



## Serious auralizations

Barteld Postma

### ► To cite this version:

| Barteld Postma. Serious auralizations. Human-Computer Interaction [cs.HC]. Université Paris Saclay (COmUE), 2017. English. NNT: 2017SACLS091 . tel-01561147

**HAL Id: tel-01561147**

**<https://theses.hal.science/tel-01561147>**

Submitted on 12 Jul 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NNT : 2017SACLS091

THÈSE DE DOCTORAT  
DE L'UNIVERSITÉ PARIS-SACLAY  
PRÉPARÉE À L'UNIVERSITÉ PARIS-SUD  
AU SEIN DU LABORATOIRE D'INFORMATIQUE POUR LA  
MÉCANIQUE ET LES SCIENCES DE L'INGÉNIEUR

Ecole doctorale n°580  
Sciences et Technologies de l'Information et de la Communication  
Spécialité de doctorat : Informatique  
par  
**M. BARTELD POSTMA**

Serious Auralizations

Thèse présentée et soutenue à Orsay, le 28 Avril 2017.

Composition du Jury :

M. M. BEAUDOUIN-LAFON	Professeur d'Informatique LRI	(Président du jury)
M. D. MURPHY	Reader Audio and Music Technology University of York	(Rapporteur)
M. S. WEINZIERL	Professor, Audiokommunikation Technische Universität Berlin	(Rapporteur)
M. J.-M. NORMAND	Research Fellow Ecole Centrale de Nantes	(Examineur)
M. P. BOURDOT	Research Directeur CNRS LIMSI	(Examineur)
M. B.F.G. KATZ	Research Directeur CNRS UPMC, Institut d'Alembert	(Directeur de thèse)



Ce document est mis à disposition selon les termes de la licence Creative Commons “Attribution - Pas d’utilisation commerciale - Partage dans les mêmes conditions 3.0 non transposé”.



Par Barteld Nicolaas Johannes Postma.

## Acknowledgments

Without the help of several persons I could not have achieved the thesis which now lays in front of you. I write these Acknowledgements in the hope that I would not forget anybody who helped me get to the point where I am able to defend this work.

First and foremost I would like to thank my advisor, Brian Katz. For too many aspects I can mention here, but among others his patience, his critical eye on my work, his co-authoring of articles etc. Too many deeds which words can describe here.

The AA group at LIMSI I via this way thank for their warm welcome in France. The team member I am most grateful and indebted to is David Poirier-Quinot. Already early on in my thesis I noticed his willingness to help others. His aid can be found throughout the thesis but is most noticeable in the multi-modal auralizations. Even after he left LIMSI his help was ongoing. Second, I would like to thank Areti Andreopolou for her overall help, teaching me how to design listening tests, and assistance during my many MatLab questions. Best of luck to you in your new job in Greece. I admire Laurent Simon for his love of the field of acoustics and thank him for his help with my many MatLab questions. I wish him all the best for his new job in Switzerland. Christophe d'Alessandro, the on-and-off head of the AA-group, I thank him for his interesting discussions about in particular French history. I wish him all the best for his new job at LAM. Albert Rilliard has helped me with denoising of anechoic files for which I am thankful but foremost for his fun discussions during lunch. I wish him all the best. I thank Marc Evrard for his friendship during this thesis, I hope you are having a fun time in Japan. During my first months I was in an office with Olivier Perrottin and Lionel Fugere. I would like to thank them for showing me the French ways from French taxes to aid in language skills. I wish Olivier all the best for his stay in Scotland and Lionel for his new job in England. I am thankful to Peter Stitt for his help with the tracking system and ambisonic renderings. I wish him all the best for his job search in France. Samuel Delalez, the last of the Mohicans of the AA group, I am grateful for his aid with Max/MSP and PRAAT. Best of luck during the last months of your thesis. Finally, I am very happy that the presented frameworks are going to be improved and further studies are going to be carried out based on my work. This will be done by David Thery, I wish him all luck and promise I am also available for all questions.

Another great help during my PhD were my interns: Daniel Furlan, Julie Meyer, and Hugo Demontis. I would like to thank Daniel for his assistance during the measurements in the Amphithéâtre, Théâtre de l'Athénée, and Saint-Germain-des-Prés, Julie for her work on the Notre-Dame model and measurement, and Hugo for his contributions to the dynamic voice auralizations.

I would like to thank the personnel of the Théâtre de l'Athénée, Notre-Dame cathedral, and Saint-Germain-des-Prés church for their assistance and patience during the measurements. Additional thanks to THALIM and LAM for their help in hosting the listening test and to all participants of the listening test for their time. Special thanks to Bengt-Inge Dalenbäck (CATT-Acoustic) for the numerous and

lengthy informative discussions, critical ear during the preparation of auralizations, and comments in preparing manuscripts for several journal articles.

Finally, I would like to thank my wife, Claudia Postma-Bawelski, for everything but foremost for her patience. It was not easy living apart for more than three years. Now, I am looking forward living together with her in Germany.

This work was funded in part by the ECHO project (ANR-13-CULT-0004, [echo-projet.limsi.fr](http://echo-projet.limsi.fr)). Partners include THALIM/ARIAS-CNRS, Bibliothèque nationale de France (BnF), and LIMSI-CNRS.

---

## Abstract (English)

Over recent decades, auralizations have become more prevalent in architectural acoustics and virtual reality. Auralizations have numerous use cases such as multi-modal virtual reality explorations, studies of the acoustical influence of renovations, and historic research. Despite their numerous use-cases, they rarely have been part of scientific studies. Therefore, the goal of this thesis was to examine the use of room acoustical auralizations based on geometrical acoustics (GA) as a scientific tool and aimed to aid in the creation of historically more accurate auralizations.

Already in the 1930s first attempts were undertaken to render audible (imaginary) sound fields. Developments of both computer techniques and room acoustic knowledge the last 80 and years have led to significant improvements of auralizations. Today, if one wishes to create auralizations, generally anechoic recordings need to be convolved with either a measured or simulated room impulse response (RIR). GA software is often employed to numerically compute the RIR of complicated geometries. Wave-based methods are computationally intensive, requiring complex geometrical models and complex input data.

This thesis aims to enhance the quality of model based auralizations in comparison to measurement based. For this purpose the calibration of room acoustical models and the inclusion of dynamic voice directivity were studied. Room acoustical measurements were carried out in four rooms and geometrical acoustics models were created of the same spaces. A methodical calibration procedure was proposed, carried out in the four rooms: the Amphithéâtre, Théâtre de l'Athénée, Notre-Dame cathedral, abbey church of Saint-Germain-des-Prés. The calibration procedure entails 9 steps. First, a set of acoustic parameters to be calibrated was defined. Subsequently, the mean error across these parameter was minimized by adjusting the absorption and scattering coefficients of large surfaces. Finally, the error variance was minimized by adjusting the surface properties of local key surfaces. This calibration procedure was first validated by means of parameter estimation comparison. Additionally, listening tests which compared measured and simulated auralizations for three spaces found that the auralizations showed no or limited significant differences for 8 tested acoustical attributes.

It should be noted that in these comparison listening test the auralization comprised omni-directional sound sources. However, real-life sources such as the human voice or musical instruments have orientations which vary due to changing radiation patterns and dynamic orientation. Therefore, a framework was presented which enabled the inclusion of dynamic voice directivity. The results of subsequent listening tests indicated that auralizations which include dynamic voice directivity are perceived more plausible (in reference to a RGB video), wider, and exhibiting a wider apparent source width than regular static source orientations.

With the calibrated model and the inclusion of the more plausible voice directivity into the auralizations it was possible to study the influence of visualizations on the acoustical experience of the auralizations, with a reasonable degree of confidence that perceived effects are also applicable in real-life situations. For this purpose, a

framework was established which enabled multi-model assessments of theater plays and concerts. Results of an audio-visual coherent multi-modal listening test confirmed that there are differences between the acoustic perception of an audio-only and uni-modal presented auralization. A second listening test employing this framework which compared incoherent to coherent aural-visual cues indicated that by means of individual statistical analysis the test population could be divided into three sub-groups: 1) participants on which increased visual source-receiver distance influenced acoustical distance perception, 2) participants who rated auralizations with greater visual source-receiver distance louder, and 3) participants who rated all tested acoustical attributes similar under different visual conditions.

## Abstract (French)

Au cours des dernières décennies, la présence des auralisations dans l'acoustique architecturale et la réalité virtuelle est devenue de plus en plus importante. De nombreuses applications en découlent, telles que les explorations de la réalité virtuelle multimodale ou les études de l'influence acoustique des rénovations et la recherche historique. Malgré ces nombreux cas d'utilisation, peu d'études scientifiques ont été réalisées sur le sujet. L'objectif de cette thèse était donc d'examiner l'utilisation d'auralisations acoustiques de salles, basées sur l'acoustique géométrique (GA) comme outil scientifique et visant à aider à la création d'auralisations historiquement exactes plus écologiquement valables.

Déjà dans les années 1930 les premières tentatives ont été entreprises pour rendre audible (de manière imaginaire) les champs sonores. Les développements des techniques informatiques et de la connaissance de l'acoustique architecturale au cours des 80 dernières années ont conduit à des améliorations significatives des auralisations. Aujourd'hui, si l'on veut créer des auralisations, l'enregistrement anéchoïque réalisé préalablement doit être convolué avec une réponse impulsionnelle ambiante, mesurée ou simulée (RIR). Les logiciels GA sont souvent utilisés pour calculer numériquement la RIR de géométries compliquées. Les méthodes basées sur les ondes sont gourmandes en calcul, nécessitant des modèles géométriques complexes et des données d'entrée complexes.

Cette thèse vise à améliorer la qualité de *auralisations entièrement calculées*. À cette fin, on a étudié l'étalonnage des modèles acoustiques des salles et l'inclusion de la directivité vocale dynamique. Des mesures acoustiques de la pièce ont été réalisées dans quatre salles et des modèles d'acoustique géométrique ont été créés des mêmes espaces. Une procédure méthodique de calibration du modèle a été proposée, réalisée dans les quatre salles: Amphithéâtre, Théâtre de l'Athénée, Cathédrale Notre-Dame-de-Paris et Abbaye de Saint-Germain-des-Prés. La procédure d'étalonnage comporte 9 étapes. Tout d'abord, un ensemble de paramètres acoustiques à étalonner a été défini. Par la suite, l'erreur moyenne dans ces paramètres a été minimisée en ajustant les coefficients d'absorption et de diffusion des grandes surfaces. Enfin, la variance de l'erreur a été minimisée en ajustant les propriétés de surface des surfaces locales. Cette procédure d'étalonnage a d'abord été validée au moyen de la comparaison d'estimation de paramètres. Deuxièmement, des tests d'écoute subjectifs comparant des auralisations mesurées et simulées pour trois espaces ont révélé que les auralisations étaient également perçues pour huit attributs acoustiques évalués.

Il convient de noter que, dans ce test d'écoute de comparaison, l'auralisation comprenait des sources de directivité omnidirectionnelles. Cependant, les sources de la vie réelle telles que la voix humaine ou les instruments de musique ont des orientations qui varient en raison de l'évolution des diagrammes de rayonnement et de l'orientation dynamique. Par conséquent, un cadre permettant d'inclure la directivité vocale dynamique a été présenté. Les résultats des tests d'écoute ont montré des différences perceptuelles entre la directivité vocale dynamique et la directivité

de source statique pour la plausibilité, l'enveloppement de l'auralisation ainsi que la largeur perçue de la source.

Avec le modèle calibrée et l'inclusion de la directivité de la voix plus plausible dans les auralisations, il était possible d'étudier l'influence des visualisations sur l'expérience acoustique des auralisations, avec un degré de confiance raisonnable que les effets perçus sont également applicables dans des situations réelles. L'amélioration de la validité écologique des auralisations a permis d'étudier l'influence des visualisations sur l'expérience acoustique, avec un degré de confiance raisonnable que les effets perçus sont également applicables dans des situations réelles. À cet effet, un cadre a été établi qui a permis des évaluations multimodales de pièces de théâtre et de concerts. Les résultats d'un test audio-visuel cohérent d'écoute multimodale ont confirmé qu'il existe des différences entre la perception acoustique d'une auralisation présentée multi-modale ou uni-modale. Un deuxième test d'écoute utilisant ce cadre qui a comparé des indices audio-visuels cohérents et incohérents indiqués que, au moyen d'une analyse statistique individuelle, la population d'essai pouvait être divisée en trois sous-groupes: 1) les participants sur lesquels la distance visuelle-source augmentée d'influence la perception de la distance l'acoustique, 2) les participants qui ont évalué les auralisations avec une plus grande distance visuelle de source-récepteur plus fort et 3) les participants qui ont évalué tous les attributs acoustiques testés similaires dans différentes conditions visuelles.

## Publications

The work in this thesis has been partly presented in the following publications:

### Journal articles

- B.N.J. Postma and B.F.G. Katz. *Creation and calibration method of virtual acoustic models for historic auralizations* Virtual Reality, vol. 19, no. SI: Spatial Sound, pp. 161-180, 2015.
- B.N.J. Postma and B.F.G. Katz. *Correction method for averaging slowly time-variant room impulse response measurements* J. Acoust. Soc. Am., vol. 140, pp. EL38-43, July 2016.
- B.N.J. Postma and B.F.G. Katz. *Perceptive and objective evaluation of calibrated room acoustic simulation auralizations* J. Acoust. Soc. Am. vol. 140(6), pp. 4326-4337, Dec 2016.
- B.N.J. Postma, H. Demontis, and B.F.G. Katz. *Subjective evaluation of dynamic voice directivity for auralizations* Acta Acustica united with Acustica. Vol. 103, pp.181-184, Jan. 2017.
- B.N.J. Postma and B.F.G. Katz. *The influence of visual distance on the room-acoustic experience of auralizations* J. Acoust. Soc. Am. (submitted).

### Conference proceedings

- B.N.J. Postma and B.F.G. Katz, *A history of the use of reflections arrival time in pre-Sabinian concert hall design* in Forum Acusticum, (Krakow), pp. 1-6, Sept. 2014.
- B.N.J. Postma, A. Tallon, and B.F.G. Katz. *Calibrated auralization simulation of the abbey of Saint-Germain-des-Prés for historical study* In Intl. Conf. Auditorium Acoustics, Paris, 2015, pp. 190-197.
- B.N.J. Postma and B.F.G. Katz. *Dynamic voice directivity in room acoustic auralizations* in German Annual Conf. on Acoustics (DAGA), pp. 352-355, Mar 2016.
- B.N.J. Postma, D. Poirier-Quinot, J. Meyer, and B.F.G. Katz, *Virtual reality performance auralization in a calibrated model of Notre-Dame Cathedral* in Euroregio, (Porto), pp. 1-10, June 2016.
- B.N.J. Postma and B.F.G. Katz. *Acoustics of Notre-Dame Cathedral de Paris* in Intl. Cong. on Acoustics (ICA), (Buenos Aires), 2016.
- D. Poirier-Quinot, B.N.J. Postma, and B.F.G. Katz, *Augmented auralization: Complimenting auralizations with immersive virtual reality technologies*, in Intl. Sym on Music and Room Acoustics (ISMRA), (La Plata), pp. 1-10, Sept 2016.



