmatrix} \text{MIC}_1(f, \phi_1, \theta_1) & \cdots & \text{MIC}_{32}(f, \phi_1, \theta_1) \\ \text{MIC}_1(f, \phi_2, \theta_2) & \cdots & \text{MIC}_{32}(f, \phi_2, \theta_2) \\ \vdots & \ddots & \vdots \\ \text{MIC}_1(f, \phi_M, \theta_M) & \cdots & \text{MIC}_{32}(f, \phi_M, \theta_M) \end{pmatrix}.$$

The matrices $\mathbf{HRTF}(f)$ and $\mathbf{MIC}(f)$ contain the complex-valued smoothed directivity patterns of the target HRTF resp. measured microphone patterns, found by evaluating the magnitude and phase spherical harmonic resynthesis (Eq. 2.5) at each direction (ϕ_m, θ_m) of an M -point evaluation grid. In the following, an evaluation grid will be chosen that coincides with the measurement grid of the microphone-on-sphere directivity patterns, as this grid puts more evaluation points around the location of the (fixed) ear positions. Further in this section, a uniform grid will be investigated as well, which allows more flexibility regarding ear positions. As it was seen that loudspeaker instabilities and phase noise led to fast varying spatial fluctuations and a decreased correlation between theoretical and measured microphone-on-sphere directivity patterns at low frequencies (Fig. 2.9-2.10), a spatial low-pass filter is applied to $\mathbf{MIC}(f)$. This filter is implemented by only taking the spherical harmonic coefficients until $N = 3$ into account for frequencies lower than 200Hz. At these frequencies, the size of the sphere is lower than $1/10^{th}$ of the wavelength, and spatial variations will be rather small.

The regularization term $\mu(f)\mathbf{I}_{32}$ in Eq. 2.20 is used to avoid over-fitting and consequently decreases the sensitivity of the fit to sensor noise and errors in e.g. microphone characteristics and other uncontrollable influences. More complex expressions for the regularization term exist, e.g. with direct inclusion of a desired white-noise gain (WNG) in the regularization term [51, 84].

Fig. 2.11 shows the performance of the HRTF pattern fitting procedure in function of frequency for different μ values. Overall rMSE ranges from -28dB at low frequencies to -5dB at 2kHz and 0dB at higher frequencies. The magnitude rMSE

gives slightly better results, with an absolute minimum value of -40dB and an increased performance at high frequencies. The phase rMSE, however, shows worse results, explaining the decrease in the overall rMSE. It should be noted that the phase rMSE shown in Fig. 2.11(d) is an adapted version of Eq. 2.13, where the modulo of the error in the numerator is taken in the same way as in the denominator.

Similar conclusions can be drawn from the spatial correlation between fitted and original magnitude and phase directivity patterns. The magnitude directivity has a spatial correlation of more than 0.95 below 2kHz, and slightly decreases at higher frequencies. The spatial phase correlation shows only good results at frequencies between 150Hz and 1.5kHz. At other frequencies, spatial correlation decreases drastically.

The figures illustrate that the best fitting results are obtained for low μ values. This is to be expected, as in these cases, the fitting procedure puts more effort in minimizing the fitting error. However, it can be seen that the effect of changing μ has influence only at frequencies lower than 2.5kHz. For $\mu < 0.01$, little difference is seen on the performance of the fitting. When μ increases above 0.1, the accuracy of the pattern fit decreases steadily.

In order to evaluate the effect of regularization on the robustness of the resulting filter coefficients, indicating the dependence of fitting performance on small errors in microphone characteristics (see e.g. [85]), and the sensitivity to microphone noise, the white-noise gain (WNG) is defined, relating the mean array gain, with inclusion of the microphone directivity patterns, to the array gain for spatially uncorrelated noise as input for each channel [84]:

$$\text{WNG}(f) = 10 \log_{10} \frac{\mathbf{W}^H(f) \mathbf{MIC}^H(f) \text{diag}(v_1 \dots v_M) \mathbf{MIC}(f) \mathbf{W}(f)}{\mathbf{W}^H(f) \mathbf{W}(f)}, \quad (2.21)$$

with $\text{diag}(v_1 \dots v_M)$ a diagonal matrix, containing the Voronoi weights of each evaluation point. Positive WNG indicates an amplification of the useful array output signal over the uncorrelated input noise.

Fig. 2.11(b) shows the WNG as a function of frequency for different values of μ . Almost no influence is seen at frequencies above 2.5kHz, where the WNG takes values around -10dB at frequencies below 10kHz. At frequencies below 2.5kHz, however, large influence of the regularization parameter is seen. When μ increases, the WNG increases proportionally from approx. -25dB for $\mu = 0.001$ to approx. -5dB for $\mu = 10$ as the focus shifts from minimizing the performance error to limiting the effort term. Positive WNG values only occur for high μ ($\mu = 100$), causing an unavoidably low fitting performance. At such high μ , only signals from microphones close to the ears are included, while the signal from other microphones will be suppressed to minimize the weight of the filter coefficients.

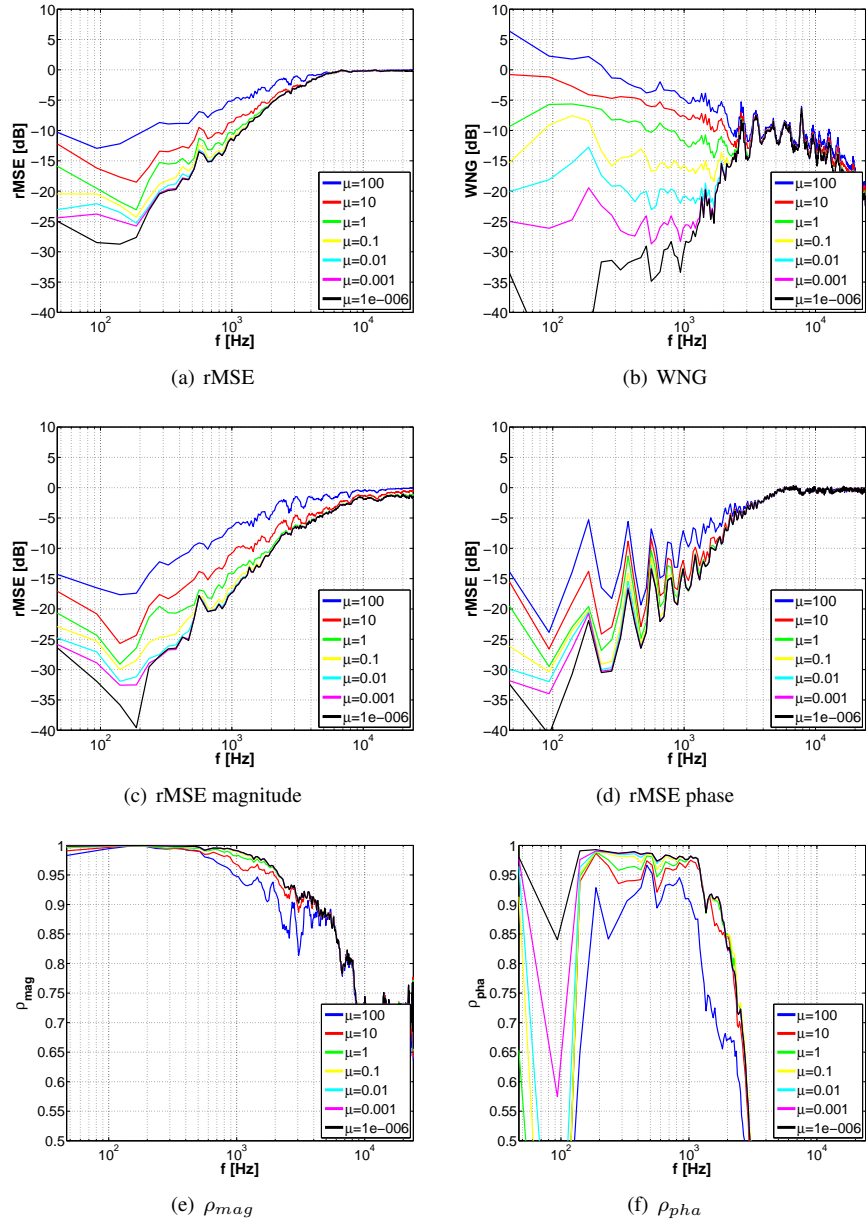


Figure 2.11: Performance of the HRTF pattern fitting for different frequency independent regularization parameters.

The foregoing demonstrates that the fitting error decreases monotonously with μ , while the WNG has a monotonous increase. This means that a trade-off should be made when choosing an appropriate μ . Ideally, μ should be chosen to maximize the WNG in order to decrease the effect of microphone noise [84]. Given the noise level of the selected microphones (estimated at approx. 37dB [81]), a value of $\mu = 1$ would be advisable to yield perceptually good results for recordings in quiet environments [84]. However, as this microphone array is to be used for the auralization of the effect of noise mitigation measures on (high) traffic noise levels (a background noise of approx. 60dB is seen in section 3.2), the WNG constraint can be loosened and more effort can be put into improving the accuracy of the pattern fit (rMSE), allowing for lower μ -values.

In order to further improve the performance, phase related errors at high frequencies, which are of little importance for the sound localization, can be minimized. Therefore, the phase of both microphone directivity patterns and target pattern are linearized for frequencies from 2kHz on, putting the focus on fitting the magnitude rather than phase.

Furthermore, the number and location of the M evaluation points for the least-square fit can be optimized. Since the spherical harmonic coefficients for magnitude and unwrapped phase of the directivity pattern of the HRTF dataset and microphone-on-sphere are known, $\text{HRTF}(f, \phi, \theta)$ and $\text{MIC}_i(f, \phi, \theta)$ can be discretized at any desired point. In the previous discussion, a 614-point grid has been used, optimized for reconstruction around the left and right ear (90° and 270°). To allow more flexibility and overall accuracy, a uniformly sampled evaluation grid could be used.

Fig. 2.12 shows the overall rMSE and WNG for the fitting with $\mu = 0.01$ based on the data with linearized phase, with evaluation grid optimized for the left and right ear, and the data with linearized phase, with uniform evaluation grid with a spacing of 10° , resulting in 614 unique evaluation points. Despite the phase linearization, little difference is seen between the fit performed on the original data and the linearized data.

Much more difference is seen when comparing the spatial distribution of the rMSE (Fig. 2.13). The effect of the L/R optimized grid is clearly seen on the rMSE when fitted to the target pattern with the original phase (Fig. 2.13(a)). Minimum rMSE values are encountered at 90° and 270° , the locations of the ears. Locations of maximal error are found at bands of high and low elevations. This can be explained both by the lack of microphones and evaluation points at these elevations. Similar conclusions can be drawn regarding the rMSE for pattern fitting with the linearized data (Fig. 2.13(b)). Little difference is expected, as the overall rMSE was hardly any different.

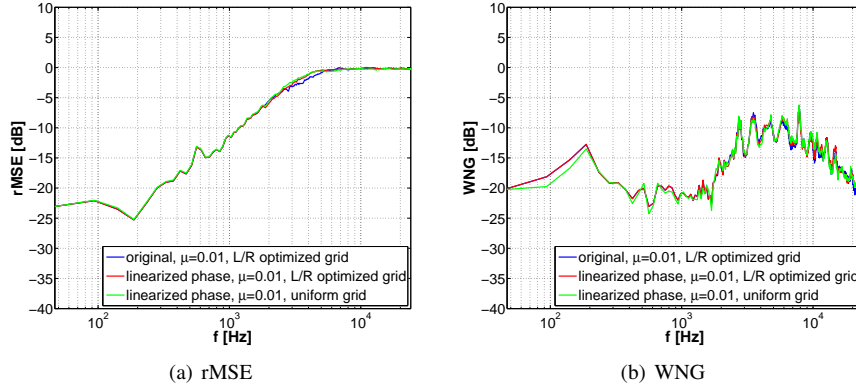
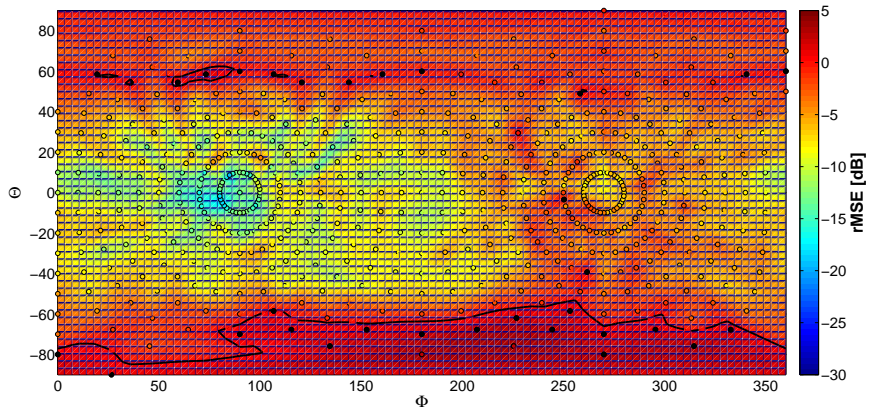


Figure 2.12: Comparison of the HRTF pattern fitting performance with the L/R optimized grid, with the data with linearized phase on the L/R optimized grid and the data with linearized phase on the uniformly sampled grid ($\mu = 0.01$).

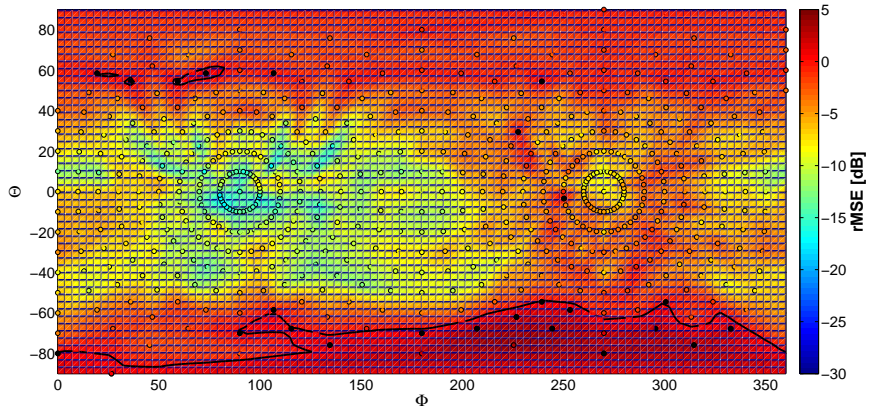
Fig. 2.13(c) shows the rMSE for fitting with linearized data on the uniform evaluation grid. The spatial rMSE error is more distributed over the full grid, as the concentration of low rMSE near the ears is smoothed out. Lower rMSE is seen near the poles.

The foregoing shows that a good approximation of the HRTF directivity pattern can be obtained at frequencies below 2kHz. At higher frequencies however, the HRTF pattern fitting performance decreases steadily due to inaccuracies in the target and measured directivity patterns and spatial aliasing caused by the microphone spacing, which is large relative to the corresponding wavelength at these frequencies.

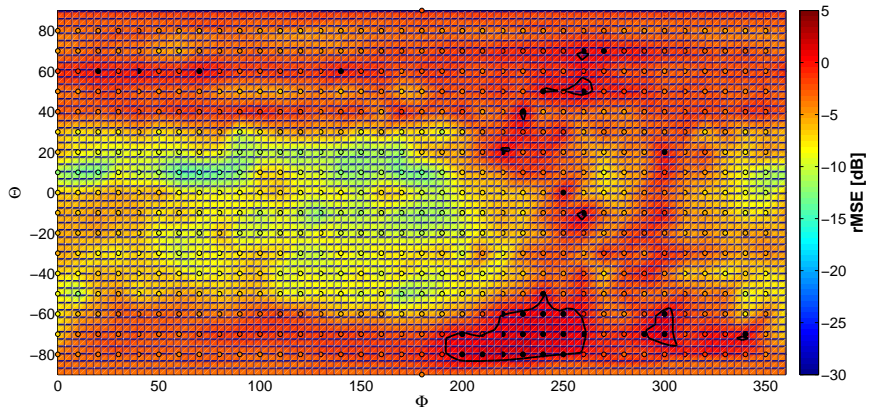
In order to provide the listener with a binaural signal, containing the full audible frequency range and directional characteristics at higher frequencies up to a certain degree, the MTB technique will be used. In practice, the signals from the microphone array will be processed in two different ways. At the low and mid frequencies, microphone signals will be filtered with the filter coefficients resulting from the HRTF pattern fitting, while at frequencies higher than a given cross-over frequency f_X only the high-passed filtered signal from the microphone closest to the left or right ear will be used. This feature is enabled by the specific array geometry, as the microphone on the hard sphere serves as a first-order approximation of the HRTFs. The cross-over frequency between both methods is to be determined by listener evaluation.



(a) rMSE - original phase, L/R optimized grid



(b) rMSE - linearized phase, L/R optimized grid



(c) rMSE - linearized phase, uniform grid

Figure 2.13: Interpolated spatial distribution of the overall rMSE for the left ear, calculated over frequencies until $f_{max} = 3\text{kHz}$ for the different fitting techniques with $\mu = 0.01$. Encircled values represent the actual rMSE values, calculated at the available measurement directions. The black contour and circles indicate an rMSE > 0dB.

2.4 Subjective validation of the binaural reconstruction methodology

In the previous sections a 32-element spherical microphone array has been developed that can be used for binaural signal extraction. Based on HRTF directivity pattern fitting, filter coefficients have been calculated for each of the 32 microphones on the sphere that ensure a good HRTF approximation at low and mid frequencies. At higher frequencies, where pattern fitting was unable to accurately approximate the HRTF, the MTB technique has been applied.

In this section, a more qualitative evaluation of the proposed binaural reconstruction technique will be presented. By means of listening tests, it will be investigated whether or not listeners can distinguish between original binaural recordings, made with a HATS, and binaurally recreated samples based on microphone array recordings. Furthermore, the influence of the regularization parameter, cross-over frequency f_X and evaluation grid will be investigated.

2.4.1 Acquisition of the audio samples

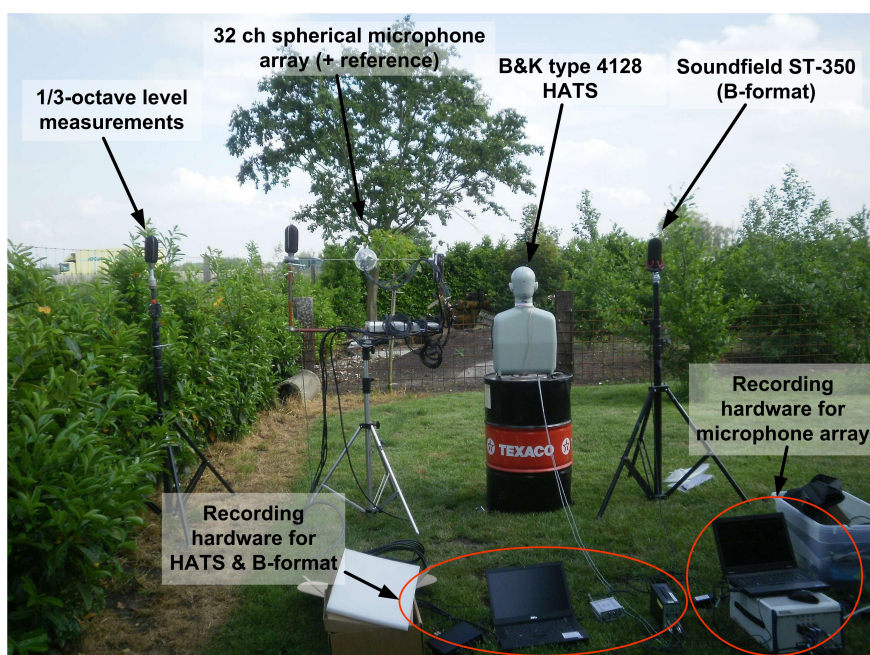
As the microphone array is mainly designed for binaural reconstruction of outdoor sound environments, and more specific for the auralization of the influence of noise mitigation measures on traffic noise (see next chapter), the sound environment at a location near the E17 highway in Kruishoutem (Belgium) is selected to evaluate the quality of the array reconstruction (Fig. 2.14(a)). This location was of specific interest as an earth mound was going to be installed to shield the residents nearby from traffic noise, which allowed us to investigate the sound environment before and after installation of the mound (see next chapter).

Apart from recordings with the 32-channel microphone array, a reference recording was done with a HATS (B&K Type 4128C) with inner ear simulator (B&K Type 4158C). Furthermore, the 1/3-octave sound pressure level was logged during recordings with a Swing measurement unit (Sinus Messtechnik GmbH) with calibrated MK-250 free-field microphone and MV-210 pre-amp (Microtech Gefell GmbH) to enable proper calibration of the reproduction level during the listening test. Fig. 2.14(b) shows the set-up of the recording equipment near the highway. An extra B-format microphone (Soundfield ST-350) was used as a back-up, but turned out to be unnecessary.

17 samples suitable for the listening test were selected from the binaural recordings and corresponding array recordings were converted to binaural signals. Therefore, array recordings are first filtered with a low-pass and high-pass filter with cross-over frequency f_X . Each of the low-pass microphone signals is filtered with the array coefficients, calculated in section 2.3, while for frequencies higher than



(a) Measurement location near the E17 highway, indicated by the red dot.



(b) Recording set-up.

Figure 2.14: Recording set-up near the E17 highway. The top figure shows the location near the E17 highway. The bottom figure shows the recording set-up.

f_X , only the high-pass filtered signal from the microphones corresponding to the left and right ear is kept. Filtering is applied by using the overlap-save technique with a block length of 8192 samples to overcome aliasing originating from circular convolution, implemented as a multiplication in the frequency domain. The high-pass MTB signal is scaled with the ratio of the rms-energy of the resulting low-pass binaural array output and the rms-energy of the low-pass filtered part of the left or right microphone, and delayed with the estimated group delay of the array coefficients of the left or right microphone.

As the listening test includes comparison with reference (static) HATS recordings, head motion is not accounted for in the binaural reproduction. Hence, only a static set of array weights is used, related to a head orientation with interaural axis parallel with the highway, while for frequencies higher than f_X only the signal from microphones A2 and A4 is taken for the left resp. right ear signal.

Different versions of the same sound fragment are created with different cross-over frequencies ($f_X = 1122\text{Hz}$, 1782Hz , 2245Hz and 4490Hz , being the upper frequency limits of the 1kHz, 1.6kHz, 2kHz and 4kHz 1/3-octave band), different regularization parameter values ($\mu = 0.0001$, 0.01 , 1 and 100) and different evaluation grids (L/R optimized grid and uniform grid with phase linearization). Furthermore, an extra version is included where the MTB technique is used for the full frequency range.

A level calibration of the reference and reproduced samples was carried out based on the low frequency 1/3-octave levels, measured by the free-field microphone, as no directivity is expected at these frequencies.

2.4.2 Listening test set-up and results

The listening test was performed with a calibrated headphone (Sennheiser HD-280 pro) and took place in the listening room at Ghent University or in a quiet room at the listener's home. In total, 40 persons took part in the test, of which three were withdrawn due to hearing problems observed from an a priori hearing test. Eventually, 14 female listeners and 23 male listeners remained, with age ranging between 21 and 57 years old. 10 listeners originated from the acoustics research group, while the others had little to no acoustical background.

In a first test it is investigated whether or not listeners can distinguish between HATS recordings and binaural reproductions from array recordings. Therefore, the listeners were presented with a list of five original binaural recordings and five binaural extractions, each generated with a different technique. Listeners were asked to indicate whether samples were true recordings or reproductions. Table 2.2 displays the indicated answers for the different samples and techniques. The right-most column indicates the probability that the null-hypothesis, assuming random

Table 2.2: Results of the first test. The columns indicate the responses given by the listeners, presented with samples recorded with the HATS or binaurally synthesized from microphone array recordings. The right column gives the probability that the null-hypothesis, assuming random choice, is accepted. ($\chi^2 = 40.13$, $p < 0.01$)

	Ind. record.	Ind. reprod.	Count	p-value
HATS recording	130	55	185	$< 0.01^{**}$
L-R array recording	19	18	37	1
Fit ext. grid, $f_X = 1122\text{Hz}$	15	22	37	0.324
Fit ext. grid, $f_X = 1782\text{Hz}$	16	21	37	0.511
Fit, $f_X = 1782\text{Hz}$	18	19	37	1
Fit, $f_X = 4490\text{Hz}$	8	29	37	$< 0.01^{**}$

choice, is accepted, resulting from a two-sided binomial test.

Table 2.2 shows that the null-hypothesis is clearly rejected for the original HATS recordings and array reproductions with L/R optimized grid and $f_X = 4490\text{Hz}$. This indicates that listeners can clearly identify the HATS recording as a true binaural recording, which proves that the listeners indeed have a good understanding of what is meant by a ‘real’ recording. The array reproduction with $f_X = 4490\text{Hz}$ is identified as reproduction. For the other array processing techniques, including the binaural reproduction solely based on microphone A2-A4 (‘L-R array recording’), the assumption of random choice could not be rejected, indicating that listeners could not identify whether or not array extracts were a true binaural recording or reproduction. The reason why some array recordings were still indicated as a reproduction probably originates from the negative array WNG, combined with the difference in microphone quality, more specific, in differences between the high quality HATS and low cost array microphone noise spectrum. In written comments made during the test and in the evaluation talk afterwards, listeners pointed out differences in (high frequency) background noise (originating from the microphone noise), which were often used to support their decision, rather than the authenticity of the car or truck passage. However, they generally indicated it was very hard to make the difference.

In a second test, the influence of the cross-over frequency f_X was investigated. In this test, the listener was presented with an original HATS recording (indicated as such), and he or she was asked to rank five binaural reproductions according to their similarity with the original recording (1 means most similar). The five samples were reconstructed with different f_X and fixed $\mu = 0.01$, but recorded at the same time as the original passage.

Table 2.3 shows the frequency count of rankings given to each reproduction technique. This table shows that the L-R microphone based reproduction, together with

Table 2.3: Influence of the cross-over frequency f_X . The columns indicate the ranking given by the listeners, presented with binaurally synthesized samples with different cross-over frequencies f_X (1 means most similar). The total sample count is 37. ($\chi^2 = 67.57$, $p < 0.01$)

	1	2	3	4	5
L-R array recording	18	3	3	5	8
Fit, $f_X = 1122\text{Hz}$	8	13	9	5	2
Fit, $f_X = 1782\text{Hz}$	5	13	9	3	7
Fit, $f_X = 2245\text{Hz}$	4	7	11	10	5
Fit, $f_X = 4490\text{Hz}$	2	1	5	14	15

Table 2.4: Influence of the regularization parameter μ . The columns indicate the ranking given by the listeners, presented with binaurally synthesized samples with different regularization μ (1 means most similar). The total sample count is 37. ($\chi^2 = 38.65$, $p < 0.01$)

	1	2	3	4	5
Fit, $\mu = 0.0001$	5	6	9	9	8
Fit, $\mu = 0.01$	6	13	9	5	4
Fit, $\mu = 1$	8	6	9	11	3
Fit, $\mu = 100$	3	4	6	9	15
L-R array recording	15	8	4	3	7

the reproductions with $f_X = 1122\text{Hz}$ and 1782Hz , are preferred over the samples with higher f_X . In order to determine a more explicit ranking and to investigate if there is a significant difference between reproductions with different f_X , a multiple comparison test is performed by applying Tukey's least significant difference procedure on the outcome of an ANOVA test to check whether or not mean rankings are significantly different. Fig. 2.15(a) shows the results from the multiple comparison test. The mean ranking is indicated for each technique, together with the confidence interval. Two reproduction techniques are significantly different when these intervals do not overlap. Clearly, the reproduction with $f_X = 4490\text{Hz}$ is ranked worst, and is significantly different from the other reproduction techniques, which is a confirmation of the results from the first test. The L-R reproduction (corresponding with $f_X = 0\text{Hz}$) and reproduction with $f_X = 1122\text{Hz}$ have the best ranking, with a (marginal) significant difference from reproductions with $f_X = 2245\text{Hz}$ and 4490Hz . The reproduction with $f_X = 1782\text{Hz}$ comes third. However, it's ranking is not significantly different than the best samples or samples with $f_X = 2245\text{Hz}$.

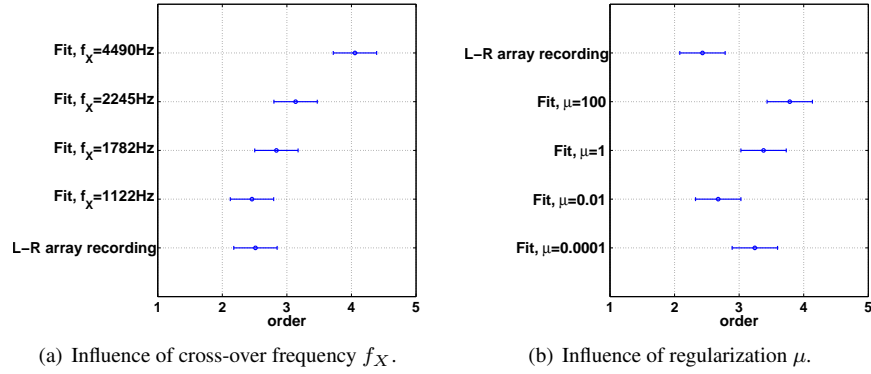


Figure 2.15: Results of the subjective ranking of the different HRTF fitting techniques. Mean ranking and 95% confidence intervals of each technique are indicated.

The influence of the regularization parameter is investigated in a third test, with similar lay-out as the second test. Again, listeners are asked to rank five binaural reproductions according to their similarity with the original recording (1 means most similar). The five samples were reconstructed with different values of the regularization parameter, but fixed $f_X = 1782\text{Hz}$. Table 2.4 shows the frequency counts of the given ranks. Worst ranking is seen for $\mu = 100$ and 1, while $\mu = 0.01$ and the L-R reproduction are ranked best. It is difficult however to draw conclusions regarding the reproduction with lowest regularization value, $\mu = 0.0001$, as frequency counts are more divided over the rank orders.

Similar as in the previous test, a multiple comparison test is performed to investigate the significance of the mean ranking, shown in Fig. 2.15(b). It can be seen that the L-R reproduction, together with the reproduction with $\mu = 0.01$, has the best ranking, which is (marginally) significantly different from reproductions with other μ . The reproduction with $\mu = 100$ has worst ranking. However, there is no significant difference from reproductions with $\mu = 1$ and 0.0001.

In a last test, the influence of the simulation grid and phase linearization is investigated with a similar procedure as test 2 and test 3. Listeners are presented with an original binaural recording and four binaurally reproduced samples. Two samples were reproduced with array weights resulting from a fit on the L/R optimized grid ($f_X = 1782\text{Hz}$ and 4490Hz , $\mu = 0.01$) and two samples were binaural reproductions with array weights from a fit on an extended (uniform) grid with linearized phase ($f_X = 1782\text{Hz}$ and 4490Hz , $\mu = 0.01$). The frequency counts of each rank order is given in Table 2.5. Apart from the observation that reproductions with high f_X score worse than other reproductions, no clear differences can be seen.

Table 2.5: Influence of the simulation grid. The columns indicate the ranking given by the listeners, presented with binaurally synthesized samples with varying simulation grid and cross-over frequency f_X (1 means most similar). The total sample count is 37.
($\chi^2 = 50.05$, $p < 0.01$)

	1	2	3	4
Fit, $f_X = 1782\text{Hz}$	14	13	7	3
Fit ext. grid, $f_X = 1782\text{Hz}$	15	15	3	4
Fit, $f_X = 4490\text{Hz}$	5	5	16	11
Fit ext. grid, $f_X = 4490\text{Hz}$	3	4	11	19

Fig. 2.16 shows the mean ranking with confidence intervals. This figure confirms that samples with $f_X = 4490\text{Hz}$ are significantly worse than samples with lower f_X . No significant difference is seen between the L/R optimized grid and extended (uniform) grid with linearized phase fitting, independent of f_X . However, it is to be expected that, when the interaural axis is changed during motion and weights need to be recalculated, the difference in performance between both methods will increase and the extended grid will yield better overall results than the L/R optimized grid, as it has a uniform distribution of evaluation points, independent from the orientation of the interaural axis.

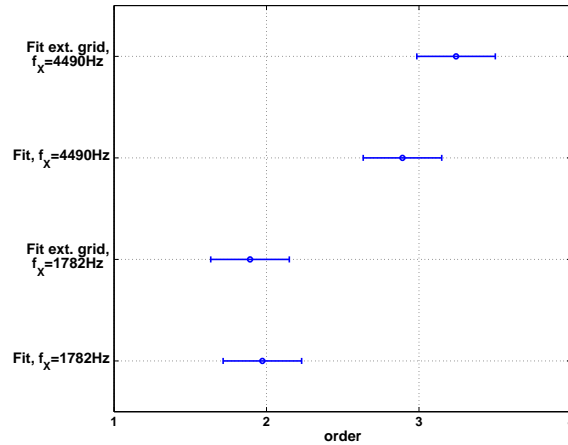


Figure 2.16: Results of the subjective ranking of the HRTF fitting technique for different evaluation grids. Mean ranking and 95% confidence intervals of each technique are indicated.

2.5 Conclusion

In this chapter an immersive audio reproduction methodology has been designed for the binaural recreation and auralization of outdoor sound environments. Here, binaural reproduction over headphones was preferred, due to the accessibility of this technique for city-planners and, by extension, the general public.

In order to overcome the limitations associated with the HATS binaural recording technique, being the inability to change head orientation after recordings have been made, and to allow direction specific attenuation of the sound (see next chapter), a 32-channel spherical microphone array has been designed that emulates a given set of HRTF directivity patterns up to a certain frequency. At higher frequencies, where HRTF patterns are subject to large spectral fluctuations, quality of the emulation decreases and the MTB technique was used.

The directivity pattern of each microphone on the sphere has been measured to account for the influence of spectral differences between microphones, positioning and calibration errors and influences of fixation mechanisms. In a next step, these measured patterns have been used as input for a pattern fitting algorithm based on a regularized least-square solution. With this algorithm, microphone weights have been estimated so that the combined directivity pattern of the array matches the target directivity pattern, in this case the equalized HRTFs as measured by Gardner. Prior to the pattern fitting, both the HRTF directivity pattern as well as the microphone-on-sphere patterns have been decomposed into a spherical harmonic representation to enable a continuous data representation, independent of the measurement direction, as well as to smooth out measurement noise in the microphone-on-sphere patterns and account for the polar gap in the measured HRTFs.

The performance of the fit was strongly dependent on the regularization parameter μ . A decreasing μ -value decreased the fitting error (improved rMSE), but had a negative effect on the WNG. Furthermore, it was seen that the overall rMSE between the resulting fit and target HRTF patterns increased with frequency. For $\mu = 0.01$, the value ranged from -23dB at low frequencies to -11dB at 1kHz. From 2kHz on, the rMSE raised above -5dB. Phase linearization at higher frequencies had little effect on the performance, as well as the use of a uniform grid rather than an L/R optimized grid, with more evaluation points around the ears.

In addition to this physical evaluation, the quality of the immersive reproduction methodology has been tested by a set of listening tests, where listeners were asked to compare true binaural recordings of traffic noise from a highway with binaurally reproduced samples, recreated from the 32-channel array recordings with different values of μ and f_X and different evaluation grids. While it was seen that listeners could identify true binaural recordings, binaurally reproduced samples with $f_X \leq 2245\text{Hz}$ were randomly categorized, indicating that listeners could hardly identify whether or not a reproduction was a true HATS recording or binau-

rally reproduced recording. In further tests, the influence of cross-over frequency f_X and regularization parameter μ on the reproduction quality was investigated. When f_X decreased, samples were typically given a better ranking. However, no significant difference between samples with $f_X \leq 1782\text{Hz}$ could be seen. No such linear relation was found between μ and the ranking of samples, and $\mu = 0.01$ was identified as good trade-off between rMSE and WNG. In a final test the influence of evaluation grid was evaluated, together with phase linearization. Nevertheless, only limited effect was seen.

Although this analysis proves that the microphone array based HRTF emulation technique is a good alternative for HATS recordings, especially because listeners indicated that it was very difficult to distinguish between the samples, it should be noted that the MTB technique for the full frequency range (by using only the microphone closest to the left and right ear) also shows to be a well appreciated binaural reproduction technique. In the next chapter, however, we will show that the auralization quality of noise mitigation measures is better when using the combined HRTF pattern fitting - MTB technique rather than the MTB technique for the full frequency range.

References

- [1] T. Lokki. *Subjective comparison of four concert halls based on binaural impulse responses*. Acoust. Sci. & Tech., 26:200–203, 2005.
- [2] S. Pelzer, B. Masiero, and M. Vorländer. *3D reproduction of room acoustics using a hybrid system of combined cross cancellation and Ambisonics playback*. In Proceedings of the International Conference on Spatial Audio, 2011.
- [3] J. Kang, Y. Smyrnova, P. Richmond, and S. Maddock. *Integration of acoustic simulation with interactive visual animation of urban environment*. In Proceedings of the EAA Euregio 2010 congress, 2010.
- [4] F. Ruotolo, L. Maffei, M. Di Gabriele, T. Iachini, M. Masullo, G. Ruggiero, and V. Senese. *Immersive virtual reality and environmental noise assessment: an innovative audio-visual approach*. Environmental Impact Assessment Review, 41:10–20, 2013.
- [5] P. Thomas, M. Boes, T. Van Renterghem, D. Botteldooren, M. Hornikx, W. Desmet, and M. E. Nilsson. *Auralisation of a car pass-by behind a low finite-length vegetated noise barrier*. In Proceedings of the 9th European Conference on Noise Control, pages 932–937. European Acoustics Association (EAA), 2012.

- [6] Lord Rayleigh. *On our perception of sound direction*. Philos. Mag., 13:214–232, 1907.
- [7] J. C. Middlebrooks and D. Green. *Sound localization by human listeners*. Annu. Rev. Psychol., 42:135–159, 1991.
- [8] J. Blauert. *Spatial hearing: The psychophysics of human sound localization*. MIT Press, 1997.
- [9] D. Wang and G. Brown, editors. *Computational auditory scene analysis*. IEEE press, 2006.
- [10] R. Algazi and R. Duda. *Headphone-based spatial sound*. IEEE Signal Processing Magazine, 28:33–42, 2011.
- [11] J. Blauert, editor. *The technology of binaural listening*. Springer - Verlag, 2013.
- [12] E. Macpherson and J. C. Middlebrooks. *Listener weighting of cues for lateral angle : The duplex theory of sound localization revisited*. J. Acoust. Soc. Am., 111:2219–2236, 2002.
- [13] H. Wallach. *On sound localization*. J. Acoust. Soc. Am., 10:270–274, 1939.
- [14] E. Wenzel, M. Arruda, D. Kistler, and F. Wightman. *Localization using non-individualized head-related transfer functions*. J. Acoust. Soc. Am., 94:111–123, 1993.
- [15] F. Wightman and D. Kistler. *Resolution of front-back ambiguity in spatial hearing by listener and source movement*. J. Acoust. Soc. Am., 105:2841–2853, 1999.
- [16] C. Jin, A. Corderoy, S. Carlile, and A. Van Schaik. *Contrasting monaural and interaural spectral cues for human sound localization*. J. Acoust. Soc. Am., 115:3124–3141, 2004.
- [17] A. Mills. *On the minimum audible angle*. J. Acoust. Soc. Am., 30:237–246, 1958.
- [18] D. Perrott and K. Saberi. *Minimum audible angle thresholds for source varying in both elevation and azimuth*. J. Acoust. Soc. Am., 87:1728–1731, 1990.
- [19] D. Grantham, B. Hornsby, and E. Erpenbeck. *Auditory spatial resolution in horizontal, vertical and diagonal planes*. J. Acoust. Soc. Am., 114:1009–1022, 2003.