- B.N.J. Postma and B.F.G. Katz, *Influence of visual rendering on the acoustic judgements of a theater auralization* in Forum Acusticum 2017, (Boston) (accepted).

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Research context . . . . .	1
1.2	Project context . . . . .	3
1.3	Contributions . . . . .	4
1.4	Outline . . . . .	4
<b>2</b>	<b>Background and Related Work</b>	<b>5</b>
2.1	Terms . . . . .	5
2.2	History of physical room-acoustical modeling . . . . .	8
2.3	Room acoustic computer modeling . . . . .	9
2.3.1	Wave based methods . . . . .	9
2.3.2	GA based methods . . . . .	13
2.3.3	Hybrid Method . . . . .	17
2.3.4	Validation . . . . .	18
2.4	From simulated energy echogram to (B)RIR . . . . .	19
2.5	RIR measurement . . . . .	21
2.6	Auralization projects . . . . .	22
2.6.1	Virtual reality simulations . . . . .	22
2.6.2	Historic research . . . . .	23
2.6.3	Architectural design . . . . .	25
2.7	Summary . . . . .	26
<b>3</b>	<b>Room acoustic measurements</b>	<b>27</b>
3.1	Introduction . . . . .	27
3.2	Studied rooms . . . . .	28
3.2.1	Amphithéâtre . . . . .	28
3.2.2	Théâtre de l'Athénée . . . . .	28
3.2.3	Saint-Germain-des-Prés church . . . . .	29
3.2.4	Notre-Dame cathedral . . . . .	29
3.3	Amphithéâtre, Théâtre de l'Athénée, and Saint-Germain-des-Prés church RIR measurements . . . . .	29
3.3.1	Protocol . . . . .	29
3.3.2	Parameter results . . . . .	32
3.4	Notre Dame RIR measurement . . . . .	32
3.4.1	Protocol . . . . .	32
3.4.2	Parameter results . . . . .	34
3.4.3	Correction for the time-variant acoustic system . . . . .	36
3.5	Summary . . . . .	40

<b>4</b>	<b>Model creation, calibration procedure, and objective validation</b>	<b>43</b>
4.1	Introduction . . . . .	44
4.2	Calibration procedure . . . . .	45
4.3	Creation of the geometrical models . . . . .	47
4.3.1	Amphithéâtre . . . . .	48
4.3.2	Théâtre de l'Athénée . . . . .	49
4.3.3	Notre-Dame cathedral . . . . .	50
4.3.4	Saint-Germain-des-Prés church . . . . .	50
4.4	Calibration of the GA models . . . . .	52
4.4.1	Simple model study on repeatability and influence of parameters	52
4.4.2	Amphithéâtre . . . . .	56
4.4.3	Théâtre de l'Athénée . . . . .	59
4.4.4	Notre-Dame cathedral . . . . .	62
4.4.5	Saint-Germain-des-Prés church . . . . .	65
4.4.6	GA models adjustments . . . . .	69
4.5	Historical studies employing the calibrated GA models . . . . .	69
4.5.1	Saint-Germain-des-Prés model . . . . .	70
4.5.2	Notre-Dame model . . . . .	73
4.6	Discussion . . . . .	74
4.7	Summary . . . . .	75
<b>5</b>	<b>Perceptual evaluation of calibrated auralizations</b>	<b>77</b>
5.1	Introduction . . . . .	77
5.1.1	Verification in the 1980s and 1990s . . . . .	78
5.1.2	Recent verifications . . . . .	79
5.2	Objective parameter results . . . . .	81
5.3	Preliminary auralization comparison listening test . . . . .	83
5.3.1	Preparation of the measured (B)RIRs . . . . .	84
5.3.2	Anechoic stimuli . . . . .	84
5.3.3	Protocol . . . . .	85
5.3.4	Results . . . . .	86
5.4	Binaural auralization comparison listening test . . . . .	90
5.4.1	Anechoic stimuli . . . . .	90
5.4.2	Protocol . . . . .	91
5.4.3	Results . . . . .	92
5.5	Discussion . . . . .	98
5.6	Summary . . . . .	99
<b>6</b>	<b>Voice-directivity: incorporation into auralizations and evaluation</b>	<b>101</b>
6.1	Introduction . . . . .	101
6.2	Source decomposition according to overlapping beam forming approach	104
6.2.1	Employed auralizations . . . . .	104
6.2.2	Protocol . . . . .	105
6.2.3	Results . . . . .	107

6.3	Evaluation of dynamic voice directivity in auralizations . . . . .	109
6.3.1	Employed auralizations . . . . .	111
6.3.2	Protocol . . . . .	112
6.3.3	Results . . . . .	113
6.4	Discussion . . . . .	115
6.5	Summary . . . . .	116
<b>7</b>	<b>Multi-modal auralizations and subjective evaluation</b>	<b>117</b>
7.1	Introduction . . . . .	117
7.2	Framework for multi-modal auralization . . . . .	120
7.2.1	Room acoustic rendering . . . . .	120
7.2.2	Visual room rendering . . . . .	120
7.2.3	Dynamic performance recording, point-cloud rendering in the VR world . . . . .	123
7.3	Coherent visual-aural experiment . . . . .	123
7.3.1	Protocol . . . . .	123
7.3.2	Results . . . . .	125
7.4	Incoherent visual-aural experiment . . . . .	129
7.4.1	Protocol . . . . .	129
7.4.2	Results . . . . .	131
7.5	Discussion . . . . .	137
7.6	Summary . . . . .	138
<b>8</b>	<b>Conclusion and future work</b>	<b>141</b>
8.1	Conclusion . . . . .	141
8.2	Recommendations . . . . .	143
8.2.1	Dynamic voice auralization framework . . . . .	143
8.2.2	Multi-modal auralization framework . . . . .	143
<b>A</b>	<b>A History of the Use of Reflection Arrival Time in Pre-Sabinian Concert Hall Design</b>	<b>145</b>
A.1	Introduction . . . . .	145
A.2	Echo theory before the discovery of Sabine's reverberation formula .	145
A.3	Venues in which these theories were used . . . . .	147
A.4	Current findings regarding early reflections . . . . .	152
A.5	Discussion & Conclusion . . . . .	153
<b>B</b>	<b>Virtual Reality Performance Auralization in a Calibrated Model of Notre-Dame Cathedral</b>	<b>155</b>
B.1	Introduction . . . . .	155
B.2	Project overview . . . . .	157
B.3	Recordings . . . . .	157
B.4	Room-acoustic model . . . . .	157
B.5	Visual model . . . . .	158

B.6	Integration of acoustics and visuals to create the VR experience . . .	159
B.7	Conclusion . . . . .	161
<b>C</b>	<b>Instructions perceptual tests</b>	<b>163</b>
	<b>Bibliography</b>	<b>177</b>

# Introduction

---

## Contents

<b>1.1</b>	<b>Research context</b> . . . . .	<b>1</b>
<b>1.2</b>	<b>Project context</b> . . . . .	<b>3</b>
<b>1.3</b>	<b>Contributions</b> . . . . .	<b>4</b>
<b>1.4</b>	<b>Outline</b> . . . . .	<b>4</b>

---

## 1.1 Research context

From as long as buildings with room acoustical demands have been constructed, architects and acousticians have attempted to predict and control their acoustics, evidenced by the work of Vitruvius [Morgan 1914]. During the 18<sup>th</sup> and 19<sup>th</sup> century the first numerical guidelines arose which were employed during the design process of several rooms [Postma 2013]<sup>1</sup>. These guidelines were based on time differences between the arrival of the direct sound and first order reflections. However, the foundation of architectural acoustics as a science was laid by Wallace Clement Sabine [Sabine 1922] at the end of the 19<sup>th</sup> century. During his research on the failed acoustics of the lecture hall in the Fogg-Art museum, he established the reverberation formula. He employed this formula in the design of the Boston Symphony Hall, which is today still regarded as having very favorable acoustics [Beranek 1996].

After the breakthrough findings of Sabine, attempts were undertaken to obtain the ‘real’ auditory impression during the construction phase of acoustic-orientated buildings. First attempts towards this application were undertaken by Spandöck in 1934 [Spandöck 1934]. He employed a scale model in combination with a gramophone technique for this purpose. An anechoic recording was performed using a wax cylinder. This recording was played in the 1:5 scale model 5 times faster than when recorded and the resulting sound was registered by another wax cylinder at the same speed. Finally, this was played back at normal speed over earphones in order to approximate the final impression of the venue. One could speculate that the results were unreliable due to the limited knowledge of the field. Over the last 80 years with the introduction of computers, increased acoustical knowledge, and gained experience, the reliability of rendering audible by physical or mathematical

---

<sup>1</sup>Appendix A discusses these guidelines in more detail.

modeling, the sound field of a source in space, in such a way as to simulate the binaural listening experience at a given position in the modeled space [Vorländer 2008] has improved. Kleiner et al. [Kleiner 1993] termed this process auralizations, in analogy to visualization.

Over recent decades, auralizations have become more prevalent in architectural acoustics and virtual reality. Auralizations are more and more accepted as a component of acoustic analysis and architectural design, obviously more so when the project targets are acoustic-oriented buildings (concert halls, theaters, etc.). Simulation results are typically presented employing objective room acoustical parameters. However, as room acoustic parameters may not have a one-to-one correlation with perceptual results it is important to verify simulation results both in terms of parameter comparison and in terms of comparison listening tests (auralization). Additionally, auralizations make the results of room acoustic design more tangible and accessible than descriptions of acoustic properties by abstract numerical quantities, making results accessible to all, from acousticians and architects to end-users. Moreover, the technique of auralization has the ability to transport the listener to another place, or even time. Despite their numerous use-cases, they rarely have been part of scientific studies. Therefore, the goal of this project is to examine the use of room acoustical auralizations as a scientific tool.

According to Kleiner et al. [Kleiner 1993] there are four basic techniques for auralization available:

1. *Fully computed auralizations* employ computed room impulse responses which are convolved with anechoic music, speech, or other suitable signals, and finally presented over a binaural system.
2. *Computed multiple-loudspeaker auralization* employs computed room impulse responses convolved with anechoic signals and presented over multiple loudspeakers.
3. *Direct acoustic scale model auralization* employs frequency scaled audio signals which are played in the scale model. Subsequently, the sound files are converted to full scale.
4. *Indirect acoustic scale-model auralization* employs room impulse responses measured in a scale model and is later on convolved with suitable signals.

Additionally, one can perform room acoustical measurements and convolve the resulting room impulse responses with anechoic signals. *Fully computed auralizations* require the simulation of the following two aspects in room acoustical software [Lokki 2008b, Vorländer 2011]:

1. 3D directivity of sound sources such as musical instruments and voice directivity.
2. Sound propagation in a 3D spaces such as concert halls, theaters, and churches.
3. 3D directivity of the receiver such as individualized Head Related Transfer Functions (HRTFs).

This thesis focuses on investigating the two first aspects of *fully computed auralizations*. Regarding the second aspect, the quality of room acoustical model can often be ensured through the use of calibration and validation procedures. However, there have been few studies examining the perceptual quality achievable by room acoustic auralizations [Lokki 2001, Lokki 2002a, Choi 2006, Yang 2007]. Such studies have highlighted potential problems in creating perceptually equivalent simulations when compared to measured auralizations. Therefore, this thesis proposes a methodical calibration procedure using room acoustical measurements as a reference. The results of this calibration procedure was studied by comparison between measured and simulated parameters and by means of two listening tests which compared measured to simulated binaural auralizations.

Regarding the first aspect of *fully computed auralizations*, in room acoustical software it is often possible to describe the directivity of an acoustic source in order to better represent the way in which a given acoustic source excites the room. However, such directivities are typically static, being defined according to source excitation as a function of frequency for the numerical simulation. While sources such as pianos vary little over the course of playing, it is known that source directivity of the human voice or other instruments such as trumpet, clarinet and french horn vary, sometimes considerably, due to both phoneme or tone dependent radiation patterns [Otondo 2004, Katz 2006] and dynamic orientation. Previous studies have performed dynamic source directivity using multi-channel anechoic recordings [Rindel 2004, Otondo 2005, Vigeant 2011]. In contrast, this thesis employs a single channel anechoic stimulus. It presents the means by which the dynamic directivity is incorporated into auralizations as well as the results of listening tests.

Parallel to auralizations, visualizations have the ability to transport a person to another space. With the calibrated auralizations in combination with visual VR models, it is possible to study the influence of visuals on the acoustical experience of auralizations, with a reasonable level of confidence that these effects are also present in real-life conditions. Multi-sensory processes are now well understood as being deeply involved in the perception of everyday events [Howard 1966, McGurk 1976]. The addition of visual feedback to an auralization, even based on static images, has been shown to impact the perception or *feel* of the room [Jeon 2008]. Therefore, a framework was developed which combined auralizations with 3D visuals of the surrounding room and acoustical sources. This enabled studies of the influence of visuals on the room acoustical experience of auralizations.

## 1.2 Project context

This thesis was carried out in the context of the ANR ECHO (Ecrire l'histoire de l'oral) project which is a cooperation between THALIM (Théorie et histoire des arts et des littératures de la modernité), LIMSI (Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur), and BnF (Bibliothèque nationale de France). This project studies the usage of voice in theaters over the previous century.



The contribution of this thesis is the auralization and visualization of present and historic configurations of the Théâtre de l'Athénée-Louis-Jouvet over its history.

### 1.3 Contributions

In this thesis a number of advances regarding the use of room acoustic auralizations are described. In particular the following contributions were made:

- A modern documentation of the acoustical parameters for the Notre-Dame de Paris cathedral, one of the most well-known places of worship, is provided.
- A method to enable averaging of RIRs which are subject to time-variant acoustic systems as a consequence of temperature changes is presented.
- A methodical calibration approach for geometrical acoustics models is presented and validated by means of comparing room-acoustical parameters between measurements and simulations.
- A study is carried out which showed to what extent simulated auralizations are similar to their measured counterparts.
- A framework is presented which enables the inclusion of dynamic voice directivity into auralizations.
- It is shown that the inclusion of dynamic voice based on orientation leads to more plausible auralizations.
- A framework is presented which couples real-time convolution based auralization and 3D visualization with the inclusion of a 3D audio-visual (3D-AV) recorded performance.
- It is shown that the inclusion of visuals affect the perceived *Distance* and *Loudness* of auralizations.