- [20] R. Algazi, C. Avendano, and R. Duda. *Elevation localization and head-related transfer function at low frequencies*. J. Acoust. Soc. Am., 109:1110–1122, 2001.
- [21] S. Perrett and W. Noble. *The effect of head rotations on vertical plane source detection*. J. Acoust. Soc. Am., 102:2325–2332, 1997.
- [22] B. Shinn-Cunningham. *Distance cues for virtual auditory space*. In Proceedings of the 1st IEEE Pacific-Rim conference on multimedia, 2000.
- [23] P. Zahorik, D. Brungart, and A. W. Bronkhorst. *Auditory distance perception in humans: a summary of past and present research*. Acta Acustica united with Acustica, 91:409–420, 2005.
- [24] A. W. Bronkhorst and T. Houtgast. *Auditory distance perception in rooms*. Nature, 397:517–520, 1999.
- [25] D. Brungart, N. Durlach, and W. Rabinowitz. *Auditory localization of nearby sources II. Localization of a broadband source*. J. Acoust. Soc. Am., 106:3589–3602, 1999.
- [26] W. Gardner. *3D audio using loudspeakers*. Kluwer Academic Publishers, 1998.
- [27] W. M. Hartmann and A. Wittenberg. *On the externalization of sound images*. J. Acoust. Soc. Am., 99:3678–3688, 1996.
- [28] S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, and F. Zotter. *Spatial Sound with loudspeakers and its perception: a review of the current state*. Proceedings of the IEEE, 101:1920–1938, 2013.
- [29] R. Duraiswami, D. Zotkin, Z. Li, E. Grassi, N. Gumerov, and L. Davis. *High order spatial audio capture and binaural head-tracked playback over headphones with HRTF cues*. In Proceedings of the 119th AES Convention, 2005.
- [30] E. Wenzel, F. Wightman, and D. Kistler. *Localization with non-individualized virtual acoustic display cues*. In Proceedings of the SIGCHI Conference on human factors in computing, 1991.
- [31] R. Algazi, R. Duda, and D. Thompson. *Motion-tracked binaural sound*. J. Audio Eng. Soc., 52:1142–1156, 2004.
- [32] R. Algazi, R. Duda, J. Melick, and D. Thompson. *Customization for personalized rendering of motion-tracked binaural sound*. In Proceedings of the 117th AES Convention, 2004.

- [33] M. Gerzon. *Width-height sound reproduction*. J. Audio Eng. Soc., 21:2–10, 1973.
- [34] M. Gerzon. *Ambisonic in multichannel broadcasting and video*. J. Audio Eng. Soc., 33:859–871, 1985.
- [35] P. Craven and M. Gerzon. *Coincident microphone simulation covering three dimensional space and yielding various directional outputs*, August 16 1977. US Patent 4,042,779.
- [36] J. Batke. *The B-format microphone revised*. In Proceedings of the 2009 Ambisonics symposium, 2009.
- [37] T. Abhayapala and D. Ward. *Theory and design of high order sound field microphones using spherical microphone array*. In Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), 2002.
- [38] B. Rafaely. *Analysis and design of spherical microphone arrays*. IEEE transactions on speech and audio processing, 13:135–143, 2005.
- [39] S. Moreau, J. Daniel, and S. Bertet. *3D sound recording with higher order Ambisonics - objective measurements and validation of a 4th order spherical microphone*. In Proceedings of the 120th AES Convention, 2006.
- [40] B. Rafaely, B. Weiss, and E. Bachmat. *Spatial aliasing in spherical microphone arrays*. IEEE transactions on signal processing, 55:1003–1010, 2007.
- [41] B. Rafaely. *Plane-wave decomposition of the sound field on a sphere by spherical convolution*. J. Acoust. Soc. Am., 116:2149–2157, 2004.
- [42] M. Park and B. Rafaely. *Sound-field analysis by plane-wave decomposition using spherical microphone array*. J. Acoust. Soc. Am., 118:3094–3103, 2005.
- [43] R. Duraiswami, Z. Li, D. Zotkin, E. Grassi, and N. Gumerov. *Plane-wave decomposition analysis for spherical microphone arrays*. In Proceedings of the IEEE workshop on ASPAA, 2005.
- [44] Z. Li, R. Duraiswami, and N. Gumerov. *Capture and reproduction of higher order 3D sound fields via reciprocity*. In Proceedings of ICAD, 2004.
- [45] Z. Li and R. Duraiswami. *Flexible and optimal design of spherical microphone arrays for beamforming*. IEEE transactions on audio, speech and language processing, 15:702–714, 2007.

- [46] M. Guillaume and Y. Grenier. *Sound field analysis based on analytical beam-forming*. EURASIP journal on advances in signal processing, 2007:189–189, 2007.
- [47] A. O'Donovan, D. Zotkin, and R. Duraiswami. *A spherical microphone array based system for immersive audio scene rendering*. Technical report, Institue for Advanced Computer Studies, Univ. of Maryland, College Park, 2008.
- [48] A. Farina, A. Capra, L. Chiesi, and L. Scopece. *A spherical microphone array for synthesizing virtual directive microphones in live broadcasting and in post production*. In Proceedings of the 40th AES Conference, 2010.
- [49] E. De Sena, H. Hacihabiboglu, and Z. Cvetkovic. *A generalized design method for directivity patterns of spherical microphone arrays*. In Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), 2011.
- [50] S. Delikaris-Manias, C. Valagiannopoulos, and V. Pulkki. *Optimal directional pattern design utilizing arbitrary microphone arrays: a continuous-wave approach*. In Proceedings of the 134th AES Convention, 2013.
- [51] E. Rasumow, M. Blau, S. Doclo, M. Hansen, S. Van de Par, D. Püschel, and V. Mellert. *Least squares versus non-linear cost functions for a virtual artificial head*. In Proceedings of meetings on acoustics, ICA 2013, 2013.
- [52] M. Noisternig, T. Musil, A. Sontacchi, and R. Höldrich. *A 3D real time rendering engine for binaural sound reproduction*. In Proceedings of ICAD, 2003.
- [53] A. Avni, J. Ahrens, M. Geier, S. Spors, H. Wierstorf, and B. Rafaely. *Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution*. J. Acoust. Soc. Am., 133:2711–2721, 2013.
- [54] K. Cuelenaere. *Design of a true-to-life 3D immersive audio environment*. Master's thesis, Ghent University, 2012.
- [55] P. Damaske. *Head-related two-channel stereophony with loudspeaker reproduction*. J. Acoust. Soc. Am., 50:1109–1115, 1971.
- [56] V. Pulkki. *Virtual sound source positioning using vector based amplitude panning*. J. Audio Eng. Soc., 45:456–466, 1997.
- [57] V. Pulkki. *Spatial sound reproduction with directional audio coding*. J. Audio Eng. Soc., 55:503–516, 2007.

- [58] J. Daniel, J. Rault, and J. Polack. *Ambisonics encoding of other audio formats for multiple listening conditions*. In Proceedings of the 105th AES Convention, 1998.
- [59] E. Benjamin and R. Lee. *Localization in horizontal-only Ambisonic systems*. In Proceedings of the 121st AES Convention, 2006.
- [60] S. Bertet, J. Daniel, L. Gros, E. Parizet, and O. Warusfel. *Investigation of the perceived spatial resolution of higher order Ambisonics sound fields: A subjective evaluation involving virtual and 3D microphones*. In Proceedings of the 30th Int. Conf. Intell. Audio Environ., 2007.
- [61] S. Braun and M. Frank. *Localization of 3D Ambisonic recording and Ambisonic virtual sources*. In Proceedings of Int. Conf. on Spatial Audio, 2011.
- [62] A. Solvang. *Spectral impairment of two-dimensional higher order Ambisonics*. J. Audio Eng. Soc., 56:267–279, 2008.
- [63] A. Berkhout, D. De Vries, and P. Vogel. *Acoustic control by wave field synthesis*. J. Acoust. Soc. Am., 93:2764–2778, 1993.
- [64] S. Spors and R. Rabenstein. *Spatial aliasing artifacts produced by linear and circular loudspeaker arrays used for wave field synthesis*. In Proceedings of the 120th AES convention, 2006.
- [65] H. Wierstorf, A. Raake, and S. Spors. *Localization of a virtual point source within the listening area for Wave Field Synthesis*. In Proceedings of the 133rd AES convention, 2012.
- [66] W. G. Gardner and K. D. Martin. *HRTF measurements of a KEMAR*. J. Acoust. Soc. Am., 97:3907, 1995.
- [67] E. Rasumow, M. Blau, M. Hansen, S. Doclo, S. Van De Par, D. Püschel, and V. Mellert. *Smoothing head-related transfer functions for a virtual artificial head*. In Proceedings of the Acoustics 2012 Nantes Conference, 2012.
- [68] H. Tokuno, O. Kirkeby, P. A. Nelson, and H. Hamada. *Inverse filter of sound reproduction systems using regularization*. IEICE Trans. Fundamentals, E80-A:809–820, 1997.
- [69] O. Kirkeby and P. A. Nelson. *Digital filter design for inversion problems in sound reproduction*. J. Audio Eng. Soc., 47:583–595, 1999.
- [70] W. Zhang, R. A. Kennedy, and T. Abhayapala. *Efficient continuous HRTF model using data independent basis functions: experimental guided approach*. IEEE transactions on Audio, Speech and Language processing, 17:819–829, 2009.

- [71] M. Pollow, K. Nguyen, O. Warusfel, T. Carpentier, M. Müller-Trapet, M. Vorländer, and M. Noisternig. *Calculation of head-related transfer functions for arbitrary field points using spherical harmonics decomposition*. Acta Acustica united with Acustica, 98:72–82, 2012.
- [72] E. H. A. Langendijk and A. W. Bronkhorst. *Fidelity of three-dimensional sound reproduction using a virtual auditory display*. J. Acoust. Soc. Am., 107:528–537, 2000.
- [73] A. Rodriguez-Morales. *Real-time sound processing for interactive auralisation of sound insulation*. In Proceedings of Forum Acusticum, 2011.
- [74] D. Kistler and F. Wightman. *A model of head-related transfer functions based on principal component analysis and minimum-phase reconstruction*. J. Acoust. Soc. Am., 91:1637–1647, 1992.
- [75] J. Chen, K. E. Van Veen, and K. E. Hecox. *A spatial feature extraction and regularization model for the head-related transfer function*. J. Acoust. Soc. Am., 97:439–452, 1995.
- [76] M. J. Evans, J. Angus, and A. Tew. *Analyzing head-related transfer function measurements using surface spherical harmonics*. J. Acoust. Soc. Am., 104:2400–2411, 1998.
- [77] W. Zhang, M. Zhang, R. A. Kennedy, and T. Abhayapala. *On high-resolution head-related transfer function measurements: an efficient sampling scheme*. IEEE transactions on Audio, Speech and Language processing, 20:575–584, 2012.
- [78] H. Na, C. Lee, and O. Cheong. *Voronoi diagrams on the sphere*. Computational Geometry, 23:183–194, 2002.
- [79] D. Zotkin, R. Duraiswami, and N. Gumerov. *Regularized HRTF fitting using spherical harmonics*. In Proceedings of the IEEE workshop on Applications of Signal Processing to Audio and Acoustics, pages 1–5, 2009.
- [80] R. Algazi, C. Avendano, and R. Duda. *Estimation of a spherical-head model from anthropometry*. J. Audio Eng. Soc., 49:472–479, 2001.
- [81] T. Van Renterghem, P. Thomas, F. Dominguez, S. Dauwe, A. Touhafi, B. Dhoedt, and D. Botteldooren. *On the ability of consumer electronics microphones for environmental noise monitoring*. Journal of Environmental Monitoring, 13:544–552, 2011.
- [82] R. Duraiswami, D. Zotkin, and N. Gumerov. *Interpolation and range extrapolation of HRTFs*. In Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), 2004.

-
- [83] F. P. Mechel, editor. *Formulas of Acoustics*. Springer - Verlag, 2nd edition, 2008.
 - [84] E. Rasumow, S. Doclo, S. Van De Par, and V. Mellert. *The impact of the white noise gain (WNG) of a virtual artificial head on the appraisal of binaural sound reproduction*. In Proceedings of the EAA joint symposium on auralization and Ambisonics, 2014.
 - [85] E. Rasumow, M. Blau, M. Hansen, S. Doclo, S. Van De Par, V. Mellert, and D. Püschel. *Robustness of virtual artificial head topologies with respect to microphone positioning*. In Proceedings of Forum Acusticum, pages 397–402. European Acoustics Association (EAA), 2011.

3

Auralization of the effect of noise mitigation

This chapter is partially based on the publication ‘Auralization of a car pass-by behind a low finite-length vegetated noise barrier’, presented during the 9th European conference on noise control (Euronoise 2012), June 10-13, 2012, Prague and the research carried out in the context of the HOSANNA-project, ‘WP 6 - Holistic acoustic design and perceptual evaluation’.



A popular measure to screen city-dwellers from road traffic noise is the introduction of sound barriers. Traditionally, required noise level reductions are used to determine the barrier dimensions and structure. With the more subtle design of soundscapes in mind, auralization of car passages behind such barriers can form a useful tool to evaluate their effectiveness and are ideal to be used as a demonstrator for the larger public.

Existing approaches are commonly based on ab initio auralization techniques. Maillard et al. developed an auralization tool to model car passages in an urban environment based on a traffic flow simulator, real-time granular source synthesis technique and moving source renderer [1, 2]. Individual source signals are generated by using an overlap-add mechanism on sound samples extracted from engine and rolling noise recordings, with the speed and acceleration from the traffic flow simulator as input [3]. The LISTEN-project [4, 5] is another example of a research project for the auralization of urban soundscapes. Here, a single car pass-by has been auralized based on individual

engine and rolling noise signals, extracted from an original pass-by recording by inversion of the propagation effects.

The perceived realism of these ab initio auralizations is typically relying on the accurateness of the description of each individual vehicle noise source, dependent on the local road surface, vehicle fleet composition and traffic flow. In this chapter, a new approach for the auralization of noise mitigation measures is suggested based on a priori recordings of the total sound environment. In this way, effects related to the traffic are readily included. Furthermore, the effect of the noise mitigation measure on the complete sound environment can be evaluated, as other sources, typical for that site, are accounted for as well, resulting in more realistic auralizations.

In this chapter two different methodologies will be designed and evaluated. The first methodology uses a binaural recording as input, while the effect of the mitigation measure is included by applying time-selective attenuation on the source signal, dependent on its relative position. This methodology is especially useful when only binaural recordings are available and the sound environment is mainly dominated by successive single car passages. The feasibility of this concept is investigated in a first case-study, where the effect of a low vegetated noise barrier in Lyon, France, is auralized. The barrier was installed in the context of the HOSANNA-project to screen people walking along the road from traffic noise [6].

The second auralization methodology is based on a priori array recordings with the 32-channel spherical microphone array developed in chapter 2. The effect of a noise mitigation measure is included by spatial-selective attenuation of the array directivity pattern, and only the sound coming from a certain direction is attenuated. Although somewhat more complex than the first methodology, this technique is very suitable for the auralization of complex-shaped noise mitigation measures in complicated sound environments, containing multiple sources at different locations. The performance of this technique is evaluated in a second case-study, when applied for the assessment of a large L-shaped mound along the E17 highway near Kruishoutem, Belgium.

3.1 Auralization of a car pass-by behind a low finite-length vegetated noise barrier

In this case-study, car passages behind a 1m high finite-length vegetated barrier have been auralized, starting from binaural recordings of single car passages before the barrier was erected. Source angle positions were deducted from the recordings, based on different binaural cues, using a trained Bayesian inference system. The insertion loss of the low and finite-length barrier was simulated for every source position with engineering models such as ISO 9613-2 and Harmonoise, and a full-wave numerical propagation model, the finite-difference time-domain technique

(FDTD). Finally, the auralization of the effect of the noise mitigation measure, estimated with the different simulation models, has been evaluated by an expert panel and opposed against the samples recorded after the barrier was erected.

3.1.1 Auralization of the effect of a small barrier on car passages

In the case-study, a finite-length vegetated barrier was erected beside a two lane one-directional road (Fig. 3.1). The height of the barrier is 1m, the length is approx. 14m. Although the road was rather busy, single car passages at low speed were frequently observed.

Binaural recordings with a head-and-torso simulator (B&K Type 4128C) were performed at approx. 3.7m from the edge of the street, at a height of 1.24m (Fig. 3.2). Recordings 20m beside the barrier were used as the starting point to create the simulated, synthetic listening samples with the barrier effect. Simultaneously, recordings behind the barrier were made to evaluate the accuracy of the auralized samples.



(a) Low vegetated sound barrier along the road.



(b) Head-and-torso simulator behind the barrier.

Figure 3.1: Sound barrier and measurement set-up behind the barrier.

The first step in the auralization methodology is to estimate the position of the car (angle ϕ) from the binaural recordings using a human-inspired source angle detector. Once the position is known, the source orientation-dependent barrier insertion loss ($IL(\phi)$) can be calculated using different methods. Given the frequency dependency of $IL(\phi)$, simulations are performed in 1/3-octave bands, from 50Hz to 16kHz. In a final step, the recorded signal, decomposed into its frequency components by a 1/3-octave band analysis filterbank, is attenuated with the calculated $IL(\phi)$. Here, inclusion of the angle dependent insertion loss is sufficient as other effects such as geometrical divergence, atmospheric absorption and Doppler effect are already included in the original recording.

The car itself is modeled by three sources (Fig. 3.2): two sources at the wheel positions and a source near the exhaust pipe. The contribution of each of the three sources is added with a scaling factor to maintain overall source power. A summary of the auralization methodology is depicted in Fig. 3.3; the different blocks are explained further on. This calculation is repeated for each 50ms interval of the sound fragment to account for changing $IL(\phi)$ due to source movement.

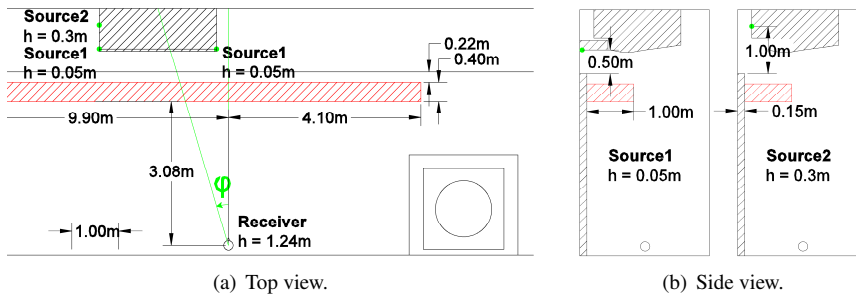


Figure 3.2: Simulation configuration with border and car geometry. Here, ϕ indicates the position of the car, rather than the source-receiver angle. Two geometries are used, depending on the source type (green dot).

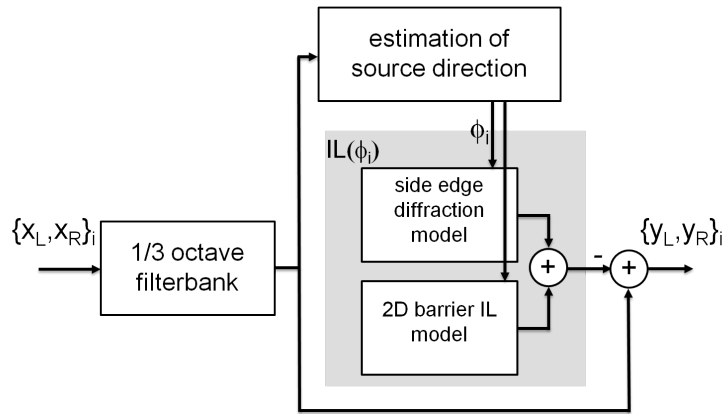


Figure 3.3: Block diagram illustrating the auralization methodology. The total barrier IL which should be applied is the energetic sum of the diffraction around the edges and the 2D IL according to the source-barrier-receiver intersection (including barrier attenuation and diffraction over the top edge).

3.1.1.1 Source angle extraction

Using a biologically inspired approach, developed by Boes et al. [7], an estimate of the source angle can be made. First, subsequent time intervals with a fixed duration of 50ms are extracted from the incoming sound sample. These are then sent through a filterbank, imitating the frequency decomposition effect of the cochlea, and half-wave rectified, simulating in a simplified way the behavior of hair cells. Next, for each of the resulting frequency channels, the response of an array of neurons sensitive to interaural time differences (ITDs) and interaural intensity differences (IIDs) is simulated. Using a Bayesian system to compare these responses to the responses on incoming sounds originating from an array of known angles, a probability distribution of angle of incidence can be calculated for each frequency channel based on the excitation of the neurons. Combining the probability distributions for each frequency channel, weighted with a characteristic car sound spectrum, an estimate of the car sound direction can be made at each time interval. Only when the resulting combined probability distribution is sufficiently sharply peaked, the estimate is considered valid.

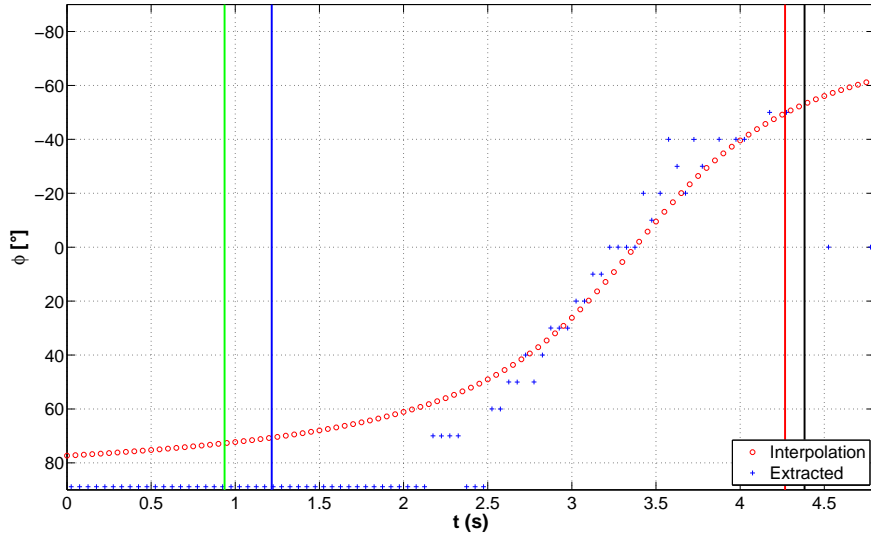


Figure 3.4: Source angle extraction from a recorded car passage. In blue (+) the extracted angles are given, in red (o) the smoothed version, assuming constant velocity. The vertical lines indicate barrier edge crossings: as the barrier has a thickness of 0.4m, the perceived barrier thickness is increasing at angles between the first and second line. Between the second and third line full barrier thickness is seen and at angles between third and fourth line, the perceived barrier thickness is decreasing. Crossing the first barrier edge takes more time than the second edge crossing due to the asymmetric receiver position.

The results are subject to fluctuations, leading to faulty backwards car movements. Therefore, a smoothing mechanism is applied, where the assumption is made that the velocity of the car is constant during the pass-by. Fig. 3.4 shows the extracted angle for a car passage and the smoothed result.

3.1.1.2 Simulation of barrier attenuation and diffraction

Different simulation methods to estimate the insertion loss of the barrier are considered. In a first step the 2D $IL(\phi)$ along the source-receiver section will be calculated with ISO 9613-2 [8], Harmonoise [9, 10] and FDTD [11]. Secondly, diffraction around the barrier side edges will be included to account for the finite length of the barrier. Here, diffraction will be modeled by means of ISO 9613-2 and contributions of the diffraction around both side edges are summed.

The simplest approach calculates the 2D $IL(\phi)$ following the ISO 9613-2 standard. Here, it is assumed that the barrier surfaces are acoustically hard and only a single diffraction over the top is taken into account (thin barrier). Geometrical details like the slightly elevated pavement near the road and the presence of the body of the car are not considered. It should be noted that the ISO 9613-2 standard explicitly assumes that no ground effects occur in the presence of the barrier, so the 2D $IL(\phi)$ of the barrier along the source-receiver section reduces to the diffraction over the barrier edge minus the ground reflection of the case without barrier.

The Harmonoise model is a more advanced engineering model than ISO 9613-2. More geometrical details can be included as e.g. the 0.15m elevated pavement. The Delany and Bazley impedance model has been applied, characterizing materials with a single parameter, namely the (effective) flow resistivity. Based on tabulated values [12], a value of 20000kNs/m⁴ was taken for the road asphalt and the pavement, while for the 0.4m thick vegetated barrier, consisting of substrate for plants, the effective flow resistivity was estimated at 200kNs/m⁴.

These engineering models both assume point sources.

A full-wave method is included as well, namely the finite-difference time-domain method, FDTD. The FDTD model is used in a 2.5D approach, by combining different cross-sections between the source positions and the receiver. Three different degrees of geometrical detail of the current situation are considered. A first geometry disregards the presence of the pavement and the body of the car (FDTD nB/nC). A second one models the presence of the elevated pavement (FDTD B/nC), while in the third approach both the pavement and the body of the car are considered (FDTD B/C) (Fig. 3.2).

The 0.4m thick barrier is modeled in FDTD as a rigid-frame porous medium with the Zwikker and Kosten model, of which parameters were fitted on measured absorption curves of a vegetated barrier (more information can be found in [13]).

A spatial discretization step of 2.5mm was taken in order to accurately simulate high frequencies. However, this will raise the calculation time, especially at large source angles. Therefore it was decided to simulate the sound field for source angle steps of 5° and make use of interpolation to estimate the attenuation for other source angles.

A comparison of the barrier attenuation for a car positioned at $\phi = 0^\circ$ is given in Fig. 3.5 for the different sources. ISO 9613-2 has a smooth course, while other simulation methods show more fluctuations in function of frequency as a higher degree of geometrical detail is included, leading to interference effects, and more detailed models are used to describe barrier (and other material) characteristics. Fig. 3.5(a) shows $IL(0^\circ)$ for the source at 0.05m. ISO 9613-2 predicts lower IL than the other methods, with a difference of 5dB to 10dB. The Harmonoise and full-wave methods are in accordance, with a maximal deviation of 5dB. From 10kHz on, the IL calculated with FDTD by including car and pavement shows a decrease in IL. In this case, the IL is limited by reflections on the body of the car, which are not accounted for in the other simulation models. Fig. 3.5(b) depicts $IL(0^\circ)$ for the source at 0.3m. Again, ISO 9613-2 estimates a lower IL than the other simulation models, which show a very similar course. All models (except ISO 9613-2) show a dip near 1kHz, caused by the reduction of the ground effect when introducing the barrier. Furthermore, it is seen that including both the body of the car and the pavement has an important effect at high frequencies (FDTD B/C).

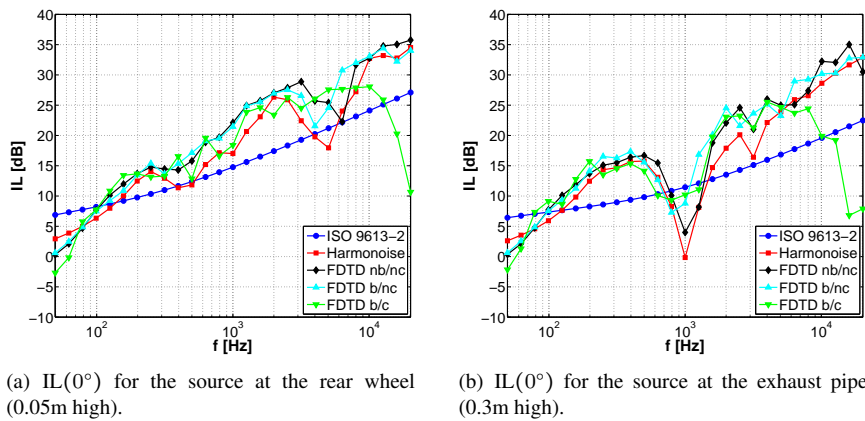


Figure 3.5: Barrier $IL(\phi)$ for a car positioned at $\phi = 0^\circ$ calculated from the energetic sum of the diffraction around the side edges and the 2D IL, estimated with different simulation techniques.

Fig. 3.6 shows the course of $IL(\phi)$ in function of angle for the different simulation methods. Predictions of $IL(\phi)$ show a smooth transition from positions beside the barrier to positions shielded by the barrier due to the inclusion of diffraction around the side edges. At low frequencies, little difference is seen between simulation models. At mid and high frequencies, differences between models increase, due to the inclusion of geometrical detail and barrier characteristics. The ISO 9613-2 model shows almost constant $IL(\phi)$ behind the barrier and predicts lowest $IL(\phi)$, with the difference between the other models increasing with increasing frequency. More variation in function of the angle is seen for the other models. Although models are generally in accordance for the source at 0.05m, somewhat higher discrepancies are seen for the source at 0.3m, caused by influence of the ground effect, the inclusion of geometrical details and differences between barrier attenuation models. Especially at 1kHz, the influence of the (absence of the) ground effect clearly results in large differences between models.

3.1.2 Validation of the auralization of the barrier effect

The quality of the different auralization methodologies can be validated by listening tests. The goal of these tests is trifold. Firstly, they will reveal whether people can distinguish between recordings behind and beside the barrier and whether this distinction can be made equally well for the auralized samples. Secondly, it is checked if the auralized samples can be distinguished from real recordings behind the barrier and thirdly, the quality of the different methodologies for simulating barrier insertion loss is compared.

3.1.2.1 Listening test set-up

In the first test, the listeners were asked for each of five car passages to indicate whether or not they had been recorded behind or beside the barrier. Two of the five samples were selected from recordings behind the barrier, two other originated from recordings beside the barrier and one sample was selected from the auralized samples. All were chosen in a random way, making sure the same car passage did not appear twice. This test was repeated four times with different samples.

In a second task, the listeners were asked to discriminate between auralized car passages (indicated as ‘simulated’) and passages recorded behind the barrier. The listener was presented with ten different car passage sounds, five of which were truly recorded behind the barrier and five auralized passages, one of each method. This test was repeated three times, with a different sample for each test.

In a third test, listeners were asked to rank five different auralizations of the same passage from one to five (one is most similar) according to their similarity with the original sample, recorded behind the barrier. As the same car passage was taken for the different auralizations, this test allows the listener to detect small

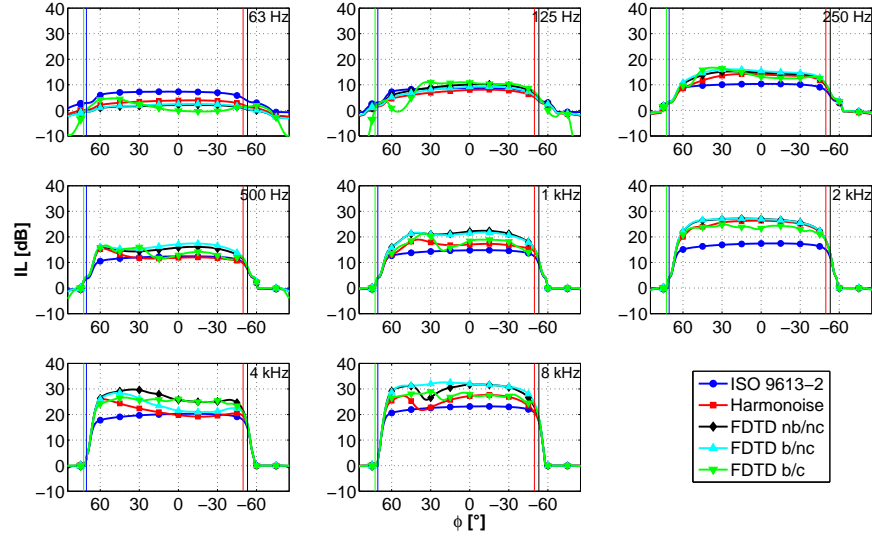
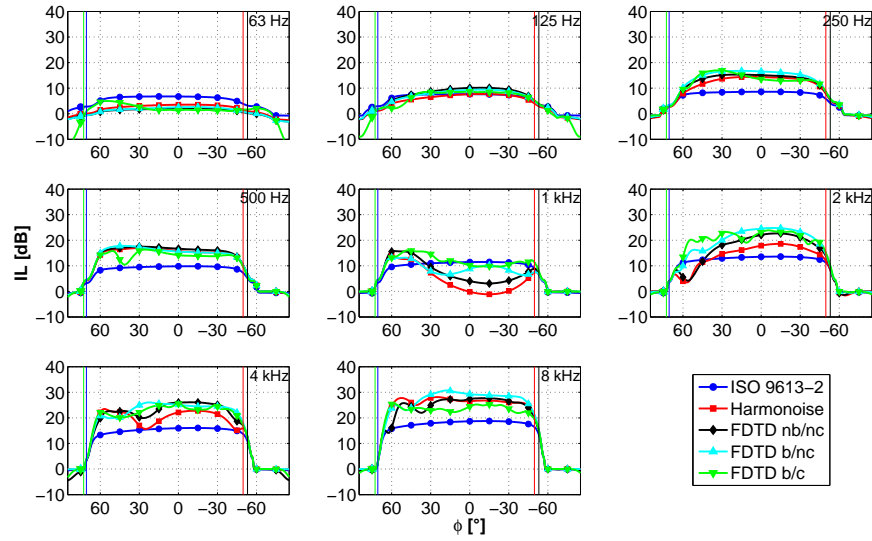
(a) $IL(\phi)$ for the source at the rear wheel (0.05m high).(b) $IL(\phi)$ for the source at the exhaust pipe (0.3m high).

Figure 3.6: Barrier $IL(\phi)$ calculated from the energetic sum of the 2D IL and diffraction around the side edges calculated with different simulation techniques for different 1/3-octave bands. The vertical lines indicate the barrier crossing.

differences between auralization methods. In this way, only the effect of the auralization methodology is evaluated, rather than differences in car spectra.

All car passages were randomly chosen from 20 samples, selected from the recordings beside the barrier. 5x20 auralizations were extracted from these recordings and the corresponding passages behind the barrier were also extracted.

3.1.2.2 Results and discussion

In total, 18 listeners took part in the test. Eight listeners were researchers of the acoustics group of Ghent University. The other listeners did not have any acoustical background. Only one female listener took part in the test, and the age ranged between 22 and 43 years old.

The results of the first test are given in Table 3.1. Listeners can clearly differentiate between car passages recorded behind or beside the barrier. In almost 80 % of the cases, the correct answer was given. As for the auralized car passage, without knowing that the passage was auralized, 90 % of the listeners categorized this auralization as being recorded behind the barrier. After the first test, listeners were able to give remarks. Here, almost all listeners indicated that their decision was mainly based on the sound level of the passages.

In the second test the ability to distinguish between recordings and auralizations was investigated. Almost 80 % of the original recordings were determined as such, while an average of 29 % of the auralized samples were identified as original recordings. Listeners commented that the auralized samples were subject to higher attenuation than the recorded samples.

Results for each individual auralization method are shown in Table 3.2. Auralizations according to ISO 9613-2 are most likely to be seen as originally recorded: 37.0 % of the samples. ISO 9613-2 is followed by Harmonoise (33.3 %), FDTD B/C (27.8 %) and FDTD B/nC and FDTD nB/nC, equally evaluated (24.1 %).

The ranking deducted from test 2 was explicitly checked in test 3, where it was asked to rank different auralizations of the same car passage according to similarity with the original recording behind the barrier. Although listeners indicated that it was very difficult to distinguish between different auralization methods, a similar ranking as in test 2 was obtained: ISO 9613-2, followed by Harmonoise, FDTD B/C, FDTD B/nC and FDTD nB/nC (Table 3.3). Auralizations based on the full-wave FDTD method got worst scores. However, it should be noted that the ranking improves when more detail (body of the car and pavement) is included in the FDTD simulation.

Table 3.1: Results of the first test. The columns indicate the responses given by the listeners, presented with original or auralized samples. The right column gives the total sample count. ($\chi^2 = 165$, $p < 0.01$)

	Indicated as behind	Indicated as beside	Count
Recorded behind	77.1 %	22.9 %	144
Recorded beside	18.8 %	81.2 %	144
Auralised behind	90.3 %	9.7 %	72

Table 3.2: Results of the second test. The columns indicate the responses given by the listeners, presented with auralized samples or samples recorded behind the barrier. The right column gives the total sample count. ($\chi^2 = 133$, $p < 0.01$)

	Indicated as recording	Indicated as simulation	Count
Recorded behind	79.3 %	20.7 %	270
ISO 9613-2	37.0 %	63.0 %	54
Harmonoise	33.3 %	66.7 %	54
FDTD nB/nC	24.1 %	75.9 %	54
FDTD B/nC	24.1 %	75.9 %	54
FDTD B/C	27.8 %	72.2 %	54

Table 3.3: Results of the third test. The columns indicate the ranking given by the listeners (1 means most similar). For each auralization method, the position with the highest occurrence is indicated in bold. The total sample count is 57. ($\chi^2 = 156$, $p < 0.01$)

	1	2	3	4	5
ISO 9613-2	66.7 %	14.0 %	3.5 %	12.3 %	3.5 %
Harmonoise	8.8 %	49.1 %	21.1 %	14.0 %	7.0 %
FDTD nB/nC	10.5 %	8.8 %	24.6 %	19.3 %	36.8 %
FDTD B/nC	1.8 %	19.3 %	17.5 %	33.3 %	28.1 %
FDTD B/C	12.3 %	8.8 %	33.3 %	21.1 %	24.6 %

3.1.3 Conclusion

In this case-study, a methodology is presented to auralize the effect of inserting a (low vegetated) noise barrier from a priori binaural recordings. To account for the time-dependent insertion loss during a car passage, the instantaneous position of the car was extracted based on a biologically inspired model. The 2D barrier IL along the source-receiver section considered was then calculated by means of ISO 9613-2, Harmonoise and FDTD models, while diffraction at the side edges was included by ISO 9613-2.

Different auralized car passages, together with original recordings beside and behind the barrier were presented to a listening panel. The difference between real recordings behind and beside the barrier was clearly heard, confirming the effect of the barrier. Almost all hidden auralized samples were labeled as ‘behind the screen’. However, comparisons between auralized and real samples indicated that listeners were still able to distinguish between them. An overestimate of the level drop when passing behind the barrier was the main reason. Still, almost 30 % of the auralized samples were recognized as real recordings. When ranking the different auralization methods, ISO 9613-2, yielding lowest attenuations, was preferred most often. The full-wave FDTD method came third.

The research carried out proves that the effect of a barrier can be auralized from a priori recordings and simulations of barrier IL. Indeed, listeners categorized the auralizations as ‘behind the screen’, when they were unaware of the fact that they were judging auralizations (test 1). Generally, it proves to be much more difficult to create auralizations that are indistinguishable in a direct comparison with real a posteriori recordings, as small differences often reveal the auralized samples. However, this does not pose a problem as the overall effect of the barrier is still present in the auralizations.

3.2 Auralization of the effect of an L-shaped mound near a highway

In the previous methodology, the effect of a barrier was included by applying time-selective attenuation on the binaurally recorded signal, based on an estimate of the momentary dominant source position. This methodology works well if only one dominant source is present at the same time, as sources other than the dominant one are equally attenuated as well. In this section, a second auralization methodology is developed for more complex sound environments. Here, the effect of the noise mitigation measure is implemented by spatial-selective attenuation of the sound sources, based on the pattern emulation technique developed in chapter 2. As such, a more complex recording technique is necessary, relying on a priori recordings with the 32-channel spherical microphone array.

The use of this auralization methodology is illustrated by a second case-study, where the noise reduction of a complex L-shaped earth mound close to the E17 highway near Kruishoutem (Belgium) is investigated. Among others, the perceived quality of this auralization mainly depends on the accuracy of the simulated IL. Therefore, in a first step, the noise reduction of the mound will be expressed quantitatively, based on simulations of the insertion loss with ISO 9613-2, the Pierce barrier diffraction model and Harmonoise model. The accuracy of the proposed simulation techniques is put to the test by comparing simulated ILs with the measured IL, based on measurements before and after installation of the mound. In a second step, the effect of the mound on the sound environment is auralized, based on the calculated IL and array recordings taken before the mound was installed. The quality of the auralization methodology is evaluated with a set of listening tests, in which auralized samples are opposed to real-life recordings after the mound is erected.

3.2.1 Description of the site

The measurement site is located at a street parallel with the E17 highway in Belgium, near the city of Kruishoutem. The L-shaped mound with dimensions 84.5m x 65.3m has an average height of 6m, while the highway is located on an embankment of approx. 1.7m high. The distance between the center of the first lane of the highway and the top center of the mound is estimated at 27m. Directly behind the mound is an open workshop, constructed with lath partitioning walls. The geometry of the measurement site is depicted in Fig. 3.7.

Four different positions of interest were chosen at the site. MIC1 is chosen just in front of the mound, serving as a reference position to be able to detect changes in the source level and spectra, in presence and absence of the mound. Two positions were selected behind the mound, MIC2 and MIC3, to measure the effect of the mound. MIC4 was chosen beside the mound, located at the same distance as MIC2 relative to the highway.

3.2.2 Developing a point-to-point acoustical model

Given the large dimensions of the measurement site, 3D full-wave simulation methods are not suitable for the calculation of the IL due to their high calculation cost. Consequently, the effect of the barrier on the soundfield will be modeled with so-called engineering models only.

When modeling the propagation from an omnidirectional sound source to a receiver, different effects need to be taken into account. In the case without mound, the sound emitted from the source is subject to propagation effects, such as spherical divergence, air absorption and ground effects due to interference between direct sound and sound reflected on the ground surface, and excess attenuation due

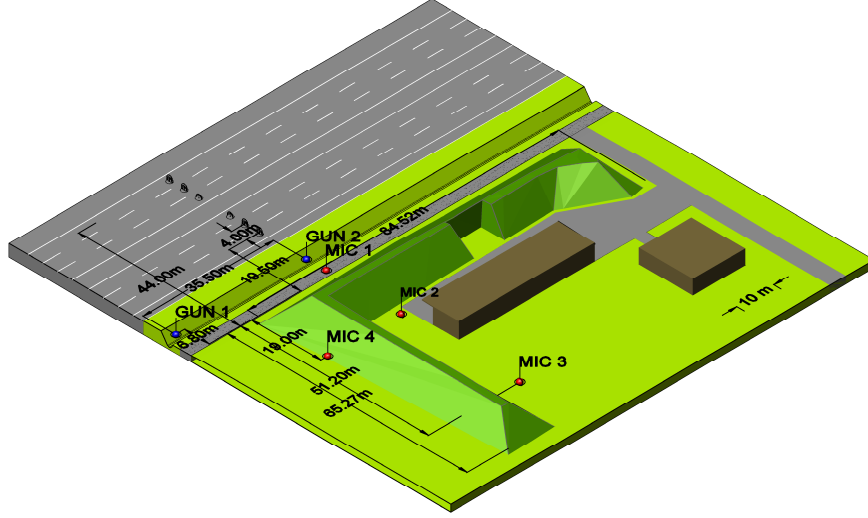


Figure 3.7: Geometry of the mound near the E17 highway. Locations of the measurement microphones are indicated in red. Locations of the shootings are indicated in blue.

to vegetation present at the measurement site. After the installation of the mound shielding should also be accounted for. This attenuation is however limited by edge diffraction and scattering on atmospheric turbulences.

Eqns. 3.1 and 3.2 summarize the different terms that should be taken into account when modeling the SPL at the receiver, before and after installation of the mound:

$$L_{p,NM} = L_w - A_{div} - A_{air} - A_{ground} - A_{vegetation} \quad (3.1)$$

$$L_{p,M} = L_w - A_{div} - A_{air} - (A_{ground} + A_{mound}) \oplus A_{scat}. \quad (3.2)$$

Here, L_w is the source power and \oplus denotes the energetic summation of the attenuation by scattering and sound attenuation by the ground and mound: $A_1 \oplus A_2 = -10 \log_{10} (10^{-A_1/10} + 10^{-A_2/10})$. In what follows, an expression for the different attenuation terms will be given based on literature.

Geometrical divergence. The geometrical divergence accounts for the spherical spreading of an omnidirectional point source in free field, related to the source-receiver travel distance. Calculation of this attenuation is based on the ISO 9613-2 standard [8], following the inverse-square law, and is taken the same for all modeling techniques discussed here.

Air attenuation. When sound propagates through air, absorption due to molecular relaxation, friction and thermal effects will occur. The amount of air attenuation

is mainly dependent on air temperature, humidity and, to a limited extent, ambient atmospheric pressure. The air attenuation of a pure tone increases linearly with the source-receiver distance, with a proportionality factor that can be calculated following ISO 9613-1, based on the aforementioned meteorological conditions and sound frequency [14]. However, as the model for IL is intended for broadband noise sources, a correction should be made. Based on [15] the air attenuation for broadband sound can be expressed as

$$A_{air} = A_0(1.0053255 - 0.00122622A_0)^{1.6} \quad [dB], \quad (3.3)$$

with A_0 the pure tone air attenuation calculated for each of the 1/3-octave center frequencies.

Ground attenuation. This term accounts for the interference between the direct sound and reflections on the ground surface.

Two models will be used here. The simplest model is based on an extended version of the calculation of A_{ground} as defined in ISO 9613-2. In ISO 9613-2 it is assumed that the ground attenuation is mainly determined by ground surfaces near the source and receiver. Therefore, three different regions are identified: a source region with size 30 times the source height h_S , receiver region (size 30 times the receiver height h_R) and, if the site is large enough, a middle region in between source and receiver regions. Simple expressions for an average attenuation in each of the zones are derived as a function of source/receiver height and projected source-receiver distance. Acoustical properties of the ground surface are included by means of a normalized ground factor G . $G = 0$ means hard ground, while $G = 1$ corresponds to a porous ground.

The use of ISO 9613-2 is however limited to flat surfaces. To overcome this, an extended version of ISO 9613-2 developed by Cho et al. [16], inspired by the French method for road traffic noise prediction (NMPB) [17], will be used for non-flat ground profiles. Here, the ground effect is calculated with ISO 9613-2 with substitution of the source and receiver height by an equivalent source $(h_S)_{eq}$, resp. receiver $(h_R)_{eq}$ height. Cho defines the mean ground plane as the average of the course of the ground profile, relative to a reference plane connecting the projection points of source and receiver on the local ground surface. Finally, $(h_S)_{eq}$ and $(h_R)_{eq}$ are found as the distance of source and receiver to this average plane.

When modeling IL according to the ISO 9613-2 standard, A_{ground} needs to be calculated only for the case without the mound, as ISO 9613-2 assumes that the effect of a ground reflection vanishes in the presence of the mound. However, when IL is modeled based on the Pierce diffraction model, we will include the ground effect, based on ISO 9613-2, on both sides of the mound: from source to first diffraction edge and from second diffraction edge to receiver, as was suggested by Jonasson [18].

In the Harmonoise simulation model [9, 10], a more elaborate calculation scheme is used to take non-flatness and varying ground characteristics explicitly into account. Basically, the influence of a ground reflection on a homogeneous flat ground is modeled by using the image source method with inclusion of a complex reflection coefficient to model ground characteristics, with an impedance based on the effective flow resistivity of the ground [19]. A coherence factor is introduced to account for fluctuations of the phase relation between direct and reflected sound due to uncertainties in source/receiver height, propagation distance and turbulence [9]. In order to deal with changing ground characteristics and non-flat surfaces, Harmonoise incorporates the Fresnel weighting technique [20]. Here, the relative contribution of a given surface element on the soundfield is determined by the intersection of the Fresnel ellipsoid, constructed between image source and receiver, and the surface element.

Mound attenuation. Three different engineering models are used to estimate the mound attenuation. In the following, it is assumed that the mound has a large density, so sound transmission can be neglected. In that case, sound from the source can only reach the receiver by diffraction around the barrier (top) edges (apart from a contribution originating from scattering in the atmosphere - see further) and calculating A_{mound} reduces to modeling the diffraction by a three-sided wedge.

In ISO 9613-2 a basic engineering method is proposed to estimate the double diffraction D_z around the two top edges of the mound. The mound is represented by two thin screens, positioned at each of the mound's upper diffraction edges, thus omitting the effect of the side slopes on the soundfield. Diffraction around the screens is expressed as a function of wavelength, source/receiver distance to the closest diffraction edge, the distance between the diffraction edges and the path length difference, defined as the difference between the shortest non-penetrating path in the 2D cross-section, determined by the vertical plane through the source and receiver, and the (virtual) direct path. When the distance between both diffraction edges becomes small relative to the wavelength, the expression for single-edge diffraction is used.

A more elaborated model for mound attenuation in absence of a ground plane is developed by Allan Pierce [21]. Based on the exact solutions derived by Sommerfeld [22], Pierce extracted expressions for the diffraction around a rigid wedge and - by extension - a three-sided semi-infinite barrier in terms of auxiliary Fresnel functions. The result is a function of the wavelength, path length difference, barrier slope angles, barrier top width and relative source/receiver position. However, in order for the approximation to be valid, the distance of the source/receiver to the closest diffraction edge (r_S and r_R) and the distance between both diffraction edges (r_W) need to satisfy $r_{S/R/w} \gg \frac{c}{2\pi f}$. For the given geometry, the restriction

on r_S and r_R are fairly met. In a review paper by Kurze [23], the condition on barrier width r_w was practically rephrased as $r_w > c/f$. For the given geometry, where the minimal barrier width r_w equals 1.7m for the section perpendicular to the barrier, the model will be applicable for frequencies above 200Hz. For lower frequencies, the single-edge diffraction model will be used.

The Harmonoise model [9, 10] is a more complex model, allowing for the inclusion of the full geometry of the measurement site. Furthermore, the acoustical impedances of the mound and various ground surfaces can be explicitly taken into account in terms of effective flow resistivities [19]. The calculation of the barrier influence is based on a recursive algorithm. A barrier is detected when diffraction edges are found lying above the source-receiver line. The most diffracting edge is then selected as the edge with the largest path length difference, and the excess attenuation is calculated as the sum of the source-edge-receiver diffraction. This procedure is then repeated recursively, calculating contributions at each side of the diffracting edge (with a new source/receiver placed on the previously detected edge). The calculation model for the diffraction around the edge is based on a numerical approximation of the diffraction integrals proposed by Deygout in function of the Fresnel number $N_F = \frac{2\delta}{\lambda}$, with δ the path length difference [9, 24]. In this work, the Harmonoise methodology will be implemented by use of the readily available point-to-point simulation library (Harmonoise Point-to-Point Propagation Model, v2.016, Dirk Van Maercke, CSTB). The result of this simulation engine provides the combined effect of ground A_{ground} and barrier attenuation A_{mound} .

Attenuation by vegetation. At the time of the reference measurements, a dense hedge parallel with the E17 highway blocked the line of sight from cars and trucks. Behind the hedge, small bushes and trees were planted at the location where the future mound was planned. A somewhat less dense hedge was located at the left side of the measurement locations (Fig. 3.8).

It has been shown that the presence of (dense) vegetation has an attenuating effect [25, 26]. In this work, the excess attenuation introduced by the vegetation belt will be estimated by using a semi-empirical model developed by Aylor [27, 28]. In his model, Aylor states that the excess attenuation is mainly caused by multiple scattering within the vegetation. By measuring the attenuation of corn fields with different densities, he found that excess attenuation increased with increasing frequency and Leaf Area Density (LAD [m^2/m^3]) [27]. Furthermore, Aylor was able to eliminate the influence of the ground effect by comparing measurements of sound attenuation before and after the removal of dense reeds in shallow water. From his measurements, an empirical model can be derived, solely quantifying the vegetational attenuation as a function of frequency, LAD, leaf width l_w and propa-



(a) Laurel hedge (and bushes behind the hedge) along the road parallel with the E17 highway.



(b) Laurel hedge and bushes at the left side of the future mound.

Figure 3.8: Vegetation at the measurement site before the installation of the mound.

gation length L through the vegetation belt. This approach successfully predicted the noise shielding provided by thick hedges [29].

Turbulent scattering. Besides direct propagation, ground reflections and diffraction at barrier edges, sound can reach the receiver by scattering on turbulences in the atmospheric surface layer. In the case without barrier, the direct sound and ground reflections contain much more energy and the turbulent scattering can be neglected. However, in the deep shadow zone of the mound, sound pressure levels are much smaller, relative to free field, and the contribution of turbulent scattering becomes more important, especially at high frequencies.

Recently, Forssén et al. [30] developed an engineering model to estimate the contribution of turbulent scattering, based on the isotropic von Kármán turbulence model. The model is based on C_v and C_T , the structure parameters of velocity and temperature fluctuations, describing the strength of the turbulences. Strong to very strong atmospheric turbulences typically result in values of $C_v^2 = 1.2 \dots 12 \text{ m}^4/\text{s}^2$ and $C_T^2 = 0.4 \dots 4 \text{ K}^2/\text{m}^{2/3}$ [30]. Other parameters of importance are the frequency and variables related to the geometry of the site.

Forssén suggested to limit the amount of scattered energy so that the scattered plus diffracted energy at the receiver does not exceed the energy measured at the receiver without barrier present.

This model will be used to predict the effect of turbulent scattering in all three IL prediction methods, and the turbulent scattering model included in the Harmonoise-point-to-point model will be disabled.

Influence of wind and temperature gradient. In this work, the influence of wind and temperature gradient will be omitted, except for the ISO 9613-2 model,

where downwind propagation conditions are inherently assumed. It is expected that these effects have only limited influence over the short range distance, i.e. 38.5m between the closest source and receiver [31]. Furthermore, calculations by Van Renterghem et al. showed that limited to no effect of the wind was seen on the IL of earth mounds (with flat top surface) in case of downwind sound propagation [32].

3.2.2.1 Determination of model parameters

In the previous section, three different simulation techniques of different complexity have been proposed to calculate the barrier IL. It is clear that the accuracy of the simulations is not only dependent on the simulation technique, but also on the correspondence between the architectural model and the as-built situation, the estimate of the surface characteristics and the measurement conditions.

Here, the architectural plans of the mound were readily used as input, as during the construction of the mound, the building plans have been carefully followed by GPS-controlled diggers. Information on the surroundings such as the ground level variations, the location of the highway and the small ditch was however not readily available, and therefore, an estimate had to be made.

Apart from the geometrical input, a different description of surface characteristics, depending on the used simulation model, is necessary. In this case-study, the parameter G and effective flow resistivities will be estimated from a parameter optimization so that the modeled IL, calculated with ISO 9613-2 resp. the Harmonoise point-to-point engine, matches the in-situ measured IL.

For the measurement of the barrier IL, a set of gunshot measurements before and after construction of the mound were conducted. Several gunshots were fired with a Bruni 96 caliber at GUN1 and GUN2 (Fig. 3.7), and the response was recorded at MIC1, just before the (future) mound, and MIC2, located behind the (future) mound by a Swing measurement unit (Sinus Messtechnik GmbH) with calibrated MK-250 free-field microphones and MV-210 pre-amps (Microtech Gefell GmbH). The gunshot measurements before the construction of the mound have been performed on June 5th 2012 around 9:00 PM ($T = 12.9^{\circ}\text{C}$, humidity = 70.8%, $v_{wind} = 2.0\text{m/s}$, wind direction = 130°). Seven gunshots have been fired from GUN1 and six from GUN2 in order to calculate an average spectrum. It was checked that the level of the gunshots is above the noise floor for frequencies higher than 150Hz. After construction of the mound, another series of gunshots have been fired. A first series of gunshots were taken on October 16th 2012 around 8:00 PM (four shots at GUN1/2, $T = 9.4^{\circ}\text{C}$, humidity = 80.9%, $v_{wind} = 2.8\text{m/s}$, wind direction = 186°), but as the SNR at 1-2kHz was only 3dB, a second series of gunshots were taken on January 11th 2013 around 9:00 PM (six shots at GUN1/2, $T = -0.5^{\circ}\text{C}$, humidity = 81.6%, $v_{wind} = 1\text{m/s}$, wind direction = 125°). During the second series, much

better SNRs were observed (at least 6dB from 200Hz on). The difference in meteo conditions between both measurement sets leads to a difference in A_{air} of maximal 0.1dB/m at 20kHz, which yields a maximal total difference of 2.5dB at 20kHz for a path length of 25m from GUN2 to MIC2.

From the shooting at GUN2, the 2D IL of the mound along the section GUN2-MIC2, perpendicular to the highway, can be extracted by subtracting the average gunshot levels at MIC2 before and after installation of the mound. The IL simulated with Harmonoise is then optimized with respect to the ground characteristics (effective flow resistivity) and properties of the vegetation. Hereby, an estimate for the turbulence strength parameters C_v^2 and C_T^2 is made based on the meteorological conditions [33, 34], yielding $C_v^2 = 1.0095\text{m}^{4/3}/\text{s}^2$ and $C_T^2 = 0\text{K}^2/\text{m}^{2/3}$, corresponding to (moderately) strong velocity turbulences. This is in agreement with the findings of Van Renterghem et al. [35], indicating that rather high values of turbulence strength are needed to account for their dominant influence on long-term equivalent sound pressure levels.

The result of the optimization yields an effective flow resistivity of 200kNs/m^4 for the ground surface near the receiver, 80kNs/m^4 for the ground underneath the future mound, 20000kNs/m^4 for the mound (based on a visual inspection of the hard tamped surface), 500kNs/m^4 for the ground between the mound and the trackway and 20000kNs/m^4 for the trackway in between the mound and the highway. The flow resistivities of the soil between the trackway and the highway and the highway itself could not be determined by this method (as it would require gunshots to be fired on the highway instead of on the trackway), and are estimated as 2000kNs/m^4 for the wet soil (small ditch) and 20000kNs/m^4 for the highway.

For the ISO 9613-2 simulation model, G is estimated as $G = 0.1$ near the source (hard ground) and $G = 0.6$ (medium porosity) for ground surfaces in the middle region and at the receiver.

Optimization of $A_{vegetation}$ yields an effective LAD of 4m^{-1} and leaf width of 10cm for the laurel hedge. Although the optimized value is higher than the actual measured LAD for a laurel hedge, fluctuating between 2m^{-1} and 0.8m^{-1} for a dense resp. less dense hedge [36], the value is similar to the optimized value of LAD (4.5m^{-1}) found by Van Renterghem et al. [29], accounting for the woody biomass as well. The propagation length L has been taken as the width of the mound (ca. 4m).

Fig. 3.9 shows the IL, calculated with the different simulation models, together with the measured IL, calculated by subtracting average gunshot levels before and after installation of the mound. Two versions of the measured IL are given, as two sets of gunshot measurements were performed after installation of the mound. It can be seen that the ILs of both measurements are in agreement at low and mid

frequencies, but above 2.5kHz the difference between measurements increases. At these frequencies, the second version of the IL yields a lower value than the initial version. This dissimilarity is to be explained by a difference in wind gradient or differences in the contribution from turbulent scattering, caused by changes in meteorological conditions, showing the strong variations of turbulent scattering over time and their effect on momentary measurements.

Fig. 3.9 shows that the general course of the Pierce simulation method agrees best with the second version of the measured IL. However, the Pierce model, as the ISO 9613-2 model, scores less on modeling the energy dip (ground effect) at 500Hz, whereas the Harmonoise model performs best in modeling spectral fluctuations. Maximal deviation is below 5dB at low and mid frequencies. Furthermore, it is seen that the IL calculated with the ISO 9613-2 model slightly overestimates measured IL at mid frequencies.

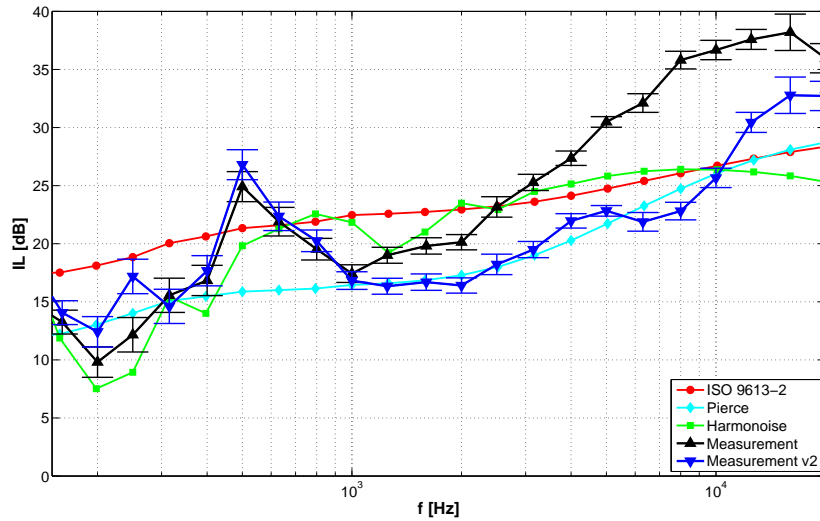


Figure 3.9: Simulation of the mound IL for a 2D section along the axis of GUN2 - MIC2. Different simulation models are compared to the measured IL. Error bars indicate the standard deviation from the mean IL.

The above methodology to estimate surface properties and model input parameters based on a priori measurements, optimizes parameters for the GUN2-MIC2 2D section, perpendicular to the direction of the highway. However, if we want to simulate the IL of the complete mound, many other sections also need to be taken into account. In order to justify the use of these optimized parameter values for all such sections, another gunshot position is chosen, GUN1, and the IL from GUN1 to MIC2 is calculated based on shootings before and after construction of

the mound. A comparison with simulated IL, based on previously derived parameters, is given in Fig. 3.10. Here, the propagation length L through the vegetation is changed to 1m, as the section GUN1-MIC2 only crosses the hedge at the side, and not the vegetation belt in front of the receiver.

Again, the general course of IL is best approximated by the Pierce model, with a limited underestimation at the highest frequencies. The Harmonoise model succeeds best in modeling spectral fluctuations, but the difference with the measurements increases at mid frequencies (maximal 10dB at 1.25kHz). The IL calculated with ISO 9613-2 deviates the most from the measured IL: at low and mid frequency bands the difference between simulated IL and measured IL can reach values up to 15dB.

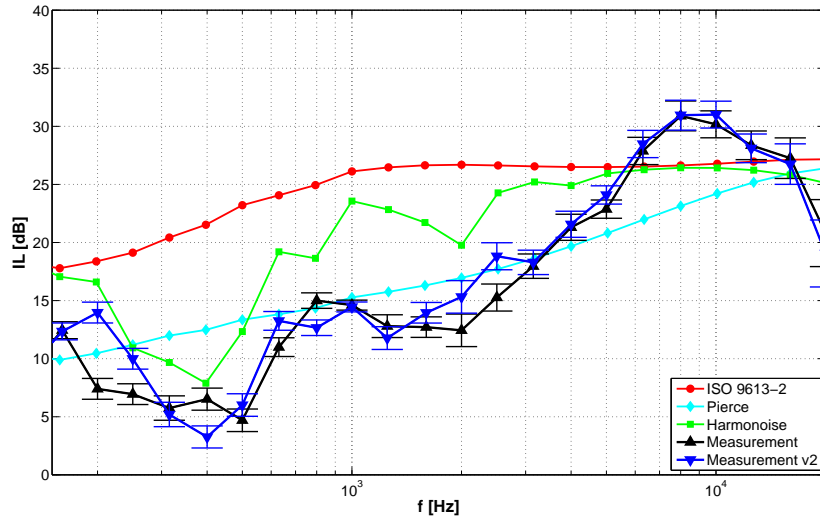


Figure 3.10: Simulation of the mound IL for a 2D section along the axis of GUN1 - MIC2. Different simulation models are compared to the measured IL. Error bars indicate the standard deviation from the mean IL.

Foregoing results show the validity of the estimated surface characteristics, atmospheric turbulence properties and characteristics of the vegetation. Especially for the Harmonoise model, a trial-and-error optimization of the effective flow resistivity is shown to yield good results regarding spectral accuracy. In the ISO 9613-2 and Pierce model only the ground parameter G near source and receiver can be changed and spectral fluctuations in the IL are less accounted for. Generally, good results can be achieved with the Pierce model, while the IL predicted by the ISO 9613-2 model suffers from a significant overestimate.

3.2.3 Modeling the highway IL of the mound

In the previous section, the performance of different acoustical models to simulate the sound propagation from a single omnidirectional source (gunshot) to a receiver behind a mound has been evaluated. In this section, the different models will be applied to simulate the $L_{eq,1h}$ at MIC2, induced by cars and trucks passing on different lanes on the highway, before and after installation of the mound. The corresponding simulated IL will then be compared with the measured IL of the mound in response to traffic noise, calculated from the average total levels $L_{eq,NM}$ and $L_{eq,M}$, measured before resp. after installation of the mound.

In what follows, the angular levels $L_{eq,NM}(\phi)$, $L_{eq,M}(\phi)$ and derived angular IL(ϕ), are defined as the levels calculated for a 2D receiver-highway intersection, with ϕ the angle between the intersection line and the normal to the highway.

3.2.3.1 Line source model

Based on the pass-by of a single source, Salomons formulated an expression for the $L_{eq,1h}$ of a constant flow of sources along a single line (incoherent line source), measured at a sufficiently large distance d from the line source [37]:

$$L_{eq,1h} = 10 \log_{10} \left(\sum_j 10^{(L_w(\phi_j) - 10 \log_{10}(4\pi d) - A_{ex}(\phi_j))/10} \delta\phi_j \right) + 10 \log_{10} \left(\frac{Q}{v} \right). \quad (3.4)$$

Here, the source model is based on the segmentation of the line source into different angular sectors, with size $\delta\phi_j$ (in radians), each represented by a single point source at the center (cf. Fig. 3.11). The total $L_{eq,1h}$ is then calculated by adding the contribution of each segment. In Eq. 3.4, $L_w(\phi_j)$ represents the sound power level of the point source (see further). $A_{ex}(\phi_j)$ is the excess attenuation, calculated for a 2D section with angle ϕ_j as the difference between the level at the receiver in free field and the level at the receiver in the real situation. A_{ex} includes the air absorption, ground interaction and attenuation by vegetation resp. mound attenuation and turbulent scattering in the case without resp. with mound. The last term in Eq. 3.4 includes properties of the traffic flow: Q is the average number of sources passing per hour and v is the speed in m/h.

In [37] an algorithm was proposed to minimize the discretization error in the calculation of $L_{eq,1h}$ by optimization of the angular segmentation $\delta\phi_j$. However, if the distance between receiver and line source is large enough, the angular segmentation can be kept constant. This constant angular segmentation results in a fine segmentation for sources close to the receiver, which should be modeled with more care as they contribute most. For large distances (at large angles) relative source power contribution is smaller and the length of the segments and, as a consequence, integrated source power is proportionally increased. In the following simulations, an angular sector of 2° will be used.

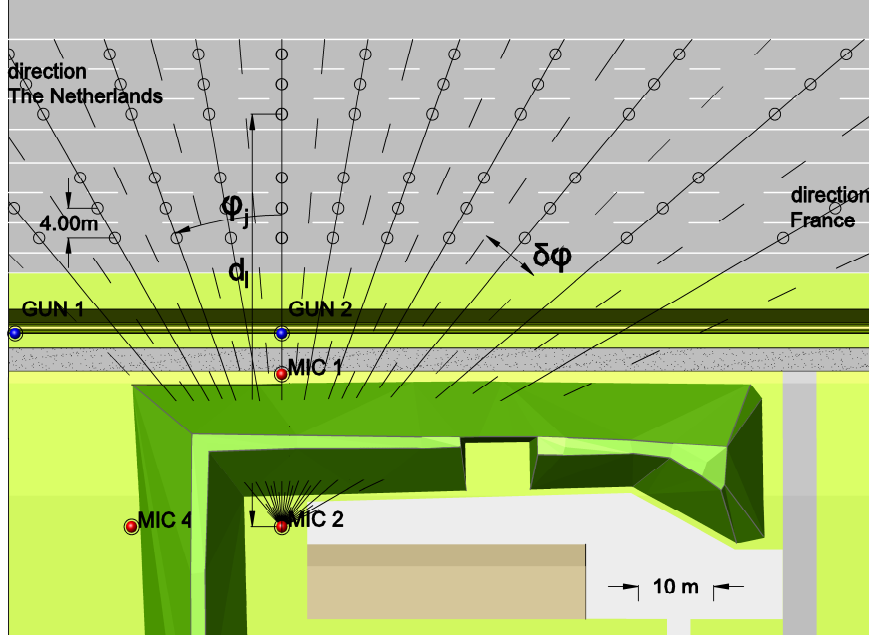


Figure 3.11: Top view of the mound near the E17 highway. The dashed lines indicate the segmentation boundaries of the different incoherent line sources, based on a constant angular sector $\delta\phi$. The equivalent point source is illustrated by the circle at the center of each segment. $L_{eq,1h}(\phi_j)$ is calculated along each 2D source-receiver intersection, with angle ϕ_j relative to the normal to the highway. For illustrative purposes $\delta\phi = 5^\circ$, for the simulations $\delta\phi = 2^\circ$ is used.

3.2.3.2 Vehicle source description

In the Harmonoise/Imagine source model each vehicle category (light vehicles, medium and heavy vehicles) is modeled as a combination of two sources at different heights [38]. For light vehicles, sources are located 0.01m and 0.3m above the ground, while for other categories sources are located at 0.01m and 0.75m. The heights are chosen corresponding with the approximate location of the tire and propulsion noise source. In reality, however, the location for tire and propulsion noise generation could deviate. The model prescribes that the tire noise $L_{w,R}(f)$ is distributed 80% resp. 20% between lowest resp. highest source, while the propulsion noise $L_{w,P}(f)$ is distributed 80% resp. 20% between highest resp. lowest source.

An expression for $L_{w,R}(f)$ and $L_{w,P}(f)$ in function of speed is given in [38] for 1/3-octave bands between 25Hz and 10kHz, at a reference temperature of 20°C and a dense asphalt concrete road surface. We will use these models for light vehi-

cles, with speed $v = 120\text{km/h}$, and vehicles of category 3, with speed $v = 90\text{km/h}$. For lane 1-3 (closest lanes) a correction is made regarding the road surface (chemically washed concrete), while for lane 4-6 the reference conditions are kept, as the road surface was unknown. Although not relevant for traffic noise, an extrapolation is made to calculate noise levels (and IL) up to the 20kHz 1/3-octave band to avoid a relative increase of the high frequency content of the recordings in the auralization process.

So far, omnidirectional sources have been assumed. In order to account for the screening of the body of the car, the horn effect (influencing the lowest source) and the screening effect of the engine compartment (influencing the highest source), a frequency dependent directivity $\Delta L_{h,cat}(\phi_j)$ is assigned to each of the different sources. It should be noted here that the calculated directivity, as defined in the Harmonoise model, has been rewritten as a function ϕ_j , the angle between the source-receiver and the normal to the highway (cf. Fig. 3.11).

Depending on the source height and vehicle category, the sound power level $L_w(\phi_j)$ of each of the sources can be expressed as

$$L_{w,h,cat}(\phi_j) = 10 \log_{10} \left(< 0.8 | 0.2 >_{h=0.01|0.3,0.75} 10^{L_{w,R,cat}/10} \right. \\ \left. + < 0.2 | 0.8 >_{h=0.01|0.3,0.75} 10^{L_{w,P,cat}/10} \right) + \Delta L_{h,cat}(\phi_j), \quad (3.5)$$

where h denotes the height of the source and cat defines the vehicle category, in this case either light vehicles (cat. 1) or heavy vehicles (cat. 3) [38]. Here, ϕ_j equals $\pm j\delta\phi$, the angle between the source-receiver and the normal to the highway (cf. Fig. 3.11).

3.2.3.3 Calculation of $L_{eq,NM|M}$ and IL

For the calculation of the noise immission at a receiver (either with or without mound), the contribution of different types of vehicles on a multitude of lanes (six in this case) has to be taken into account. In this work we are not only interested in the total sound immission at a given receiver point (total $L_{eq,NM}$ or $L_{eq,M}$), but also in the angular levels originating from a given direction ϕ_j . Therefore, the sound energy of each source from the same angular sector ϕ_j (cf. Fig. 3.11), corresponding to different lanes and different vehicle categories, is summed:

$$L_{eq,1h}(\phi_j) = 10 \log_{10} \left(\sum_{l=1}^6 \sum_{cat=1,3} \sum_{h=0.01,0.03|0.75} \frac{Q_{l,cat}}{v_{cat}} 10^{(L_{w,cat,h}(\phi_j))/10} \right. \\ \left. \times 10^{(-10 \log_{10}(4\pi d_l) - A_{ex,l,h}(\phi_j))/10} \delta\phi \right). \quad (3.6)$$

Eq. 3.6 incorporates a summation over 20 sources: two sources representing cat. 1 vehicles on each of the six lanes and two sources representing cat. 3 vehicles on

Table 3.4: Yearly averaged number of vehicles per hour, $Q_{l,cat}$, passing over the three different lanes in each direction. The light vehicles are distributed according to 20% (left lane), 60% (middle lane), 20% (right lane), while the heavy traffic is distributed as 85%, 15% and 0% [39].

	direction France			direction The Netherlands		
	l_1	l_2	l_3	l_4	l_5	l_6
cat. 1	291.70	875.10	291.70	227.47	682.40	227.47
cat. 3	491.09	86.66	0	0	56.98	322.86

lane 1, 2, 5 and 6 (no heavy traffic is assumed on the left lanes). However, as not all sources equally contribute to the angular level, each source is weighted with the average number of vehicles per unit length ($Q_{l,cat}/v_{cat}$). From traffic counts, the average number of vehicles in category 1 and 3 in each direction of the highway was known. Based on these average values, an estimate is made for the distribution of vehicles over the three lanes in each direction (cf. Table 3.4).

With Eq. 3.6, the sound immission from the highway, originating from a given direction ϕ_j , can be calculated in the case without and with mound ($L_{eq,NM}(\phi_j)$ resp. $L_{eq,M}(\phi_j)$). The angular IL ($IL(\phi_j)$) is defined as the difference between both.

In a similar way, the total IL is based on the simulation of the total level before and after installation of the mound. These levels can be readily extracted from the angular levels:

$$L_{eq,NM|M} = 10 \log_{10} \left(\sum_j 10^{L_{eq,NM|M}(\phi_j)/10} \right). \quad (3.7)$$

The total IL is then defined as the difference between $L_{eq,NM}$ and $L_{eq,M}$.

In Fig. 3.12-3.14 the resulting $L_{eq,NM}(\phi)$ and $L_{eq,M}(\phi)$ are depicted together with the $IL(\phi)$, calculated with the three different simulation models for $A_{ex}(\phi)$. Generally, it can be concluded that both the ISO 9613-2 model and the Harmonoise model yield similar results regarding $L_{eq,NM}(\phi)$. Levels at low and mid frequencies fluctuate around 40dB. Starting from 1kHz, the calculated levels drop significantly, with a somewhat higher drop for the Harmonoise model. At frequencies higher than 10kHz, levels drop below 0dB for both models. It is to be expected that errors in the auralization of the effect of the mound on traffic noise will be of little consequence at such frequencies, given the low amount of energy in these high-frequency bands.

Simulations based on ISO 9613-2 depend only in a limited way on ϕ , which is in contrast with the Harmonoise model, where sound levels at mid-frequencies decrease with approx. 9dB between $\phi = 0^\circ$ and $\phi = 84^\circ$. At -45° , a discrete jump is seen in the angular level, emerging for all three models. This can be explained by the changing contribution of $A_{vegetation}$, estimated with Aylor's model, as at that angle the thickness of vegetation suddenly decreases from 4m (frontal vegetation) to 1m (hedge near the side).

When comparing $L_{eq,M}(\phi)$, a stronger difference is seen between simulation models. The highest mound attenuation is predicted by the ISO 9613-2 model, which estimates levels for $L_{eq,M}$ between 20dB and 10dB at mid frequencies (200Hz - 1kHz), where the other models predict levels between 30dB and 20dB. At low frequencies, the difference between models is less pronounced. For angles between -70° and 60° , a discrete jump of about 5dB is seen along the frequency axis for the ISO 9613-2 model at approx. 200Hz. This jump is caused by the switching between the single-edge approximation for the mound (at low frequencies) and double-edge approximation. This switch-over between single edge - double edge is much less radical when using the Pierce model, while no discrete jump at all is seen for the Harmonoise model.

Given the fact that little difference between models was seen for the $L_{eq,NM}(\phi)$, differences regarding the $IL(\phi)$ are to be explained by fluctuations in the $L_{eq,M}(\phi)$. The highest $IL(\phi)$ is seen for the ISO 9613-2 model, ranging between 10dB and 25dB at mid frequencies, with a jump at 200Hz. $IL(\phi)$ is approx. 10dB lower when calculated with the Pierce and Harmonoise model. Furthermore, it should be noted that the Harmonoise model shows more fluctuation depending on frequency as e.g. a dip of 4dB is seen around 250Hz.

Fig. 3.15 shows a comparison between the different models for the total $L_{eq,NM|M}$ and total IL . Results are in line with what should be expected from previous conclusions.

At low and mid frequencies, the total $L_{eq,NM}$ is very similar for the ISO 9613-2 and Harmonoise simulation model. At low frequencies, the difference is max. 1.5dB. At high frequencies, starting from 3kHz on, the Harmonoise model predicts somewhat lower levels and the difference between both models increases with frequency.

More dissimilarity is seen for the total $L_{eq,M}$. As was expected, the ISO 9613-2 model predicts much lower levels at low and mid frequencies than the other models (10dB difference at 200Hz). Highest values for $L_{eq,M}$ are predicted by the Harmonoise model. The Pierce model has a more limited deviation from the Harmonoise model (max. 4dB at 200Hz). At high frequencies, all three models yield comparable levels.