### 1.4 Outline

The remainder of this thesis is organized as follows. Background information and related work on room acoustical auralizations are described in Chapter 2. Chapter 3 outlines the measurements carried out in order to serve as a reference for the subsequent calibrations, described in Chapter 4. Chapter 5 discusses listening tests which compared calibrated simulated to measured auralizations. Chapter 6 addresses the use of dynamic voice directivity in auralizations and Chapter 7 studies the impact of the inclusion of a visual model on the acoustic experience of auralizations. Finally, Chapter 8 provides conclusions and suggestions for future work.

# Background and Related Work

---

## Contents

<b>2.1</b>	<b>Terms</b>	<b>5</b>
<b>2.2</b>	<b>History of physical room-acoustical modeling</b>	<b>8</b>
<b>2.3</b>	<b>Room acoustic computer modeling</b>	<b>9</b>
2.3.1	Wave based methods	9
2.3.2	GA based methods	13
2.3.3	Hybrid Method	17
2.3.4	Validation	18
<b>2.4</b>	<b>From simulated energy echogram to (B)RIR</b>	<b>19</b>
<b>2.5</b>	<b>RIR measurement</b>	<b>21</b>
<b>2.6</b>	<b>Auralization projects</b>	<b>22</b>
2.6.1	Virtual reality simulations	22
2.6.2	Historic research	23
2.6.3	Architectural design	25
<b>2.7</b>	<b>Summary</b>	<b>26</b>

---

In order to establish the context for the remainder of this thesis, this chapter summarizes the related work in the field of auralization. First the definitions employed throughout this thesis are provided. Sec. 2.2 discusses the history of room-acoustical modelling. In Sec. 2.3 the possible room acoustic calculation algorithms are described. As some room acoustic algorithms provide results in an energy echogram and auralization can only be created on the basis of room impulse responses, this transition is described in Sec. 2.4. Sec. 2.5 discusses methods which can be employed to obtain measured room impulse responses. Finally, previous auralization projects are described (see Sec. 2.6).

## 2.1 Terms

Due to the interdisciplinary nature of multi-modal virtual reality research, and the range of terms used for similar concepts throughout the literature, a brief list of pertinent terms frequently used in the following manuscript is provided here for clarity.

*Geometrical model* The geometrical model consisting solely of an ensemble of defined surfaces. The *geometry* of a room is a relatively straightforward concept, with the main variable being the complexity of the model.

*Geometrical Acoustics (GA) model* The model comprising the geometrical model with the inclusion of acoustic material properties (*absorption* and *scattering coefficients*) assigned to each surface.

*Geometrical Acoustics (GA) software* The prediction software which implements various GA algorithms used to carry out the numerical simulations based on the GA model as input. Throughout this study TUCT v1.1a/v.2.0, CATT-Acoustic v.9.0.c/v.9.1, was employed. This software allows for various physical feature definitions in the GA model consisting of the geometry, sound absorption, and scattering. These properties can be considered analogous to the geometry, RGB color values, and diffusion material properties in visualization software. The employed GA software is detailed in Sec. 2.3.

#### *Perceptual parameters*

- *Reverberance* Refers to the quality of a sound that persists in a room after a played tone is abruptly stopped [Beranek 1996].
- *Clarity* The degree to which discrete sounds stand apart from each other [Beranek 1996], relating to perceived definition and speech intelligibility.
- *Spaciousness* In an acoustical sense spaciousness consists of the perceptual components *Apparant Source Width (ASW)* and *Listener envelopment (LEV)* [Bradley 1995a, Bradley 1995b]. *ASW* describes the perceived horizontal extent of the acoustic image. The sources may sound ‘narrow’ (in the extreme case it is as if the sound is coming from a point). On the contrary the source can also sound very ‘wide’. *LEV* describes the sensation of being surrounded by the sound and room. Higher *LEV* means a more uniform distribution, while less *LEV* means a more localized or directional reverberant sound.

#### *Objective parameters related to cited perceptual parameters*

- *Reverberation time* Reverberance can be quantified in simple rooms by several acoustic parameters (T30, T20, and EDT), which vary as a function of frequency. T30 is the reverberation time (the time for a sound to decay by 60 dB) extrapolated from the decay between −5 and −35 dB. T20 is similarly obtained from the level decay between −5 and −25 dB. EDT is the early decay time, obtained as T30 but extrapolated from the decay between 0 and −10 dB. EDT is related to perceived “running reverberance”, while T30 and T20 are associated with the perceived “late reverberant” decay.

- *C50 and C80* Clarity correlates to the acoustic parameters C50 and C80 which are logarithmic ratios between early and late arriving energy, which vary as a function of frequency, calculated respectively for a time division of 50 (for speech) and 80 ms (for music).
- *Inter-Aural-Cross-Correlation* ASW correlates to the acoustic parameter Inter-Aural-Cross-Correlation<sub>early</sub> (IACC<sub>e</sub>) which is determined according to formulas 2.1 and 2.2 with  $t_1 = 0$  s and  $t_2 = 0.08$  s [ISO 2009]. LEV correlates to the acoustic parameter IACC<sub>late</sub> (IACC<sub>l</sub>) which is determined with  $t_1 = 0.08$  s and  $t_2 = 1.00$  s [ISO 2009].

$$IACF_{t_1 t_2}(\tau) = \frac{\int_{t_1}^{t_2} p_l(t) \times p_r(t + \tau) dt}{\sqrt{\int_{t_1}^{t_2} p_l^2(t) dt \int_{t_1}^{t_2} p_r^2(t) dt}} \quad (2.1)$$

where

$p_l(t)$  = the impulse response at the entrance of the left ear canal.

$p_r(t)$  = the impulse response at the entrance of the right ear canal.

The inter-aural cross correlation coefficients, IACC, are then given by Equation 2.2.

$$IACF_{t_1 t_2}(\tau) = \max \left| IACF_{t_1 t_2} \right| \text{ for } -1 \text{ ms} < \tau < +1 \text{ ms} \quad (2.2)$$

*Just Noticeable Difference (JND)* The minimal perceivable difference for a subjectively evaluated parameter. For simplicity and uniformity in this thesis, the ISO 3382 standard's [ISO 2009] JNDs (JND<sub>EDT</sub> = 5%, range = 1.0 – 3.0 s; JND<sub>C80</sub> = 1 dB, range = ±5 dB) were selected as calibration target thresholds for the different parameters. It is noted that these JND values vary as a function of room usage [Seraphim 1958, Cox 1993, Bradley 1999, Ahearn 2008], but they are chosen as a base model tolerance reference value for the purpose of this study. In a specific instance, a more suitable threshold should be used which is specifically appropriate to the room's function.

*Room Impulse Response (RIR)* is the sound pressure in a room resulting from a very brief sound pulse measured as a function of the time at a receiver position. From the RIR room acoustical parameters such as T30, T20, EDT, C50, and C80 can be obtained. When one employs a dummyhead microphone during the room acoustical measurement a binaural RIR (*BRIR*) can be obtained from which additionally spatial room acoustical parameters IACC<sub>e</sub> and IACC<sub>l</sub> can be acquired.

*Signal-to-noise ratio (SNR)* is the measure used to compare the RIR's maximum level to the level of background noise. It is defined as the ratio of signal power to the noise power in dB. In order to obtain a valid T30, T20, and EDT parameter result one needs to obtain a SNR of 45 dB, 35 dB, and 20 dB respectively.

## 2.2 History of physical room-acoustical modeling

With the definitions described, attention is focused on room-acoustical modeling. Room acoustic assessment tools can roughly be sorted in two categories: physical models and computer simulations. The earliest found example of physical modeling was based on investigating the distribution of acoustic energy [Davioud 1878]. In 1878 the Palais du Trocadero was inaugurated (see Fig. 2.1). The architects, Davioud and Bourdais, used the assumption that echoes arose when the path length of first order reflections was 34 m more than the direct sound to come to their acoustic concept<sup>1</sup>. The concert hall was “horse shoe” shaped, 62 m in length and 55 m in height. Therefore, numerous surfaces were  $> 17$  m from the center of the stage and measures had to be taken. The architects covered all surfaces  $< 17$  m from the center of the stage with plaster and surfaces located  $> 17$  m from the center of the stage were covered with painted silk. Additionally, 100 reflectors were placed on the back wall of the stage. Each reflector directed sound towards a different  $1/100^{th}$  of the audience. This design was tested with an optical acoustic scale model. The sound source was modeled using a light bulb, the absorbing surfaces with embossed copper and the reflective areas with polished silver. The light bulb was placed in the center of the stage, and it was noted how the light was distributed over the model. The outcome of the first test was that the “sound” was evenly distributed over the concert hall. These results however did not satisfy the architects as they thought the most remote parts in the hall needed more reflections. Therefore, the reflectors of the back wall were realigned and the test was repeated. Despite this acoustic study employing the optical scale model, the acoustic properties of the Palais du Trocadero turned out to be poor [Gournay 1985].

<sup>1</sup>Appendix A discusses ‘echo theory’, a design method which influenced the design of several rooms with room acoustic demands during the 19<sup>th</sup> and early 20<sup>th</sup> century.

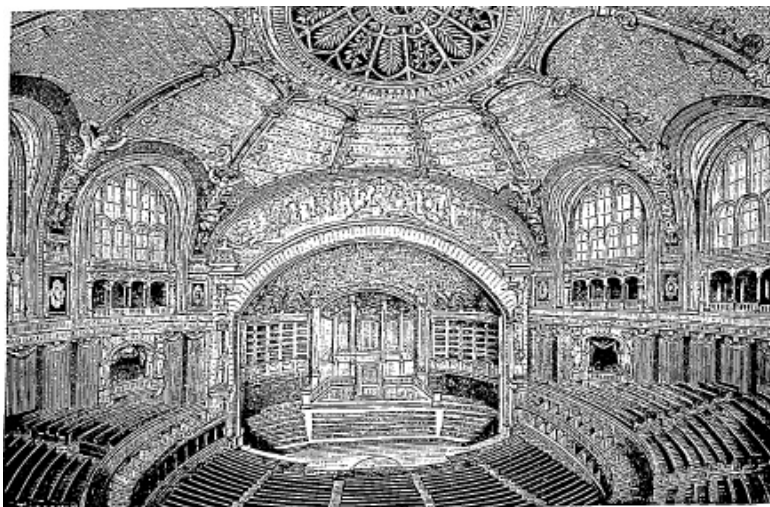


Figure 2.1: Overview of the Palais du Trocadero ([Davioud 1878] p.45).

Subsequently, physical models were developed based on the use of high-frequency acoustic waves propagating in 2D sections of scaled models, coupled with Schlieren photography to observe local variations in wave front density [Sabine 1913]. A similar method used ripple tanks [Davis 1927] where vibrating sources were used to create wavelets in shallow water to observe the acoustic properties of a given room geometry. Other phenomenological approaches were designed such as investigation of the distribution of acoustic energy [Vermeulen 1936] or the development of microphones accurate enough to enable scale model based techniques as used in modern acoustic design. Scale models allow for full 3D study of room acoustics [Spandöck 1934, Katz 2015b], scaling all physical dimensions of the room, including sound wavelengths. Techniques based on RIR recordings progressively replaced in-situ listening tests to assess rooms acoustical quality [Jordan 1941]. A continuous effort was since made at developing scale model techniques that would allow work with smaller scale factors [Barron 1979].

## 2.3 Room acoustic computer modeling<sup>2</sup>

Room acoustical simulations started with the landmark article of Schroeder et al. [Schroeder 1961] which simulated the transmission of sound in rooms using speech and music as signals. When one currently wishes to create simulated RIRs for auralizations, generally there is a choice between GA based and wave based methods. GA software are often employed to numerically compute the RIR of complicated geometries. Wave-based methods are computationally more intensive and therefore typically used for smaller rooms. Software which combine multiple calculation algorithms are called *Hybrid Methods*. Validation of room acoustic software is important in order to ensure that simulated auralizations approximate their measured counterpart.

### 2.3.1 Wave based methods

Wave based methods attempt to solve the actual wave equation numerically [Siltanen 2010].

The wave equation:

$$\nabla^2 p - \frac{1}{c^2} \frac{\delta^2 p}{\delta t^2} = 0 \quad (2.3)$$

where  $p$  is the sound pressure,  $c$  is the velocity of sound. This equation is valid for simple rectangular geometries, however more complicated room shapes require numerical approaches. Wave-based methods can be separated in spatial and/or temporal discretization. For brevity, wave-based algorithms discussed here are Finite Difference Methods (*FDM*), Finite Element Methods (*FEM*), and Boundary Element Methods (*BEM*). One can regard [Hamilton 2016] for an overview of other wave-based methods.

---

<sup>2</sup>This section greatly benefits from [Calamia 2009, Savioja 2015, Hamilton 2016], who give extensive overviews of room acoustic modeling techniques.



Typically wave-based methods divide the space into small elements or nodes [Elorza 2005]. These elements or nodes interact with each other according to the basics of wave movement phenomena. The size of these elements or nodes have to be much smaller than the wavelength (at least six  $\times$ ) for every particular frequency, causing increasing computational cost with increasing frequency. The number of natural modes in a room increases approximately with the  $3^{rd}$  power of the frequency, which means that for practical use, wave models are typically restricted to low frequencies and small spaces [Rindel 2000]. As these methods simulate propagation effects, detailed geometry and material properties (complex impedance) must be defined.

### 2.3.1.1 Finite difference methods (FDM)

*FDM* can be subdivided in first-order systems of equations (conservation equations) and second-order equations (e.g., the wave equation) [Hamilton 2016]. Both techniques employ finite difference operators that are centered around junctions in time or space of interest. The pressure time history at these junctions can be modeled through alternating updates of pressure and particle velocity at the grid points [Karjalainen 2005] (see Fig. 2.2). This can be accomplished for first-order systems by placing different discrete field components on Cartesian staggered space grids. The most popular staggered grids method today is probably Yee's scheme of Maxwell's equations, later on termed "finite difference time domain" (*FDTD*) method. The use of finite differences have developed at first according to these methods based on the work of Botteldoorn [Botteldoorn 1995] (first-order form) and Van Duyne [Van Duyne 1993] (second-order form i.e. digital waveguide/finite difference mesh). Eventually it was realized that these two methods were equivalent for the calculation of interior acoustic fields [Bilbao 2001]. Finally, it should be noted that *FDM*, whether expressed in first-order staggered form or second-order unstaggered form, have become synonymous with *FDTD* methods. *FDM* are typically employed for simple geometries.