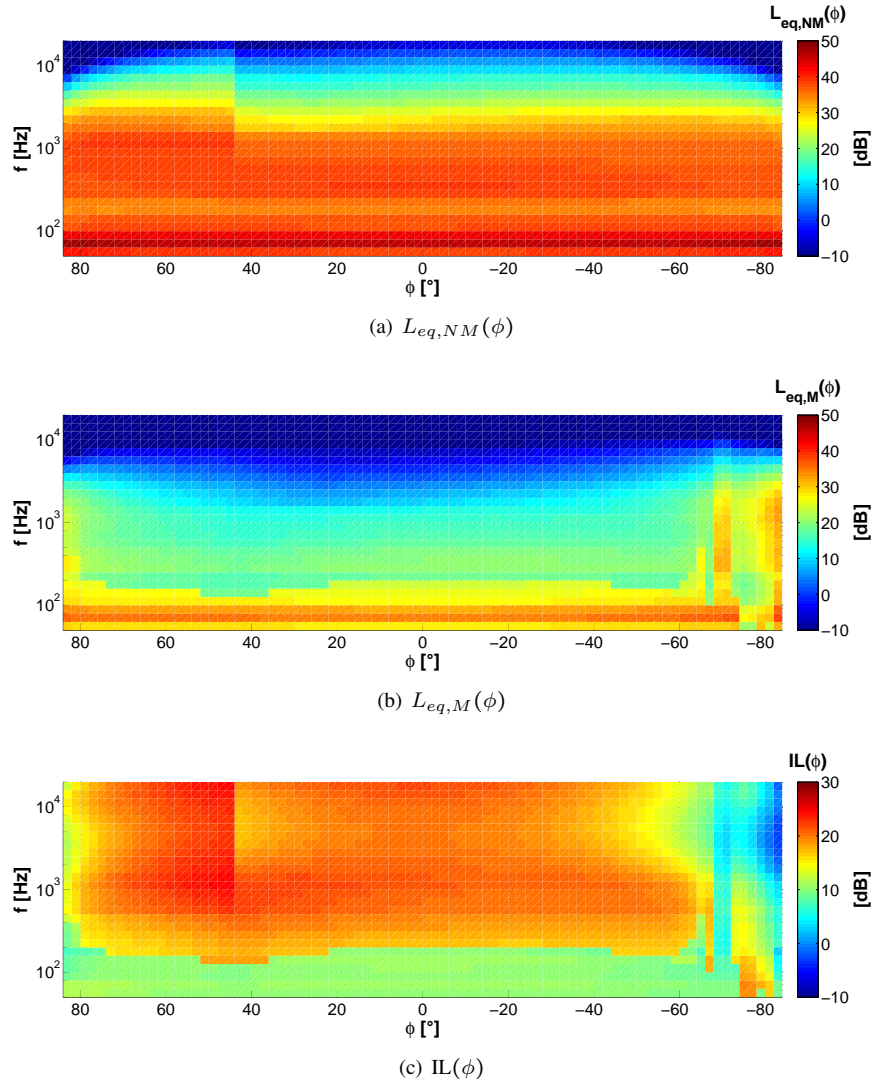


Figure 3.12: Angular level $L_{eq,NM}(\phi)$ and $L_{eq,M}(\phi)$ and resulting $IL(\phi)$ for 1/3-octave bands between 50Hz and 20kHz. ϕ ranges between -84° and 84° and a floor at -10dB has been set for the angular level. $A_{ex}(\phi)$ is calculated according to ISO 9613-2.

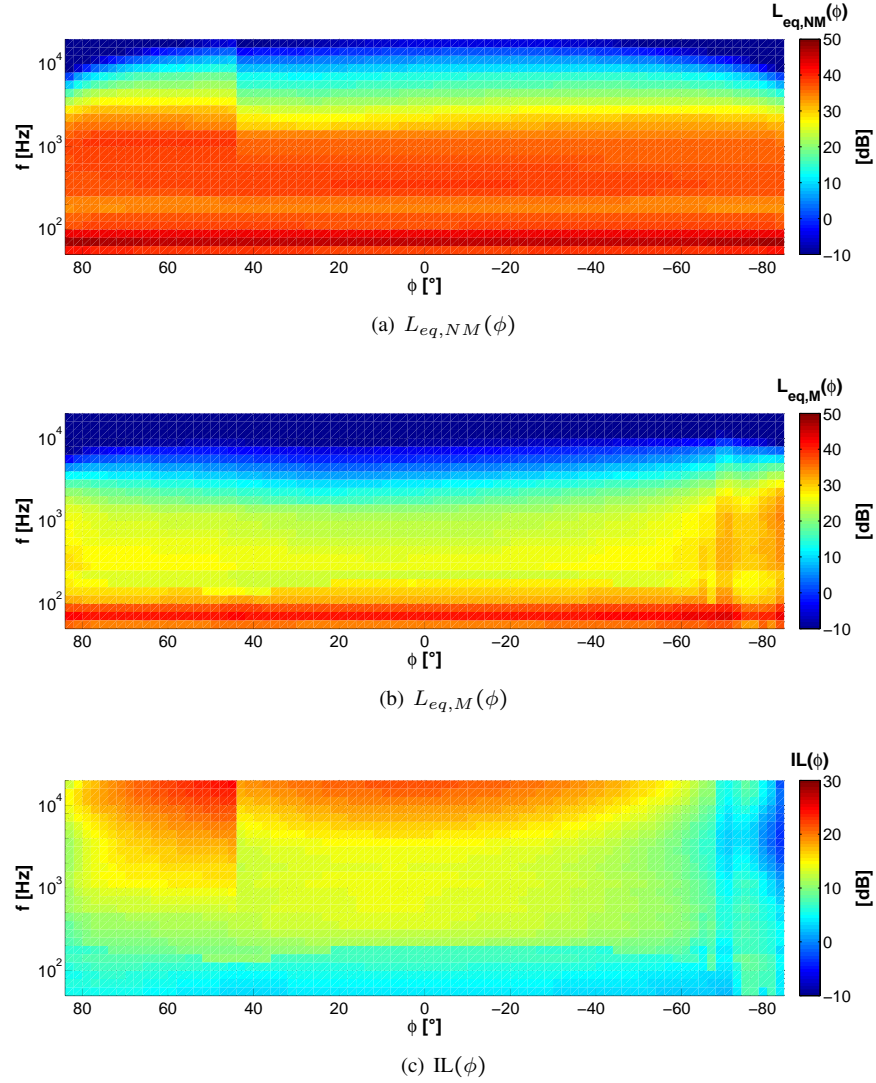


Figure 3.13: Angular level $L_{eq,NM}(\phi)$ and $L_{eq,M}(\phi)$ and resulting $IL(\phi)$ for 1/3-octave bands between 50Hz and 20kHz. ϕ ranges between -84° and 84° and a floor at -10dB has been set for the angular level. Diffraction by the mound $A_{mound}(\phi)$ is calculated according to Pierce, other terms of $A_{ex}(\phi)$ are calculated by using ISO 9613-2.

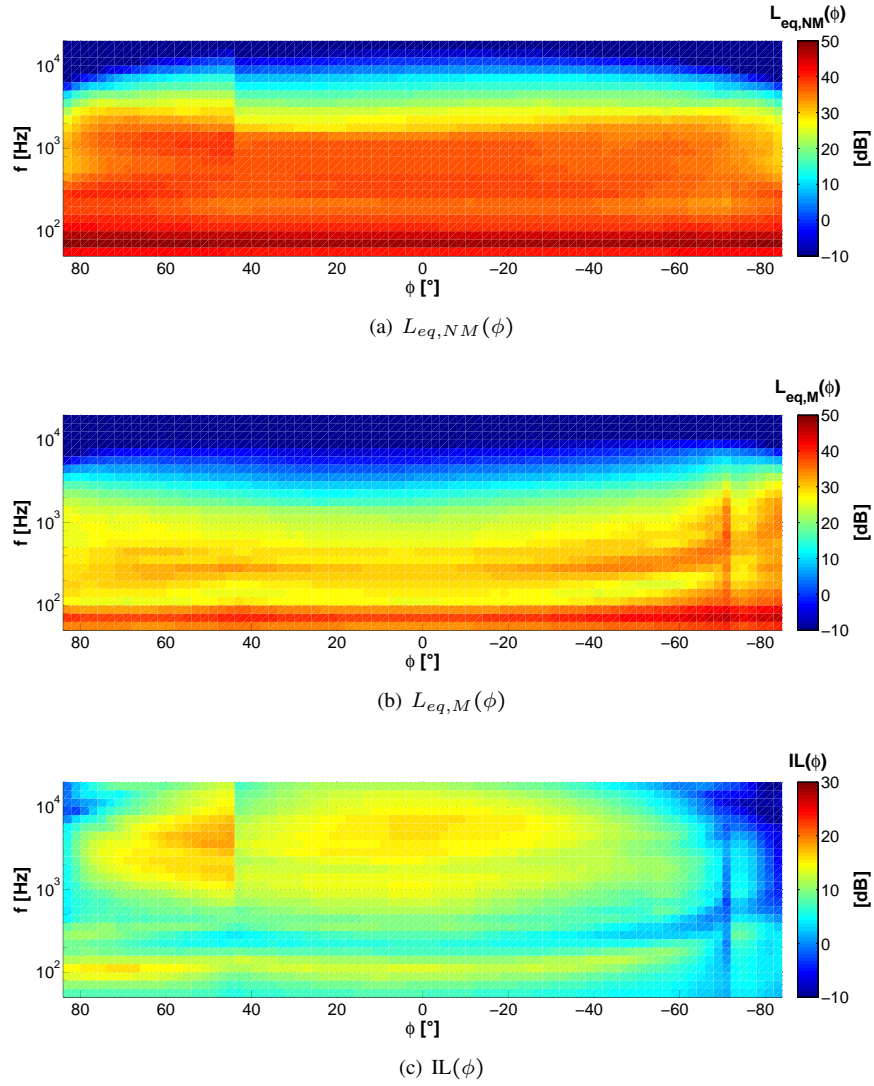


Figure 3.14: Angular level $L_{eq,NM}(\phi)$ and $L_{eq,M}(\phi)$ and resulting $IL(\phi)$ for 1/3-octave bands between 50Hz and 20kHz. ϕ ranges between -84° and 84° and a floor at -10dB has been set for the angular level. $A_{ex}(\phi)$ is based on the Harmonoise point-to-point model.

The total IL of the mound fluctuates around 10dB for the different simulation models. The ISO 9613-2 model yields the highest IL, with a maximum of 15dB at mid frequencies and reaching 20dB at high frequencies. The Harmonoise model predicts a dip for the mid frequency range (approx. 7dB), and limited IL (10dB) at high frequencies. The Pierce model calculates the lowest IL at low frequencies, but keeps moderate values at mid frequencies (10dB). At high frequencies, the course of the ISO 9613-2 model is followed.

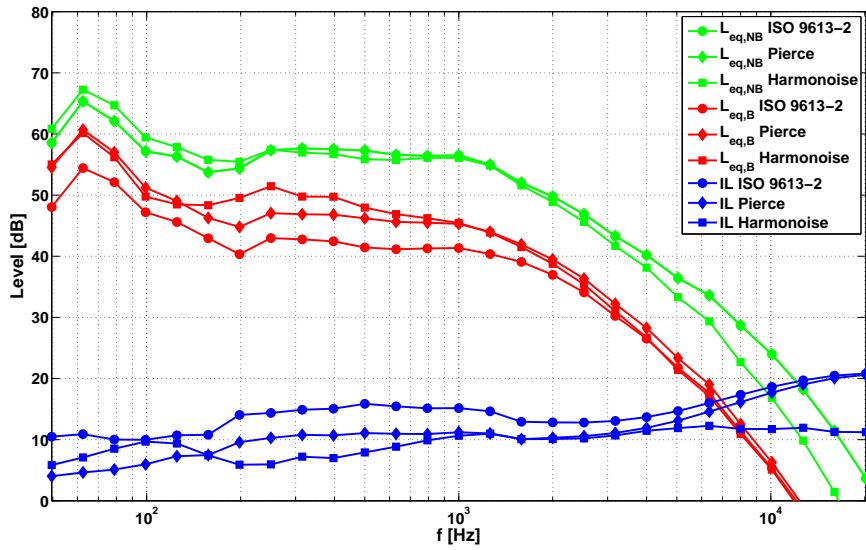


Figure 3.15: Comparison between total levels $L_{eq,NM|M}$ and total IL, calculated based on the ISO 9613-2, Pierce and Harmonoise simulation models. $L_{eq,NM}$ resp. $L_{eq,M}$ is given in green resp. red. The blue lines indicate the total IL.

3.2.4 Model performance and validation

In this section, results are compared with field measurements in order to evaluate the accuracy of the different simulation models. As it is impossible to measure the angular IL with omnidirectional microphones, a validation will only be made for the total IL. However, (dis)agreement between models and results can serve as an indicator for the accuracy of the angular IL.

3.2.4.1 Measurement set-up and data clean-up

In order to measure the highway IL of the mound, the total levels $L_{eq,NM|M}$ should be determined. Therefore, long-term level measurements have been performed

before and after construction of the mound. At four different microphone positions, two positions before the (future) mound and two positions behind the mound (cf. Fig. 3.7), a free-field microphone (microphone capsule MK-250 and pre-amp MV-210, Microtech Gefell GmbH) was installed and connected to a Swing measurement unit (Sinus Messtechnik GmbH). $L_{eq,0.2s}$, analyzed in 1/3-octave band sound pressure levels from 20Hz until 20kHz, was continuously logged from June 1st 2012 until July 6th 2012 (36 days in total) for the case without mound but with original vegetation.

After construction of the mound, the measurement set-up was reinstalled and the level was logged from October 16th 2012 until November 23th 2012. However, due to a power failure only 29 days of data could be collected.

Given the long measurement period, covering several weeks, post-processing on the raw measurement data is needed to reduce data and filter out invalid measurements. In a first step, $L_{eq,0.2s}$ -values are aggregated to $L_{eq,1h}$ -values, enabling a large data reduction.

In a next step, a selection is made of $L_{eq,1h}$ -values to remove non-representative meteorological conditions. Here, the data selection is done on an hourly basis as the meteo data was not collected at the same site and thus correlation can not be guaranteed on a shorter basis. In this work it is decided to keep $L_{eq,1h}$ -values when the hourly averaged wind speed was lower than 6m/s and humidity lower than 95%. Furthermore, when the amount of precipitation in the given timeslot was more than 0.05mm, measurements of that hour, and the following two hours were also discarded [40].

3.2.4.2 Extraction of $L_{eq,NM|M}$

From the filtered long-term measurements 459 resp. 338 $L_{eq,1h}$ -values have been retained for the calculation of $L_{eq,NM}$ resp. $L_{eq,M}$. It is expected that the emission by the road is very similar at the same hour, irrespective of the day of the week (only little difference between weekdays and weekends is seen). However, there is much more difference between levels at different hours of the day due to changing traffic conditions. Therefore, all $L_{eq,1h}$ values at a specific hour are linearly averaged over the full measurement period to reduce measurement error, since these values can be seen as different observations of a single parameter. In this way, an average $L_{eq,1h}$ can be calculated for each hour of the day. The total $L_{eq,NM|M}$ values are then calculated by energetically averaging all 24 $L_{eq,1h}$ -values over a complete day. The resulting levels are given in Fig. 3.16.

A typical traffic noise spectrum can be easily recognized at all measurement locations: the propulsion noise level peaks at low frequencies (63Hz - 80Hz) and the rolling noise causes a peak at about 1kHz. At frequencies higher than 10kHz,

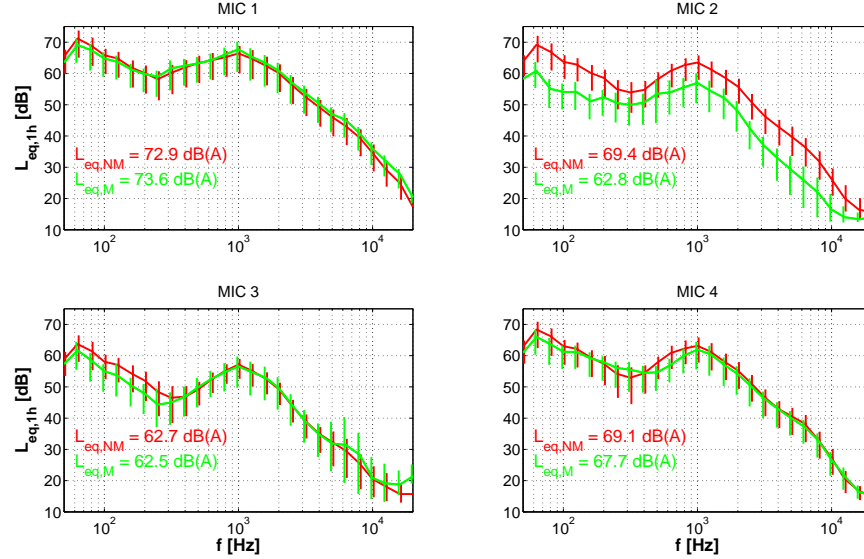


Figure 3.16: $L_{eq,NM|M}$, energetically averaged over a full day before (red) and after (green) the construction of the mound, for each measurement position. The total A-weighted $L_{eq,NM|M}$ is indicated in the graph. The length of the vertical bars indicates the difference between minimum and maximum levels over an average day.

the measurement equipment noise floor is reached and little useful information can be extracted. Generally, it can be observed that the fluctuation of noise levels over the course of a day, shown as vertical bars, keeps a nearly constant length of $10\text{dB} \pm 2\text{dB}$ over the considered frequency range (except at high frequencies $> 10\text{kHz}$). This supports the assumption that the traffic noise spectrum changes little over the different hours of the day, but only differs in absolute level, corresponding with an increase or decrease of the traffic flow.

The highest levels of traffic noise (approx. 73dB(A)) are measured at MIC1, closest to the highway. At MIC2, 19m from MIC1, a total A-weighted $L_{eq,NM}$ of 69.4dB(A) is measured, corresponding to a decrease of 3.5dB , while at MIC3, at 52m from MIC1, total A-weighted $L_{eq,NM}$ decreases further to 62.7dB(A) .

When comparing $L_{eq,NM}$ and $L_{eq,M}$, it can be noted that the largest effect of the mound is seen at MIC2, directly behind the mound. A total level drop of 6.6dB(A) is caused by the introduction of the mound. At MIC3, further behind the mound, almost no effect of the mound is seen on the total A-weighted level. However, at frequencies lower than 300Hz , a spectral level drop of approx. 3dB is seen. At that location, the mound offers almost no shielding for sound coming from the right side of the highway.

Not unexpectedly, at measurement locations in between the mound and highway only limited influence of the installation of the mound can be noticed. The level at MIC4 shows a decrease of 1.4dB compared to the situation without mound, although its location beside the mound. Similar levels as at MIC2 are measured before construction of the mound. However, when the mound is put in place, some shielding of the noise coming from the right side of the highway is noticed.

At MIC1 the overall level increases with approx. 0.7dB. Possible explanations of this (small) increase can be given by a change in emission, differences in the meteorological conditions or backscatter of the sound on the slope of the mound.

3.2.4.3 Estimation of the measured highway IL of the mound

Based on the measured $L_{eq,1h}$ before and after construction of the mound, an estimate for the highway IL of the mound can be calculated. Theoretically, the IL is calculated as the difference between the sound pressure level at a given receiver location before and after construction of the mound. However, this definition assumes that measurement conditions do not change between the before and after situation. In reality, as the measurements before resp. after construction took place during summer resp. autumn, the IL does not only include the effect of the mound, but also differences in the traffic flow composition, road surface temperature, meteorological conditions and microphone positioning.

In this work, three different approaches are illustrated to estimate the IL of the mound. The most straightforward approach is to calculate the difference between the total $L_{eq,NM}$ and $L_{eq,M}$, calculated as the energetical average of 24 hourly values (cf. previous section). This approach does not account for seasonal differences and puts more emphasis on time slots with high levels.

A second approach is to calculate the average hourly IL, derived as the difference of the average hourly levels before and after, and estimate the IL as the average of all 24 hourly IL-values. This approach proves to be a more robust way of calculating the IL, as it allows averaging over multiple observations and fluctuations over different times of the day can be equally taken into account.

A third approach is to estimate the IL based on the hourly transmission loss (TL) between the receiver location and a fixed reference microphone (MIC1). First, a linearly averaged TL for each hour of the day is calculated before and after construction of the mound. The IL is then given by the average of the hourly IL, estimated by the difference of the TL before and after construction.

With this calculation scheme the dependency on the shape of the source spectrum can be eliminated. However, this method is not error proof, as it now relies on the ability of only one reference position to capture all effects, and cannot account for the positional errors made when reinstalling the microphones in the after situation. Furthermore, the effect of backscatter of sound on the slope of the mound is ig-

nored: as the TL-based IL automatically implies 0dB IL at the reference position, level changes will be distributed to the IL calculated at the other positions.

In Fig. 3.17 the different estimates of the IL are compared. It can be seen that the general course of the different estimates is similar at all measurement positions, with a maximum deviation of 3dB only at high frequencies. The IL, calculated by averaging hourly ILs, shows limited to no difference with the IL, calculated by subtracting total levels. Except at MIC3, a small deviation is seen between 5kHz and 10kHz.

A somewhat larger difference is seen between the IL, calculated by subtracting TLs, and the other approaches. It should be noted that the frequency dependent difference between this and the second approach is exactly the same at all measurement positions and originates from the fact that the IL at the reference position is not exactly zero. At low frequencies, the TL-based IL estimates lower values (maximal 2dB difference), while at mid and high frequencies, a higher estimate is made (cf. IL at MIC1). The difference is limited, but slightly increasing with frequency at all measurement positions.

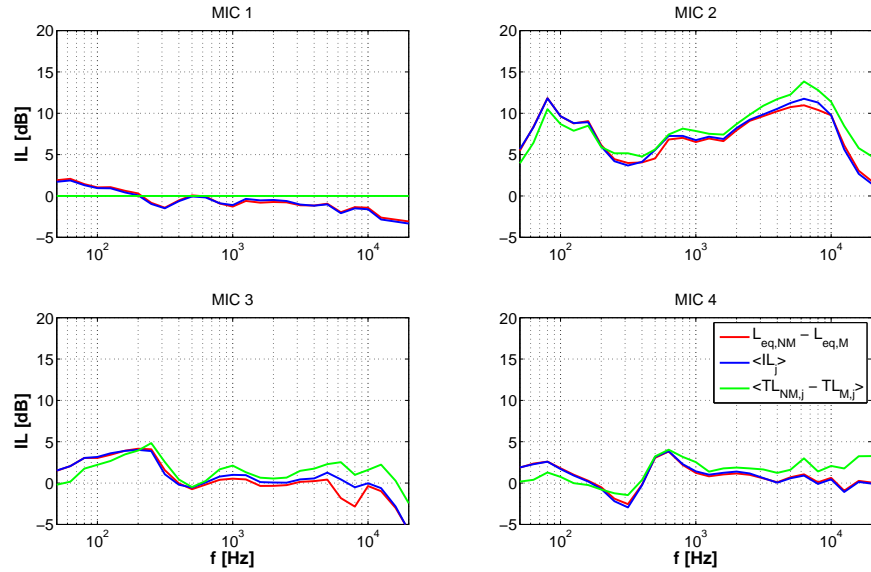


Figure 3.17: Measured IL at different microphone positions. Red: the IL, calculated by subtracting total $L_{eq,NM}$ and total $L_{eq,M}$ (first approach). Blue: the IL calculated by averaging different hourly ILs, extracted by subtracting the hourly linearly averaged sound level before and after construction of the mound (second approach). Green: the IL, calculated by averaging the difference between averaged hourly TLs before and after construction (third approach). MIC 1 is taken as a reference.

From Fig. 3.17 a more quantitative analysis for the IL at different locations can be made as well. As could be expected from the discussion of total levels, the highest IL is seen at MIC2, directly behind the mound. Values range between 4dB and 15dB. At 300Hz, a dip is seen for the IL, caused by the elimination of the ground effect by the introduction of the mound. A similar effect has been noted by calculations by Van Renterghem et al. [32].

At MIC3, somewhat further behind the mound, the measured IL is approx. 3dB at low and mid frequencies, but much less at higher frequencies. As explained above, this is caused by the L-shape of the mound, shielding only a part of the traffic from the highway. As the receiver distance from the highway increases, the contribution from the non-shielded part of the highway also rises, diminishing the sound reduction capabilities of the mound. In extremis, only one side of the highway is shielded by the mound, yielding a theoretically maximum achievable IL of 3dB.

At MIC4, at the same distance from the highway as MIC2, but beside the mound, the IL ranges between -1.5dB and 3dB. At this location, the mound shields half of the highway at most, while line-of-sight is kept with the other part of the highway. At 300Hz a small negative IL is seen, as the introduction of the mound decreases the magnitude of the ground effect for sources at the right side of the location.

3.2.4.4 Validation of the simulation models

Based on the measurements of the IL at MIC2, a comparison can be made with the simulated IL to evaluate the performance of each of the three models.

From Fig. 3.18 it can be seen that the Harmonoise model succeeds best in predicting the measured highway IL of the mound. The deviation from the TL-based IL is within the range of 3.5dB at frequencies below 10kHz.

The IL calculated by the Pierce model corresponds less with the measured IL than the IL calculated with the Harmonoise model. At low frequencies, the IL is underestimated, while at mid frequencies an overestimate of approx. 5dB is made.

The IL based on the ISO 9613-2 model shows the largest deviations from the measured IL. At mid frequencies, the simulated IL reaches values of 16dB, an overestimate of approx. 10dB. Furthermore, almost no correspondence with the spectral shape of the IL is seen. At high frequencies, the IL calculated with the Pierce and ISO 9613-2 model further increases, whereas the Harmonoise model tends to stabilize at 10dB. However, it should be noted that at such high frequencies, levels of traffic noise will be close to the noise floor, reducing the impact of modeling errors.

It is concluded that the more advanced Harmonoise model has the best performance regarding magnitude of IL and spectral shape. While the Pierce model still gives acceptable results, the ISO 9613-2 model shows the worst performance, with

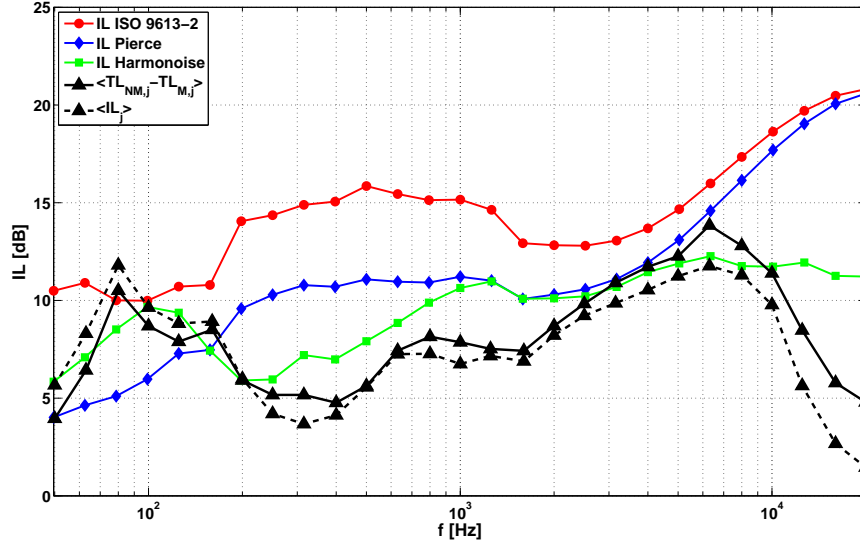


Figure 3.18: Comparison of the IL, calculated based on ISO 9613-2 (circle), Pierce (diamond) and Harmonoise (square), with the measured IL at MIC2 (black). Two estimates of the IL are given, the full line indicates the TL-based IL, the dashed line shows the hourly averaged IL.

a significant IL overestimate and pronounced differences in spectral shape. In this study, no explicit validation could be made regarding angular IL(ϕ). However, we can assume that for the Harmonoise and Pierce model good agreement for the angular IL is most likely, as the models were validated for both an omnidirectional point source and a traffic line source.

3.2.5 Auralization of the effect of the mound

This section describes the methodology to auralize the effect of the mound on the sound environment, based on simulated angular ILs and 32-channel array recordings before construction of the mound, and uses the pattern emulation technique developed in chapter 2 for the binaural recreation of (urban) soundfields.

At frequencies below a certain cross-over frequency, the effect of the mound will be auralized by emulation of a directivity pattern that corresponds with the HRTF directivity pattern for the left resp. right ear, attenuated with the angular IL of the mound. In this way a binaural signal is created where each source is subject to the angle-dependent mound attenuation.

At higher frequencies, only the signal from the microphone on the hard sphere at a location as close as possible to the position of the ear is used, as the sphere is assumed to approximate the human head. The effect of the mound is integrated by attenuating the signal with the IL, integrated over the left or right hemisphere.

Alternative auralization techniques could make use of a beamformer implementation of the microphone array. In this way, the contribution from each direction could be separately identified and processed. However, to create a binaural signal, extracted signals require filtering with the HRTFs, corresponding with each of the originating source directions, which can rapidly increase for a complex sound environment. Therefore, the pattern emulation technique is opted for now.

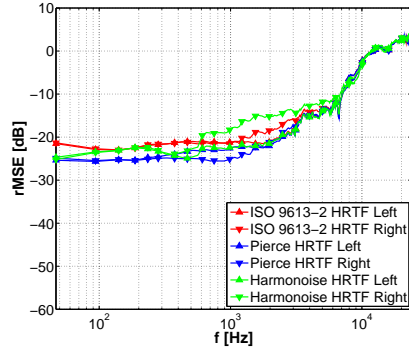
3.2.5.1 Auralization based on directivity pattern emulation

In a first step, the target directivity pattern should be defined. To account for the binaural aspect, this directivity pattern is based on the frequency dependent HRTF directivity patterns, $\text{HRTF}(f, \phi, \theta)$, equalized for the ear canal and reproduction headphones (cf. section 2.3.1). In order to integrate the effect of the mound, each of the impulse responses $\text{HRIR}(t, \phi, \theta)$, corresponding with the equalized HRTFs, are decomposed into 1/3-octave band components $\text{HRIR}_i(t, \phi, \theta)$ by using a similar 1/3-octave band analysis filterbank as in the first case-study. Each of the frequency bands is then attenuated with the corresponding 1/3-octave band angular $\text{IL}_i(\tilde{\phi})$:

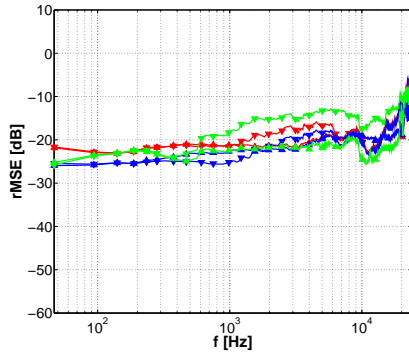
$$\text{HRIR}_{att}(t, \phi, \theta) = \sum_i \text{HRIR}_i(t, \phi, \theta) 10^{-\text{IL}_i(\tilde{\phi})/20}, \quad (3.8)$$

with $\text{HRIR}_{att}(t, \phi, \theta)$ the attenuated impulse response, i the index of the 1/3-octave band and $\tilde{\phi}$ the angle closest to the azimuth angle ϕ of the HRIR. The elevational dependency of the angular IL is not included here, as the human hearing has only limited capabilities to distinguish the elevation of the source (see section 2.1). In reality however, the sound will no longer originate from the highway, but from over the top of the mound. Furthermore, a discrete jump in the magnitude directivity pattern arises at $\phi = \pm 85^\circ$, as the angular $\text{IL}(\phi)$ has non-zero values for $-84^\circ \leq \phi \leq 84^\circ$ and zero values at $86^\circ \leq \phi \leq 274^\circ$. At the poles, the directivity pattern is attenuated with the frontal $\text{IL}(0^\circ)$.

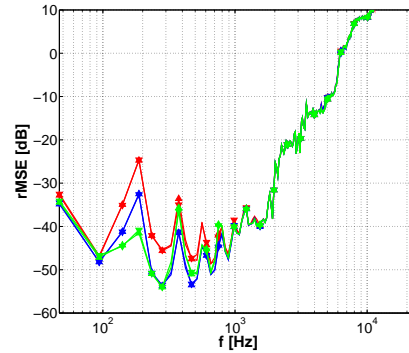
To create a smooth continuous representation of the attenuated HRTF directivity patterns without discrete jumps near the edges, an 18th-order spherical harmonic decomposition is made. Fig. 3.19 shows the performance of this decomposition in terms of rMSE for the left and right ear directivity patterns, with the head oriented towards the mound. At frequencies below 2kHz, the overall rMSE of the complex-valued HRTFs takes a value between -25dB and -15dB. However, from 2kHz on, the performance decreases to an rMSE of approx. 0dB at 10kHz.



(a) rMSE of the complex-valued HRTFs.



(b) rMSE of the HRTFs magnitude.



(c) rMSE of the HRTFs phase.

Figure 3.19: rMSE of the recomposed HRTFs, attenuated with the angular IL, calculated with the ISO 9613-2, Pierce and Harmonoise method ($N = 18$).

From the magnitude and unwrapped phase rMSE, it can be seen that this increase in overall rMSE is mainly due to the decrease in performance of the unwrapped phase decomposition at higher frequencies. While the magnitude rMSE increases only limited with frequency, with a value between -25dB and -13dB at frequencies below 10kHz, the unwrapped phase rMSE initially takes low values (approx. -40dB) and increases significantly from 2kHz on. A significant difference in performance can be seen when comparing the accuracy of the spherical harmonic representation of the attenuated HRTFs with the decomposition performance of the original non-attenuated HRTFs, where an rMSE lower than -40dB could be noted for the mid frequency range (Fig. 2.2).

The decrease in accuracy for the attenuated HRTFs can be explained when looking at the spatial distribution of the overall rMSE for the left ear attenuated directivity pattern, Fig. 3.20. Apart from a decrease in performance at $\phi = 270^\circ$, which has

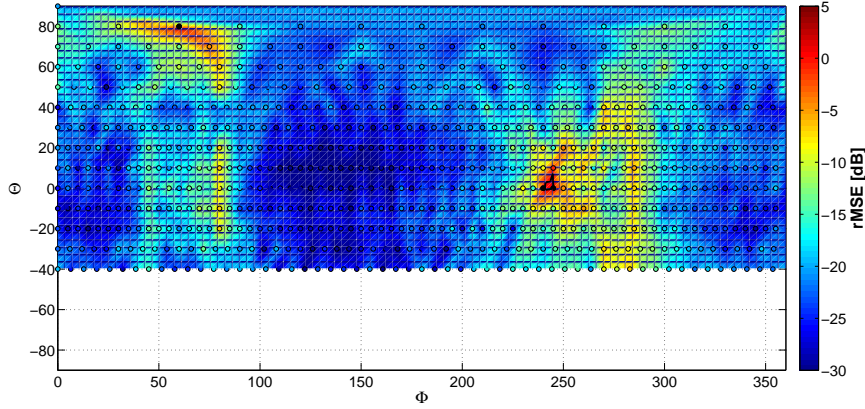


Figure 3.20: Interpolated spatial distribution of the overall rMSE, calculated over frequencies until $f_{max} = 5\text{kHz}$, for the left ear HRTF directivity pattern attenuated with the IL, calculated with Harmonoise. The left ear is located near $\phi = 90^\circ$ and the nose at $\phi = 0^\circ$, oriented towards the highway. Encircled values represent the actual rMSE values, calculated at the available measurement directions. The black contour and circles indicate an rMSE > 0dB.

been seen for the non-attenuated left ear directivity pattern as well, a rise in rMSE is seen at azimuth angles around $\phi = 90^\circ$. This decrease typically results from the reconstruction error of the discrete transition between the attenuated HRTFs in the frontal plane and the non-attenuated HRTFs in the back plane. Furthermore, the rMSE for the attenuated directions in the frontal plane ($0^\circ \leq \phi \leq 90^\circ$ and $270^\circ \leq \phi \leq 360^\circ$) will typically be higher than in the back plane ($90^\circ \leq \phi \leq 270^\circ$), as (small) errors in reproduction have relatively more influence at these (highly) attenuated directions. Fig. 3.19 furthermore proves that this results not only in a model dependent, due to differences in angular IL, but also ear dependent rMSE, as the left and right ear attenuated directivity patterns are not symmetrical due to the L-shape of the mound.

In the final step, a set of frequency dependent weights is found from a regularized least-square solution with the measured microphone-on-sphere directivity patterns as input and the attenuated directivity pattern as target pattern (cf. chapter 2). It should be noted that at frequencies below 200Hz, the spherical harmonic representation has been truncated until $N = 3$. Based on the quantitative (rMSE and WNG) and qualitative (listening tests) evaluation in chapter 2, a uniform 614 unique point grid is chosen for the evaluation of the least-square algorithm with a regularization parameter set to $\mu = 0.01$. A phase linearization procedure has been used at frequencies higher than 2kHz.

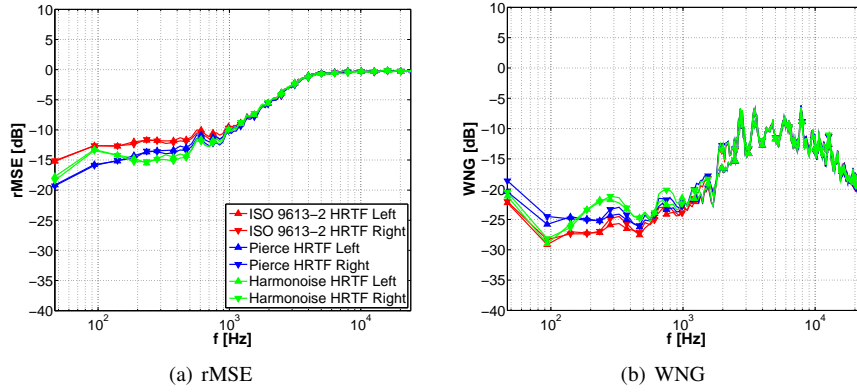


Figure 3.21: Comparison of the attenuated HRTF pattern fitting performance for the different IL models. The fitting has been performed with a uniformly sampled grid (614 unique points) and data with linearized phase ($\mu = 0.01$).

Fig. 3.21 shows the performance of the pattern fit in terms of overall rMSE and WNG for the left and right ear emulated directivity pattern, attenuated by using the different IL models. The rMSE ranges between -20dB at the lowest frequencies and -10dB at 1kHz, indicating a lower bound which is 5dB higher than the lower bound of the rMSE of the non-attenuated HRTF fit, ranging between -25dB and -12dB (Fig. 2.12). At higher frequencies, the rMSE of both attenuated and non-attenuated HRTF fit coincides, reaching -5dB at 2kHz.

The WNG ranges between -30dB at low frequencies and -20dB at 1kHz, while the WNG of the non-attenuated HRTF fit takes values between -25dB and -15dB. At higher frequencies, both the attenuated and the non-attenuated fit have similar WNG levels, fluctuating around -10dB.

It should be noted that, although the reproduction performance of the target directivity pattern showed to be dependent on left or right ear, no or limited ear-specific influence is found on the rMSE or WNG. However, a difference in performance related with the IL simulation model is still present. At mid frequencies, the pattern fit to the HRTF attenuated with the Harmonoise IL model shows the best performance regarding both rMSE and WNG. The fit with the Pierce model performs slightly worse, while the fit with the ISO 9613-2 model performs the least. The difference in performance between the fit with the Harmonoise model and the ISO 9613-2 model is the largest near 250Hz (3.5dB for the rMSE and 5dB for the WNG), but decreases when the frequency rises.

Fig. 3.22 shows the spatial rMSE for frequencies until 3kHz for the fitting performance with different IL models. Spatial rMSE values are typically much higher for the frontal plane than for the back plane and the 0dB value is frequently ex-

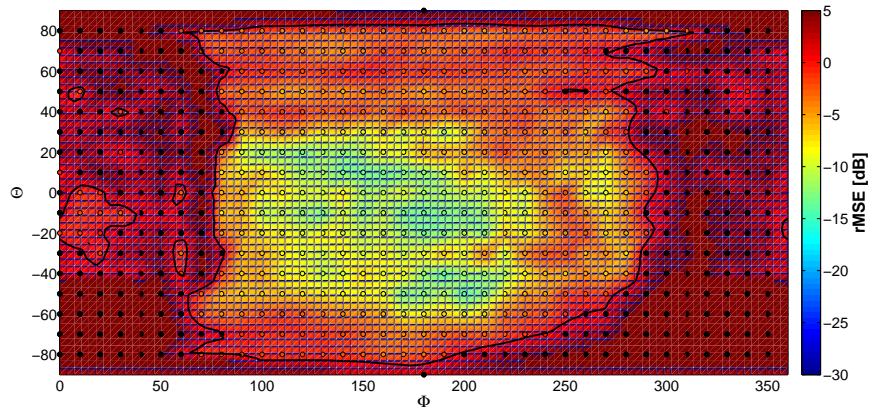
ceeded. Due to the attenuation of the mound, the target pattern takes low values in the frontal plane and (small) fitting errors will have relatively more influence than in the back plane. However, it should be noted that the rMSE might not be the best performance indicator, as errors for these already (highly) attenuated directions are perceptually less important.

Furthermore, a difference in fitting performance can be seen, depending on the used IL model. It is found that the rMSE of the Harmonoise and Pierce attenuated pattern has (much) lower rMSE in the frontal plane than when the pattern is attenuated with ISO 9613-2. A possible explanation is that the ISO 9613-2 model predicts much higher angular IL than the other models and as such has an increased error sensitivity in the frontal plane. Furthermore, the IL predicted with the ISO 9613-2 model has a more abrupt transition at $\pm 84^\circ$ (Fig. 3.12), giving rise to larger approximation errors, while the IL simulated with the Harmonoise model is already close to zero near the edges (Fig. 3.14), ensuring a much smoother transition. In the back plane all the fitted patterns show similar performance.

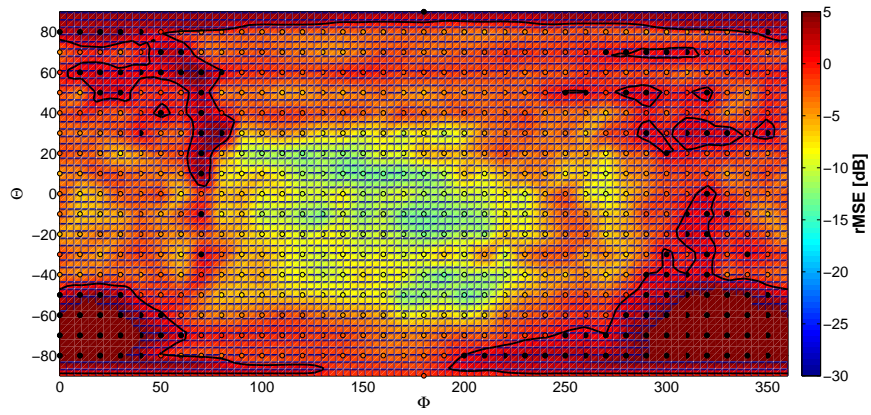
3.2.5.2 Auralization based on the MTB technique

From the previous section it could be seen that the pattern fitting technique only has a good performance up to a certain cross-over frequency. At higher frequencies, the auralization of the mound will be based on the MTB technique.

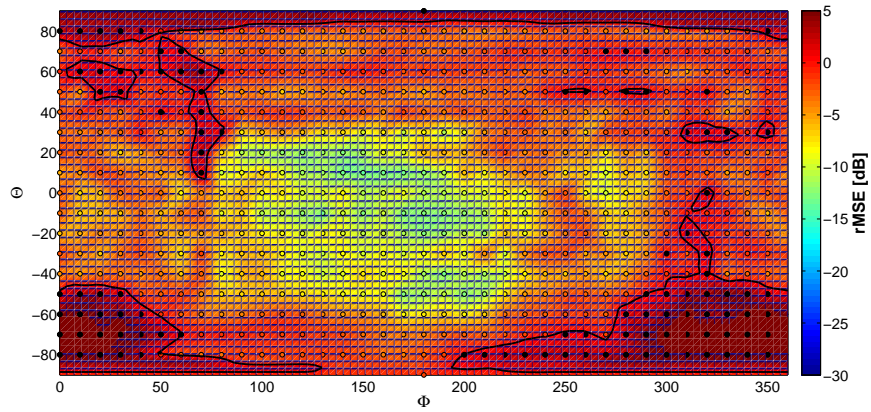
This technique is implemented by taking the signal from the microphone on the sphere closest to the location of the ear, corresponding to a first order spherical head approximation. As directional dependent mound attenuation with the angular IL is no longer possible, a more basic approach is followed to include the effect of the mound. Therefore, the signal is attenuated with the IL calculated for the complete left resp. right frontal hemisphere ($0^\circ \leq \phi \leq 90^\circ$ resp. $270^\circ \leq \phi \leq 360^\circ$ for the head oriented towards the highway). To calculate the IL for these hemispheres, the simulated angular levels $L_{eq, NM|M}$ are energetically summed for the corresponding hemisphere, and the IL is calculated as the difference between both. For sources in the corresponding hemisphere, the attenuation will be reasonably correct, but sources in the opposing hemisphere will be subject to the same attenuation levels as well. However, it should be noted that sources in the opposing hemisphere will have relatively less influence, as they are shielded by the hard sphere.



(a) rMSE - ISO 9613-2



(b) rMSE - Pierce



(c) rMSE - Harmonoise

Figure 3.22: Interpolated spatial distribution of the overall rMSE for the left ear, calculated over frequencies until $f_{max} = 3\text{kHz}$ for the different IL models ($\mu = 0.01$). Encircled values represent the actual rMSE values, according to the original measurement directions. The black contour and circles indicate an rMSE $> 0\text{dB}$.

3.2.6 Qualitative evaluation of the auralization of the mound

In the previous section, a methodology for the auralization of the effect of a mound on the sound environment has been developed. To evaluate the quality of this auralization methodology and investigate whether or not the specific IL model of the mound has an influence, a qualitative evaluation is made by means of a series of listening tests, similar as in the first case-study.

3.2.6.1 Audio sample acquisition and auralization

For the acquisition of audio samples, a similar set-up as in chapter 2 has been used. Recordings before installation of the mound have been performed on July 7th 2012 with the 32-channel microphone array and a HATS (B&K Type 4128C) with inner ear simulator (B&K Type 4158C). To enable comparison with the real sound environment after installation of the mound, microphone array recordings have been carried out on November 20th, 2012. However, due to a hardware failure no HATS recordings could be made on that date and the recordings with the HATS have been repeated on January 11th, 2013. Furthermore, the 1/3-octave sound pressure level was logged with a Swing measurement unit (Sinus Messtechnik GmbH) with calibrated MK-250 free-field microphone and MV-210 pre-amp (Microtech Gefell GmbH) to enable proper calibration of the reproduction levels of the a priori recordings. All recordings are made at the MIC2 position and the position of the recording equipment is reproduced as accurately as possible for each measurement campaign.

In the a priori case, 17 binaural samples have been extracted from the microphone array recordings and the corresponding samples from the HATS recording have been selected as well. For the case after installation of the mound, 17 samples resp. 11 samples have been extracted from the array resp. HATS recordings. To allow a proper evaluation, samples were approx. 30s long and contained several distinct truck passages. Some samples also contained other sounds, not related to traffic (eg. barking dog). The auralization methodology from section 3.2.5 has been applied on the a priori array recordings, by using the three different models for the IL. The pattern fitting has been performed on a uniform grid (614 unique points), with data with linearized phase (from 2kHz on) and a regularization $\mu = 0.01$. The cut-off frequency start of the MTB technique is used here as a parameter. An extra auralization has been created as well, by using the MTB technique with hemispherical IL over the full frequency range.

3.2.6.2 Listening test set-up and results

To evaluate the quality of the auralizations, four different tests have been designed. These tests were taken in the same run as the listening test from chapter 2. As such,

Table 3.5: Influence of the cross-over frequency f_X . The columns indicate the ranking given by the listeners, presented with samples auralized with Harmonoise for different cross-over frequencies f_X (1 means most similar). The total sample count is 37. ($\chi^2 = 70.81$, $p < 0.01$)

	1	2	3	4	5
Array sim., $f_X = 1122\text{Hz}$	10	11	9	6	1
Array sim., $f_X = 1782\text{Hz}$	9	9	12	6	1
Array sim., $f_X = 2245\text{Hz}$	7	10	13	5	2
Array sim., $f_X = 4490\text{Hz}$	3	4	3	12	15
L-R array sim.	8	3	0	8	18

the analysis is based on the test results of 14 female and 23 male listeners, with age ranging between 21 and 57 years old.

In a first test it is investigated whether or not the value of the cross-over frequency has an influence on the performance of the auralization. Therefore, the listener is presented with a reference sample, recorded with a HATS after construction of the mound, and five auralized samples. The auralized samples are all based on the same a priori array recording, and as such have different content than the a posteriori reference sample. One sample is created by using the MTB technique over the full frequency range, while the other four samples are generated with the pattern emulation technique with four different cross-over frequencies, $f_X = 1122\text{Hz}$, 1782Hz , 2245Hz and 4490Hz . The Harmonoise model has been used to model the IL in all five samples. The listener is asked to rank the simulated samples according to their similarity with the sound environment after installation of the mound (1 means most similar). It is pointed out to the listener that the reference and auralized samples have different content.

The frequency count of rankings given to each of the auralizations are displayed in Table 3.5. Based on these results, it can be seen that when f_X increases, the auralized sample typically gets worse rank-order.

In order to investigate whether or not mean rankings are significantly different, a multiple comparison test is performed by applying Tukey's least significant difference procedure on the outcome of an ANOVA test. Fig. 3.23 shows the mean ranking, together with the intervals of significance. Samples with lowest f_X are ranked as best. However, it should be noted that auralizations with $f_X \leq 2245\text{Hz}$ have no significantly different ranking. The pattern emulation technique with $f_X = 4490\text{Hz}$ has the worst ranking, but it is seen that the ranking is not significantly different from the full-range MTB technique. Both are however significantly different from the best performing techniques.

The observations from this test indicate that the choice of $f_X = 1782\text{Hz}$, suggested

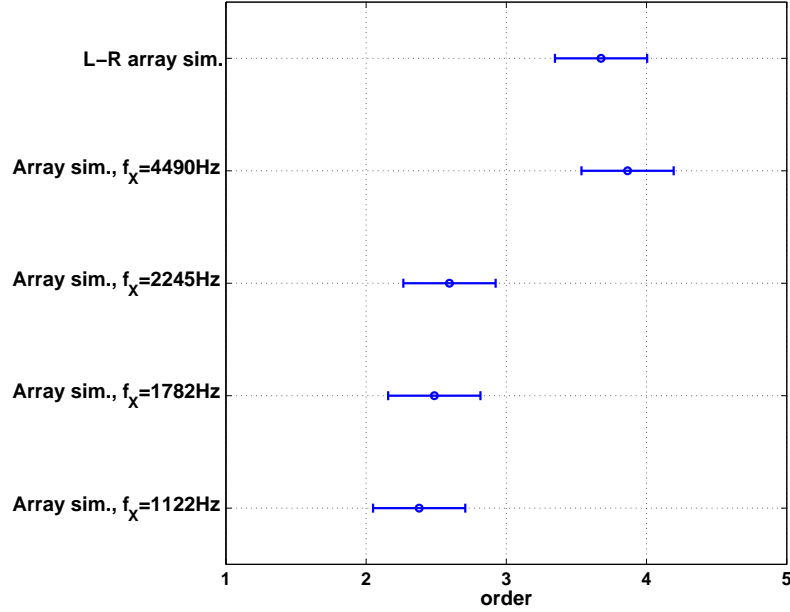


Figure 3.23: Results of the subjective ranking of the samples auralized with Harmonoise in function of f_x . Mean ranking and 95% confidence intervals of each technique are indicated.

by the listening tests in chapter 2, is justified for the given auralization methodologies and this cross-over frequency will be used in the following tests.

In a second test, it is checked whether or not listeners can distinguish between recordings made before installation of the mound and recordings made after installation. Both array recordings and HATS recordings as well as some auralized samples are integrated, and listeners were asked to indicate whether or not the fragments were recorded before or after construction of the mound.

The results of this test are given in Table 3.6. The columns indicate whether or not listeners indicated the samples as recorded before or after installation of the mound. The rightmost column gives the probability that the null-hypothesis, assuming random choice, is accepted, resulting from a two-sided binomial test.

The results of this test clearly indicate that listeners can indeed distinguish between fragments recorded before and after construction of the mound, whether or not fragments have been binaurally synthesized from a microphone array recording or recorded with a HATS. Furthermore, almost all auralizations made with the array recording technique are categorized as recorded after installation of the mound, independent of the used IL simulation model.

Table 3.6: Results of the second test. The columns indicate the responses given by the listeners, presented with auralized samples and samples recorded before or after installation of the mound. The right column gives the probability that the null-hypothesis, assuming random choice, is accepted. ($\chi^2 = 303.81$, $p < 0.01$)

	Ind. before	Ind. after	Count	p-value
HATS rec. before	74	0	74	$< 0.01^{**}$
Array rec. before	73	1	74	$< 0.01^{**}$
HATS rec. after	5	32	37	$< 0.01^{**}$
Array rec. after	1	36	37	$< 0.01^{**}$
Array sim. Harmonoise	1	36	37	$< 0.01^{**}$
Array sim. Pierce	0	37	37	$< 0.01^{**}$
Array sim. ISO 9613-2	0	37	37	$< 0.01^{**}$

Table 3.7: Results of the third test. The columns indicate the responses given by the listeners, presented with auralized samples and samples recorded after installation of the mound. The right column gives the probability that the null-hypothesis, assuming random choice, is accepted. ($\chi^2 = 17.31$, $p < 0.01$)

	Ind. record.	Ind. simulat.	Count	p-value
HATS rec. after	84	64	148	0.118
Array sim. Harmonoise	15	22	37	0.324
Array sim. Pierce	19	18	37	1
Array sim. ISO 9613-2	13	24	37	0.099
L-R sim. Harmonoise	19	18	37	1
L-R sim. Pierce	23	14	37	0.188
L-R sim. ISO 9613-2	10	27	37	$< 0.01^{**}$

In a third test, the fidelity of the auralized samples was investigated. Therefore, the listener is presented with ten audio fragments and he or she was asked to indicate if the given sample was a true recording or a simulation, created from recordings made before installation of the mound. From these samples, four samples were true HATS recordings made after installation of the mound, while six samples were auralizations, generated with the different IL models, of which three samples were created by using the MTB technique with hemispherical IL model for the full frequency range ('L-R sim.').

Table 3.7 indicates whether listeners categorize the fragment as a recording or a simulation. Again, the rightmost column indicates the probability that the null-hypothesis, assuming random choice, is accepted. The data in this table shows that listeners have great difficulty to distinguish whether samples have been auralized

Table 3.8: Influence of the simulation methodology. The columns indicate the ranking given by the listeners, presented with samples auralized with different simulation techniques (1 means most similar). The total sample count is 37. ($\chi^2 = 116.92$, $p < 0.01$)

	1	2	3	4	5	6
Array sim. Harmonoise	8	4	11	8	5	1
L-R sim. Harmonoise	6	5	3	8	11	4
Array sim. Pierce	8	10	9	5	3	2
L-R sim. Pierce	6	5	6	11	8	1
Array sim. ISO 9613-2	7	13	5	2	6	4
L-R sim. ISO 9613-2	2	0	3	3	4	25

or recorded after installation of the mound. 43% of the a posteriori HATS recordings have been categorized as simulations. Only for the samples auralized with use of the full-range MTB technique with the ISO 9613-2 IL model the assumption of random choice is rejected, as almost 73% of the samples was indicated as simulated. For auralizations based on array pattern emulation with the ISO 9613-2 model attenuated pattern, an increased number of samples (65%) have been identified as simulations, however, the assumption of random choice is still valid. For the other auralization techniques, at least 40% of the samples is categorized as a recording.

In order to further investigate whether or not a difference in performance can be seen between auralization methods, an additional test is performed. In this test, the listener is presented with a reference sample, i.e. a true HATS recording made after installation of the mound, and six auralizations. Three auralizations are made with the array pattern emulation technique with each of the three different IL models, while the other three are created with the full-range MTB technique with inclusion of the hemispherical IL. The auralized samples all originate from the same array recording made before the barrier, and as such have different content than the a posteriori reference sample. The listener is then asked to rank the simulated samples according to their similarity with the sound environment after installation of the mound (1 means most similar). Again, it is pointed out that the reference and auralized samples have different content.

Table 3.8 shows the frequency count of rankings given to each of the auralization methodologies. From this table, it can be clearly seen that the full-range extended MTB technique with ISO 9613-2 IL model has the worst ranking. However, it is more difficult to draw conclusions regarding the other techniques.

In order to compare the different techniques and investigate if the mean ranking of the samples is significantly different, a multiple comparison test, similar as in

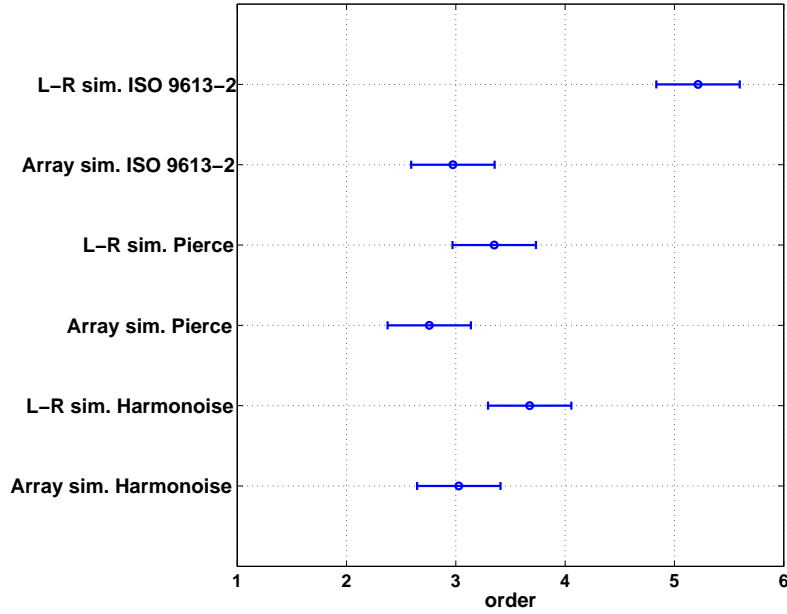


Figure 3.24: Results of the subjective ranking of the different auralization techniques. Mean ranking and 95% confidence intervals of each technique are indicated.

the previous test, is performed. Fig. 3.24 shows the results from this test. The array emulation technique with the Pierce IL model performs best based on absolute ranking, although no significant difference is seen with the pattern emulation techniques with other IL models. However, this technique is significantly better than the full-range MTB techniques with the Harmonoise and ISO 9613-2 IL model. Typically, the pattern emulation technique outperforms the full-range MTB technique for a given IL model, although the difference is only marginally significant.

3.2.7 Conclusion

A new methodology has been introduced to evaluate the effect of a noise mitigation measure in a complex sound environment by means of a priori microphone array recordings and simulations of the angular IL of the noise mitigation measure. Auralized samples have been created by weighting the array signals so that the resulting pattern approximates the left or right ear HRTF directivity patterns, with each direction attenuated according to the corresponding angular IL. At high frequencies, the MTB technique has been used, where each signal is attenuated with the hemispherical IL.

The validity of this technique has been evaluated in a case-study, where it has been investigated to what extent the effect of a large L-shaped mound on the sound environment can be predicted. Therefore, both a quantitative analysis, in terms of highway mound IL, as well as a qualitative analysis, based on the newly developed auralization methodology, has been performed.

Three different techniques have been used to estimate the IL: the ISO 9613-2 propagation model, the Pierce barrier diffraction model and the Harmonoise point-to-point model. The physical accuracy of these engineering models has been put to the test. By comparison with the measured highway IL of the mound, it has been found that the measured IL was best approximated by the IL calculated with the Harmonoise model, which accounted best for spectral fluctuations. The IL calculated with the ISO 9613-2 model showed worst performance, with an overestimate of approx. 10dB at mid frequencies.

Based on the calculated angular IL and recordings made before the installation of the mound, an auditive preview of the effect of the mound has been created and the performance of the auralization methodology has been investigated by means of listening tests. From these tests it was seen that the effect of the mound was clearly heard by the listeners, while hidden auralizations were recognized as a posteriori recordings, adequately incorporating the effect of the mound. Furthermore, the validity of the auralization methodology was shown, as it was seen that listeners could not distinguish between recordings and auralizations, except for the full-range MTB technique with ISO 9613-2 IL model.

When ranking the different auralization techniques according to their correspondence with the a posteriori sound environment, the full-range MTB technique with ISO 9613-2 IL model was ranked worst. The pattern emulation techniques performed best, but showed no significant difference even when different IL models were used. After the tests, listeners indicated that, although a ranking was difficult to make, samples based on the full-range MTB technique were somewhat easier to detect, due to the large difference between the left and right ear level.

The results of this case-study prove the validity of the array auralization methodology and show its potential in simulating the sound environment prior to the installation of noise mitigation measures.

References

- [1] J. Maillard and J. Jagla. *Auralization of non-stationary traffic noise using sample based synthesis - comparison with pass-by recordings*. In Proceedings of the 41st international congress and exposition on noise control engineering (Inter-Noise 2012), 2012.

- [2] J. Maillard and J. Jagla. *Real time auralization of non-stationary traffic noise - quantitative and perceptual validation in an urban street*. In Proceedings of AIA-DAGA, 2013.
- [3] J. Jagla, J. Maillard, and N. Martin. *Sample-based engine noise synthesis using an enhanced pitch-synchronous overlap-and-add method*. J. Acoust. Soc. Am., 132:3098–3108, 2012.
- [4] J. Forssén, T. Kaczmarek, J. Alvarsson, P. Lundén, and M. E. Nilsson. *Auralization of traffic noise within the LISTEN project - preliminary results for passenger car pass-by*. In Proceedings of the 8th European conference on Noise Control, 2009.
- [5] A. Peplow, J. Forssén, P. Lundén, and M. E. Nilsson. *Exterior auralization of traffic noise within the LISTEN project*. In Proceedings of Forum Acusticum, pages 665–669, 2011.
- [6] HOSANNA. www.greener-cities.eu.
- [7] M. Boes, B. De Coensel, D. Oldoni, and D. Botteldooren. *A biologically inspired model adding binaural aspects to soundscape analysis*. In Proceedings of the 40th international congress and exposition on noise control engineering (Inter-Noise 2011), 2011.
- [8] ISO 9613-2:1996, *Acoustics - Attenuation of sound during propagation outdoors - Part 2: General method of calculation*. International Organization for Standardization, Geneva, Switzerland, 1996.
- [9] D. Van Maercke and J. Defrance. *Development of an analytical model for outdoor sound propagation within the Harmonoise project*. Acta Acustica united with Acustica, 93:201–212, 2007.
- [10] E. Salomons, D. Van Maercke, J. Defrance, and F. De Roo. *The Harmonoise sound propagation model*. Acta Acustica united with Acustica, 97:62–74, 2011.
- [11] D. Botteldooren. *Acoustical finite difference time domain simulation in a quasi-cartesian grid*. J. Acoust. Soc. Am., 95:2313–2319, 1994.
- [12] R. Nota, R. Barelds, and D. Van Maercke. *Engineering method for road traffic and railway noise after validation and fine-tuning*. Technical report har32tr-040922-dgmr20, 2005.
- [13] T. Van Renterghem, M. Hornikx, P. Jean, Y. Smyrnova, J. Kang, D. Botteldooren, and J. Defrance. *Road traffic noise reduction by vegetated low noise barriers in urban streets*. In Proceedings of the 9th European Conference on Noise Control, 2012.

- [14] *ISO 9613-1:1993, Acoustics - Attenuation of sound during propagation outdoors - Part 1: Calculation of the absorption of sound by the atmosphere.* International Organization for Standardization, Geneva, Switzerland, 1993.
- [15] P. D. Joppa, L. C. Sutherland, and A. J. Zuckervar. *Representative frequency approach to the effect of bandpass filters on evaluation of sound attenuation by the atmosphere.* J. Noise Control Eng., 44:261–273, 1996.
- [16] D. S. Cho, J. H. Kim, T. M. Choi, B. H. Kim, and D. Manvell. *Highway traffic noise prediction using method fully compliant with ISO 9613: comparison with measurements.* Applied Acoustics, 65:883–892, 2004.
- [17] G. Dutilleul, J. Defrance, D. Ecotière, B. Gauvreau, M. Bérengier, F. Besnard, and E. Le Duc. *NMPB-Routes-2008: The revision of the French method for road traffic noise prediction.* Acta Acustica united with Acustica, 96:452–462, 2010.
- [18] H. G. Jonasson. *Sound reduction by barriers on the ground.* J. Sound. Vib., 22:113–126, 1972.
- [19] M. E. Delany and E. N. Bazley. *Acoustical properties of fibrous absorbant materials.* Applied Acoustics, 3:105–116, 1970.
- [20] D. C. Hothersall and J. N. B. Harriot. *Approximate models for sound propagation above multi-impedance plane boundaries.* J. Acoust. Soc. Am., 97:918–926, 1995.
- [21] A. Pierce. *Diffraction of sound around corners and over wide barriers.* J. Acoust. Soc. Am., 55:941–955, 1974.
- [22] A. Sommerfeld. *Mathematische Theorie der Diffraction.* Math. Ann, 47:317–374, 1896.
- [23] U. J. Kurze. *Noise reduction by barriers.* J. Acoust. Soc. Am., 55:504–518, 1974.
- [24] J. Deygout. *Multiple knife-edge diffraction by microwaves.* IEEE Transactions on Antennas and Propagation, 14:480–489, 1966.
- [25] T. Van Renterghem, D. Botteldooren, and K. Verheyen. *Road traffic noise shielding by vegetation belts of limited depth.* J. Sound. Vib., 331:2404–2425, 2012.
- [26] K. Horoshenkov, A. Khan, and H. Benkreira. *Acoustic properties of low growing plants.* J. Acoust. Soc. Am., 133:2554–2565, 2013.

- [27] D. Aylor. *Noise reduction by vegetation and ground*. J. Acoust. Soc. Am., 51:197–205, 1972.
- [28] D. Aylor. *Sound transmission through vegetation in relation to leaf area density, leaf width and breath of canopy*. J. Acoust. Soc. Am., 51:411–414, 1972.
- [29] T. Van Renterghem, K. Attenborough, M. Maennel, J. Defrance, K. Horoshenkov, J. Kang, I. Bashir, S. Taherzadeh, B. Altreuther, A. Khan, Y. Smyrnova, and H. Yang. *Measured light vehicle noise reduction by hedges*. Applied Acoustics, 78:19–27, 2014.
- [30] J. Forssén, M. Hornikx, D. Botteldooren, W. Wei, T. Van Renterghem, and M. Ögren. *A model of sound scattering by atmospheric turbulence for use in noise mapping calculations*. Acta Acustica united with Acustica, 100:810–815, 2014.
- [31] K. Rasmussen. *Outdoor sound propagation under the influence of wind and temperature gradients*. Journal of Sound and Vibration, 104:321–335, 1986.
- [32] T. Van Renterghem and D. Botteldooren. *On the choice between walls and berms for road traffic noise shielding including wind effects*. Landscape and Urban Planning, 105:199–210, 2012.
- [33] J. Defrance, E. Salomons, I. Noordhoek, D. Heimann, B. Plovsing, G. Watts, H. Jonasson, X. Zhang, E. Premat, I. Schmich, F. Aballea, M. Baulac, and F. De Roo. *Outdoor sound propagation reference model developed in the European Harmonoise project*. Acta Acustica united with Acustica, 93:213–227, 2007.
- [34] D. Wilson and V. Ostashev. *A reexamination of acoustic scattering in the atmosphere using an improved model for the turbulence spectrum*. Technical report, New Mexico state univ., 1998.
- [35] T. Van Renterghem, W. Wei, J. Forssén, M. Hornikx, M. Ögren, D. Botteldooren, and E. Salomons. *Improving the accuracy of engineering models at shielded building facades : experimental analysis of turbulence scattering*. In Proceedings of the 42nd international congress and exposition on noise control engineering (Internoise 2013), 2013.
- [36] Van Renterghem et al. *WP3 : Acoustical shielding by hedges and shrubs*. Technical report, www.greener-cities.eu, 2013.
- [37] E. Salomons, H. Zhou, and W. Lohman. *Efficient numerical modeling of traffic noise*. J. Acoust. Soc. Am., 127:796–803, 2010.

-
- [38] H. G. Jonasson. *Acoustical source modeling of road vehicles*. Acta Acustica united with Acustica, 93:173–184, 2007.
 - [39] *Traffic data, yearly average by workday and by hour for 2010*. Verkeerscentrum Vlaanderen, Flemish Government, Department of Mobility.
 - [40] C. Ribeiro, D. Ecoti re, P. Cellard, and C. Rosin. *Uncertainties of the frequency response of wet microphone windscreens*. Applied Acoustics, 78:11–18, 2014.

4

Using room acoustical parameters for evaluating the quality of urban squares for open-air rock concerts

P. Thomas, T. Van Renterghem and D. Botteldooren

Published in *Applied Acoustics*, 72(4):210-220, November 2011



Sound quality research of urban squares used for open-air (rock) concerts is very scarce. In contrast to the study of (classical) concert halls, little is known about useful design parameters. For the design of the amplification system, the sound engineer currently often takes into account the desired sound pressure level only.

In this study, the ability of existing room acoustical parameters to characterize urban squares acoustically is investigated. An independent parameter set is identified for specific use on such squares. Besides the distribution of the sound pressure level over the square, different impulse response related parameters such as the clarity, center time, reverberation time and bass ratio were considered. In addition, binaural measures were included to measure qualities related to human spatial hearing.

This study is based on a measurement campaign, performed at five squares in Belgium before and during live rock concerts. Special attention was paid to the signal processing methodology, given the significant amount of environmental noise often found at such squares during measurements. The variation of these parameters is investigated in relation to the square geometry, the amplification set-up and the presence of delay-lines. Parameters like C_{80} , T_{30} , $IACC_{E3/L3}$ and $\Delta L_{eq,A/C}$ were shown to be very useful when characterizing the sound field at urban squares.

4.1 Introduction

Over the last few decades the number of open-air concerts has grown a lot. Because of their historical character and social function, public squares are frequently used for concerts, despite the fact that these squares were not designed to hold such concerts and their acoustical quality is mostly far from ideal.

The difference with concert halls is significant [1]. Most squares have a typically large irregular shape and are bordered with acoustically quite rigid surfaces, whose degree of scattering varies to a large extent. The shape and location of scatterers is usually not adapted to the location of sources and listeners during concerts. Also the absence of a roof results in a high energy loss. In practice, this means that only a limited amount of reflections reach the listener. The influence of the audience is also apparent. In contrast to a concert hall, the public on a square is standing or even seated on the ground with varying density along the square. This results in a fluctuating absorption coefficient [2, 3].

In order to cope with these shortcomings, sound engineers have to design an amplification system that optimizes the sound field of the square, since the geometrical and acoustical properties are fixed. This amplification system mostly consists of different speakers or line-arrays, each speaker trying to optimize the sound field in a part of the square. For this purpose, delay-lines are used. These are line-arrays placed along the square, reproducing a delayed version of the sound signal.

For years now, the preferred and only design parameter for these amplification systems has been the sound pressure level SPL [4, 5]. However, this parameter only describes the loudness of the sound and proves to be insufficient in giving full information on how the listener experiences the sound field. For classical concert halls, many studies have identified different sets of statistically independent parameters that determine the acoustical quality. The Goettingen school found the interaural cross-correlation IACC, reverberation time T_{30} and clarity C_{50} to be the dominating factors [6]. Cremer et al. determined the strength G , bass ratio BR and C_{80} , related with loudness, timbre and clarity as a set of representative parameters [7]. In 1983, Ando proposed four objective orthogonal factors that are related to subjective preferences, which he combined into a rating scale for concert halls [8, 9]. These four factors consist of three temporal factors; the initial

time-delay gap ITDG (related to intimacy), G (related to loudness), T_{sub} (related to reverberance) and a spatial factor namely the interaural cross-correlation IACC (related to the difference between left- and right-ear signals). In addition to these factors, Beranek used two more statistically independent parameters; the surface diffusion index SDI and bass ratio BR [10, 11]. By studying the acoustics of 35 concert halls, Gade showed in 1991 that the early decay time EDT, T_{30} and C_{80} are highly correlated, but orthogonal to the lateral fraction LF and G [12]. In a more recent study of 2009, Cerda measured the acoustical parameters at many locations in different types of performance halls in order to investigate the correlation of different acoustical parameters [13]. By using factor analysis he obtained a set of three orthogonal factors; the reverberation-intelligibility-clarity factor, the spatiality-subjectivity factor and the strength factor, respectively represented by T_{30} , Beranek's binaural quality index BQI [10] and G .

Until now, no such studies have been undertaken to identify a statistically independent set of parameters for square-acoustics. It is therefore useful to investigate the existing parameters as mentioned in the ISO 3382 standard [14] and extract those which can deliver more information on the sound field. Parameters such as the clarity C_{80} and center time T_s give us insight in how the received acoustical energy is distributed over time. Other parameters such as the early decay time EDT and reverberation time T_{30} are useful when describing the sound energy decay on the square over somewhat longer time scales. These parameters are extended with binaural parameters related to a subjective experience of apparent source width ASW and listener envelopment LEV to include the spatial impression of the sound field. The initial time-delay gap ITDG [15] is, in the particular settings of outdoor amplified rock concerts, strongly determined by the (distributed) loudspeaker system, rather than by the acoustics of the space. As such, it correlated very strongly with other parameters measuring the fraction of early energy such as C_{80} . Therefore, this parameter was not considered explicitly. Other parameters such as the speech transmission index STI, lateral fraction LF and strength factor G could be useful as well, but were not considered in this study. However, an alternative formulation of the strength factor is proposed by using a relative sound pressure level.

In this chapter, a set of statistically independent parameters for specific use on urban squares during rock-festivals is identified by correlation analysis. Their dependence on the geometry of the square and the used amplification system is then investigated with special attention for the influence of delay-lines on the parameters. This will be accomplished by comparing results of an extensive measurement campaign at four urban squares during and before rock concerts. The influence of the delay-lines is investigated by measurements at an open festival site in order to eliminate the influence of the geometry. A study of the subjective attributes and listener preferences related to these objective parameters is outside the scope of this research.

The measurements of the impulse response and related parameters are based on the ISO 3382-1 and ISO 18233 standard [14, 16]. Special attention is given to the choice of test signal and signal (post-) processing.

4.2 Parameter definitions

The parameters which will be investigated are mainly extracted from the ISO 3382 standard [14] and already well-known in the world of (concert hall) acoustics [10]. This section gives a short overview, highlighting some slight variations in their definition.

4.2.1 Energy related parameters

In the introduction the strength factor G was mentioned. The parameter G measures the ratio of the sound field, generated by an omnidirectional loudspeaker on stage, at different locations in the room to a reference value recorded in an anechoic chamber at 10m distance. In outdoor rock concerts, the directional speaker set-up has to be considered as part of the acoustical source. Therefore, the strict definition of G based on an omnidirectional loudspeaker on stage and a reference point at 10m makes little sense. As an approximation, we define the relative sound pressure level $\Delta L_{eq,oct}$, in octave bands, as the difference between the A- or C-weighted sound pressure level at a position of interest and the weighted level at a certain reference point, usually chosen near the sound engineer. $\Delta L_{eq,oct}$ has values for each octave band. To have a single-number value, $\Delta L_{eq,A/C}$ is defined as the difference between the frequency summed A- or C-weighted sound pressure level $L_{eq,A/C}$ at the measurement position and $L_{eq,A/C}$ at the reference point. It can be seen that $\Delta L_{eq,oct}$ and $\Delta L_{eq,A/C}$ are determined by the combined effects of frequency dependent path loss, audience absorption and reflection and absorption at the square edges. These parameters also incorporate the influence of the speaker radiation patterns and delay-lines.

In the ISO 3382 standard, three other important parameters are listed that are related to how the sound energy is divided over the early and late part of the sound signal. These are the Deutlichkeit D_{50} , clarity C_{80} and center time T_s .

The Deutlichkeit and clarity are defined by the ratio of the early-to-late arriving energy. However, the Deutlichkeit will not be included in this work, as it applies more to speech conditions. C_{80} is more suitable for music signals, as the early energy is defined as the sound energy arriving within the first 80ms. There is however some discussion about the time which separates the early reflections from the late ones. Literature review shows that this separation time has values between 50ms and 200ms. In their work, Hidaka and Beranek chose the separation

limit for the early sound field to be at 80ms, based on own research and findings from Barron and Marshall [17–19]. Bradley and Soulodre used the same separation time [20]. However, in later research, Soulodre found the separation time to be frequency dependent and proposed a set of perceptually motivated integration limits, based on the temporal integration properties of the hearing system [21]. In a 2008 paper by Ryu and Jeon, the correlation between C_x is investigated for $10\text{ms} \leq x \leq 200\text{ms}$ [22]. It is found that C_{80} is significantly correlated with C_x for $60\text{ms} \leq x$, and 80ms is chosen as separation limit. Based on these findings and to make comparisons with measured data from concert halls possible, it is decided in this work to opt for the ISO 3382 definition of C_{80} [23].

This discussion can be circumvented when using center time T_s . T_s can be interpreted as the point of gravity of the acoustical energy over time, so no division of the impulse response between early and late sound has to be made.

4.2.2 Reverberation related parameters

The reverberation time RT is most often used by acousticians to describe the energy decay of the sound field. It is defined as the amount of time it would take for the sound field to attenuate 60dB, given the slope of the level function. This level function is determined by Schroeder as the backwards integrated squared impulse response [24]:

$$L(t) = 10 \log_{10} \int_t^{\infty} h^2(x) dx + C_{te} \quad [dB], \quad (4.1)$$

where $h(x)$ is the measured impulse response and C_{te} accounts for the reference pressure and signal power. C_{te} will be omitted in further discussions. Since this level function can have several slope changes, alternative definitions of the reverberation time were postulated. In this study, the early decay time EDT and T_{30} are used.