*FDM* comprises several advantages and disadvantage:

- advantages:
  1. *FDM* are computationally less intensive than *FEM* and *BEM*. The computational intensity is reduced due to the simple implementation of the Dirichlet and boundary conditions [Murphy 2000, Chen 2006].
  2. *FDM* extracts standardized room-acoustics parameters, such as reverberation time, from RIRs rather than transfer functions [Calamia 2009].
  3. The *FDM* algorithm can be applied when the air in the room is non-still and inhomogeneous [Hargreaves 2005].
- disadvantage:
  1. Errors can occur during simulations with *FDM* [Hamilton 2016]. When the grid does not line up with the room boundaries a stair-stepped or

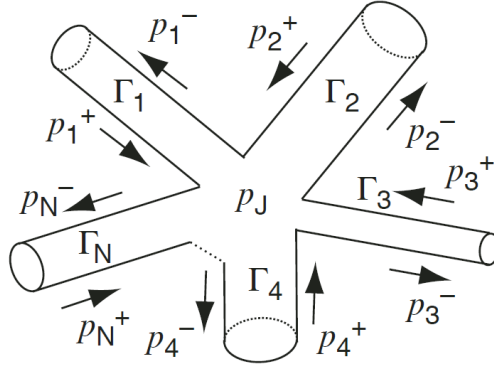


Figure 2.2: A scattering junction of connected acoustic tubes. Traveling pressure waves are denoted by ‘+’ for incident and by ‘-’ for scattered wave components (from [Karjalainen 2005]).

pixelated approximation is required [Luizard 2013]. At long wavelengths, the excess scattering should not substantially affect the results, but it is a potential source of error. Additionally, the boundary conditions are simulated by reflection coefficients [Luizard 2013]. These values are defined for normal incidence reflection. However, oblique incidence occurs during simulations and these effective values are slightly different.

2. One of the significant limitations is dispersion error, being both direction and frequency dependent, that limits the useful bandwidth of the results.

### 2.3.1.2 Finite Elements Methods (FEM)

In *FEM*, the spatial domain is discretized into polygonal or polyhedral elements—typically triangles or tetrahedra—and localized basis functions are chosen to span these elements [Hamilton 2016]. The solution is projected onto the basis elements and some chosen residual is minimized, resulting in a linear system of equations whose solution (the weighting coefficients) leads to an approximation to the unknown variable of interest (e.g., the pressure field). *FEM* are typically used for irregular rooms, especially when boundary conditions involve derivatives.

*FEM* comprise various advantages and disadvantages:

- advantages:

1. *FEM* employs more easily unstructured grids than *FDM*, allowing for adjustment to boundary surfaces that do not align with grid axes [Hamilton 2016].
2. Since the properties of each element are evaluated separately different material properties for each element can be incorporated [Desai 2001]. Consequently almost any degree of non-homogeneity can be included.
3. *FEM* can be applied when the air in the room is non-still and inhomogeneous.



- disadvantages:

1. In comparison to regular-grid *FDM* the implementation in practice of *FEM* is more complicated, as commonly a pre-processing meshing step is necessary [Hamilton 2016].
2. As all the air in the room must be modelled *FEM* is computationally intensive for larger volumes [Hargreaves 2005]. With a doubling of frequency or room size the number of elements or nodes are multiplied by eight.

### 2.3.1.3 Boundary Element Methods (BEM)

*BEM* also performs spectral discretization and attempts to numerically solve an appropriate wave equation on the surfaces of the modelled room [Siltanen 2010]. The surfaces are divided into elements on which the acoustic field is described aided by some basic functions. Finally, the interactions between the elements according to the wave equation and the boundary conditions are calculated which are imposed on the surfaces. For boundary conditions the acoustic admittance, which is inverse of acoustic impedance, is assigned to each surface element, and its value is determined for each one-third octave band from a corresponding absorption coefficient (consequently these values need to be complex numbers) [Luizard 2013].

*BEM* comprises several advantages and disadvantages:

- advantages:

1. *BEM* provide ease of use over *FDM* and *FEM* [Hamilton 2016]. Where *FDM* and *FEM* employ a 3D element mesh of the whole space, *BEM* uses only 2D elements on the surfaces which are the material interfaces or assigned boundary conditions (Dirichlet or Neumann). Especially in two dimensions where the boundary is a curve this allows for very simple data input and storage methods [Costabel 1987].
2. *BEM* employs more easily unstructured grids than *FDM*, allowing for adjustment to boundary surfaces that do not align with grid axes [Hamilton 2016].

- disadvantages:

1. In some case the computational intensity is high. The reduction of dimension means that the degrees of freedom in the numerical approximation is decreased significantly [Hamilton 2016]. However, the appearing system of equations could result in a full matrix, therefore the linear system solution can become computationally intensive.
2. In *BEM* simulations the acoustic admittance values are defined for normal incidence reflection [Luizard 2013]. However, oblique incidence occurs during the simulation and the effective values are slightly different.
3. The interactions between *BEM* elements take longer to calculate than *FEM* elements [Hargreaves 2005]. A room must be of a certain size before computational cost swings in the favour of *BEM*.

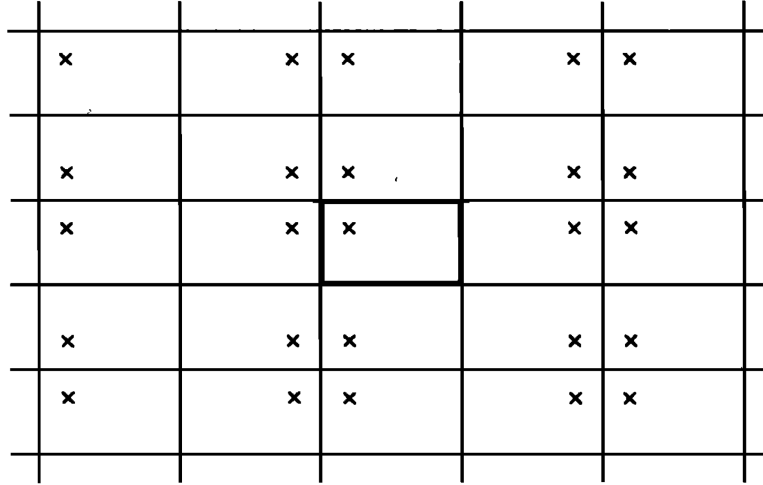


Figure 2.3: Conceptual schematic detail of the *ISM* in a rectangular room (from [Borish 1984]). The highlighted box in the center represents the real room, and the surrounding boxes depict the virtual rooms each containing a single virtual source.

### 2.3.2 GA based methods

Geometric models of complex surfaces can also be simplified relative to wave based models, with the application of scattering or diffusion properties to surfaces, as is performed by GA based methods. A period of great growth in GA modeling research occurred in the early 1990s [Savioja 2015]. Most current GA software, such as CATT-Acoustic, Odeon, and EASE, date back to that period. GA software considers, in analogy with light, sound as rays which transport energy along straight lines, which allows for simplified models of propagation, reflection, and scattering. Consequently, it provides a good approximation when the wavelength is small compared to the dimensions of the considered room. GA calculation algorithms comprise: *Image-Source Method*, *Ray and Particle Tracing*, *Beam and Cone Tracing*, and *Acoustic Radiosity*.

#### 2.3.2.1 The Image-Source Method

Image-source method (ISM) involves the recursive mirroring of a sound source about the reflecting planes in a virtual environment to find valid specular reflection paths between that source and one or more receiver positions [Calamia 2009] (see Fig. 2.3). This process is continued up to a prescribed order of the image sources. A final step examines which of the obtained image sources are “visible”, since a meaningful ray path can be assigned only to “visible” sources. Fig. 2.4 provides an example of an “invisible” source; as the virtual source from surface a is blocked by surface b this first-order specular reflection is non-existent and needs to be omitted from further analysis [Funkhouser 2014].

ISM-like applications have been used as long as room-acoustics have been regarded for the construction of buildings with acoustical demands (e.g. see [Langhans 1810]). However, in pre-Sabinian times only first order reflections were considered. In 1930 Eyring [Eyring 1930] employed multiple reflection order ISM to establish his reverberation-time formula. In the 1970s, ISM techniques were included into computer models [Gibbs 1972, Allen 1979].

Limitations of ISM techniques include the extension of the method to arbitrary shaped rooms [Borish 1984], computational efficiency [Kirszenstein 1984, Lee 1988], and simulating interference effects through the use of complex superposition [Suh 1999]. As [Allen 1979] noted, ISM techniques are only entirely correct for ideal (Neumann or Dirichlet) boundary conditions and certain geometries. Furthermore, the number of significant image sources grows exponentially with the length of the RIR.

In comparison to other GA calculation algorithms, *ISM* is restricted to only model specular reflections [Kuttruff 1993]. However, despite this limitation *ISM* ensures that all perceptually important specular early reflections are accounted for. Additionally, *ISM* is the only GA algorithm which can take phase information into account (see Sec. 2.4), although, in practice, the calculations are often done only with energy [Siltanen 2007]. *ISM* techniques are readily used in GA software, usually in combination with other calculation algorithms (see Sec. 2.3.3).

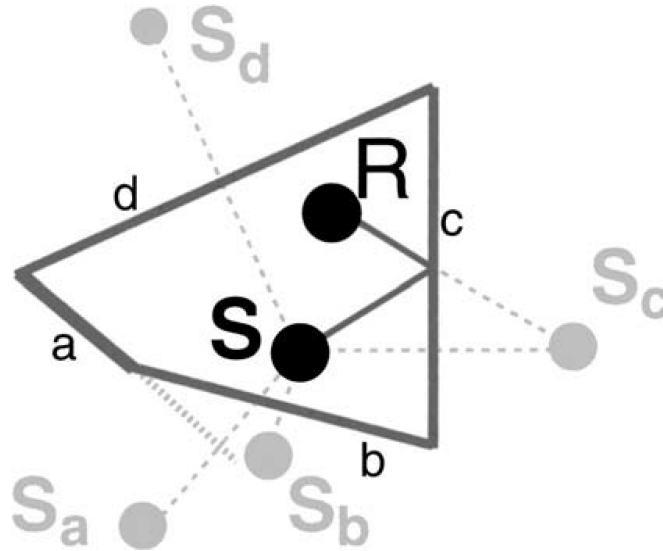


Figure 2.4: Example situation of an “invisible” source employing *ISM* (From [Funkhouser 2014]). The specular first-order reflection of surface a is non-existent.

### 2.3.2.2 Ray and Particle Tracing

A second GA calculation algorithm is *Ray and Particle tracing*. *Ray and particle tracing* methods find propagation paths between a source and receiver by generating rays emanating from the source position and following them through the environment until a set of rays has been found that reach the receiver (see Fig. 2.5) [Funkhouser 2014].

Early work on *ray and particle tracing* employed only specular reflections as evidenced by Allred and Newhouse's work [Allred 1958a, Allred 1958b]. They employed computation of *ray tracing* techniques with specular reflection to randomly choose initial ray directions in order to find the mean free path. They used this implementation to evaluate the room's shape on the reverberation time and the effectiveness of absorbers in the room. [Krokstad 1968] studied *ray tracing* techniques with specular reflections on the distribution of early reflected sound over the audience areas in concert halls, especially with respect to the shape of halls. In contrast [Schroeder 1970] studied the entire RIR using specular and diffuse reflections in 2-D enclosures:

1. He found discrepancies between existing reverberation-time formulas and decay rate based on the shape of the room.
2. He laid the foundations for simulated auralizations.
3. He simulated the frequency and spatial response of stationary sound fields and their statistical properties.

A final fundamental step was made by [Wayman 1977] who included *ray tracing* techniques based on Schroeder's work into 3-D models in order to estimate reverberation decay rates.

In comparison to *ISM* techniques, *ray tracing* models provide the ability to include diffuse reflections [Hodgson 1991]. However, results are less suitable for auralizations, as *ray and particle tracing* methods are unable to model strong

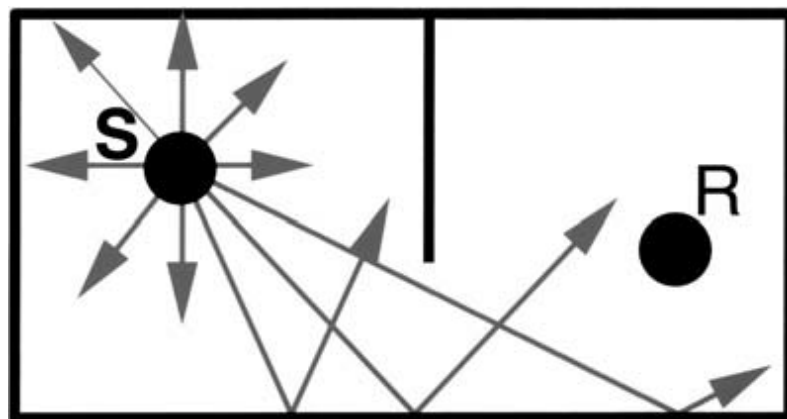


Figure 2.5: Conceptual schematic detail of *Ray and Particle tracing* (from [Funkhouser 2014]).

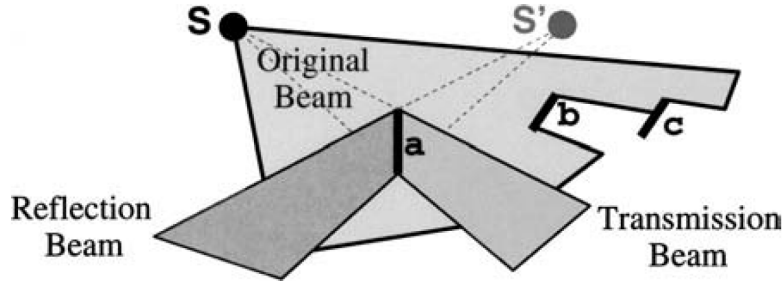


Figure 2.6: Conceptual schematic detail of *Beam and Cone tracing* (from [Funkhouser 2014]).

and isolated components caused by specular reflections from smooth and extended walls [Kutruff 1993]. Additional, disadvantages are the run-to-run variance between simulations due to both the stochastic implementation of the ray directivity and number of used rays [Kulowski 1982]. A final disadvantage is the potential detection errors related to finite-volume receivers and inadequate spatial sampling [Lehnert 1993].

### 2.3.2.3 Beam and Cone Tracing

A third GA calculation algorithm is *beam and cone tracing*. *Beam and cone tracing* methods classify propagation paths from a source by recursively tracing pyramidal beams (i.e., sets of rays) through the environment (see Fig. 2.6). Briefly, for each beam, polygons in the environment are considered for intersection with the beam in front-to-back visibility order (i.e., such that no polygon is considered until all others that at least partially occlude it have already been considered) [Funkhouser 2014].