In general, the sound field of a concert hall is conditioned in such a way that the level function is linearly descending in time, so the EDT and T_{30} have more or less the same value [1]. On an open square this is hardly the case. On these squares, there are only reflections coming from the sidewalls. Because of the size of the squares, only the direct sound and discrete early reflections contain most of the energy, resulting in a kind of sagging decay. The measured EDT will prove to be more location dependent. T_{30} , however, has a more or less constant value over the square, since this parameter is determined by later parts of the reverberant sound field.

An important derived descriptor of the reverberation time is the bass ratio (BR), which allows to investigate the frequency dependency of the reverberation time RT:

$$BR_{RT} = \frac{RT_{125Hz} + RT_{250Hz}}{RT_{500Hz} + RT_{1000Hz}} \quad [-]. \quad (4.2)$$

4.2.3 Binaural related parameters

With the studies of Barron and Marshall the importance of spatial impression was introduced; it describes how the listener experiences the spaciousness of the sound field [19]. They identified factors that relate to the spatial impression, composed of two components: the apparent source width and listener envelopment. Since then, the correlation of the subjective quantities ASW and LEV with variants of objective parameters such as LF, IACC, G and LG has been investigated in many studies [17–21, 25–27].

4.2.3.1 Objective measures for apparent source width ASW

The apparent source width ASW is a measure for the auditory width of the sound field at the square. Early reflections, arriving from the sides, tend to broaden the source, so the stage gains auditory width.

Objective parameters for ASW include LF_E , $[1 - IACC_E]$ and G_{low} [25, 26]. In this study we will use $[1 - IACC_{E3}]$ as a representation for ASW. Here, the subscript $E3$ means that the IACC is calculated from the early sound field (first 80ms) at the three mid-frequency octave bands (500Hz, 1kHz and 2kHz). Beranek called this factor the binaural quality index BQI and it was found to be a better measure of ASW than LF_{E4} [23].

4.2.3.2 Objective measures for listener envelopment LEV

The listener envelopment LEV describes the directionality of the reverberant field and expresses how the listener feels emerged in the sound field. It is known that the listener prefers to be enveloped in the sound field [28].

Different measures for LEV can be found. From Bradley and Soulodre it is known that LEV is mostly influenced by the level and direction of arrival of late reflections, arriving 80ms after the direct sound [20]. It was concluded that LEV correlates well with $IACC_L$ and LF_L , with respect to the influence of angular distribution of the late sound. To combine the influence of the level with angular distribution, the late lateral sound level LG_{L4} was defined. In later work, Soulodre found that it was better to give separate weightings to both the level component and spatial component [21]. Furthermore, he introduced perceptually motivated integration limits for the time separation between early and late arriving sound, as discussed in section 4.2.1. A new measure GS_{perc} , based on a linear combination of both measures and using the separation times, was proposed [21].

In this research we will examine $[1 - IACC_{L3}]$, with the subscript $L3$ restricting the IACC to the late sound (starting from 80ms) and the three mid-frequency octave bands. A second measure, based on an adaptation from Beranek on Soulodre's work, will also be used. Beranek uses the following formula to calculate an

objective value for LEV:

$$\text{LEV}_{calc} = 0.5G_{L3} + 10\log_{10}(1 - \text{IACC}_{L3}) \quad [dB], \quad (4.3)$$

with $G_L = G - 10\log_{10}(1 + \log_{10}^{-1} C_{80}/10)$ [23]. However, since G could not be measured in this study, $\Delta L_{eq,oct}$, as defined in section 4.2.1, will be used instead.

4.3 Measurement methodology

For the measurement of the different quantities, numerous measurement techniques have been developed. However, since most of the described parameters are extracted from the impulse response (by using a HATS or omnidirectional microphone), it is worthwhile to optimize this impulse response measurement. Different methods have been described [29] and standardized [16]. However, for the choice of the test signal, deconvolution technique and post-processing, it should be kept in mind that these measurements will be conducted at large open squares, where the levels of background noise are often high and the transfer functions are not time invariant.

4.3.1 Test signal and deconvolution technique

In order to measure the impulse response, a well-known test-sequence is emitted by the loudspeakers. The commonly used set of test signals can be divided into two groups. A first group uses noise sequences, which excite the full audio frequency range at the same time. White or pink noise sequences and Maximum-Length Sequences MLSs could be used. However, this category seems unfit to use at an open square. Since the levels of background noise are expected to be high at a square, high signal levels should be produced. Furthermore, an ensemble average should be taken, so many impulse response measurements are required. The use of a MLS overcomes these problems, but this signal is very sensitive to time variations in the transfer function (wind, temperature changes, small movements...) and will not be used.

The second class of test signals are the frequency sweeps, linear or exponential. The main advantage is that a sweep excites only one frequency at the time, so the available power is not spread over the full frequency range, compared to the noise test-sequences. This results in an excellent signal-to-noise ratio (SNR). Prolongation of the sweep increases the SNR. Repetition of the sweep is however not advisable, since the corresponding frequency components of each repeated sweep are spread in time and undergo different changes due to the time varying character of the system. An exponential sweep of 12s with frequencies ranging from 10Hz to 16kHz is chosen. This signal duration corresponds to three times the estimated reverberation time. The exponential sweep is chosen over the linear sweep, so that

each octave band is excited with an equal amount of energy.

The extraction of the impulse response out of the recorded sweep (deconvolution) is done by spectral division of the recorded spectrum and the source spectrum, measured at the mixing console. However, since the source signal does not contain any frequency components between 16kHz and 22.05kHz (half of the sampling frequency), care should be taken to exclude this frequency range from the analysis.

4.3.2 Measurement set-up

In order to investigate the variation of the sound field along the square, it is essential that the described acoustical parameters are measured at different locations. The position of these locations, typically five or six, is based on the ISO 3382 standard [14]. Here it is stated that the measurement locations should be chosen randomly over the square, at a minimal distance of a quarter-of-wavelength from each other (min. 2m). Also a minimum distance of 7m from the speakers will be maintained.

The measurements at each selected position are performed in two phases. First, the impulse response measurements and binaural measurements are done (Fig. 4.1(a)). Because it is impossible to play a test-sequence like a sweep during a concert, these measurements took place before the start of the concert. At that moment the background noise was lower. For the determination of the impulse response, the exponential sweep is recorded with an omnidirectional 1/2-inch free-field microphone, type 4189 with a preamplifier type 2669C of Brüel&Kjær. This microphone is wind-shielded and its height is about 1.8m. The measurement of the binaural parameters is done at the same time, by recording the test-sequence with a head-and-torso simulator type 4128C, in which two ear simulators type 4158C are mounted, also of Brüel&Kjær. The signal of these three microphones is recorded simultaneously with the original sweep, as it was sent to the speaker system by the mixing console.

In a second phase, the relative sound pressure level is measured (Fig. 4.1(b)). These measurements took place during the concert in order to include the influence of the audience. For this, a fixed reference SPL measurement is carried out simultaneously with a position dependent SPL measurement. For the reference measurement, a calibrated SVANTEK SVAN 959 soundanalyzer is used with a free-field microphone (microphone capsule MK-250 and SV12L preamplifier). This measurement is carried out near the mixing console, where an open space of about 5m² is available. The position dependent SPL measurement is carried out with a calibrated SVANTEK SVAN 949 soundanalyzer with a similar microphone and preamplifier. At each position the SPL is logged during approx. 10min with an integration time of 1s. Out of each measurement only the useful parts, where



(a) Equipment used for the IACC and impulse response measurement.



(b) Equipment used for the relative SPL measurement.

Figure 4.1: Measurement set-up before and during the concert.

the music is dominant, are selected. After A- or C-weighting of the individual samples, the average energy per octave band is computed. An identical procedure is applied to the corresponding time intervals of the reference measurement. $\Delta L_{eq,oct}$ is then calculated as the difference between the weighted sound levels per octave band at measurement and reference position, while $\Delta L_{eq,A/C}$ is calculated by subtracting $L_{eq,A/C}$ of both positions (see section 4.2.1).

4.3.3 Impulse response post-processing

4.3.3.1 Octave band filtering

After extraction of the impulse response by spectral division, further signal processing is needed in order to extract reliable quantities. First, the impulse response has to be decomposed into its different octave bands, ranging from the 63Hz to the 8kHz-band. For this purpose a filterbank with $2N^{th}$ -order Butterworth octave bandfilters is used, following the ANSI S1.11-2004 standard [30]. The use of 1/3-octave bandfilters will be discarded. Because of their small bandwidth at lower frequencies (11.5Hz for the 50Hz 1/3-octave band), the corresponding filterresponse length will be long, since the bandwidth-time product is constant at a

given filter order. This can pose a problem because at low frequencies the measured impulse responses can be short, especially at large open squares with high losses. Since octave band filters have a larger bandwidth (45 Hz at the 63 Hz octave band) and thus a shorter filter response, they are more suitable. For the choice of the filter order, the bandwidth-time product but also the filter characteristics have to be considered. A $2N^{th}$ -order Butterworth filter has a maximal flatness in the pass-band. In the stopband this filter attenuates according to $Att_S[dB] \approx 20N \log_{10} \omega$, or equivalent, $6N$ dB/octave. It is found that an 8^{th} -order Butterworth octave filter ($N = 4$) has a bandwidth-time product that is small enough for the given octave bands and pass/stop characteristics that meet our requirements.

4.3.3.2 Impulse response truncation

The measured impulse responses will be inevitably contaminated with noise. When calculating the level function and other related parameters, this noise will pose a lower limit to the signal decay. Therefore, the impulse response is processed to select only the useful part. First, the starting point, t_0 , of each band-response is determined. This is the time when the level function of the band-filtered impulse response starts to decay. This starting point can differ for each octave band, since not all frequency bands are reproduced by the same loudspeakers. The endpoint, t_1 , can be found by using a truncation method described by Lundeby et al. [31]. This algorithm is based on an estimation of the crosspoint of the decaying slope and a constant noise level of the time-averaged band-response. Originally, this procedure was designed to operate automatically. In some cases however, the estimation will be adapted manually, since peaks in the background noise can disturb the automatic procedure.

For the calculation of the clarity and center time, the upper limit of the integral will be restricted by the finite length of the truncated band-response. Correction coefficients are introduced to account for this restriction [31]. These are based on the assumption that a regression line can be fitted to the exponential decay of the late energy. Knowing the slope A and the offset B at $t = 0$ of the energy decay $h^2(t) = E(t) = Be^{At}$, the compensation coefficients are calculated in Eq. 4.4, with t_1 the truncation point:

$$E_{comp} = B \int_{t_1}^{\infty} e^{At} dt = -\frac{B}{A} e^{At_1}, \quad (4.4a)$$

$$T_{comp} = B \int_{t_1}^{\infty} t \cdot e^{At} dt = -\frac{B(At_1 - 1)}{A^2} e^{At_1}. \quad (4.4b)$$

Using Schroeder's integrated impulse method [24], the level function can now be calculated from this restricted band-response. However, before the integration of $h^2(t)$, the mean-square value of the noise is subtracted from the squared restricted band-response to cope with faulty estimations of the truncation point [32]. For the

binaural measurements no such corrections are applied, and only the truncated ear responses will be used.

4.4 Case-studies

4.4.1 Description of the squares

In this section the acoustical quality of four different squares in Belgium, Europe, is measured. All four squares are situated in the city-center and are surrounded by different types of facades. Two squares are located in Leuven, where measurements were performed during ‘Marktrock 2008’ at the Oude Markt and the Vismarkt. At both squares, the amplification system was very elaborate. Besides speaker-arrays and subwoofers near the stage, different delay-lines were placed along the square. The Oude Markt is pictured in Fig. 4.2(a). This elongated square is bordered by high 19th-century facades, and has dimensions of 160m \times 85m. The Vismarkt (Fig. 4.2(b)) is a smaller square with dimensions 85m \times 40m and big trees near the edges.

Two other squares were chosen in Mechelen during ‘Maanrock 2008’, namely the Grote Markt and the Vismarkt. No distributed speaker set-ups were used. All speakers, including subwoofers, are located beside or under the stage. The Grote Markt (Fig. 4.2(d)) is the largest square, having dimensions of 100m \times 85m. It is bordered by two-storey buildings near the side, while the backside of the square is partially open. The Vismarkt is the smallest investigated square (Fig. 4.2(c)). The dimensions are 60m \times 25m. At one side it is completely open due to its location near the river Dijle.

Measurements at a fifth square were used to illustrate the influence of the delay-lines on the acoustical parameters. This was possible since this square was very open with almost no reflecting boundaries. The square is situated near the city of Lokeren, where an artificial festival site was created for the ‘Lokerse Feesten’ (Fig. 4.2(e)). The site was approximately 90m \times 70m.

For easy notation the names of the squares will be abbreviated in the remainder of this chapter: LO for Oude Markt (Leuven), LV for Vismarkt (Leuven), MG for Grote Markt (Mechelen) and MV for Vismarkt (Mechelen).

4.4.2 Correlation analysis

Before characterizing the sound field at the different squares, it is useful to define an independent set of parameters. Therefore, we perform an overall correlation analysis of the measured quantities at LO, LV, MG and MV. The parameters measured at the fifth open square will not be included, as this square is not representative for an enclosed urban square and is mainly investigated to characterize the influence of the delay-lines.

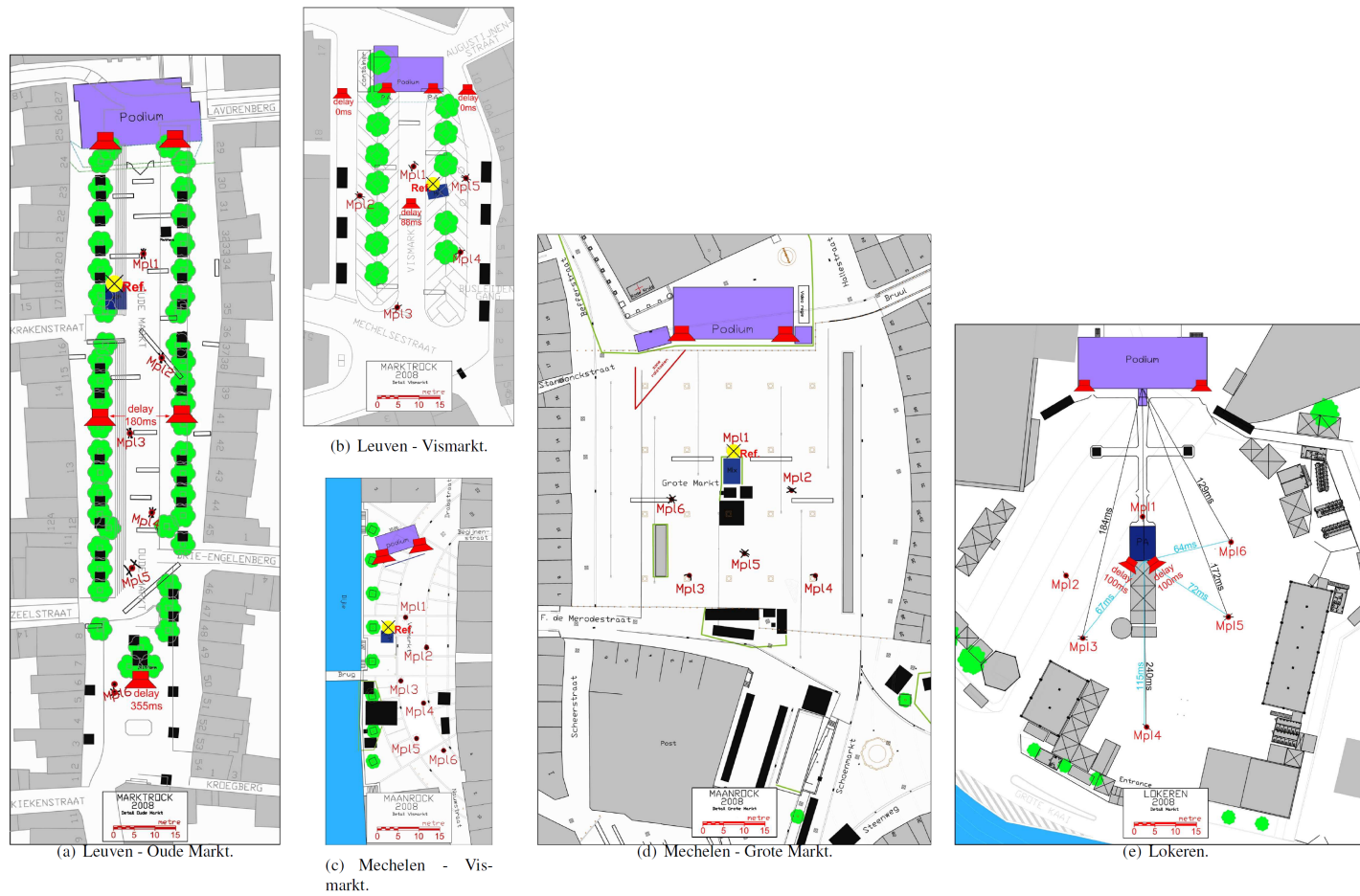


Figure 4.2: Five squares under test: two squares in Leuven during ‘Marktrock 2008’ ((a) and (b)), two squares in Mechelen during ‘Maanrock 2008’ ((d) and (c)) and one square in Lokeren during the ‘Lokerse Feesten’ ((e)). In red the measurement positions and speakers with delay-time are indicated. For the ‘Lokerse Feesten’, the estimated path delay is also given. The yellow dot with black cross indicates the reference position for the relative SPL measurement. The pink rectangle gives the stage’s location.

As already mentioned, the included parameters are the early response parameters C_{80} and T_s , reverberation related parameters EDT, T_{30} and BR, binaural parameters $[1 - IACC_{E3}]$, $[1 - IACC_{L3}]$ and LEV_{calc} , and parameters related to the relative sound pressure level $\Delta L_{eq,oct}$, $\Delta L_{eq,A}$ and $\Delta L_{eq,C}$. However, as C_{80} , T_s , EDT, T_{30} and $\Delta L_{eq,oct}$ are determined for each octave band, we will reduce the data by only correlating their single number averages, calculated from the mid frequency bands (500Hz - 1kHz), based on the ISO 3382 standard [14].

In Table 4.1, it can be seen that $C_{80,avg}$, $T_{s,avg}$ and EDT_{avg} are highly correlated ($p < 0.01$). It is concluded that these parameters have similar fluctuations over the square, related to the delay-lines and geometry. $T_{30,avg}$, on the other hand, is correlated with EDT_{avg} , but does not correlate with $C_{80,avg}$ and $T_{s,avg}$. BR is uncorrelated with the considered parameters. Furthermore, apart from a correlation between LEV_{calc} and $\Delta L_{eq,avg}$, which are correlated by construction, the different binaural parameters are uncorrelated, not only with each other but also with the other objective parameters. For the parameters related to $\Delta L_{eq,oct}$, only $\Delta L_{eq,C}$ is correlated with the reverberation related parameters and $\Delta L_{eq,A}$ is highly correlated with $\Delta L_{eq,avg}$. It should be noted however, that there is little correlation between $\Delta L_{eq,A}$ and $\Delta L_{eq,C}$. From this it is concluded that both measures will have a different distribution over the square. Thus, including the lower frequencies by C-weighting, alters the behavior of the parameter significantly.

From the earlier defined parameters, we can now extract a statistically independent set of parameters, by using C_{80} , T_{30} , BR, $[1 - IACC_{E3/L3}]$, LEV_{calc} and $\Delta L_{eq,A/C}$. This is more or less in accordance with literature, except for T_{30} , which is much more correlated with C_{80} and T_s in performance halls [12, 13].

4.4.3 Influence of the delay-lines at the Lokerse Feesten

To investigate the influence of the delay-lines apart from the influence of the square geometry, measurements were performed at an open festival site, with little to no reflecting boundaries. The speaker set-up consisted of two speaker-arrays at each side of the stage and a group of subwoofers beneath the stage. Two delay-lines were placed at the center of the site, which reproduced the sound with a delay of approximately 100ms. All four line-arrays reproduced the mid and high frequency octave bands but did not contribute to the lowest two octave bands (63Hz and 125Hz). Impulse response measurements were done at six locations along the square. At four of these positions, additional measurements were performed. First, all speakers were turned on, including the delay-lines at the center, with the delay set at the actual value. During the second measurement, only the speakers near the stage were used.

Table 4.1: Correlation coefficients r between the different parameters, as calculated from measured values from measurement positions at LO, LV, MG and MV. Correlations where $|r| \geq 0.6$ are indicated in bold (* indicates $p < 0.05$, ** indicates $p < 0.01$).

	$C_{80,avg}$	$T_{s,avg}$	EDT_{avg}	$T_{30,avg}$	BR_{EDT}	$1 - IACC_{E3}$	$1 - IACC_{L3}$	LEV_{calc}	$\Delta L_{eq,avg}$	$\Delta L_{eq,A}$	$\Delta L_{eq,C}$
$C_{80,avg}$	-										
$T_{s,avg}$	-0.92**	-									
EDT_{avg}	-0.84**	0.81**	-								
$T_{30,avg}$	-0.49*	0.49*	0.68**	-							
BR_{EDT}	-0.06	0.04	-0.02	-0.30	-						
$1 - IACC_{E3}$	0.16	-0.01	-0.31	-0.41	0.18	-					
$1 - IACC_{L3}$	0.48*	-0.34	-0.40	-0.30	0.20	0.38	-				
LEV_{calc}	-0.54*	0.59**	0.54**	0.59**	0.32	-0.15	-0.05	-			
$\Delta L_{eq,avg}$	0.25	-0.09	-0.10	0.22	0.30	-0.09	0.19	0.65**	-		
$\Delta L_{eq,A}$	0.44*	-0.28	-0.32	0.00	0.26	-0.04	0.28	0.46*	0.94**	-	
$\Delta L_{eq,C}$	0.55**	-0.48*	-0.73**	-0.73**	0.31	0.37	0.21	-0.25	0.23	0.49*	-

Results of the impulse response measurements are depicted in Fig. 4.3, with the actual time-delay highlighted in the figure. For comparison, the estimated path travel-time, measured from the center of the stage, is given in Fig. 4.2(e). Comparing both impulse responses shows that the time difference between the arrival of the sound of the main speakers and the delay-lines strongly depends on the position. At the 3th and 4th position, the sound of the delay-lines arrives earlier (10ms at MPL3 and 20ms at MPL4). At the 6th position, the first sound to arrive is coming directly from the stage. 50ms later a delayed peak in the impulse response is seen, clearly weakened because of the off-axis position from the directive delay-lines. At the 5th position, the arrival of the sound from both the stage and delay-lines is simultaneously.

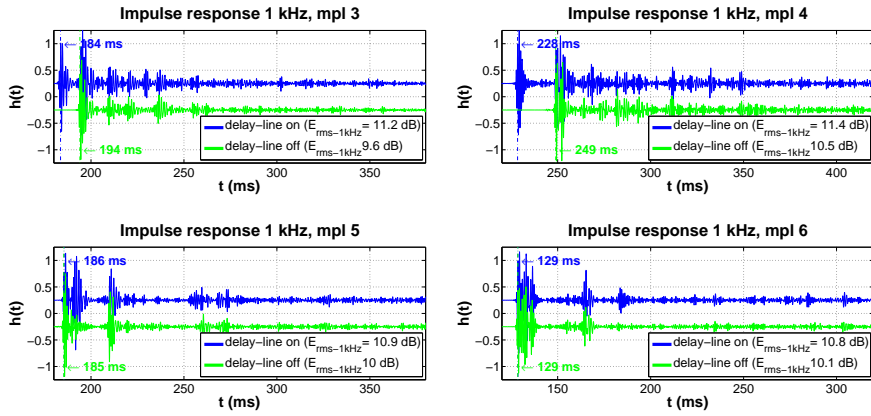


Figure 4.3: Impulse responses, with (blue) and without (green) delay-lines, at four measurement positions at the ‘Lokerse Feesten’ (Fig. 4.2(e)). For both impulse responses the time of arrival of the first sound is indicated, together with the E_{rms} -value of the 1kHz octave band.

This observation is due to have an effect on the acoustical parameters. Fig. 4.4 shows the effect of the delay-lines on C_{80} as a function of frequency at the four selected locations on the square. At the lowest octave bands little or no effect is seen because these frequencies are produced at the same place, namely by the subwoofers beneath the stage and not by the delay-lines. At higher frequencies, the differences when using delay-lines become clear. At MPL3 and MPL5 the clarity increases slightly. These are the two positions close to the line-arrays which have a minimal mismatch in time. At the other positions with a bigger mismatch, MPL4 and MPL6, the clarity becomes lower than in the absence of delay-lines. Analogous conclusions can be drawn from the extraction of the center time and early decay time (not shown).

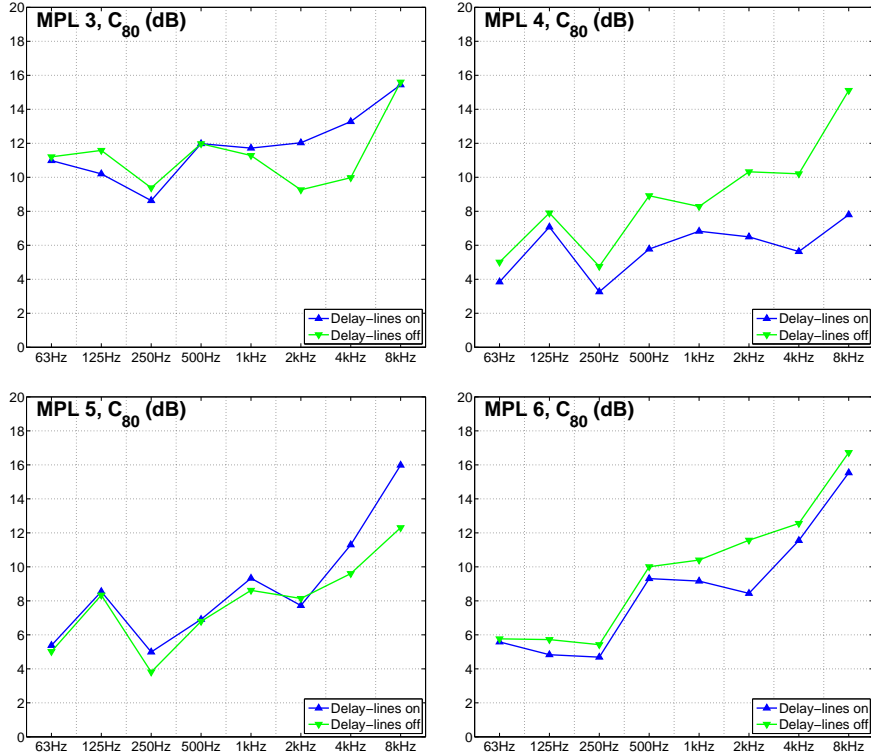


Figure 4.4: C_{80} for each octave band for different measurement positions at the ‘Lokerse Feesten’ (Fig. 4.2(e)), with (blue) and without (green) delay-lines.

Concerning T_{30} , differing conclusions are drawn. As is clear from previous research [33], T_{30} changes only little over space. Fig. 4.5(a) shows that this is valid as well for a festival site, since the deviation from the median value is relatively small. When delay-lines are used, only a limited decrease of T_{30} in the mid frequency bands is observed.

The binaural quality index BQI or $[1 - \text{IACC}_{E3}]$ is more influenced by the delay-lines. It is shifted to higher values because of the broadening effect of the delay-lines, directed to the sides of the square. For $[1 - \text{IACC}_{L3}]$, no change is noticed since this measure is completely determined by the late reverberant sound field, which is not influenced by the delay-lines. This is proved when looking at the impulse responses and T_{30} .

For LEV_{calc} and $\Delta L_{eq,A/C}$ no investigations were made at this square.

This example shows that the introduction of delay-lines can improve the sound quality at the square. However, it should be accounted for that at more distant po-

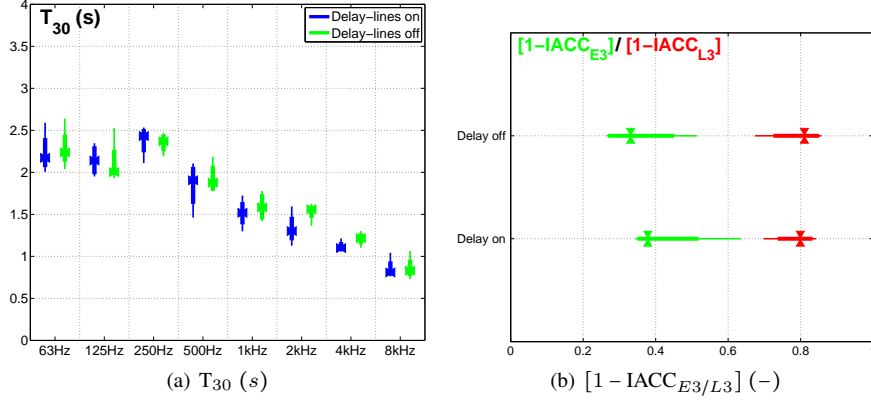


Figure 4.5: T_{30} and $[1 - IACC_{E3/L3}]$ for MPL3-6 at the 'Lokerse Feesten' (Fig. 4.2(e)) plotted in a boxplot in relation to the variation of the parameter over the square. The central mark is the median, the edges of the box are the 25th and 75th percentiles and the whiskers extend to the most extreme data points.

sitions, the quality can worsen as the difference in arrival time grows. In between the main line-arrays and the delay-lines, the disturbing effects will be perceived less, because of the frontal directivity of the speakers, and this in spite of the larger time mismatch.

4.4.4 Comparison of four urban squares

With the knowledge of the influence of the delay-lines at the open site, it is now possible to describe the sound field at urban squares, (partially) surrounded by facades. The chosen squares differ by geometry and amplification system (with or without delay-lines).

The following comparison is based on the statistically independent acoustical parameters measured at each square. Some of these acoustical parameters are function of the measurement position and octave band. To have a global view, the results will be grouped in a boxplot for an easy inter-square comparison. In this boxplot, the central mark indicates the median, the edges of the box are the 25th and 75th percentiles and the whiskers extend to the most extreme data points.

4.4.4.1 Description by clarity C_{80}

The first analysis is based on the clarity factor C_{80} (Fig. 4.6(a)) and, related with this factor, EDT (Fig. 4.6(b)). In the discussion below, a distinction is made between the low octave bands (63Hz-250Hz), mid octave bands (500Hz-2kHz) and high octave bands (4kHz-8kHz).

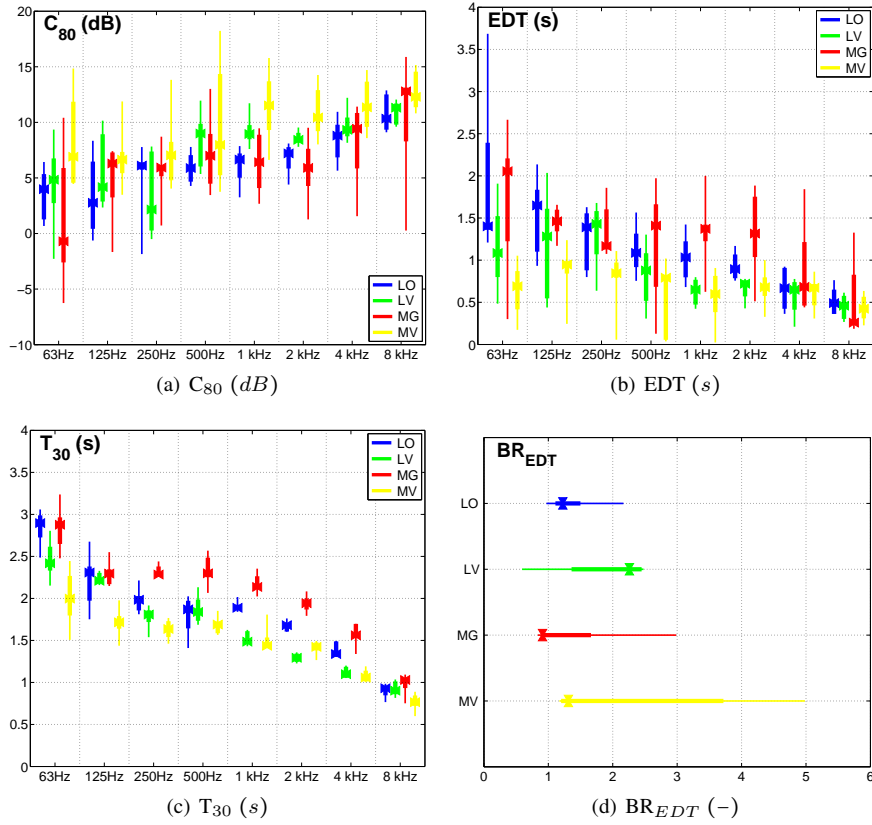


Figure 4.6: C_{80} , EDT, T_{30} and BR_{EDT} at the four different squares plotted in a boxplot in relation to the spatial variability of the parameter. The central mark is the median, the edges of the box are the 25th and 75th percentiles and the whiskers extend to the most extreme data points. LO stands for Oude Markt (Leuven), LV for Vismarkt (Leuven), MG for Grote Markt (Mechelen) and MV for Vismarkt (Mechelen).

At low octave bands, large deviations of the measured quantities are seen at each square. This is a consequence of the centralized subwoofer system beneath the stage: the energy confined to the early sound decreases with the distance to the stage, while the reverberant energy increases due to the more frequent late reflections. In a later section, special attention is given to MG, where the parameters at the 63Hz band vary the most.

At mid and high octave bands the range in measured values becomes smaller. Especially the quantities measured at LO and LV are quite constant over the square. At both of these squares delay-lines are used, which can ensure less spatial fluctuation of the parameters, even at a very elongated square such as LO. In Meche-

len, no delay-lines were used, causing larger fluctuations in the measured quantities. However, at higher frequencies the deviations from the median value become smaller at MV, the smallest square. This is not the case at MG, the largest square, where obstacles like screens shield measurement locations from high frequencies (MPL5). Since this is not compensated by the use of delay-lines, large variations in the observed parameters were found. The highest values for EDT are found at this square, not only at places shielded from direct sound, but also further away from the speakers.

From these data it can be seen that the smallest squares, MV and LV, have the best behavior for electronically controlled music reproduction. C_{80} is high and EDT low, relative to the other squares. Early reflections contain more energy, since they suffer from less path loss at a smaller square. The use of delay-lines at the large LO helps to ameliorate the parameters at places with inferior acoustics. By this, a less fluctuating parameter course is obtained, in contrast to the largest square at Mechelen.

4.4.4.2 Description by T_{30} and BR_{EDT}

As described earlier, T_{30} is independent of measurement position and speaker placement. It is thus an ideal measure to make a comparison of the squares, solely based on their geometry. Generally it is seen that T_{30} increases with the size of the square (Fig. 4.6(c)). The size of LV and MV is comparable (with MV adjacent to a river) and this is reflected in a similar magnitude of T_{30} . At LO T_{30} is higher, because of the length of the square. The highest values for T_{30} are found at MG. This is not only caused by its dimensions, but also because this square is adjacent to another open area.

The broad range of BR_{EDT} , although proven to be uncorrelated with EDT, is related to the range of EDT fluctuations. No strong conclusions can be drawn, but it seems that the variation of bass ratio is smallest at squares with a delay-line. The bass ratio extracted out of T_{30} shows to be concentrated around the median (not shown). Somewhat higher values are observed at both squares in Leuven, but a relation with the shape and size of the square is not easy to find.

4.4.4.3 Description by $[1 - IACC_{E3/L3}]$ and LEV_{calc}

In Fig. 4.7, the different binaural parameters are shown. Fig. 4.7(a) depicts $[1 - IACC_{E3}]$, which is related to the apparent source width, together with $[1 - IACC_{L3}]$, describing the angular dependency of the listener envelopment.

For $[1 - IACC_{E3}]$ (or BQI) the difference between the squares in Leuven and the squares in Mechelen (no delay-lines) is clear. $[1 - IACC_{E3}]$ is higher at both squares in Leuven because of the broadening effect of the delay-lines. $[1 - IACC_{L3}]$, on the other hand, is quite high and constant at all squares considered,

because the high diffusivity of the facades causes the reverberant sound to reflect in various directions.

Fig. 4.7(b) shows that LEV_{calc} differs much more between squares than $[1 - IACC_{L3}]$. The discriminating factor for LEV_{calc} is not the angular distribution, which is already quite high at all squares, but the difference in relative sound pressure level and C_{80} (Eq. 4.3). LEV_{calc} is highest at MG, the largest square without delay-lines. This can be explained by the large amount of energy still contained in the reverberant field (low C_{80} subtracted from $\Delta L_{eq,oct}$ in Eq. 4.3). At the smallest square, MV, LEV_{calc} is lowest (high C_{80}). However, at both squares with delay-lines, the median value of LEV_{calc} is more or less equal.

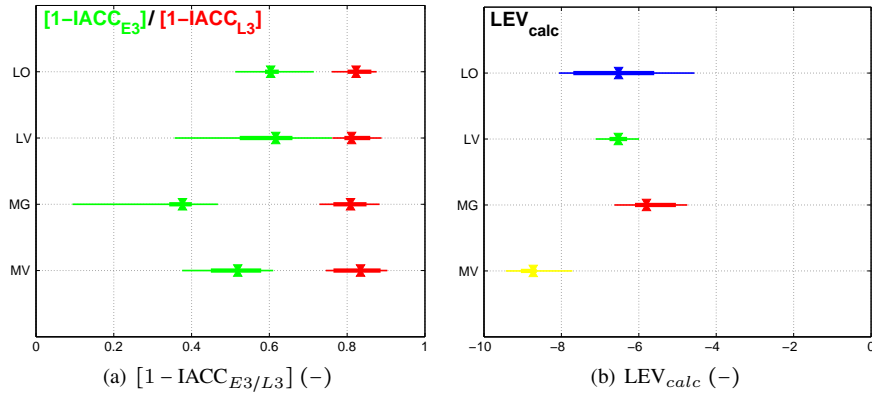


Figure 4.7: $[1 - IACC_{E3/L3}]$ and LEV_{calc} at the four different squares expressed by boxplots in relation to the variation of the parameter over the square. The central mark is the median, the edges of the box are the 25th and 75th percentiles and the whiskers extend to the most extreme data points. LO stands for Oude Markt (Leuven), LV for Vismarkt (Leuven), MG for Grote Markt (Mechelen) and MV for Vismarkt (Mechelen).

4.4.4.4 Description by $\Delta L_{eq,A/C}$

The last parameters to describe the sound field at the different squares are based on the relative sound pressure level. In contrast to previously discussed parameters, these measurements are performed during the concert and thus include the influence of the audience. In Fig. 4.8 the difference in L_{eq} between different positions at the square and the reference position is plotted for both A- and C-weighting for a real-life rock music spectrum. At LO, LV and MG measurements were done during two ten-minute periods, of which only the music samples were selected for calculation of L_{eq} , and the line indicates the average taken over these two measurements. The position of the delay-lines is indicated by the vertical lines.