*Beam and Cone tracing* arrived during the mid-nineties with the work of Lewers [Lewers 1993], Farina [Farina 1995b, Farina 1995c, Farina 1995d, Farina 2000a], and Stephenson [Stephenson 1996]. As beams and cones cover the entire spatial extent they allow for the employment of point receivers instead of receiver spheres (see *ray and particle tracing*). Additionally, since *Beam and Cone tracing* techniques rapidly identify available series of reflecting boundaries, they can be used to limit the calculation time of *ISM* technologies. An important limitation with *Beam and Cone tracing* is the inability to cover the full sphere around a source without gaps or overlapping cones. The former results in possible missed propagation paths and the latter in multiple detections, although this can be addressed with a Gaussian weighting across the cone faces to account for the overlap [Maercke 1993]. Another weakness of *Beam and Cone tracing* is a ray dependent late reflection loss which can be corrected for by an automatic reflection growth extrapolation [Dalenbäck 1996].

### 2.3.2.4 Acoustical Radiosity

A final GA calculation algorithm is *Acoustical Radiosity* (*Radiosity*). The governing equation (see Equation 2.4) needs to be solved numerically, which involves discretization of the boundary into elements [Koutsouris 2013]. In this way every surface is subdivided into small areas and the reflections are assumed to take place at predefined points on the surfaces, in contrast to ray tracing and ISM.

$$P_{diff}(x) = \beta(x)/\pi \quad (2.4)$$

where  $\beta$  is the reflectivity coefficient of the material at point  $x$ . This simplifies the room acoustic rendering equation by omitting the directional dependence. Then, discretizing the geometry  $G$  into patches transforms the room acoustic rendering equation into a matrix form [Siltanen 2007].

*Radiosity* is an energy-based method originated from computer graphics. The initial formulation of the method was performed by Kuttruff [Kuttruff 2000] in the early 1970s, and in the subsequent decade Miles [Miles 1984] applied it to a rectangular room. In the mid-nineties, improvements on the basic algorithm were carried out by Tsingos and Gascuel [Tsingos 1997] who implemented hierarchical patch subdivision.

As *Radiosity* assumes diffuse reflections it is effective in predicting the late part of the RIR where nearly all energy is diffusely reflected. A disadvantage is that *Radiosity* has higher computational cost than other GA calculation algorithms [Nosal 2004].

### 2.3.3 Hybrid Method

All discussed room acoustical simulation algorithms have their specific advantages and disadvantages. Therefore, several room acoustical tools combine multiple calculation algorithms (i.e. hybrid method). Crossover between calculation algorithms can be based on time or reflection order, spectral content, or reflection nature (e.g. specular and diffuse reflections).

Hybrids based on time or reflection order employ the assumption that early reflections are majorly specular as well as limited in number and late reflections are majorly diffuse as well as more numerous [Dalenbäck 1996]. As the temporal and directional structure of the direct sound and early reflections is of crucial importance for the sensation of loudness, clarity, spaciousness etc. [Cremer 1982], the correct simulation of this part is critical for perceptual accuracy. As such, *ISM* (possibly accelerated with *beam and cone tracing* techniques), a method which without errors predicts all spectral reflections, is typically employed to model direct sound and early reflections. The remaining part of the RIR consists of numerous overlapping reflections. Due to the randomizing effect of the enclosure this part becomes increasingly uniform and thus lacks information on the directional structure of the sound field. In some halls the sound field can be regarded as diffuse after 100–150 ms [Junius 1959]. Therefore, the correct simulation of individual late reflections is less important and

can be modeled statistically or with *ray tracing*. The transition between calculation algorithms has to be based on the simulated room; for instance for an acoustically mixing geometry the transition can be earlier in the acoustical response than for non-mixing geometries.

The main consideration for hybridization according to spectral content is computational intensity. When one employs wave based applications memory requirements grow cubically as a function of frequency. In contrast, GA based algorithms are not well-suited to model low frequencies as phase and wave effects are more prominent in these octave bands. Therefore, proposals of hybrids have been made which combine low frequency BEM with high frequency GA based calculation [Summers 2004, Southern 2013] as well as high frequency ray tracing with low frequency FDTD [Van Mourik 2014] or DWM algorithms [Schiettecattè 2003].

Besides hybridization based on time or octave band, separation of algorithms can also be based on specular and diffuse reflections. Various calculation algorithms, such as *ISM* and *Beam and Cone tracing*, are only able to simulate specular reflections. Therefore, several commercial software combine these algorithms for specular reflections with *ray-tracing* or *Radiosity* for diffuse reflections [Vorländer 1989, Drumm 2000, James 2001].

### 2.3.4 Validation

Validation of room acoustical software has been carried out through comparison between parameter results (e.g. Reverberation time and Clarity) of measured and simulated RIRs. The most pertinent validation was carried out by the *Physikalisch-Technische Bundesanstalt (PTB)* Braunschweig [Vorländer 1995, Bork 2000, Bork 2005a, Bork 2005b]. Three “Round Robins” on Room Acoustical Computer Simulation addressed a different space. For the first “Round Robin” multiple room acoustical computer software were employed to create a simulation model of the PTB auditorium in two phases. For both phases, plans and sections provided the reference for the geometry of the space. The first phase provided a description of the surface materials while the second distributed the absorption and scattering data. The second “Round Robin” compared multiple room acoustical simulation software for the *Elmia* hall in Jönköping (Sweden) a multipurpose hall. This “Round Robin” consisted of the same phases as the previous. “Round Robin” 3 consisted of three phases which compared a recording studio. Phase 1 provided a simple model consisting of seven walls with equal absorption and scattering for all walls across octave bands. For Phase 2 a more detailed model was provided where the fine structure of the diffusing elements of the ceiling and the wooden wall is neglected and represented as two planes with frequency dependent absorption and scattering coefficient. Phase 3 provided the same model, but with geometrical details for the diffusing areas. For all three phases the geometry data and absorption and scattering coefficients were specified and supplied. As “Round Robin 3” was carried out most recently, conclusions from that study are repeated here:



- GA software showed a good coincidence concerning the change in absorption as well as following the local dependency of the calculated parameters.
- The limit at low frequency performance can only be broken by introducing different algorithms which should also take into account the phase nature of sound propagation including also diffraction and complex surface impedances.
- Simulation results did not improve by increasing the level of geometric detail.

As wave-based calculation take into account wave effects like diffraction in contrast to GA based calculations [Mak 2015] they are more accurate. However, as the use of wave based techniques for room acoustical simulation is limited due to their computational intensity, validation attempts are scarcer. A website [Sakuma 2002a] presents benchmark problems for a wave based software round robin [Sakuma 2002b]. However, a widespread test of these benchmark problems did not occur thus far [Calamia 2009]. More benchmark cases have been provided by Hornickx et al. [Hornickx 2015].

Another comparison considered GA vs. wave based techniques. Luizard et al. [Luizard 2013] compared both GA (CATT-Acoustic and Odeon) and wave (BEM and FDTD) based simulated RIRs to measured RIRs obtained in a scale model. Comparisons were made for a specific coupled volume configuration, for 1 source and 4 receiver positions. The authors stated that one should be careful in overgeneralizing the results as it was one specific situation. However, the differences were quantified and GA software seemed to predict more accurate time-energy decay curves than custom codes for FDTD and BEM methods. Odeon predicted the late decay time more accurately while CATT-Acoustic better predicted the bending point level. Both wave based methods underestimated the reverberation time, however FDTD seemed to be globally closer to measurements than BEM.

## 2.4 From simulated energy echogram to (B)RIR

Due to the architectural size of the spaces under consideration in the current study, we will focus the rest of the discussion on GA methods. Where wave-based simulations produce RIRs, GA based simulation typically produce energy echograms [Dalenbäck 1996] (see Fig. 2.7). In order to serve for auralizations, these energy echograms need to be converted into RIRs. This conversion is not trivial as energy echograms contain less information than RIRs, such as phase.

The algorithmic representation of the energy echogram calculations can be written as follows,  $M$  denoting the number of beams and  $N$  the total number of reflections [Maercke 1993]:

$$h(t) = \sum_{i=1}^M \sum_{j=1}^N \frac{W Q_i}{(4\pi d_{ij}^2)(1 - \alpha_i) \exp(-2\alpha d_{ij}) D(\theta_{ij})} \quad (2.5)$$

where  $W$  is the power of the source,  $Q_i$  is the directivity function corresponding to the initial ray direction,  $d_{ij}$  is the distance from the source to the receiver, measured



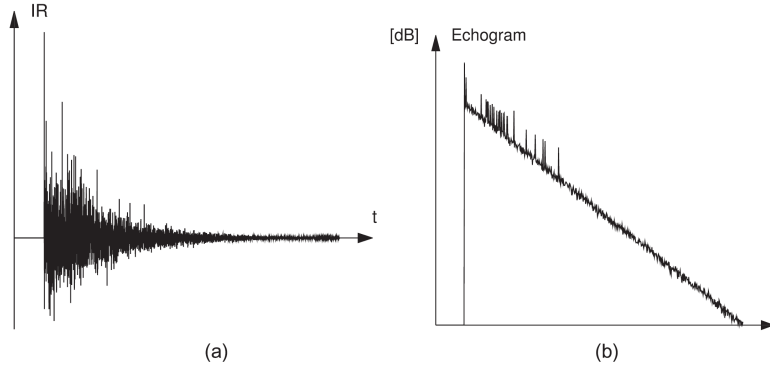


Figure 2.7: (a) An example room impulse response and (b) an example energy echogram (from [Savioja 2015]).

along the ray path,  $\alpha_{ij}$  is the absorption on the walls,  $\alpha$  is the air absorption coefficient, and  $D(\theta_{ij})$  is the weighting function of the beam as a function of  $\theta_{ij}$ , the angle between the ray's direction and the straight line connecting the receiver to the image source. The time distribution of the incoming energy is obtained by dividing the time axis into a given number of intervals of equal length and by adding each elementary contribution into the interval corresponding to its arrival time.

An approach to reinterpret the energy as pressure was presented by Kuttruff [Kuttruff 1993]. He performed this by converting the ray-trace's energy echogram data into equivalent RIRs. A simulation was run for a number of separate frequency ranges which results in a continuous function of frequency and time  $E(f, \tau)$ . A RIR with the same energy spectrum was obtained by taking the square root and Fourier transformation of  $E(f, \tau)$ . This required knowledge of the phase spectrum which is not contained in  $E(f, \tau)$ . Therefore, a sequence of Poisson-distributed points  $t_n$  on the time axis was selected by generating exponentially distributed random numbers. The equivalent RIR in the time interval  $\Delta t_i$  was obtained by:

$$g(t) = \sum_n g_i(t - t_n) \quad (2.6)$$

According to the author the randomization of the phase spectrum did not perceptually affect the auralizations [Becker 1993]. Thus the phase spectrum could be chosen in any reasonable way convenient from a computational point of view.

Another approach was described by Dalenbäck [Dalenbäck 1992, Dalenbäck 1995] who stated that the synthesis process basically consists of converting the energy echogram's octave band magnitude data to a narrow band representation, and constructing a corresponding phase function. He separated the RIR into two parts: 1) Direct sound & early reflections and 2) reverberation tail. As the direct sound & early reflections were simulated using *ISM* phase was included by means of Hilbert transform relations:

$$\arg(H_i) = H[\ln(|H_i|)] \quad (2.7)$$

With  $H_i$  being the  $i^{th}$  wall reflection filter (angle of incidence dependent). With this procedure, the reflection path transfer function is minimum-phase and yields a well-defined distinct impulse. Subsequently, a simplified synthesis was employed in order to construct the early part of the RIR. The RIR's reverberation tail was derived from a simplified model (virtual shoebox) as human hearing cannot distinguish separate late reflections. Image sources were calculated up to very high orders and the same procedure as for the early part was applied. In order to create a convincing diffuseness 5-6 sources surrounding the listener were applied.

As binaural auralizations provide a more natural reconstruction of the listening experience, a final requisite is to take into account the diffraction of sound around a listener's head [Dalenbäck 1996]. This is typically accounted for by including head related transfer functions (HRTFs) into the auralization filter. The result is a RIR for both ears (BRIR).

## 2.5 RIR measurement

As shown in Sec. 2.3, RIR measurements can provide a reference for GA model calibration. RIR measurements can be carried out using short impulsive sources such as exploding balloons and air shot guns. However, these sources exhibit problems in reproducibility [Pätynen 2011]. Therefore, omni-directional sound sources with more elegant excitation signals in combination with a deconvolution process are generally preferred:

- *Maximum Length Sequence (MLS)*. With this technique the room is excited with a sequence of pseudo random white noise signals [Stan 2002]. The RIR is obtained by circular cross-correlation between the measured output and the determined input.
- *Inverse Repeated Sequence (IRS)*. IRS employs two identical MLS sequences, however the second sequence being inverted [Farina 1997]. The deconvolution process is performed through circular cross-correlation as well.
- *Time stretched pulses*. This method employs a time expanded impulsive signal [Aoshima 1981, Suzuki 1995]. Due to the expansion process more sound energy is emitted than for an impulsive signal with the same amplitude resulting in an increased SNR. Obtaining the RIR from the measured signal is performed using a compression filter.
- *Sine swept method*. This technique employs a sine signal with an increasing frequency over time [Farina 2000b] (see Fig. 2.8). This method was originally proposed by Berkhout et al [Berkhout 1980]. The RIR is obtained from deconvolving the input from the measured signal resulting in separate impulse responses corresponding to the harmonic distortion orders considered [Rébillat 2011].

The sine swept method presents several advantages over the other techniques for indoor room acoustic measurements [Müller 2001]. The MLS, IRS, and Time-

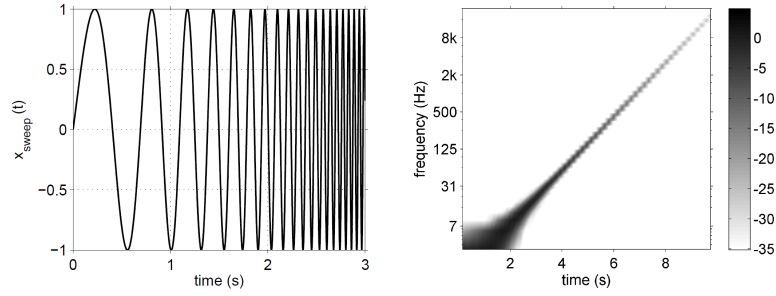


Figure 2.8: (Left) The first 3 s of a swept sine in the time domain and (right) a spectrogram of the swept sine (in dB) (from [Van Dorp Schuitman 2011]).