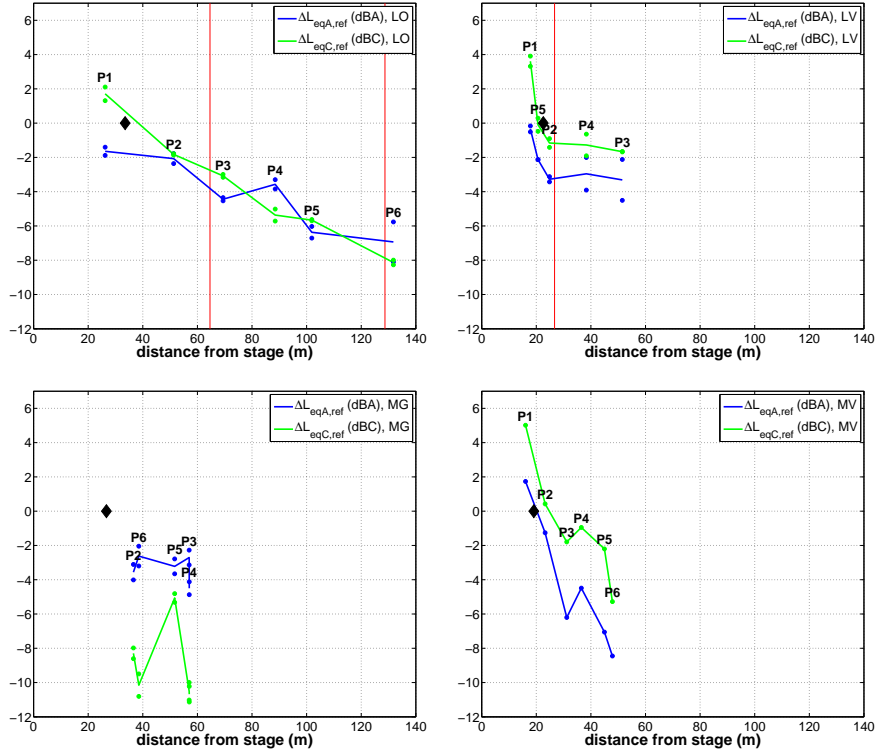


Figure 4.8: Relative sound pressure level ($\Delta L_{eq,A}$ (blue) and $\Delta L_{eq,C}$ (green)) as a function of the distance from the stage. The measurement positions are indicated as P1-6 and the reference position is represented by a black diamond-shaped marker. The vertical lines indicate the positions of the delay-lines. LO stands for Oude Markt (Leuven), LV for Vismarkt (Leuven), MG for Grote Markt (Mechelen) and MV for Vismarkt (Mechelen).

As is clear from Fig. 4.8, using delay-lines results in a more uniform distribution of the sound pressure level over the square. The positive effect of the delay-line at LV is very pronounced. At LO, a linear decrease of level with distance is observed. At the squares in Mechelen, delay-lines are not present. Although MV is a small square, ΔL_{eq} decreases rapidly with distance; the half-openness of the square is accounting for this. The distribution of the sound pressure level at MG however jumps out. The fluctuations in $\Delta L_{eq,A}$ are caused by the non-symmetrical square geometry. Since the back of the square is only halfway closed, as it borders another square, more energy is lost at this side, compared to the energy loss at the closed side. This accounts for low levels at MPL2 and MPL4, in contrast to MPL3 and MPL6, which are located close to reflective structures.

From the correlation analysis, it was concluded that the correlation between $\Delta L_{eq,A}$ and $\Delta L_{eq,C}$ is small. This is reflected in differences in distribution of both parameters over the squares. For LO, LV and MV the course of $\Delta L_{eq,A}$ and $\Delta L_{eq,C}$ as a function of the distance from the stage contains limited, but noticeable, differences. At MG however, high differences are observed when applying different frequency weighting schemes. In A-weighted levels, acoustical energy in the lower octave bands is less important. When looking at $\Delta L_{eq,C}$ as a function of the distance from the stage, distinct peaks are found at MPL2 and MPL5, in contrast to the A-weighted level. These findings were confirmed during a second sound pressure level measurement campaign at MG in 2010 (not shown). An explanation can be found in the specific situation at this square. Points MPL1/REF, MPL2 and MPL5 are situated above an underground parking lot, which could act as a kind of coupled resonator. The other points are less sensitive to these vibrations since these are at the border of the parking lot. Indeed, many bystanders were complaining about annoying low frequency vibrations. These vibrations could even be felt on the square surface during music passages with a large amount of low frequency energy.

Based on these findings, it is clear that frequency weighting should not be restricted to A-weighting. The use of C-weighted sound pressure levels gives added value to the evaluation of the sound field, given the often prominent low frequencies at rock concerts.

4.5 Conclusion

In this study, an extensive measurement campaign has been set-up to characterize the sound field at urban squares. Different acoustical parameters were measured before and during rock concerts held at such squares. From this data, different statistically independent factors were identified by correlation analysis. It was found that C_{80} , T_s and EDT are highly correlated, but not correlated with T_{30} , in contrast to findings in performance halls [12, 13]. At rock concerts held on urban squares, $\Delta L_{eq,A}$ and $\Delta L_{eq,C}$ were also found to have a low correlation. The investigated independent parameter set contained C_{80} , T_{30} , BR, $[1 - IACC_{E3/L3}]$, LEV_{calc} and both $\Delta L_{eq,A}$ and $\Delta L_{eq,C}$. This selection of parameters is obviously influenced by the set that was initially selected. Other researchers [8, 9] started from different sets and found comparable dimensionality. In particular T_{sub} , IACC and listening level are closely related to T_{30} , $IACC_{E/L}$ and relative sound level respectively.

In order to estimate the influence of the delay-lines on the sound field, measurements were done at an open festival site. It is clearly seen that the effect on the parameters, except for T_{30} , strongly depends on the considered measurement positions. Near the line-arrays, an improvement is seen. However, at distant locations, the quality decreases due to large time-shifts between the arrival of sound

coming from different sources. This became clear when evaluating C_{80} . As for $[1 - IACC_{E3}]$, an increase is seen at all locations when delay-lines are introduced. T_{30} is found to be an interesting parameter to compare the sound field at different squares. This parameter is strongly dependent on the square geometry. Furthermore, it was shown that T_{30} is only little dependent on the location at the square and is almost not influenced by the amplification system and the presence of delay-lines.

A comparison between four urban squares is made, based on statistically independent acoustical parameters extracted from impulse response measurements (before the concerts) and sound pressure level measurements (during the concerts). In absence of distributed speaker-arrays, the sound quality at smaller squares is higher than at larger squares. The introduction of delay-lines can greatly improve the acoustics, even at large elongated squares. For the design of these speaker systems, the sound engineer can make use of different parameters. C_{80} tends to be most dependent on the considered location and speaker set-up. Squares with delay-lines also have higher $[1 - IACC_{E3}]$, related to ASW and BQI. However, concerning LEV_{calc} , the determining factor is $G_L (\Delta L_{eq,oct})$ and not the angular distribution, which is already high at all squares. Relative sound pressure level measurements are useful to describe the level decrease over the square. The use of C-weighted levels was shown to indicate potential problems with strong low frequency passages during concerts. The distribution of the acoustical energy over time cannot be derived when considering relative sound pressure levels only. Nevertheless, this parameter is nowadays the only commonly used design parameter.

In this chapter, room acoustical parameters were measured at urban squares in order to characterize them acoustically. Statistically independent parameters that are sensitive to the specific square geometry and amplification system were identified. Further research should link these parameters to a subjective evaluation of the sound quality at different squares and give more insight in the preferred parameter values for rock music [34, 35]. Related studies have already been undertaken for rock concert halls [3, 36] and classical concert halls [8, 11], but not yet for open-air concerts at public squares.

References

- [1] D. Paini, A. Gade, and J. Rindel. *Is reverberation time adequate for testing the acoustical quality of unroofed auditoriums?* In Proceedings of the 6th international conference on auditorium acoustics, 2006.
- [2] L. Beranek. *Audience and seat absorption in large halls.* J. Acoust. Soc. Am., 32:661–670, 1960.

- [3] N. Adelman-Larsen, E. Thompson, and A. Gade. *Acoustics in rock and pop music halls*. In Proceedings of the 112th AES Convention, 2007.
- [4] *L'Acoustics - Soundvision*. www.l-acoustics.com.
- [5] *Meyer - MAPP Online Pro*. www.meyersound.com.
- [6] M.R. Schroeder, D. Gottlob, and K.F. Siebrasse. *Comparative study of European concert halls: correlation of subjective preference with geometric and acoustic parameters*. J. Acoust. Soc. Am., 56:1195–1201, 1974.
- [7] L. Cremer and H. Mueller. *Principles and applications of room acoustics*. Hirzel, Stuttgart, 1978.
- [8] Y. Ando. *Calculation of subjective preference at each seat in a concert hall*. J. Acoust. Soc. Am., 74:873–887, 1983.
- [9] M. Tohyama, H. Suzuki, and Y. Ando. *The nature and technology of acoustic space*. Academic Press, 1995.
- [10] L. Beranek. *Concert halls and opera houses*. Springer - Verlag, 2004.
- [11] L. Beranek. *Subjective rank-orderings and acoustical measurements for fifty-eight concert halls*. Acta Acustica united with Acustica, 89:494–508, 2003.
- [12] A.C. Gade. *Prediction of room acoustical parameters*. J. Acoust. Soc. Am., 89:1857(A), 1991.
- [13] S. Cerda, A. Gimenez, J. Romero, R. Cibrian, and J.J. Miralles. *Room acoustical parameters: A factor analysis*. Applied Acoustics, 70:97–109, 2009.
- [14] *ISO 3382-1:2009(E), Acoustics - Measurement of room acoustic parameters - Part 1: Performance spaces*. International Organization for Standardization, Geneva, Switzerland, 2009.
- [15] Y. Ando and D. Gottlob. *Effects of early multiple reflections on subjective preference judgments of music sound fields*. J. Acoust. Soc. Am., 65:524–527, 1979.
- [16] *ISO 18233:2006(E), Acoustics - Application of new measurement method in building and room acoustics*. International Organization for Standardization, Geneva, Switzerland, 2006.
- [17] T. Hidaka, L. Beranek, and T. Okano. *Interaural cross-correlation, lateral fraction and low- and high-frequency sound levels as measures of acoustical quality in concert halls*. J. Acoust. Soc. Am., 98:988–1007, 1995.

- [18] M. Barron. *The subjective effects of first reflections in concert halls - The need for lateral reflections*. J. Sound. Vib., 15:475–494, 1971.
- [19] M. Barron and A.H. Marshall. *Spatial impression due to early lateral reflections in concert halls*. J. Sound. Vib., 77:211–232, 1981.
- [20] J. S. Bradley and G. A. Soulodre. *Objective measures of listener envelopment*. J. Acoust. Soc. Am., 98:2590–2597, 1995.
- [21] G.A. Soulodre, M.C. Lavoie, and S.G. Norcross. *Objective measures of listener envelopment in multichannel surround systems*. J. Audio Eng. Soc., 51:826–840, 2003.
- [22] J.K. Ryu and J.Y. Jeon. *Subjective and objective evaluations of a scattered sound field in a scale model opera house*. J. Acoust. Soc. Am., 124:1538–1549, 2008.
- [23] L. Beranek. *Concert hall acoustics - 2008*. J. Audio Eng. Soc., 56:532–544, 2008.
- [24] M.R. Schroeder. *New method of measuring reverberation time*. J. Acoust. Soc. Am., 37:409–412, 1965.
- [25] T. Hidaka, T. Okano, and L. Beranek. *Interaural cross correlation (IACC) as a measure of spaciousness and envelopment in concert halls (A)*. J. Acoust. Soc. Am., 92:2469, 1992.
- [26] T. Okano, L. Beranek, and T. Hidaka. *Relations among interaural cross-correlation coefficient, lateral fraction, and apparent source width in concert halls*. J. Acoust. Soc. Am., 104:255–265, 1998.
- [27] M. Morimoto and Z. Maekawa. *Effects of low frequency components on auditory spaciousness*. Acustica, 66:190–196, 1988.
- [28] L. Beranek. *Concert and opera halls*. Acoust. Soc. Am., 1996.
- [29] G. Stan, J. Embrechts, and D. Archambeau. *Comparison of different impulse response measurement techniques*. J. Audio Eng. Soc., 50:249–262, 2002.
- [30] ANSI S1.11-2004, *Specification for octave-band and fractional-octave-band analog and digital filters*. Acoust. Soc. Am., Melville, New York, 2009.
- [31] A. Lundeby, T.E. Vigran, H. Bietz, and M. Vorländer. *Uncertainties of measurements in room acoustics*. Acustica, 81:344–355, 1995.
- [32] W. T. Chu. *Comparison of reverberation measurements using Schroeder's impulse method and decay-curve averaging method*. J. Acoust. Soc. Am., 63:1444–1450, 1978.

-
- [33] D. Paini, J. Rindel, A. Gade, and G. Turchini. *The acoustics of public squares/places: a comparison between results from a computer simulation program and measurements in situ*. In Proceedings of the 33rd international congress and exposition on noise control engineering (Inter-Noise 2004), 2004.
 - [34] Y. Ando, T. Okano, and Y. Takezoe. *The running autocorrelation function of different music signals relating to preferred temporal parameters of sound fields*. J. Acoust. Soc. Am., 86:644–649, 1989.
 - [35] S. Kaieda, K. Kawai, T. Yano, and Y. Ando. *A study on measures of timbre of electric guitar sounds in terms of power spectrum and auto correlation function*. J. Temporal Des. Arch. Environ., 9:43–46, 2009.
 - [36] N. Adelman-Larsen, E. Olmos, and A. Gade. *Acoustics in halls for rock music*. In Proceedings of the Joint Baltic-Nordic Acoustics Meeting 2004, 2004.

5

Reverberation-based urban street sound level prediction

**P. Thomas, T. Van Renterghem, E. De Boeck, L. Dragonetti and
D. Botteldooren**

Published in the Journal of the Acoustical Society of America, 133(6):3929-3939,
June 2013



Street reverberation can strongly increase sound pressure levels in urban streets. In noise mapping models on the other hand, including a sufficient number of reflections is computationally very costly. A simple regression model is developed in this study, based on typical street width and facade height or facade roughness in old city centers of many European cities. This model is based on measurements with a mobile set-up in 99 streets in the city of Ghent, Belgium, representative for common European city morphology. The reflection ratio (RR) is used as a parameter to determine the energy contained in the reverberant part of the sound field. The model has been successfully validated by detailed measurements at different source-receiver distances in an urban street canyon.

5.1 Introduction

Sound propagation in urban street canyons is strongly influenced by multiple reflections in between building facades, leading to an important amplification of the sound relative to open terrain. This effect is important since the acoustic properties of the materials appearing in streets (concrete, bricks, windows, etc.) are close to acoustically rigid.

As a consequence, this leads to increased sound pressure levels, and to increased reverberation times [1, 2]. Both aspects can be linked to an increased negative perception of noise. Street reverberation can be noticed especially for the case of a single passage of a motorized vehicle.

Realistic street geometries can be rather complex. It is expected that when streets are very wide, or the amount of open spaces in between buildings bordering the street is large, this enhancement will be much smaller. Furthermore, the effect of the finite length of streets, and the presence of cross-streets could potentially influence this reverberation gain. Also, scattering at building facade irregularities, building up diffuse sound fields in urban streets, is known to be an important aspect of urban street acoustics [3–8]. For narrow street canyons, interference effects play a significant role [9].

Noise maps are nowadays an important management tool for abating noise. However, producing urban noise maps comes at a high computational cost, mainly due to the large spatial extent one usually aims at. The use of simplified models, like e.g. ISO 9613-2 [10], does not relax this to a sufficient degree. As a practical, but inaccurate solution, the number of reflections is commonly reduced to only a few. In this study, it is investigated to what extent the gain in sound pressure level due to the multiple reflections between building facades (and on the street surface) can be approached by simplified models, taking into account street dimensions and street characteristics only. The final goal, which is not the subject of this chapter, is to use such a model to produce urban noise maps: only direct contributions from the source to the receiver points should then be considered, and the street amplification is added afterwards. The potential application of such a model is illustrated by means of a set of measurements performed in a real street.

A ray-tracing based approach has been used by Heutschi [11] to derive a simplified model of ‘building correction’ for a number of typical street geometries. Relevant parameters were the height of the facades, width of the street canyon (distance between the facades), the absorption coefficient of the facades, the degree of diffusion for reflections from the facades, and the source and receiver position. A correction was proposed for interrupted rows of buildings bordering a street. Here, a study with a similar purpose has been performed, which is now based on a large set of in-situ measurements in urban streets with a dedicated measurement set-up. A model of similar complexity as in Heutschi’s work is aimed at.

This chapter is organized as follows. In section 5.2, the measurement set-up to determine the acoustical parameters in street canyons is described. The reflection ratio will be defined in section 5.3: a prediction model for this parameter will be given, based on geometrical and architectural characteristics. With the reflection ratio, an estimate of the course of the sound pressure level along the length of a street canyon can be made. A description of the model and validation in a real street canyon is given in section 5.4.

5.2 Measurement set-up

In order to be able to draw statistically relevant conclusions regarding the acoustical behavior of street canyons as a function of geometry, a vast amount of different streets needs to be investigated. This requires fast and easy impulse response measurements that do not disturb the traffic for too long. Therefore, the measurement set-up was mounted on the roof of a car (Fig. 5.1). This set-up contains an omnidirectional source with twelve speakers positioned on a dodecahedron. The source is mounted on the middle of the roof. Two free-field microphones of Brüel&Kjær, Type 4189, are placed at each side of the source, at a distance of 2.48m. In this way, the measurement set-up becomes very mobile, while two measurements can be done simultaneously to introduce a minimum of spatial averaging. It further allows error detection, and investigating the influence of the body of the car on the measurements.

5.2.1 Impulse response extraction

As measurements will take place in noisy outdoor environments, the choice of the excitation signal is critical. Therefore, a 30s-long exponential sweep with frequency span covering the 63Hz to 16kHz octave band was used. Since this type of signal excites only one frequency at a time, full sound power can be applied for each frequency, leading to improved signal-to-noise ratios compared to noise-based techniques, where the sound power is distributed over all frequencies [12]. Furthermore, as the increase of frequency is exponential, each octave band contains an equal amount of energy. To increase the signal-to-noise ratio even more, the emission levels can be raised and averaging techniques can be applied.

The impulse response is calculated from the spectral division between the recorded and emitted sweep. Here, we will focus on an octave band analysis of the parameters. For the octave band decomposition, a filterbank is used with 8th-order Butterworth octave band filters, following the ANSI S1.11-2004 standard [13].

As the band-filtered impulse responses are likely to be contaminated with noise due to the outdoor environment, degenerating the quality of extracted parameters, start and endpoints need to be estimated [14]. The start point of the response is

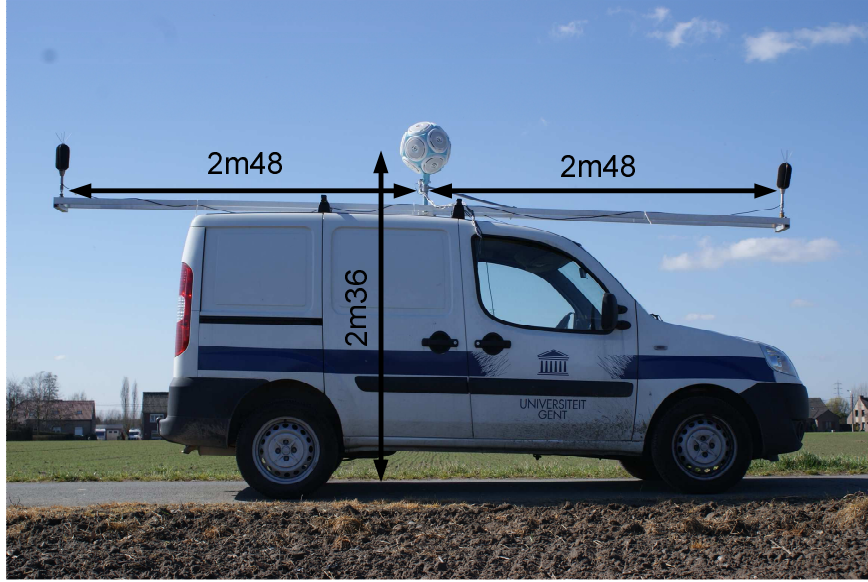


Figure 5.1: Measurement set-up with car for impulse response measurements. An omnidirectional source is placed on the middle of the roof. A free-field microphone (B&K Type 4189) is placed at 2.48m on each side of the source.

given by the time when the level function of the impulse response, calculated by using the backwards integration technique [15], starts to decay. The endpoint or truncation point is estimated as the intersection between the decaying slope and averaged background noise level of the time-averaged squared band response [16].

5.2.2 Reflection ratio $RR(t_0)$

From the band-filtered responses, several (room) acoustical parameters can be extracted. Also, in an urban environment, such parameters were shown to be useful to characterize the sound field [17, 18]. In this work however we will focus on the reflection ratio $RR(t_0)$, a parameter which will prove to be useful for simplified noise map calculation:

$$RR(t_0) = 10 \log_{10} \left(\frac{\int_{t_0}^{\infty} h^2(\tau) d\tau}{\int_0^{t_0} h^2(\tau) d\tau} \right) \quad [dB]. \quad (5.1)$$

The reflection ratio is the ratio of the sound energy of reflections arriving at the listener between time t_0 and $+\infty$ and the energy of the sound arriving from start to time t_0 and is extracted from the impulse response $h(t)$. The parameter is very useful when the sound intensity level of the direct sound or early sound field,

$L_{early}(t_0)$, is known, since the level of the remaining part of the sound field (in [dB]) can be calculated as $L_{reflections} = RR(t_0) + L_{early}(t_0)$, with the dividing parameter t_0 indicating the span of the early sound field.

Fig. 5.2 depicts the typical course of the reflection ratio in function of t_0 for the different octave bands in a 5.4m wide street. The curve decays each time a reflection arrives at the listener, raising the energy in the denominator of $RR(t_0)$. In the early part of the reflection ratio, a stepwise decay is clearly visible, as the early sound field is built up out of discrete reflections from the facades. For higher t_0 , the decay of $RR(t_0)$ is smoother as multiple reflections in the reverberant sound field arrive more continuously.

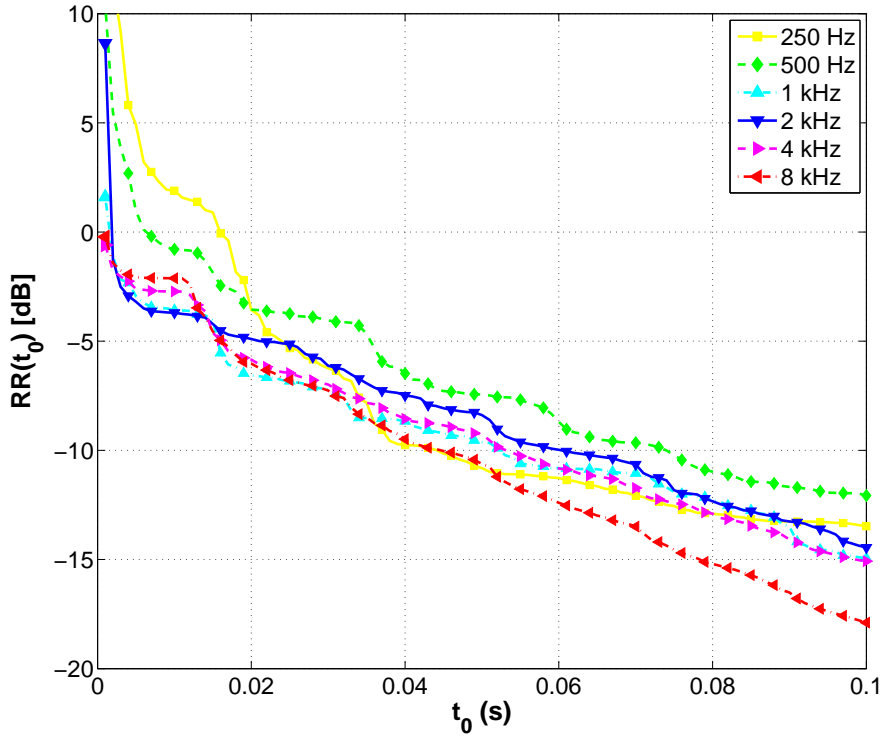


Figure 5.2: Reflection ratio $RR(t_0)$ for the 250Hz to 8kHz octave band in a 5.4m wide street. The arrival of early reflections, relative to the direct sound, can be noticed by sudden drops in the curve at multiples of the street width travel-time (16ms).

Assuming an image source model for geometrically reflecting facades [5], it can be deduced that the steps of RR-decay are approximately located at multiples of the travel-time between two opposing facades. For a simple model, with source and receiver at the same height in between two reflecting facades, the energy of

the first reflection is given by

$$E_{1st} = \frac{1 - \alpha}{4\pi(w^2 + \Delta_{ref}^2)}, \quad (5.2)$$

with α the absorption coefficient of the facades, w the street width and Δ_{ref} the source-receiver distance. This expression can be generalized for the n^{th} reflection as

$$E_{n^{th}} = \frac{(1 - \alpha)^n}{4\pi((nw)^2 + \Delta_{ref}^2)}. \quad (5.3)$$

The travel-time needed for the n^{th} reflection to arrive at the receiver is given as $t_{n^{th}} = c^{-1}\sqrt{(nw)^2 + \Delta_{ref}^2}$, with c the speed of sound. The source-receiver distance Δ_{ref} in the measurement set-up is (much) smaller than nw , so $t_{n^{th}}$ can be approximated as $t_{n^{th}} \approx c^{-1}nw$.

5.2.3 Validation in half-open space

All impulse response measurements and derived parameters are valid only if the influence of the measurement car is much lower than the influence of the surroundings. In order to investigate this, impulse response measurements have been performed in an open rural area, without reflecting surfaces, except for the soil and the car itself. These reference measurements were conducted repeatedly at the start and the end of each measurement campaign in the city, in order to check the measurement set-up. In total, 57 measurements in open field have been carried out over nine campaigns.

The average reflection ratio at the front end and rear end, together with the standard deviation, is plotted in Fig. 5.3 as a function of dividing time t_0 . At the lowest octave band (63Hz), the standard deviation becomes too large (approx. 5dB and even larger than 7dB for $t_0 > 80ms$). This is due to the source, which is not suited to reproduce such low frequencies. Measurements at this octave band will not be taken into consideration.

At the other octave bands, the standard deviation for front and rear measurements remains limited to maximal 3.5dB, even for high values of t_0 . In general, both front and rear reflection ratios are alike and the small offset between front and rear (approx. 1.2dB) remains within the standard deviation. However, for measurements in the 500Hz and 1kHz octave band, the difference between the average front and rear reflection ratio becomes higher (2.2dB at 500Hz and 4dB at 1kHz). It is concluded that the shape of the car has some (small) influence at the 1kHz (and 500Hz) octave band. For further analysis this will not be accounted for and the averaged reflection ratio over front and rear reflection ratios will be used. Furthermore, as will be shown, the reflection ratio measured at the open field proves to

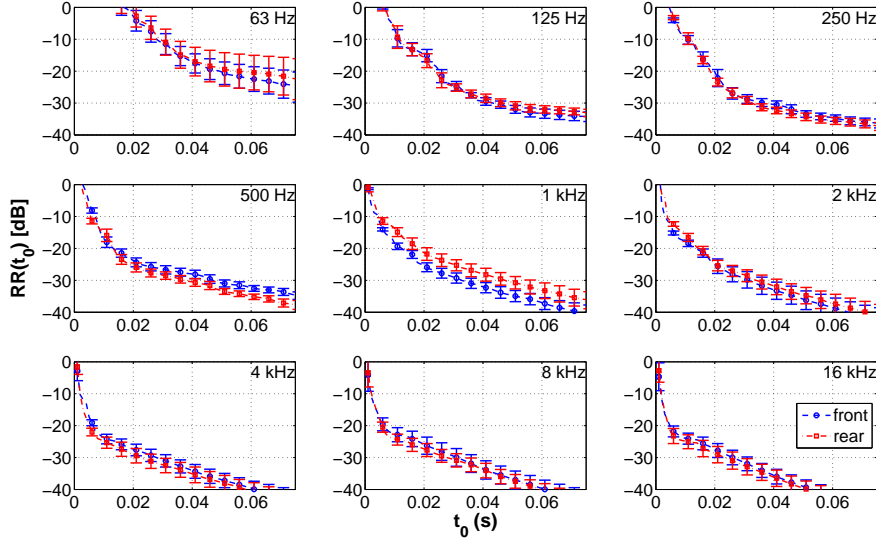


Figure 5.3: Mean reflection ratio for 57 two-sided measurements in an open field, performed on different days. $RR(t_0)$ is depicted as a function of the dividing time t_0 for the 63Hz - 16kHz octave band. Round markers (\circ) indicate the course of $RR(t_0)$ at the front end of the car and square markers (\square) indicate the course of $RR(t_0)$ at the rear end, with error bars indicating the standard deviation from the mean value.

be much lower than the different reflection ratios measured in urban streets. Thus, the influence of the car and measurement set-up is small compared to the influence of the acoustics of urban streets and it is assumed that the measured reflection ratio is mainly determined by the acoustics of the street.

5.3 Street reflection ratio measurements

Impulse responses at 99 different locations in the city of Ghent, Belgium, are measured, while different geometrical aspects of the location were noted. Where possible, the car is positioned in the center of the street. Measurements were repeated at least two times at each location, and averaged afterwards. Fig. 5.4 shows a map with the different locations where measurements were performed.

For the reflection ratio, $RR(t_0)$, the dividing time t_0 is chosen so that only the direct sound is contained in the denominator, while the energy of all reflections is contained in the numerator. In this way, the energy of the reflections can be derived from the energy of the direct sound. A dividing time t_0 of 12ms is chosen as a compromise between being able to measure $RR(t_0)$ in small streets and being able to ignore free field effects of the measurement set-up. The chosen t_0 corresponds

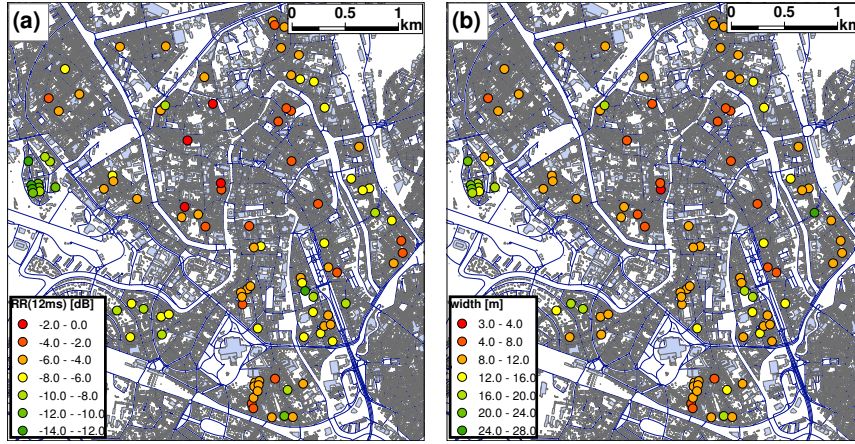


Figure 5.4: Map indicating the location of the 99 measurement positions in the city of Ghent, Belgium. (a) Measured $RR(12ms)$ for the 1kHz octave band. (b) Street width at each measurement location.

to the travel time of the first reflection of a 4m wide street. Yet $RR(12ms)$ of the measurement set-up remains lower than -15dB at frequencies above 500Hz.

5.3.1 Creating a model for $RR(12ms)$

In this section, a model based on different geometrical and architectural properties of the streets will be developed. The main geometrical characteristics include the street width (ranging between 3m and 25m) and the average height of the facades (ranging between 5.5m and 15m). In order to include the architectural properties, a visual categorization of the typical roughness of the street facades has been carried out. Initially, four categories were distinguished (Fig. 5.5). Streets with facades without regressions or protrusions relative to the facade plane (apart from where the windows and doors are located), are placed in a first category. The second category contains streets with facades with little relief and small ornaments, while in the third category streets with facades containing one balcony are classified. Streets where buildings have more than one bay window or balcony are classified in a fourth category. After location selection, large differences in category number count were noticed, so it was decided to keep category 1 (57 counts), renamed as category A, and combine category 2, 3 and 4 (42 counts) into category B. Although this is a rather crude approach, it allows studying whether including facade scattering increases correlation coefficients of the models.

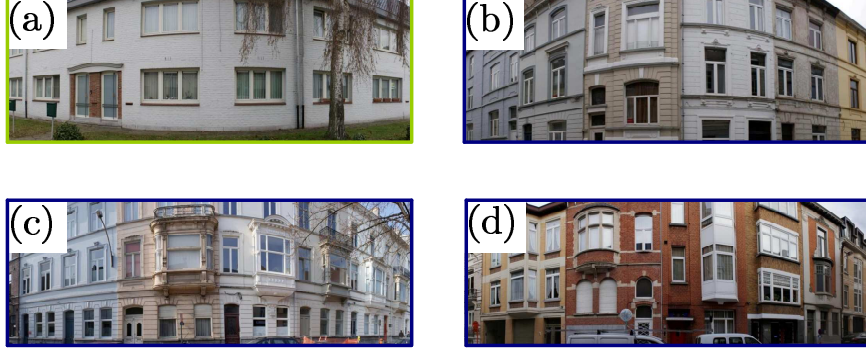


Figure 5.5: Categorization of streets based on the typical roughness of the facades. Four categories are distinguished, ranging from flat facades to facades with many balconies and relief (a)-(d). For the final model, category 2 (b), 3 (c) and 4 (d) were merged to category B (dark frame), while category 1 (a) was renamed as category A.

5.3.1.1 Regression analysis based on street width

A very strong correlation between street width and reflection ratio was noticed. In order to maximize model fit, it is opted not to use the raw street width directly as input variable in a linear regression, but instead to use a simplified theoretical model for the reflection ratio ($RR_{1st}(w)$), assuming geometrically reflecting boundaries:

$$\widehat{RR}(12ms) = A \cdot RR_{1st}(w) + B. \quad (5.4)$$

For the simplified model of $RR_{1st}(w)$, only the direct sound and first reflections on both facades (given the measurement car is located in the center of the street) are taken into account:

$$RR_{1st}(w) = 10 \log_{10} \frac{2/(4\pi(\Delta_{ref}^2 + w^2))}{1/(4\pi\Delta_{ref}^2)} \quad (5.5)$$

$$= 10 \log_{10} \frac{2\Delta_{ref}^2}{\Delta_{ref}^2 + w^2}, \quad (5.6)$$

with Δ_{ref} the distance between the source and receiver (2.48m) and w the street width.

Fig. 5.6 displays the measured $RR(12ms)$ for each street in function of $\log_{10}(\text{width})$. $RR(12ms)$ measured in the open field is indicated as well; the horizontal line gives the mean value, the horizontal dotted line is the standard deviation. In the 125Hz and 250Hz octave band, a small overlap between open field

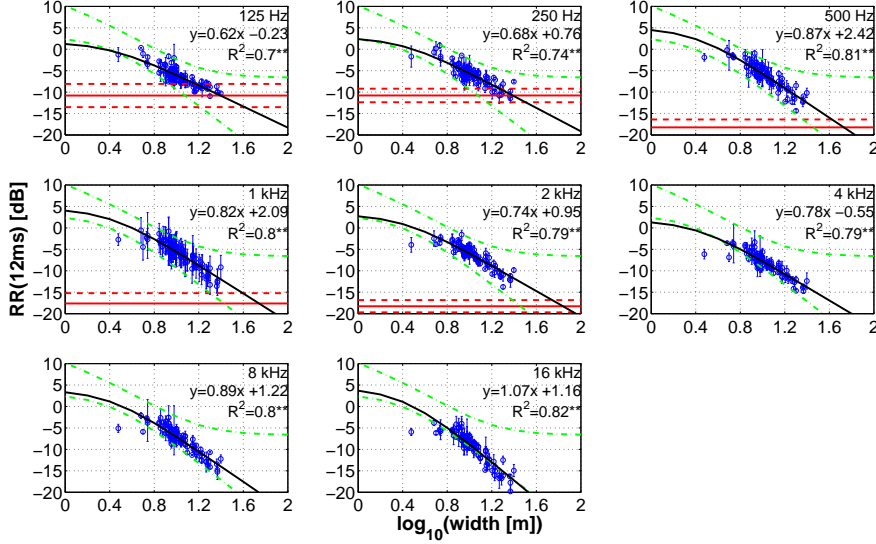


Figure 5.6: Regression analysis of $RR(12ms)$ on 99 locations at different octave bands. The course of $\widehat{RR}(12ms) = A \cdot RR_{1st}(w) + B$ (full line) and its lower and upper bound (dashed-dotted line) is plotted as a function of $\log_{10}(w)$. Measured $RR(12ms)$ is indicated by circles (\circ), with error bars indicating the standard deviation from the mean value at each location. $RR(12ms)$ of the open field is also indicated (mean value: horizontal line, mean value \pm standard deviation: horizontal dashed line). The regression coefficients are indicated together with the squared correlation coefficient. * indicates $p < 0.05$, ** indicates $p < 0.01$.

and street reflection ratio is seen only when streets are wider. On average however, $RR(12ms)$ of the open field remains lower than the measured street values. Starting from the 500Hz octave band, the overlap between street and open field measurements vanishes.

The linear regression model shows strong correlation ($R^2 \approx 0.8$ on average) with the logarithm of the street width. At low and mid frequency bands, $RR_{1st}(w)$ itself is a lower bound for the measured reflection ratio: higher order reflections still have an important contribution. At high frequencies (8kHz and 16kHz) the simplified model is already sufficiently accurate in modeling the measured reflection ratio as higher order reflections are highly attenuated by air and facade absorption. An upper bound for $RR(12ms)$ can be calculated based on the image source model for geometrically reflecting boundaries. In order to calculate an upper limit RR_{max} , the energy contained in all reflections, including the ground reflections, needs to be summed. Eq. 5.3 describes the energy contained in the n^{th} reflection, calculated by only considering the street facades. In a similar way, the energy

contained in the n^{th} ground reflection can be calculated as

$$E_{n^{th},gr} = \frac{(1 - \alpha')(1 - \alpha)^n}{4\pi((nw)^2 + 4H_{ref}^2 + \Delta_{ref}^2)}, \quad (5.7)$$

with $H_{ref} = 2.36\text{m}$, the height of the source and receiver and α' the absorption coefficient of the ground surface. With Eq. 5.3 and Eq. 5.7, the upper bound can be modeled as

$$\begin{aligned} \text{RR}_{max} &= 10 \log_{10} \left(\frac{2 \sum_{n=1}^{+\infty} E_{n^{th}} + 2 \sum_{n=1}^{+\infty} E_{n^{th},gr} + E_{0^{th},gr}}{1/4\pi\Delta_{ref}^2} \right) \\ &= 10 \log_{10} \left(\frac{2 \sum_{n=1}^{+\infty} \frac{(1-\alpha)^n}{4\pi((nw)^2 + \Delta_{ref}^2)} + 2 \sum_{n=1}^{+\infty} \frac{(1-\alpha')(1-\alpha)^n}{4\pi((nw)^2 + 4H_{ref}^2 + \Delta_{ref}^2)} + (1-\alpha')/4\pi(\Delta_{ref}^2 + 4H_{ref}^2)}{1/4\pi\Delta_{ref}^2} \right). \end{aligned} \quad (5.8)$$

$$(5.9)$$

If perfectly reflecting facades are assumed ($\alpha = \alpha' = 0$), this equation can be written as

$$\begin{aligned} \text{RR}_{max} &= 10 \log_{10} \left(\left(\frac{\pi\Delta_{ref}}{w} \coth \left(\frac{\pi\Delta_{ref}}{w} \right) - 1 \right) \right. \\ &\quad \left. + \frac{\Delta_{ref}^2}{g_{ref}^2} \left(\frac{\pi g_{ref}}{w} \coth \left(\frac{\pi g_{ref}}{w} \right) - 1 \right) + \frac{\Delta_{ref}^2}{g_{ref}^2} \right), \end{aligned} \quad (5.10)$$

with $g_{ref}^2 = \Delta_{ref}^2 + 4H_{ref}^2$.