Stretched Pulses methods rely on the assumption of perfect linearity and time-invariance of the measurement system. The deconvolution process of the sine swept method separates source non-linearity within the measurement system [Stan 2002]. An additional advantage of this deconvolution process is that the harmonic distortions appear prior to the linear impulse response resulting in a higher SNR. A final pro is the higher reproducibility of the RIR. Summarizing, the sine swept method is very robust to minor time-variance of the measurement system, to mismatch between the sampling clock of the signal generation and recording, while obtaining a RIR with a high SNR [Farina 2000b].

## 2.6 Auralization projects

There are numerous cases in which auralizations are employed. This section will describe its usages in virtual reality simulations, historic research, and architectural design. This list is by no means exhaustive. In the description of these use-cases the emphasis is laid on the difference in calibration requirements.

### 2.6.1 Virtual reality simulations

In virtual reality simulation a set of RIRs is generated allowing the user to move around in the space typically combined with a visualization. Visualization can be presented on desktop screens, however more immersive options are Head Mounted Displays (HMDs) and cave-like applications. The perceptual validity of the auralization is dependent on the nature of the VR applications. For instance for game engines calibration may be of a lesser importance. Taylor et al. [Taylor 2010a, Taylor 2010b] created audio-visual scenes employing the game engine *Gamebryo* and rendered in real-time the auralizations. As it was a real-time implementation the early part of the RIR was based on simulating the specular and diffuse reflections and the late part on statistical means.

When one aims to realistically recreate an existing audio-visual scene calibration requirements become higher. For instance Lindebrink et al. [Lindebrink 2015] employed a software platform combining the game engine TyrEngine and the room

acoustical software BIM/CAD. RIRs were calculated and convolved with anechoic recordings offline. When progressing through the visual scene, the audio rendering was performed by playing the sound file of the nearest neighbor.

### 2.6.2 Historic research

Typically in historic research auralizations are employed to bring back to life the room acoustic conditions of spaces which do not exist anymore. The perceptual validity of these auralization is of a great importance. When the space is demolished or destroyed the calibration of the room acoustical model has to be based on historic sources. For instance Takala et al. [Takala 2014] and Niemi et al. [Niemi 2014] attempted to create realistic auralizations of theaters in Finland which no longer exist. Geometrical models were created according to surviving building plans. The absorption coefficients were selected according to visual inspection of the venues' surviving photos and scattering coefficients according to [Zeng 2006]. For the anechoic recordings *Mi tradi quell'alma ingrata* from Mozart's *Don Giovanni* was used, as it was performed in these venues at the time. Validation of the resulting auralizations was performed according to reviews of the acoustical qualities in historic sources such as newspapers.

Vissilantonopoulos et al. [Vissilantonopoulos 2001] presented historic auralizations of spaces which no longer existed as well. Instead of the mean goal being the accurate acoustic representation of the disappeared rooms, they were interested in simulating acoustic characteristics described in literature. For the Acheron Necromancy, a purpose-built temple in which rituals requiring high speech intelligibility occurred, they established it had 'dry' acoustics allowing good intelligibility for most receiver positions. The Olympia echo hall, famous for its seven time repeating echo, was found to give sufficient speech intelligibility only for listeners positioned within 5 meters of the speaker. The temple of Zeus, which was used for speeches during the Olympic games, was found to have a high level of EDT relative to T30 and 'good' Clarity values indicating positive speech conditions. Auralizations of all these simulated spaces were made available online.

A final example where the room no longer exists is Weinzierl et al. [Weinzierl 2010]. They studied the room-acoustic characteristics of the three historic Leiziger Gewandhaus concert halls by means of virtual reconstruction. The room-acoustic models were created with the GA software EASE 4.3. The geometrical models were based on architectural plans and images. Auralizations were created employing anechoic recordings of Mozart's *Quartett in G KV 80*.

When the space still exists, but has undergone (severe) modifications, a room acoustic model can be created and calibrated according to the current conditions and subsequently the model is reverted to the previous state in order to represent its former acoustic condition. Numerous studies can be found which employ this process; several examples are mentioned here without the intention to be exhaustive. In the ERATO project (2003-2006) various ancient Roman theaters and roofed odeums were virtually restored in terms of visual and acoustical

models [Gade 2004, Lisa 2004, Lisa 2006, Rindel 2006]. The virtual restitution integrated the visual and acoustical simulations, based on the most recent results of research in archaeology, theatre history, clothing, theatre performance, and early music. Visual models were created using *3dsMax*. Room acoustical models were created in Odeon and calibrations were performed according to measured parameter results. In parallel, anechoic recordings were made of contemporary scant, instrumental performances, and plays. These recordings were convolved with RIRs from GA simulation models.

A study of the Fogg Art Museum in which Wallace C. Sabine did his first tests on reverberation produced a historic room acoustically accurate model [Katz 2005]. Calibration was performed according to balloon measurements in a later configuration of the hall. Subsequently, the model was reverted to the state in which Sabine did his tests. The final result was a visual and acoustical virtual walk through in the room which saw the foundation of architectural acoustics as a science.

Pedredo et al [Pedrero 2014] virtually restored the sound of Hispanic rite. A group of 5 Spanish pre-Romanesque churches was simulated using virtual acoustic reality technologies in its current configuration. These models were calibrated with room acoustical parameter results from in-situ measurements. Subsequently, changes in the room acoustical models were made to revert them to previous configurations. The acoustic response of these churches was convolved with anechoic recordings of original Mozarabic Chant repertoire [Asensio 2005]. In parallel, visual models were created using *SketchUp 8*. The final result was a visual and acoustic virtual walk through in the churches while avatars were singing Hispanic rite(s).

Garcia et al. [Garcia 2014] studied the change in acoustic conditions of the “Misteri d’Elx” which was performed in the Basilica “Santa Maria de Elche”. The model of the Basilica was calibrated based on in-situ measurements. It was set out that the average simulated T30 needed to differ less than 5% of the measured values at 500 and 1000 Hz. Absorption coefficients were selected from existing databases. The calibration was performed by adjusting scattering and absorption coefficients of the pews and adjusting the scattering calculation algorithm. Finally, an anechoic recording was performed of the “Misteri d’Elx”, which enabled a virtual recreation of the mystery play in the “Santa Maria de Elche”.

In 2015, Weinzierl et al. [Weinzierl 2015] studied the room-acoustical properties of three renaissance theaters in Italy. Impulse response measurements were the basis for the subsequent room-acoustic model calibration. Models were created and simulations were ran in the GA software Odeon. The calibration concerned the theater’s recent configurations in the commercial GA software Odeon. RIR envelopes and room acoustics parameters were compared between measurement and simulation. Post-calibration two of the three models were reverted to the original state, as the last room was preserved this model remained unchanged. Auralizations were made available online.

In the context of architectural heritage, Murphy et al. [Murphy 2016] created and calibrated a GA model of the St Margaret’s church in York. The simulations were performed in Odeon and CATT-Acoustic. First, the absorption coefficient of

the main walls were based on the early reflections energy measured with Soundfield measurement. Scattering coefficients were based on literature. A set of auralizations was made available online.

All these studies compared measured to simulated parameter estimations. However, as the auralizations with similar parameter estimations can still sound very different preference should be given to auralization comparison by means of listening tests. The CAHRISMA project (2000-2003) focused on the hybrid architectural heritage which covered both acoustic and visual features. The examples studied in the project were 6<sup>th</sup> century Byzantine churches and 16<sup>th</sup> century Turkish Mosques were studied. [Weitze 2001, Weitze 2002, Rindel 2003] created auralizations of these spaces. Both measured parameters and auralizations were compared to their simulated counterparts.

### 2.6.3 Architectural design

In architectural design auralization have two incentives: 1) aid in the design process of a room with particular acoustical demand and 2) investigate and demonstrate the acoustic implications of modifications to the architectural space. In both of these applications the auralization's ecological validity is also of great importance. However, in the first case calibration is troublesome as the room does not exist yet. Azevedo et al. [Azevedo 2014] presented numerous cases in which the presentation of various design options via auralizations led to candid discussion of the different acoustical treatment options or the sound of the entire space. Two cases are summarized here to give an idea of these discussions:

1. For the design of a variable chamber music hall in the University of Massachusetts, auralizations presented several different types of performance heard at three audience positions, with various curtain configurations. The auralization helped the university music faculty to gain confidence in the design, particularly its acoustical variability.
2. For an amphitheater-like lecture hall in the Olin Business School, auralizations were performed to show the effect of five varying levels of acoustical treatments. The school administrators were able to make the connection between the architectural design, the numerical descriptions of the room acoustic parameters, and perceive the described environment through auralizations.

In the case of architectural modifications calibration of the room acoustic model is possible. [Azevedo 2014] also presented two examples which give an idea of this implementation:

1. In the renovation of the Margery Milne Battin Hall, the users were unsatisfied with the sound reinforcement system and the HVAC system was said to be too loud. The principals were presented with calibrated auralizations of different solutions for the PA systems and HVAC-noise.

2. Auralizations were employed to ascertain the noise impact on the animals in the tank of the New England Aquarium, as well as exploring potential room treatments to reduce the overall noise level for the comfort of their staff and guests. A model was created and calibrated according to measurements and many sounds were recorded within the space. The auralizations showed that the most effective measure for quieting the exhibit was to control the visitors' number.

## 2.7 Summary

This chapter provided an overview of related studies in the field of architectural auralizations. The advantages and disadvantages of various approaches of room-acoustic modeling were reviewed. As this thesis studies rooms with considerable size, it employs GA software. Because GA based methods typically result in echograms, the process of the addition of phase in order to create RIRs was described. As measured acoustical parameters can be employed as a reference for the GA model calibration, various room-acoustic measurement techniques were described and compared. The perceptual validity of the auralizations depends on its use-case. Therefore, chapters 3-5 of this thesis are devoted to the creation of perceptually valid architectural auralizations employing GA models. In several previous auralization projects source directivity of the human voice was simulated with a static sound radiation, to improve the ecological validity of the sound source Chapter 6 studies the inclusion of dynamic source directivity. As in these previous projects auralization were combined with visuals Chapter 7 studies the influence of visuals on the room acoustic experience of auralizations.

# Room acoustic measurements<sup>1</sup>

## Contents

<b>3.1</b>	<b>Introduction</b>	<b>27</b>
<b>3.2</b>	<b>Studied rooms</b>	<b>28</b>
3.2.1	Amphithéâtre	28
3.2.2	Théâtre de l'Athénée	28
3.2.3	Saint-Germain-des-Prés church	29
3.2.4	Notre-Dame cathedral	29
<b>3.3</b>	<b>Amphithéâtre, Théâtre de l'Athénée, and Saint-Germain-des-Prés church RIR measurements</b>	<b>29</b>
3.3.1	Protocol	29
3.3.2	Parameter results	32
<b>3.4</b>	<b>Notre Dame RIR measurement</b>	<b>32</b>
3.4.1	Protocol	32
3.4.2	Parameter results	34
3.4.3	Correction for the time-variant acoustic system	36
<b>3.5</b>	<b>Summary</b>	<b>40</b>

## 3.1 Introduction

The 2<sup>nd</sup> aspect of *fully computed auralizations* is sound propagation in the 3D space. In order to serve as a reference for the modeling of this aspect RIR measurements were carried out. Four test rooms were selected, which are described in Sec. 3.2. Sec. 3.3 describes the measurements of the Amphithéâtre, the Théâtre de l'Athénée, the Saint-Germain-des-Prés church and Sec. 3.4 describes the measurement in the Notre-Dame cathedral.

<sup>1</sup>This work was partly presented in:

- B.N.J. Postma and B.F.G. Katz. *Correction method for averaging slowly time-variant room impulse response measurements* J. Acoust. Soc. Am., vol. 140, EL38-43, July 2016.
- B.N.J. Postma and B.F.G. Katz. *Acoustics of Notre-Dame Cathedral de Paris* in Intl. Cong. on Acoustics (ICA), (Buenos Aires), 2016.
- B.N.J. Postma, A. Tallon, and B.F.G. Katz. *Calibrated auralization simulation of the abbey of Saint-Germain-des-Prés for historical study* In Intl. Conf. Auditorium Acoustics, Paris, 2015, pp. 190-197.





Figure 3.1: Overview photos of the (a) Amphithéâtre (18-09-2014), (b) Théâtre de l'Athénée (22-07-2014), (c) Notre-Dame cathedral (from [Murray a]), and (d) Saint-Germain-des-Prés church (from [Murray b]).

## 3.2 Studied rooms

### 3.2.1 Amphithéâtre

The first room to be considered was a small unfinished 200-seat Amphithéâtre (French for lecture hall). As a test case, it was chosen for its accessibility, general simple geometry & materialization, and considerable level of reverberation.

### 3.2.2 Théâtre de l'Athénée

The Parisian Théâtre de l'Athénée is a 4-floor 570-seat theater. It was designed by architect Stanislas Loison in one of the foyers of the Éden-Théâtre. The Théâtre de l'Athénée opened in 1893 and was closed again on 9 May 1895 after two bankruptcies.

In 1896 the Athénée Comique took possession of this room which was renamed at 28 December 1898 the Comédie Parisienne and at 25 October 1899 renamed again to the Théâtre de l'Athénée. When Louis Jouvet became director of this theater in 1934 it changed its name for the last time and was renamed Théâtre de l'Athénée-Louis-Jouvet. This space was selected as it was part of the ECHO project.

#### 3.2.3 Saint-Germain-des-Prés church

The third considered space is the abbey church of Saint-Germain-des-Prés. It is approximately 67 m long, 27 m wide, and 18 m high. The large volume in combination with its vast painted plaster and marble surfaces lead to long reverberation times. The abbey church was begun in the 11<sup>th</sup> century, with major modifications undertaken in the 12<sup>th</sup> and again 17<sup>th</sup> centuries which resulted in changes in the acoustic conditions [Tallon 2016]. This space was chosen for its historical interest and combined study with the art department of Vassar college.

#### 3.2.4 Notre-Dame cathedral

The final room is the Cathédrale Notre-Dame de Paris. This cathedral is amongst the most well-known worship spaces in the world. This medieval cathedral is widely considered to be one of the finest examples of French Gothic architecture. It is approximately 130 m long, 48 m wide, and 35 m high. The large volume in combination with its vast exposed limestone and marble surfaces lead to long reverberation times. This room was selected as it was part of the BiLi-project in which it was employed for a VR demonstration (see Appendix B).

### 3.3 Amphithéâtre, Théâtre de l'Athénée, and Saint-Germain-des-Prés church RIR measurements

As the measurement protocol and equipment usage were similar for the Amphithéâtre, the Théâtre de l'Athénée, the Saint-Germain-des-Prés church these are described in a single section.

#### 3.3.1 Protocol

The room-acoustic measurements in the Amphithéâtre, Théâtre de l'Athénée, and Saint-Germain-des-Prés church were performed according to the following details:

- *Measurement positions* - Fig. 3.2 shows the measurement plans, indicating source and receiver position which were chosen to represent typical usage. In addition to RIR measurement for calibration, BRIRs were also measured for the subsequent listening test for the Théâtre de l'Athénée and Saint-Germain-des-Prés church, their positions are indicated by the numerated positions. No omnidirectional microphones were placed on the positions of the dummyhead.

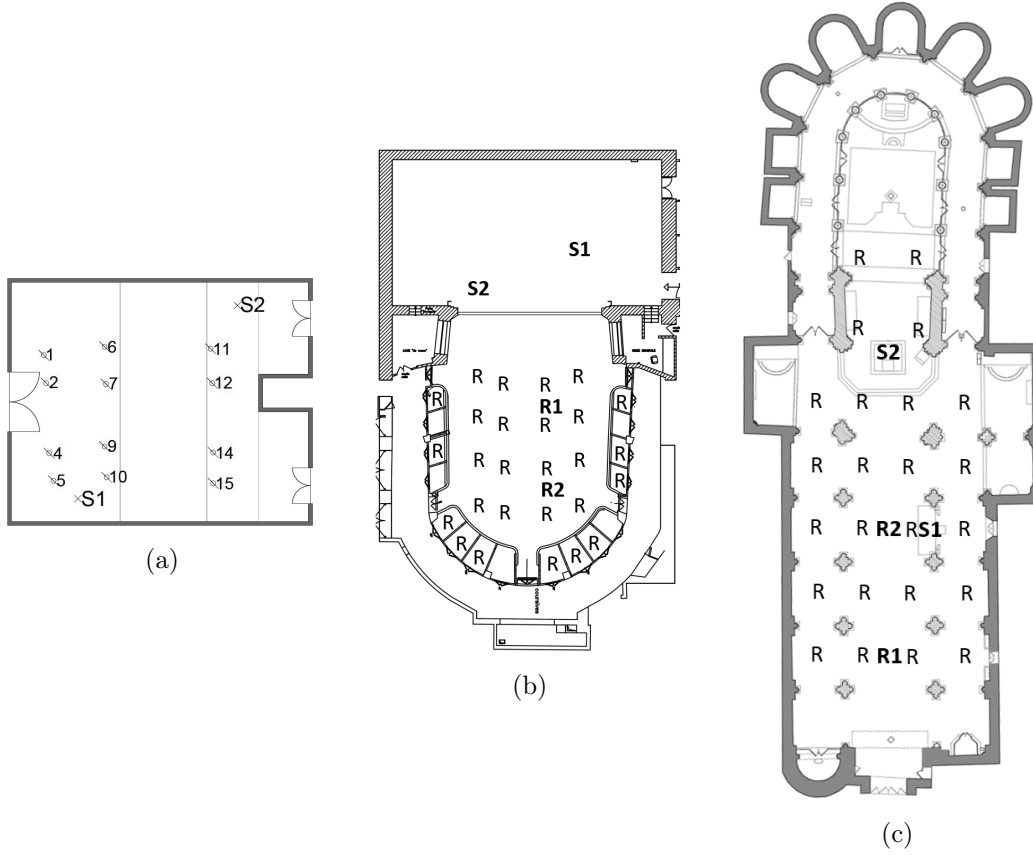


Figure 3.2: Measurement plans of (a) Amphithéâtre, (b) Théâtre de l'Athénée, (c) Saint-Germain-des-Prés church. **S** and **R** represent source and receiver positions (**S**# and **R**# were employed in the listening test).

- *Signal* - The Exponential Swept Sine method was employed. The sweep frequency went from 20 Hz to 20 kHz, duration of 10 s.
- *Measurement equipment output* - The stimuli signal was sent to an amplifier (Servo 120a, SAMSOM) and sequentially to miniature (diameter = 0.12 m) dodecahedral omnidirectional sound sources (model 3D-032, Dr-Three; variation in directivity within  $\pm 3$  dB up to 4 kHz, considered per  $1/3^{rd}$  octave band).
- *Measurement equipment input* - Four omnidirectional microphones (model 4006, DPA) were employed to cover audience areas. It should be noted that the dummy head (Neumann KU 80, equipped with model 4060, DPA) orientation in the Saint-Germain-des-Prés church was towards **S2** and in the Théâtre de l'Athénée towards the center of the stage.
- *Post-processing* - Post measurement deconvolutions were carried out using

Table 3.1: Mean and Standard Deviation (SD) over different source-receiver combinations of the EDT, T20, C50, and C80 for the Amphithéâtre (24 s-r combinations), Théâtre de l’Athénée (Athénée) (88 s-r combinations), and Saint-Germain-des-Prés church (SGdP) (48 s-r combinations).

Frequency band (Hz)	Mean T20 (SD)		
	Amphithéâtre	Athénée	SGdP
125	2.11 (0.26)	2.06 (0.20)	-
250	2.33 (0.16)	1.74 (0.10)	7.36 (0.23)
500	3.06 (0.09)	1.53 (0.08)	7.00 (0.20)
1000	3.35 (0.07)	1.31 (0.05)	6.04 (0.19)
2000	3.08 (0.05)	1.17 (0.03)	4.57 (0.15)
4000	2.33 (0.03)	1.03 (0.03)	3.16 (0.14)
	Mean EDT (SD)		
125	1.82 (0.34)	1.91 (0.36)	7.31 (1.34)
250	2.30 (0.29)	1.81 (0.24)	7.15 (0.79)
500	2.96 (0.20)	1.56 (0.17)	6.72 (0.67)
1000	3.27 (0.12)	1.30 (0.16)	5.76 (0.62)
2000	2.99 (0.15)	1.20 (0.13)	4.38 (0.58)
4000	2.20 (0.13)	0.98 (0.11)	2.99 (0.43)
	Mean C50 (SD)		
125	-0.70 (2.45)	1.71 (2.43)	-5.49 (2.95)
250	-2.06 (1.87)	1.04 (1.75)	-7.32 (3.36)
500	-4.13 (2.01)	1.08 (1.25)	-7.15 (3.04)
1000	-4.90 (1.74)	1.63 (1.46)	-7.07 (3.09)
2000	-4.75 (1.67)	1.93 (1.79)	-7.07 (3.15)
4000	-3.20 (1.89)	2.57 (1.70)	-4.27 (2.82)
	Mean C80 (SD)		
125	1.72 (1.98)	2.32 (2.27)	-3.97 (2.86)
250	-0.23 (1.83)	1.73 (1.49)	-5.83 (2.99)
500	-2.15 (1.65)	2.40 (1.13)	-5.70 (2.65)
1000	-2.93 (1.35)	3.45 (1.32)	-5.38 (2.64)
2000	-2.69 (1.42)	3.28 (1.65)	-5.09 (2.49)
4000	-1.04 (1.60)	4.99 (1.53)	-2.36 (2.24)

MatLab<sup>2</sup>. The resulting RIRs were analyzed using LIMSI’s in-house MatLab impulse response analysis (IRA) toolkit. For the purpose of this study, six standard [ISO 2009] parameters were calculated: T20, EDT, C50, C80, IACC early ( $IACC_e$ ), and IACC late ( $IACC_l$ ).

### 3.3.2 Parameter results

Table 3.1 presents the obtained average and standard deviation (SD) of the T20, EDT, C50, and C80 of the Amphithéâtre, Théâtre de l'Athénée, and Saint-Germain-des-Prés church. As can be expected the Saint-Germain-des-Prés church showed longer reverberation parameters and lower clarity parameters than the other two spaces. Consequently, the Just Noticeable Difference (JND) of reverberation time for the Saint-Germain-des-Prés is higher as it is a percentage of this parameter.

## 3.4 Notre Dame RIR measurement

The Notre-Dame measurements were carried out in conjunction with an installation setup typically used for concert recordings. Additionally, it was found that the temperature slightly changed over the measurements creating a time-variant acoustic system. Finally, the results are compared to a 30-year old measurement. For these reasons this measurement is separately described. Despite the fame of the Notre-Dame cathedral, there are few examples of published data on the acoustical parameters of this space. Hamayon [Hamayon 1996] presented reverberation time estimations as a function of octave bands [125-4000 Hz: 8.5, 8.0, 7.5, 6.0, 4.5, 2.7 s]. Mercier [Mercier 2002] presented slightly different reverberation time values [125-4000 Hz: 8.5, 8.2, 6.5, 6.2, 4.7, 2.5 s]. Both studies presented simply the reverberation times without any measurement protocol information, nor with information about other room-acoustical parameters.

Therefore, an acoustical measurement (2015) employing sine-sweeps and balloon bursts as excitation signals was carried out. Additionally, archival recordings (1987) were recovered which included balloon bursts [Castellengo 1987]. The measurement protocol of these two measurements are presented in Sec. 3.4.1 as well as their parameter results in Sec. 3.4.2.

During the 2015 measurements, 4 S(ource)  $\times$  8 R(eceiver) combinations recorded 4 sets of 2 sine sweeps over the course of 0.5 hour. It was identified that averaging of RIRs resulted in reverberation time underestimations. Sec. 3.4.3 studies the cause for this error.

### 3.4.1 Protocol

The RIR measurements in the Notre-Dame cathedral were performed according to the following details:

#### 3.4.1.1 1987 measurement

- *Measurement positions* - Fig. 3.3a shows the measurement plan for the 1987 measurements with S-R positions. While a variety of techniques using different stimuli were employed, only balloon burst sources were exploitable due to

---

<sup>2</sup><http://www.mathworks.com>

lack of anechoic stimuli details. 3 balloon bursts from source position 1 were recorded as well as 1 balloon burst from source position 2.

- *Measurement equipment input* - The sound was recorded with 13 omnidirectional microphones which were connected to a multitrack recorder (Tascam).

#### 3.4.1.2 2015 measurement

- *Measurement positions* - Fig. 3.3b shows the measurement plan highlighting S-R positions for the 2015 measurement. 3 measurement sets of 2 sine-sweeps during which microphones 1-8 changed positions were carried out (the changing positions of these microphones are represented by the letters behind the measurement position). Due to excessive exterior noise, the first measurement repetition was carried out twice, resulting in 4 measurement sets. Microphones 9-16 hang from the ceiling (7 m above floor level), thus remained at the same position and consequently recorded eight similar RIRs. After the last sweep measurement, a balloon burst at every source position was recorded with the receivers at the final position to provide comparable stimuli to the 1987 measurements.
- *Signal* - The Exponential Swept Sine method was employed. The sweep frequency went from 20 Hz to 20 kHz, duration of 20 s.
- *Measurement equipment output* - The output measurement equipment was equal to the Amphithéâtre, Théâtre de l'Athénée, and Saint-Germain-des-Prés measurements.
- *Measurement equipment input* - The input signal was recorded by two measurement chains as the measured session was carried out in conjunction with a concert recording installation. 1) The sweep was recorded at a sample rate of 44.1 kHz by 5 omni-directional microphones (4 DPA model 4006 (1-4) and 1 Schoeps model MK5 omni, (5) ) and transferred to a sound card (RME Fireface 800). 2) The sweep was recorded at a sample rate of 48 kHz by the other 11 omni-directional microphones (6 DPA model 4006 (11-16), 5 Schoeps model MK5 omni, (6-10)) and using a sound card (RME Micstasy). The dummyhead microphone (Neuman KU 80, equipped with model 4060, DPA) was positioned on 1a, b, and c.

#### 3.4.1.3 General

- *Post-processing* - Subsequent deconvolution, sample rate conversion, and post-processing steps were performed in MatLab. Other steps were equal to the Amphithéâtre, Théâtre de l'Athénée, and Saint-Germain-des-Prés measurements.



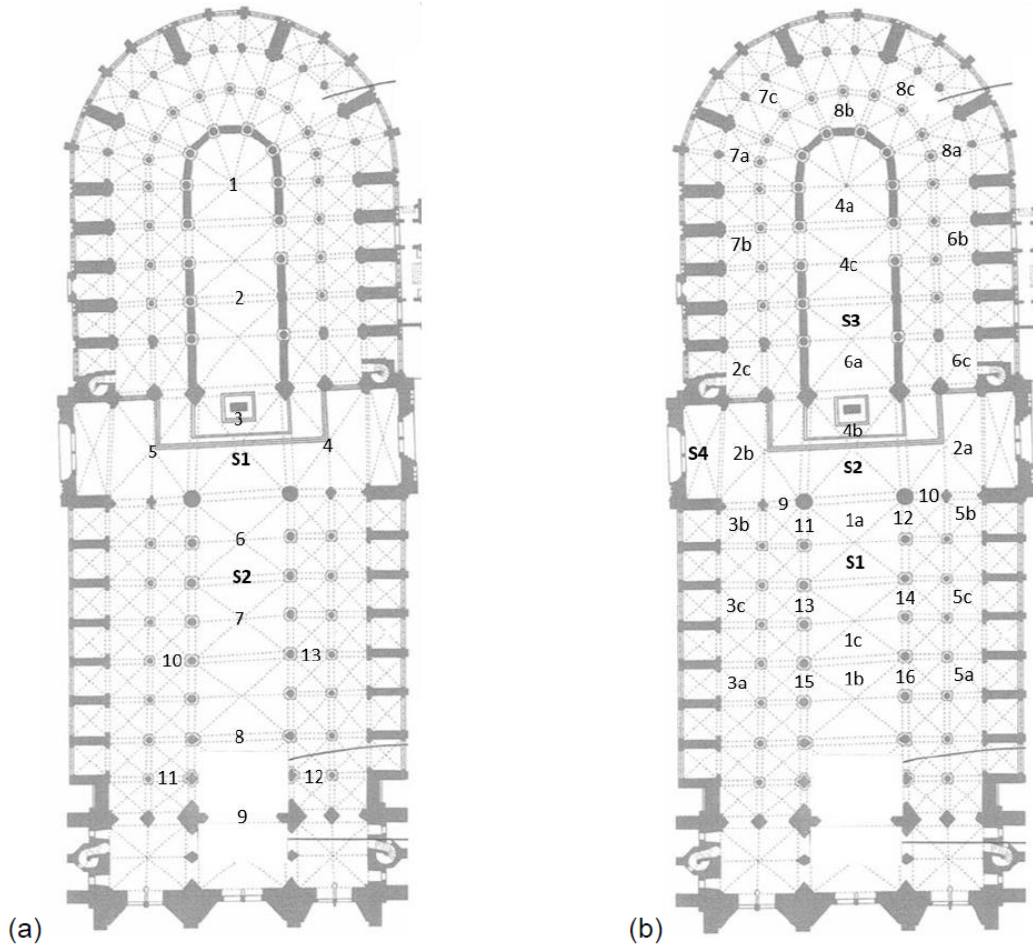


Figure 3.3: Measurement plans (a) 1987 and (b) 2015 of the Notre-Dame cathedral.