At low and mid frequency octave bands, RR_{max} , together with RR_{1st} , define a close interval to which $\text{RR}(12\text{ms})$ is confined. At high frequency octave bands, reflections are more attenuated by (air) absorption and RR_{max} becomes much higher than the measured values. Furthermore, it is seen that RR_{max} tends to saturate for increasing street width as the energy contained in the sums of Eq. 5.9 becomes smaller, relative to the energy contained in the direct sound and first ground reflection $E_{0^{th},gr}$. In reality however, this first ground reflection will be less pronounced due to the presence of the car, but is included to show the upper limit.

5.3.1.2 Multiple regression analysis

In the previous section, only street width was used in the model. Although this model was already very satisfactory, this section will investigate if inclusion of

Table 5.1: Results of the multiple regression analysis on the measured reflection ratio, with $RR_{1st}(w)$ and the facade roughness as input variables. The squared correlation coefficient of the model is also given (* indicates $p < 0.05$, ** indicates $p < 0.01$).

f_{oct}	Model coefficients				R^2
	C^{ste}	$RR_{1st}(w)$	roughness CAT		
125Hz	-0.111	0.615**	-0.429		0.71**
250Hz	0.673	0.688**	0.295		0.74**
500Hz	2.327**	0.876**	0.346		0.81**
1kHz	1.933**	0.831**	0.550*		0.81**
2kHz	0.828*	0.749**	0.418		0.80**
4kHz	-0.726	0.784**	0.619*		0.80**
8kHz	0.986*	0.901**	0.840**		0.82**
16kHz	0.915	1.086**	0.865**		0.83**

other parameters (height of the street and facade roughness) can lead to further improvements.

Analysis of variance showed that height and facade roughness were not statistically independent (F-value 22.5, $p < 0.01$). This is not surprising, as in urban streets buildings are more likely to have more balconies (and higher degree of facade roughness) when building height increases. For this reason, both parameters cannot be used simultaneously as input variable in a multiple regression model. However, it is interesting to investigate the dependency of reflection ratio on building height and facade roughness. As both parameters are not necessarily known, it is decided to create two models, based on multiple regression analysis on the measured reflection ratio, with the simplified reflection ratio $RR_{1st}(w)$ (including street width) and one of both dependent parameters as input variables. Based on available data (either height or facade roughness), the appropriate model can then be selected.

In Table 5.1 the resulting coefficients of the multiple regression analysis with $RR_{1st}(w)$ and the facade roughness as input parameters are displayed. Since the facade roughness is a categorical variable consisting of two categories, of which only one is statistically independent, this parameter is transformed into one independent dummy variable prior to inclusion in the model. The model then becomes

$$\widehat{RR}(12\text{ms}) = A \cdot RR_{1st}(w) + B \cdot \text{roughnessCAT} + C^{ste}, \quad (5.11)$$

with $\text{roughnessCAT} = 1$ if the typical roughness of the facade is in category B and $\text{roughnessCAT} = 0$ in the other case.

At frequencies below the 1kHz octave band, facade roughness has no significant

Table 5.2: Results of the multiple regression analysis on the measured reflection ratio, with $RR_{1st}(w)$ and the average height as input variables. The squared correlation coefficient of the model is also given (* indicates $p < 0.05$, ** indicates $p < 0.01$).

f_{oct}	Model coefficients			R^2
	C^{ste}	$RR_{1st}(w)$	average height	
125Hz	1.209	0.627**	-0.164**	0.72**
250Hz	0.420	0.683**	0.038	0.74**
500Hz	1.741*	0.868**	0.078	0.81**
1kHz	0.414	0.816**	0.190**	0.82**
2kHz	-0.412	0.737**	0.154**	0.80**
4kHz	-1.864**	0.770**	0.149*	0.80**
8kHz	-0.442	0.882**	0.189**	0.81**
16kHz	-0.576	1.066**	0.197*	0.83**

influence ($p > 0.05$). At higher frequencies, the positive contribution becomes statistically significant. Although small (max. 0.9dB at 16kHz), inclusion of the facade roughness leads to some improvement, compared to the linear regression model solely based on $RR_{1st}(w)$. However, $RR_{1st}(w)$ remains the dominant factor over all octave bands: coefficients hardly differ compared to the previous regression model. The main reason for the small effect of facade roughness on the sound field in the streets is the presence of receivers very close to the source, which is a drawback of the employed measurement set-up. By comparing the difference in SPL for a street canyon with geometrically reflecting facades (simulated by means of an image source method) with fully diffusely reflecting facades (simulated with the radiosity method), Kang [5] showed that diffuse reflections, caused by facade irregularities, indeed have an influence on the soundfield. For increasing source-receiver positions, an increase of attenuation was noted, whereas for positions close to the source, a small increase in SPL was seen due to backscattering effects. The tendencies found here are however statistically significant at frequencies where this is expected.

Table 5.2 gives the resulting coefficients from the multiple regression analysis for each octave band with $RR_{1st}(w)$ and the average height as input variables:

$$\widehat{RR}(12\text{ms}) = A \cdot RR_{1st}(w) + B \cdot \text{average height} + C^{ste}. \quad (5.12)$$

This model has similar performance as the previous model. As average height and the categorization of facade roughness proved to be related, similar conclusions can be drawn. Starting from the 1kHz octave band, the contribution of the average height becomes significant ($p < 0.05$). At all octave bands, except at 125Hz,

the model implies that the average height has a positive influence on the reflection ratio. Indeed, increasing the height of the street enables more reflections to backscatter, raising $RR(12\text{ms})$.

5.3.2 Model performance

This section describes the performance of the derived model by comparing the predicted reflection ratio with the measured reflection ratio. The model discussed here, is the one described in Table 5.2, based on multiple regression analysis of $RR_{1st}(w)$ and the average height. This model is chosen as the average height is based on objective measurements, in contrast to the more subjective categorization of facade roughness. A comparison based on research from Heutschi [11] will also be made.

To incorporate the effect of the buildings, Heutschi defined the building correction BC_{SR} as the ratio of total energy to the energy of the direct sound for a given source-receiver position. Based on extended ray tracing simulations, he derived a set of simple formulas, depending on the degree of diffusion, to estimate BC_{SR} . For general use however, a diffusion coefficient of $d = 0.2$ was proposed [11]. Under this condition, BC_{SR} is modeled as follows:

$$BC_{SR}|_{d=0.2} = \left(3.0 + 1.3 \log_{10} \left(\frac{h}{w} \right) \right) \frac{d_{SR}}{w} \quad [dB], \quad (5.13)$$

with h the facade height, w the street width and d_{SR} the source-receiver distance. From BC_{SR} we can easily calculate the corresponding reflection ratio:

$$RR_{Heutschi} = 10 \log_{10} \left(10^{\frac{BC_{SR}}{10}} - 1 \right) \quad [dB]. \quad (5.14)$$

Fig. 5.7 depicts a comparison of the different models and the measured reflection ratio. The measured reflection ratio (\circ) at all 99 locations is plotted against the model based on $RR_{1st}(w)$ and average height at these locations. If the model is accurate, data should be on a straight line (1:1 mapping). Fig. 5.7 shows that this condition is fairly met. Only a limited part of the data points deviate from this line, but the highest deviation is less than 5dB.

When comparing the reflection ratio derived from Heutschi with the modeled reflection ratio, it is seen that $RR_{Heutschi}$ slightly underestimates the modeled and measured reflection ratio at frequencies below the 2kHz octave band (approx. 2dB). In the 16kHz octave band, $RR_{Heutschi}$ delivers slightly higher prediction for the reflection ratio at low values. It should however be noted that the model based on $RR_{1st}(w)$ and average height incorporates frequency dependencies, while this is not the case in $RR_{Heutschi}$.

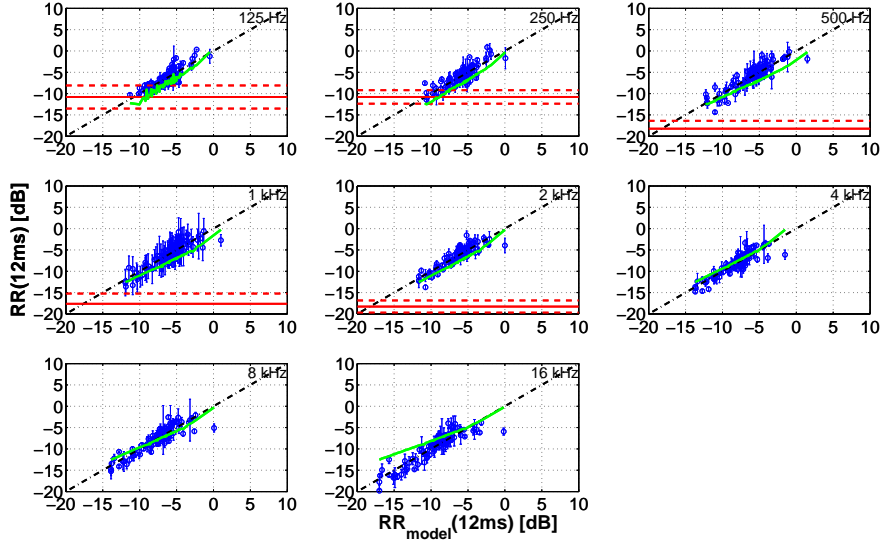


Figure 5.7: Comparison between the predicted and measured reflection ratio. The x -coordinate gives the value of the modeled reflection ratio, based on the multiple regression model described in Table 5.2. The corresponding measured reflection ratio is shown by discrete circles (\circ), with error bars indicating the standard deviation from the mean value at each location. The full line shows the reflection ratio derived from Heutschi, plotted against the modeled reflection ratio based on the same geometrical data. The condition that should be met (1:1 mapping) is indicated by the dashed-dotted line. $RR(12ms)$ of the open field is also indicated (mean value: horizontal line, mean value \pm standard deviation: horizontal dashed line).

5.4 Predicting sound pressure level based on the reflection ratio

In the previous section, a model for the reflection ratio has been created, based on an extensive measurement campaign. Here, it will be shown how the model can be used to predict sound pressure levels in a street canyon.

5.4.1 General model for the SPL in a street canyon

The energy of the direct sound (free-field conditions) can be expressed as

$$E_{dir} \propto \frac{1}{4\pi r^2}, \quad (5.15)$$

with r the source-receiver distance. Furthermore, in a street canyon, the reverberant energy can be approached by

$$E_{rev} \propto \frac{1}{2\pi r_{xz}}, \quad (5.16)$$

with r_{xz} the projected source-receiver distance on the plane along the canyon, parallel with the facades. Propagation from a point source in between reflecting planes (here facades) can be considered as the geometrical divergence of a (badly discretized) line source. This corresponds to an energy decay of one over the distance at sufficient distances from the source.

Assuming that all distance dependence is captured by Eq. 5.15 and Eq. 5.16, the ratio of proportionality factors A_0 becomes distance independent, yet not frequency independent. A_0 can be related to $RR(12\text{ms})$ by calculating the ratio of reverberant to direct field at the reference distance:

$$RR(12\text{ms}) = 10 \log_{10} \frac{E_{rev}}{E_{dir}} \quad (5.17)$$

$$= 10 \log_{10} \left(A_0 \frac{1/(2\pi r_{xz})}{1/(4\pi r^2)} \right). \quad (5.18)$$

The reflection ratio at the reference source-receiver distance ($\Delta_{ref} = 2.48\text{m}$) can be estimated from the multiple regression model, based on the street width and average height. Given the reference source-receiver position along the canyon axis ($r = r_{xz} = \Delta_{ref}$) and \widehat{RR}_{model} , an estimate for A_0 can be extracted:

$$\widehat{A}_0 = \frac{10 \frac{\widehat{RR}_{model}}{10}}{2\Delta_{ref}}. \quad (5.19)$$

With Eq. 5.19, the course of the total sound pressure level in a street canyon is calculated as

$$SPL \propto 10 \log_{10} \left(E_{dir} + \widehat{A}_0 \frac{b}{2\pi r_{xz}} \right) \quad [dB], \quad (5.20)$$

with E_{dir} the energy of the direct sound and b a correction factor, related to the source position relative to the ground. When the source is close to the ground, free-field conditions are no longer valid, and an important ground reflection will be contained within the scope of the direct field ($t < 12\text{ms}$). This ground reflection is not covered in the reverberant energy. Therefore, the direct energy, as defined by Eq. 5.15, will be extended with an additional ground reflection when the source is close to the ground. By using an image source for the ground reflection, the early soundfield can be calculated for each frequency as

$$p_{dir}(f) = \frac{1}{\sqrt{4\pi r^2}} e^{-jk|r|} + \frac{1}{\sqrt{4\pi r'^2}} e^{-jk|r'|}, \quad (5.21)$$

with k the wave number, r the source-receiver distance and r' the distance between the mirrored source and receiver. E_{dir} is then calculated by summing the energy over the frequency bands: $E_{dir} = \sum_f |p_{early}(f)|^2$.

The proximity of the ground also influences the energy contained in the reverberant

field. Therefore, a correction factor b is introduced, ranging between one (free-field conditions) and two (source located at groundlevel and in case of a rigid street surface). A value in between one and two might be used in case of absorbing street surfaces.

5.4.2 Model validation in a real street canyon

In order to validate the proposed model, measurements of the sound pressure level in a real street canyon have been performed. Measurements in this street were not considered to build up the regression model. The selected street canyon is the ‘Gustaaf Magnelstraat’, a 105m long and 10.7m wide street in Ghent, Belgium. The average height is estimated at 17m. The street was completely blocked for traffic during the measurements (Fig. 5.8). Only a few parked cars were present.

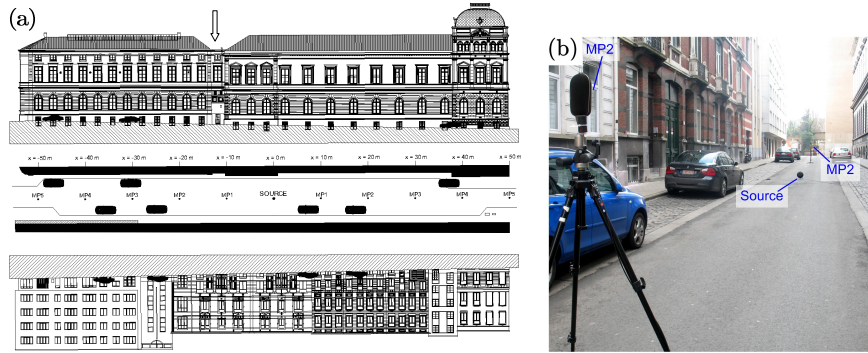


Figure 5.8: Impulse response measurements in the ‘Gustaaf Magnelstraat’, a 105m long and 10.7m wide street canyon in Ghent, Belgium. The different measurement positions (MPx) along the canyon axis (x-axis) are indicated, as well as the source position (a). The arrow indicates the location of the facade depression. The center of the source was located 30cm above the asphalt, the measurement microphones were at 1.8m height (b).

An omnidirectional source was positioned in the center of the street, 30cm above the asphalt, emitting a single 12s-long exponential sweep. Receiver locations were chosen at multiples of 10m from the source, measured along the canyon axis, at 1.8m height. Two MK250 omnidirectional microphones, one at each side of the source, were used to record the sweep signal. The resulting SPL at each position is extracted from the truncated impulse response, calculated as described in section 5.2.1. At least four repeated measurements have been carried out at each location to enable averaging to further increase the signal-to-noise ratio.

Fig. 5.9 shows the course of the measured SPL along the street canyon axis together with the predicted SPL (Eq. 5.20, $b = 2$), based on the $\bar{R}\bar{R}$ estimated with the

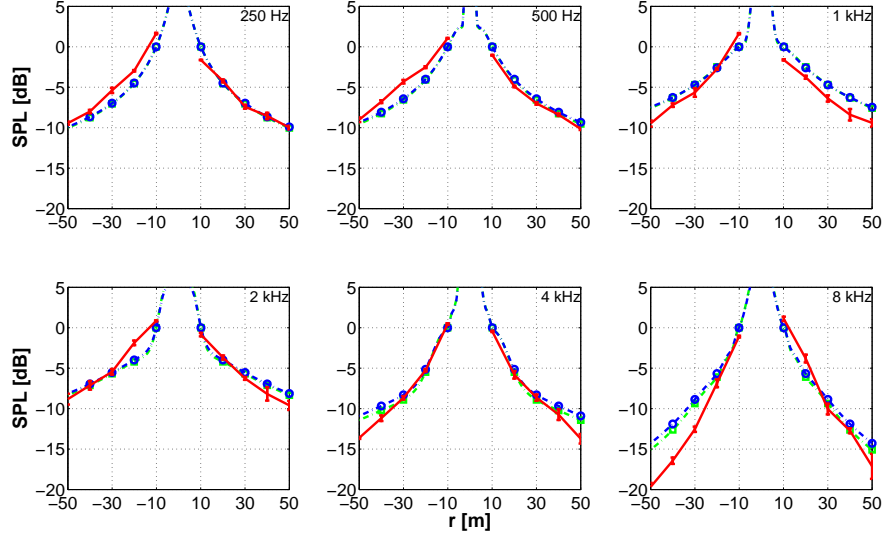


Figure 5.9: Measurement and prediction of SPL in a street canyon in function of source-receiver distance for each octave band. The measured SPL is given in red, with error bars indicating the standard deviation from the mean value at each location. Square green markers (\square) indicate the SPL, modeled with the simple model for \widehat{RR} , solely based on the street width. Round blue markers (\circ) indicate the SPL based on the extended model for \widehat{RR} , including the average height.

simple model (Eq. 5.4), solely based on the street width, and the extended model (Eq. 5.12), including the average height. Here, a simple model for the atmospheric absorption has been included, based on ISO 9613-1 [19] (measured meteorological parameters during the experiment: temperature = 5.0°C, humidity = 80% and atmospheric pressure = 997hPa). As a reference, the averaged SPL at $x = -10\text{m}$ and $x = 10\text{m}$ is taken to be 0dB for each model. The difference in course between the predicted SPL, based on the two different models for \widehat{RR} , is negligible after normalization. Only at the high frequency octave bands, a small difference (<1dB) in SPL is seen. However, it must be noted that a frequency dependent offset exists between both SPL predictions (which is not seen due to the normalization), caused by a different estimate of \widehat{RR} . At low frequencies, these estimates differ approx. 0.3dB, while at high frequencies a difference of approx. 1.5dB exists.

In contrast to the predicted SPL, the measured SPL has a non symmetrical course, with higher values for the left end of the street ($x \leq -10\text{m}$), that might be caused by some parked cars and a large facade depression near $x = -12\text{m}$ (Fig. 5.8). The larger density of the cars parked at both sides of the left could virtually decrease the street width, giving rise to an increase of the measured SPL (see section 5.4.3). This effect is not seen at the 8kHz octave band, as the SPL is much lower than at

the right end. At these high frequencies, the cars could act as large scattering objects, scattering part of the sound energy out of the street, consequently lowering the SPL. Differences up to 3.5dB were noted between positions left and right from the source at the same source-receiver distances.

Despite the asymmetric course of the measured sound pressure level, the predicted SPL still proves to be a good approximation. At low and mid frequency octave bands, the differences between the predicted SPL and measured SPL are lower than 2dB. At high frequency octave bands (4kHz and 8kHz), the measured SPL is slightly overpredicted near the ends of the street only.

5.4.3 Influence of street width and height on predicted SPL

This section investigates the dependency of the proposed SPL-model on the street width and average height. Based on Eq. 5.20, Eq. 5.19 and Eq. 5.12, the fluctuation of the total energy in function of street width can be calculated mathematically:

$$\frac{\partial \left(10 \log_{10} \left(E_{dir} + \frac{b \hat{A}_0}{2\pi r_{xz}} \right) \right)}{\partial w} = \frac{10}{\ln(10)} \frac{1}{E_{dir} + \frac{b \hat{A}_0}{2\pi r_{xz}}} \frac{b}{2\pi r_{xz}} \frac{\partial \hat{A}_0}{\partial w} \quad (5.22)$$

$$= \frac{\frac{b 10^{RR/10}}{4\pi r_{xz} \Delta_{ref}}}{E_{dir} + \frac{b 10^{RR/10}}{4\pi r_{xz} \Delta_{ref}}} \frac{\partial RR}{\partial w} \quad (5.23)$$

$$= - \frac{20A}{\ln(10)} \frac{1}{\frac{4\pi r_{xz} \Delta_{ref} E_{dir}}{b 10^{RR/10}} + 1} \frac{w}{\Delta_{ref}^2 + w^2}. \quad (5.24)$$

If we approximate the energy of the direct soundfield by $E_{dir} = \frac{b}{4\pi r^2}$ (no interference effects) and take $r = r_{xz}$ for source and receiver in the middle of the street, the derivative can be simplified to

$$\frac{\partial \text{SPL}}{\partial w} = - \frac{20A}{\ln(10)} \frac{1}{\frac{\Delta_{ref}}{r_{xz} 10^{RR/10}} + 1} \frac{w}{\Delta_{ref}^2 + w^2}. \quad (5.25)$$

Since the coefficient A is always positive (Table 5.2), Eq. 5.25 is always negative, which proves that the SPL is inversely proportional with the street width: increasing the width reduces the SPL. Furthermore, the change of total energy in function of width is bigger for changes relative to small street widths and positions further from the source as the magnitude of the derivative increases with decreasing street width and increasing distance r_{xz} .

In an analogue way, the variation of the SPL on the average height (H) can be calculated as well:

$$\frac{\partial \left(10 \log_{10} \left(E_{dir} + \frac{b \hat{A}_0}{2\pi r_{xz}} \right) \right)}{\partial H} = \frac{\frac{b 10^{RR/10}}{4\pi r_{xz} \Delta_{ref}}}{E_{dir} + \frac{b 10^{RR/10}}{4\pi r_{xz} \Delta_{ref}}} \frac{\partial RR}{\partial H} \quad (5.26)$$

$$= \frac{B}{\frac{4\pi r_{xz} \Delta_{ref} E_{dir}}{b 10^{RR/10}} + 1}. \quad (5.27)$$

Again, E_{dir} is approximated by $E_{dir} = \frac{b}{4\pi r^2}$ and $r = r_{xz}$, simplifying the derivative:

$$\frac{\partial SPL}{\partial H} = \frac{B}{\frac{\Delta_{ref}}{r_{xz} 10^{RR/10}} + 1}. \quad (5.28)$$

Eq. 5.28 proves that an increase of average height generally leads to an increase of SPL. The change of the SPL as a function of the average height is bigger at positions further from the source than at positions close to the source. The same conclusion can be drawn when changing the average height: relative to large heights, the change in SPL will be bigger than for low values (because of the factor $1/10^{RR/10}$ in the denominator). However, the effect of the average height on the derivative will be rather small, given the model coefficient $B < 0.2$.

In order to give a more quantitative estimation of the influence of the width and average height on the predicted sound pressure level, a simulation is made for the source-receiver set-up of the Magnestraat with virtual street widths and heights. Fig. 5.10 shows the resulting course of the SPL, for street widths starting from 5m to 20m (full lines with triangular markers) and average heights starting from 5m to 20m (dashed lines with square markers). It is clear from the figure that the variation in SPL is much more dependent on variations of the street width than on variations of the average height. At low and mid frequency octave bands, changing the average height has almost no influence on the SPL (approx. 2dB difference for the SPL at the 1kHz octave band between the extreme ranges of the average height at $x = 30$ m). For higher frequency octave bands, the influence of the average height becomes somewhat higher (approx. 3.5dB difference at the 16kHz octave band).

Changing the width however has a larger influence on the course of the SPL. Increasing the street width from 5m to 20m results in a decrease of 6.5dB at $x = 30$ m for the 1kHz octave band. This decrease is however not linearly proportional with the increase of street width, as a larger change in SPL is seen for changes at already low street width values. Furthermore, it is seen from the figure that a change in street width has a bigger effect on the SPL at receiver positions further from the source. At $x = 30$ m, the difference between the SPL at the extreme values of the street width is 6.5dB, which increases to 7.2dB at $x = 50$ m.

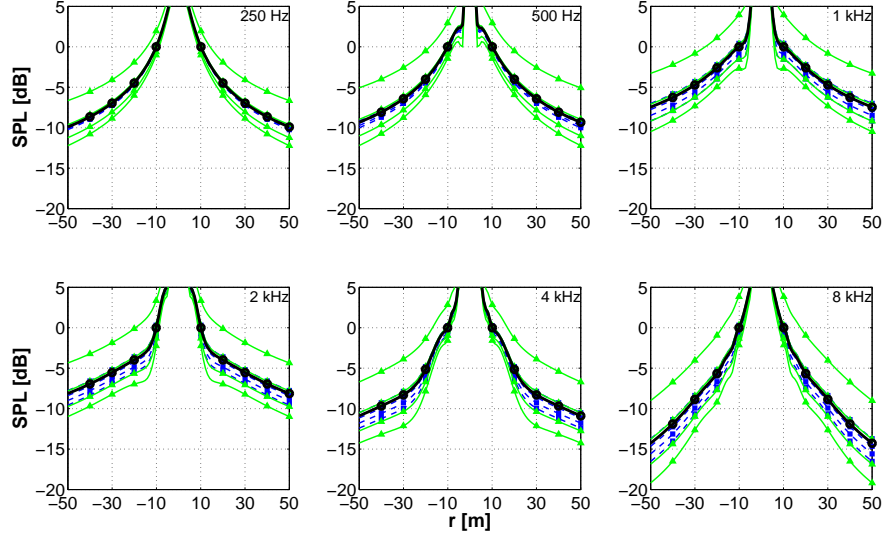


Figure 5.10: Influence of the width and average height of the street canyon on the prediction of the SPL as a function of the source-receiver distance for each octave band. The predicted SPL is given for the geometry of the Magnelstraat (width equals 10.7m and average height equals 17m), plotted in full thick line with round markers (\circ). The full green lines with triangular markers (\blacktriangle) indicate the influence of the variation of street width on the SPL. The width is increased from 5m (uppermost curve) to 20m (lowest curve) in steps of 5m. The dashed blue lines with square markers (\blacksquare) show the variation of SPL in function of the average height. The height is varied in steps of 5m from 5m (lowest curve) to 20m (uppermost curve).

5.5 Conclusion

In this chapter, a new model is developed to predict sound pressure levels in urban streets. This model is based on the reflection ratio, defined as the reverberant-to-direct sound energy ratio. An extended measurement campaign has been set up to measure the reflection ratio in 99 streets in the city of Ghent, Belgium. A dedicated measurement set-up has been designed, mounted on the roof of a car to enable fast and mobile measurements. The body of the car was shown to have negligible influence on this parameter. The latter has been assessed by measurements in the open field. The current measurement set-up was able to quantify street reverberation starting at the octave band with center frequency of 125Hz, at a sufficient repeatability.

Based on a few geometrical parameters such as street width, average height and facade roughness, a prediction model is proposed to estimate the reflection ratio. It is found that the reflection ratio highly correlates with the street width (i.e. the simplified reflection ratio $RR_{1st}(w)$) with $R^2 \approx 0.8$. Including the average street

height or the building roughness slightly improves the proposed model. Resulting coefficients have been given and are tested with the real measured data and the model proposed by Heutschi. Good accordance is found.

To illustrate the use of the reflection ratio, a model for the total sound pressure level in a street canyon has been developed, where the ratio between direct sound energy and reverberant energy is extracted from the reflection ratio. The model has been verified with additional measurements in a city street canyon for different source-receiver positions. Although sound pressure measurements indicated that the distribution of sound energy was not symmetric around the street center, the model was still able to roughly predict the course of the sound pressure levels, solely based on reflection ratio. Additionally, the influence of changes in the street width and average building height on the predicted SPL has been investigated.

In this work the potential of the reflection ratio for the calculation of urban noise maps has been illustrated. Once the direct sound energy contribution is known, the increase by street reverberation can be estimated. In this way calculation cost can be strongly reduced, without loss of accuracy. Furthermore, the reflection ratio can be easily predicted, based on simple geometrical street characteristics only.

References

- [1] J. Kang. *Urban sound environment*. Taylor & Francis, London, 2007.
- [2] J. Picaut, T. Le Pollès, P. L'Hermite, and V. Gary. *Experimental study of sound propagation in a street*. *Applied Acoustics*, 66 (2):149–173, 2005.
- [3] R. Lyon. *Role of multiple reflections and reverberation in urban noise propagation*. *J. Acoust. Soc. Am.*, 55 (3):493–503, 1974.
- [4] J. Kang. *Numerical modelling of the sound fields in urban streets with diffusely reflecting boundaries*. *Journal of Sound and Vibration*, 258:793–813, 2002.
- [5] J. Kang. *Sound propagation in street canyons: Comparison between diffusely and geometrically reflecting boundaries*. *J. Acoust. Soc. Am.*, 107:1394–1404, 2000.
- [6] J. Picaut. *Numerical modeling of urban sound fields by a diffusion process*. *Applied Acoustics*, 63:965–991, 2002.
- [7] T. Van Renterghem, E. Salomons, and D. Botteldooren. *Parameter study of sound propagation between city canyons with coupled FDTD-PE model*. *Applied Acoustics*, 67 (6):487–510, 2006.

- [8] M. R. Ismail and D. J. Oldham. *A scale model investigation of sound reflection from building facades*. *Applied Acoustics*, 66:123–147, 2005.
- [9] K. K. Iu and K. M. J. Li. *The propagation of sound in narrow street canyons*. *J. Acoust. Soc. Am.*, 112 (2):537–550, 2002.
- [10] *ISO 9613-2:1996, Acoustics - Attenuation of sound during propagation outdoors - Part 2: General method of calculation*. International Organization for Standardization, Geneva, Switzerland, 1996.
- [11] K. Heutschi. *A simple method to evaluate the increase of traffic noise emission level due to buildings, for a long straight street*. *Applied Acoustics*, 44:259–274, 1995.
- [12] *ISO 18233:2006(E), Acoustics - Application of new measurement method in building and room acoustics*. International Organization for Standardization, Geneva, Switzerland, 2006.
- [13] *ANSI S1.11-2004, Specification for octave-band and fractional-octave-band analog and digital filters*. *Acoust. Soc. Am.*, Melville, New York, 2009.
- [14] P. Thomas, T. Van Renterghem, and D. Botteldooren. *Using room acoustical parameters for evaluating the quality of urban squares for open-air rock concerts*. *Applied Acoustics*, 72:210–220, 2011.
- [15] M.R. Schroeder. *New method of measuring reverberation time*. *J. Acoust. Soc. Am.*, 37:409–412, 1965.
- [16] A. Lundeby, T.E. Vigran, H. Bietz, and M. Vorländer. *Uncertainties of measurements in room acoustics*. *Acustica*, 81:344–355, 1995.
- [17] E. Schröder. *Nachhall in geschlossenen bebauten Strassen*. *Lärmbekämpfung*, 17:11–18, 1973.
- [18] P. Thomas, E. De Boeck, L. Dragonetti, T. Van Renterghem, and D. Botteldooren. *Creating an urban street reverberation map*. In *Proceedings of Forum Acusticum*, pages 397–402. European Acoustics Association (EAA), 2011.
- [19] *ISO 9613-1:1993, Acoustics - Attenuation of sound during propagation outdoors - Part 1: Calculation of the absorption of sound by the atmosphere*. International Organization for Standardization, Geneva, Switzerland, 1993.

6

Conclusions and perspectives

The last few years more and more attention has been given to the quality of the urban sound environment. In their design, urban city-planners often include silent areas and restoration zones to enable inhabitants to recover from the urban hubbub, even while commuting. Noise mitigation measures such as sound barriers are frequently used for this. Up to now, the design and evaluation of such measures has been mainly based on quantitative predictions of the sound level.

In this work an improved set of parameters is suggested that allows a characterization of the urban sound environment in a more subtle and comprehensive way. Furthermore, two different auralization methodologies have been designed that enable a more perceptive assessment of the effect of noise mitigation measures.

The first auralization methodology developed in this work is based on time-dependent attenuation of binaural HATS recordings of individual car passages, made before installation of the measure, with the angle dependent IL. The momentary source direction is determined by a human-inspired source angle detector. This technique has been applied to a low vegetated noise barrier. A quantitative prediction of the source-dependent barrier IL has been made with different engineering models (ISO 9613-2 and Harmonoise) and a full-wave model (FDTD - with different degrees of geometrical accuracy). Based on these calculations and a priori recordings of individual car passages, the effect of the barrier has been auralized to enable a qualitative assessment. Listening tests proved that the barrier, although limited in height, could be aurally perceived by the participants. However, auralizations could still be identified when compared with a posteriori recordings.

Samples auralized with the ISO 9613-2 and Harmonoise IL model were typically considered to be more realistic than samples auralized with the full-wave method.

A second auralization methodology is based on a direction-dependent attenuation of the sound environment. For this purpose, a flexible binaural reproduction technique has been designed (see chapter 2). Based on a 32-channel array recording, a binaural signal has been extracted by linear combination of each of the microphone signals, so that the resulting array directivity pattern emulates the left and right ear HRTF directivity pattern. Microphone weights have been determined by a regularized least-square technique with the measured microphone-on-sphere directivity patterns as input and the target HRTF pattern as output. At high frequencies, the fitting performance strongly reduced and the MTB technique was used instead. The quality of the binaural reproduction technique has been assessed by comparing binaural reproductions with original binaural recordings in a listening test. Although listeners were able to identify HATS recordings as true binaural samples, they could not discriminate between binaurally reproduced array recordings and true binaural recordings.

Based on the array pattern emulation technique, the effect of a mound has been auralized by emulation of the left and right ear HRTF directivity pattern, attenuated with the predicted angular dependent IL. The advantage of this methodology is that, in contrast to the binaural auralization methodology used in the first case-study, multiple sources (vehicles) can now be taken into account.

The use of this methodology has been demonstrated in a second case-study for the assessment of the effect of an L-shaped mound near the E17 highway. For the quantitative prediction of the (angular) IL of the mound, three different models have been considered, i.e. the Harmonoise point-to-point model, the Pierce barrier diffraction formula and the ISO 9613-2 propagation model. Comparison with long-term measurements showed that the Harmonoise model succeeded best in predicting overall IL, while the ISO 9613-2 showed the largest deviation.

The quality of the auralization methodology has been subjectively evaluated by a series of listening tests and it has been found that listeners had great difficulty in differentiating between true a posteriori recordings and auralized samples. When auralized with the pattern emulation technique, no significant difference was seen between IL models, although the Pierce barrier IL model was judged slightly better.

In the second part of this work, a set of quantitative room acoustic inspired parameters was proposed for the analysis of square and street acoustics, complementary to the spatial sound pressure level distribution. In chapter 4, it has been investigated to what extent existing room acoustic parameters, related to energy distribution, reverberation and binaural quality can be used to characterize urban squares, used

for rock concerts. Different parameters have been measured at five squares and a statistical independent set, containing the clarity C_{80} , reverberation T_{30} , binaural parameters $[1 - \text{IACC}_{E3/L3}]$ and relative sound pressure level $\Delta L_{eq,A/C}$ has been created by means of correlation analysis. While C_{80} and $[1 - \text{IACC}_{E3}]$ were strongly dependent on the measurement location and speaker set-up, T_{30} showed to be mainly correlated with square dimensions. Furthermore, the relevance of considering both $\Delta L_{eq,A}$ as well as $\Delta L_{eq,C}$ has been shown.

In chapter 5 a new model was developed to estimate the sound pressure level in an urban street canyon, solely based on the knowledge of the direct energy and the reflection ratio RR, defined as the ratio of reverberant-to-direct energy. It has been investigated to what extent the RR can be linked to known geometrical and architectural properties of the street. Therefore, the RR has been measured in 99 different urban streets with a dedicated measurement set-up, and a prediction model has been designed. It was found that the RR is mainly dependent on the street width, while the accurateness of the model can be improved when either facade roughness or average street height is included. The model has been validated by comparison with sound pressure level measurements along a real street canyon.

6.1 Perspectives

The auralization methodology, designed in the first part of this work, mainly focuses on the auralization of noise mitigation measures such as noise barriers. By emerging listeners in a virtual sound environment, they were able to evaluate the effect of such measures. However, the visual aspect has only been included by means of pictures. As the visual aspect might have a (positive) influence on the perception when evaluating (vegetated) noise mitigation measures, further research could investigate to what extent the (3D) visual aspect should be accounted for and how it influences the complete perception of the city-dweller. As such, a holistic tool can be created to aid city-planners in evaluating their design and investigate the correlation between acoustical and visual perception.

The pattern emulation technique used for the auralization methodology showed promising results for the binaural reproduction of the sound environment by directional source attenuation. For now, only static positions of the listener were allowed. Attempts have been made for the inclusion of a head-tracker for dynamic sound recreation, but the available processing power did not allow for real-time pattern emulation.

The potential of this technique in other applications is manifold, directive beam-forming for conferencing is one. Furthermore, a combination of many of such arrays can be used to perform spotforming, enabling a location-dependent rather than direction-dependent sound environment processing. However, if to be com-

mercially used, an effort should be made to create a more practical real-time implementation of the 32-channel microphone array emulation technique and an alternative should be found for the expensive and bulky multi-channel data-acquisition hardware. The recently developed digital MEMs microphones seem to offer a solution, having the advantage of being low-cost with improved phase-matching characteristics and low noise floor.

The acoustical parameters presented in the second part of this work enabled a more extensive quantitative characterization of the acoustical environment on squares and street canyons. It might be investigated if these parameters, describing the acoustics on a square, can be linked with the perception, and which of these is most determining for the perceived audio quality. As is already done for the acoustics in concert halls, research should be conducted towards a model that enables to allocate a score to a given urban sound environment based on these objective parameters.

Although not shown explicitly in this work, a clear advantage of using the reflection ratio for the construction of noise maps is expected. Traditionally, such maps are calculated by accounting for a limited set of reflections only. By introduction of the reflection ratio, calculation of street reverberation, strongly affecting sound pressure levels, could be expedited. However, questions remain on how to calculate sound pressure levels near crossings and half-open streets, and how to include multiple sources in such reflection ratio based noise maps.