### 3.4.2 Parameter results

Table 3.2 presents the T20, EDT, C50, and C80 results of these measurements. As expected, reverberation parameters were within one JND between 2015 sine-sweeps and balloon burst. Subsequently, 1987 and 2015 balloon bursts were compared under two assumptions: 1.) parameter averaging over multiple S-R combinations located throughout the space provides sufficiently comparable results, and 2.) averaging over repeated balloon bursts compensates for variations in individual burst emissions [Pätynen 2011]. Results show a decrease of 1-3 JND in T20 for the octave bands 250-2000 Hz (limited data for the 125 Hz octave band due to poor SNR). Additionally, the 2015 T20 measurements show better resemblance to values presented in the previous 1996 and 2002 studies [Hamayon 1996, Mercier 2002] than the 1987 measurements. EDT showed similar differences, decreasing for octave bands 125-1000 Hz by 1-3 JND.

As expected, clarity parameters between 2015 sine sweeps and balloon bursts

Table 3.2: Mean and SD of the EDT, T20, C50, and C80 of all measured S-R combinations for the 1987 balloon burst, 2015 balloon burst, and 2015 sine-sweeps.

	Balloon burst 1987	Balloon burst 2015	sine-sweep 2015
Octave band (Hz)	Mean T20 (SD)		
125	9.93 (0.39)	-	-
250	9.62 (0.30)	8.22 (0.18)	8.41 (0.32)
500	7.93 (0.23)	7.58 (0.20)	7.38 (0.18)
1000	6.56 (0.29)	5.89 (0.21)	6.08 (0.24)
2000	5.04 (0.21)	4.69 (0.16)	4.61 (0.18)
4000	3.25 (0.22)	3.13 (0.19)	3.04 (0.20)
	Mean EDT (SD)		
125	9.20 (0.89)	8.34 (1.11)	8.57 (1.17)
250	8.75 (1.03)	7.95 (0.93)	8.20 (1.01)
500	7.59 (0.93)	7.19 (1.07)	7.42 (0.94)
1000	6.25 (0.85)	5.86 (0.75)	6.14 (0.83)
2000	4.86 (0.69)	4.73 (0.63)	4.67 (0.70)
4000	3.06 (0.46)	3.07 (0.51)	3.20 (0.64)
	Mean C50 (SD)		
125	-5.12 (2.48)	-6.98 (2.84)	-3.75 (5.08)
250	-6.22 (4.38)	-8.28 (2.81)	-7.60 (3.46)
500	-7.40 (3.09)	-8.08 (2.81)	-8.03 (3.41)
1000	-6.33 (2.94)	-7.23 (3.05)	-7.18 (3.81)
2000	-6.07 (3.00)	-6.27 (3.31)	-7.04 (4.16)
4000	-3.90 (2.55)	-4.09 (3.11)	-4.21 (4.55)
	Mean C80 (SD)		
125	-4.26 (2.53)	-5.95 (2.97)	-2.44 (4.74)
250	-4.95 (4.02)	-6.97 (2.71)	-6.32 (3.27)
500	-6.17 (2.90)	-6.64 (2.69)	-6.56 (3.18)
1000	-5.02 (2.75)	-5.64 (2.86)	-5.67 (3.57)
2000	-4.68 (2.95)	-4.67 (3.14)	-5.38 (3.89)
4000	-2.25 (2.51)	-2.36 (2.95)	-2.46 (4.40)

showed good resemblance, except in the 125 Hz octave band where the sine sweeps were more than 3 JND higher, though the meaning of clarity and the validity of the ISO JND values in such a low frequency octave band may be questionable. As clarity parameters are dependent on S-R distance, these parameters were compared for similar S-R combinations between 1987 and 2015 balloon bursts (1987: **S1R2**<sup>3</sup>, **S1R7**, **S2R2**, and **S2R7**; and 2015: **S2R4c**, **S2R1c**, **S1R4c**, and **S1R1c** respectively). The mean clarity parameter (500-1000 Hz) are within 1 JND for these position (C50 - 1987: -6.3, 2015: -6.9. C80 - 1987: -5.1, 2015: -5.2.).

<sup>3</sup>The notation employed here denotes the source-receiver combination of Source 1 and Receiver 2. This notation is employed throughout this thesis.



Table 3.3: Mean EDT and T20 (s) and SNR (dB) results of the 16 S-R combinations: Numerical average of the 8 individual sweeps and their averaged RIR before and after correction and their associated *error ratios*.

	250	500	1000	2000	4000	8000
$\overline{T20(RIR)}$	8.24	7.37	6.00	4.55	2.94	1.46
$T20(\overline{RIR})$	7.87	6.82	5.32	3.59	2.18	1.10
<i>error ratio</i>	4.5%	7.5%	11.3%	21.1%	25.9%	24.7%
$\overline{T20(RIR_{corr})}$	8.02	7.13	5.87	4.47	2.89	1.43
<i>error ratio</i>	2.7%	3.3%	2.2%	1.8%	1.7%	2.1%
$\overline{EDT(RIR)}$	7.61	6.78	5.55	4.12	2.57	1.23
$EDT(\overline{RIR})$	7.52	6.57	5.20	3.60	2.17	1.01
<i>error ratio</i>	1.2%	3.1%	6.3%	12.6%	15.6%	17.8%
$\overline{EDT(RIR_{corr})}$	7.59	6.72	5.50	4.07	2.54	1.21
<i>error ratio</i>	0.3%	0.9%	0.9%	1.2%	1.2%	1.6%
$\overline{SNR(RIR)}$	35.4	36.5	40.7	43.8	49.5	53.7
$SNR(\overline{RIR})$	43.8	46.5	49.2	50.6	55.9	62.2
$\overline{SNR(RIR_{corr})}$	43.8	45.2	48.7	50.9	56.4	62.2

### 3.4.3 Correction for the time-variant acoustic system

As the Notre-Dame cathedral has a large volume (ca. 84,000 m<sup>3</sup>), obtaining a sufficiently high SNR was more troublesome than in smaller volumes, such as theatres or concert halls. A sufficiently high SNR is necessary in order to obtain reliable room acoustical parameter estimations. Averaging of RIRs can be performed under the assumption of a time-invariant acoustic system in order to increase the SNR. This assumption can be invalidated by factors such as air turbulence or temperature variations between repetitions resulting in a time-variant acoustic system. RIRs averaged under time-variant conditions exhibit shorter reverberation time estimations, especially in higher octave bands.

An averaging was performed over the S-R combinations which were repeated 8 times. Figs. 3.4a and 3.4b clearly show that the average results for EDT and T20 are longer than comparable results calculated on the averaged RIR. Table 3.3 shows that the mean percentage difference (*error ratio*) for EDT rose from 1.2% to 17.8%, mean SD of 2.9%. The mean *error ratio* for T20 rose from 4.5% to 25.9%, mean SD of 2.3%. These *error ratios* render the averaged RIRs unreliable for parameter analysis and auralizations.

The reason for such errors was therefore investigated. The effect of averaging RIRs under time-variant conditions on reverberation estimations was studied by Satoh et al. [Satoh 2002, Satoh 2007], who performed measurements in five different sound fields with varying temperatures and air turbulence. It was concluded that the *error ratio* of reverberation estimations in octave bands 4000–8000 Hz was more than 10% when temperature changes between repetitions exceeded 0.05°C. Secondly, it was found that the acceptable temperature change for the MLS method, defined

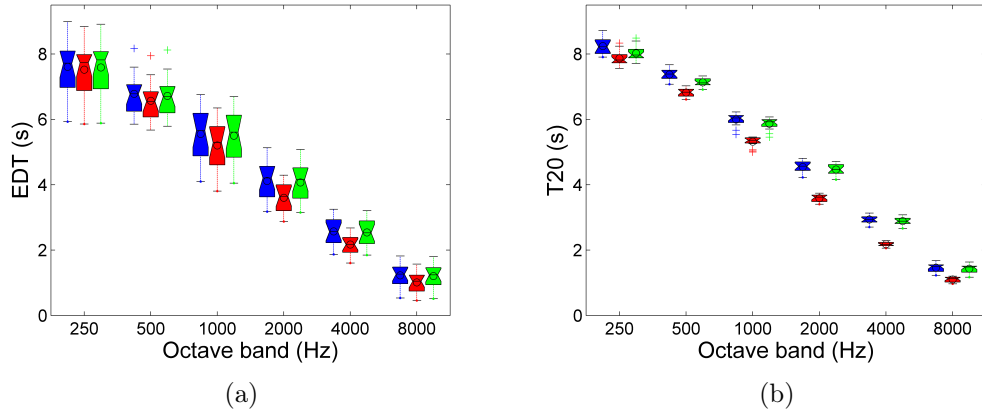


Figure 3.4: Summary of results for (a) EDT and (b) T20 of the 16 considered configurations. (Blue): Numerical average of parameters calculated individually on the 8 RIRs. Parameters calculated on averaged RIR before (red) and after (green) time-stretching correction.

in ISO 18233 as  $\Delta T \leq 200/(f \times RT)$ , where:  $f$  = center frequency and  $RT$  = reverberation time, was also appropriate for the swept sine method. Additionally, a method was presented to categorize whether the time-variance was caused by either temperature changes or air turbulence. For this purpose, a cross-correlation between a reference and considered RIR pair at 2.5 ms steps using a 10 ms window, examining the argument of the maximum versus time (time lag or  $\tau$ ) was performed. When  $\tau$  evolved linearly, the time variance was considered to be due to temperature changes. In contrast, random fluctuations of  $\tau$  were attributed to air fluctuation. The described studies did not attempt to correct for the observed time-variances.

When one qualitatively regards two repeated RIRs measured at different temperatures, one can observe that, due to the changed speed of sound, RIRs become progressively more out of sync. In the context of sound insulation measurements, Wang [Wang 2014] compensated for this effect by time-stretching IRs to a standard temperature. The time-stretching factor was estimated by maximizing the cross-correlation between the considered IR and reference using an iterative process.

### 3.4.3.1 Erroneous reverberation time estimations due to averaging

In order to investigate the reason for the shorter reverberation time estimations in the Notre-Dame measurements, a time-variant system caused by slight temperature changes was hypothesized and simulated. As temperature differences affect the speed of sound, three similar exponential sines with a slightly different phase (equal to a maximum temperature difference of ca.  $0.08^\circ\text{C}$ ) were generated ( $RIR = e^{-0.8t} \times \sin(1000\pi t^\delta)$  where  $t$  = time and  $\delta = 0.9999, 1.0000, \text{ or } 1.0001$ ). Low level Gaussian white noise was added to simulate background noise in the RIR and finally a synchronous averaging was performed. Due to the slight tempera-

ture difference the averaged RIR contains less energy than the three original RIRs (Fig. 3.6a). This is confirmed by inspection of corresponding Schroeder decay curves, depicted in Fig. 3.6b. While the Schroeder decay curves of the three synthesized RIRs are equal, that of the averaged RIR is steeper and concave. This translates to increasingly more energy loss when the RIRs become increasingly out of sync, resulting in shorter reverberation estimations. Figs. 3.6c and 3.6d show the results of the measured RIRs mirroring the effect observed in the simulated case (Figs. 3.6a and 3.6b).

The cross-correlation differences between the measured RIRs were studied in order to determine whether reverberation time estimation errors were caused by air turbulence or temperature differences. Fig. 3.5a depicts  $\tau$  determined at 10 ms steps between the first and last sweeps ( $10\times$  up-sampled) as well as the least-squares linear-fit between  $t_0 + 0.2$  s and  $t_0 + 2.5$  s. As  $\tau$  follows a linear slope, temperature changes were deemed to be the cause of the time-variant acoustic system. Since  $\tau$  fluctuates somewhat around the linear fit, some air turbulence is also considered as contributing.

### 3.4.3.2 Correction method

To correct for the observed energy loss, a time-stretching based on interpolation of the considered RIRs to a reference was performed (the first measured RIR was used as the reference). Firstly, the RIRs were up-sampled  $10\times$  in order to reduce discretization errors, following Wang [Wang 2014]. Secondly, the slope  $m$  of the linear least-squares best-fit function, over the range  $t = t_0 + 0.2$  s and  $t = t_0 + 2.5$  s, of  $\tau$  (calculated in 10 ms steps, with a 10 ms window) was calculated. Subsequently, a re-sampling of each RIR was performed using the re-sample factor ratio:  $f_{s,ref}/(f_{s,ref} + m)$ . Finally, an anti-aliasing filter (Butterworth filter,  $10^{th}$  order,  $f_c = 20$  kHz) was applied.

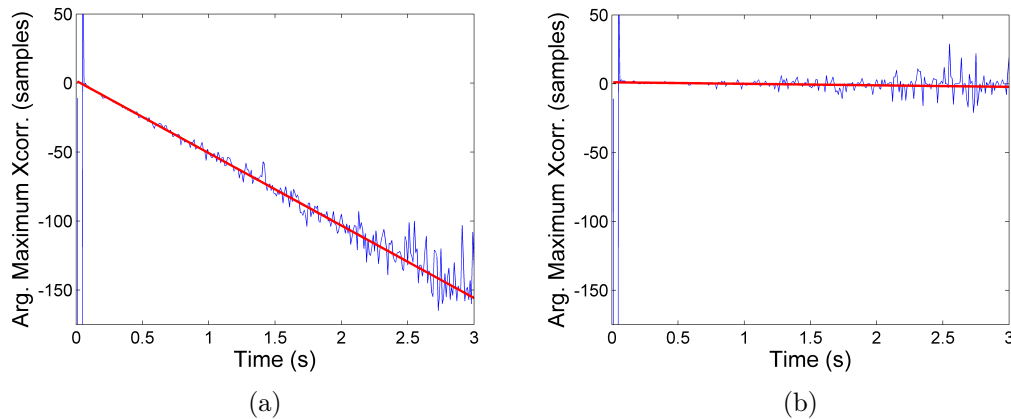


Figure 3.5: (—) Time lag ( $\tau$ ) between the first and last RIRs and (—) linear fit to (a) before and (b) after correction for configuration S1R8 ( $f_s = 480$  kHz).

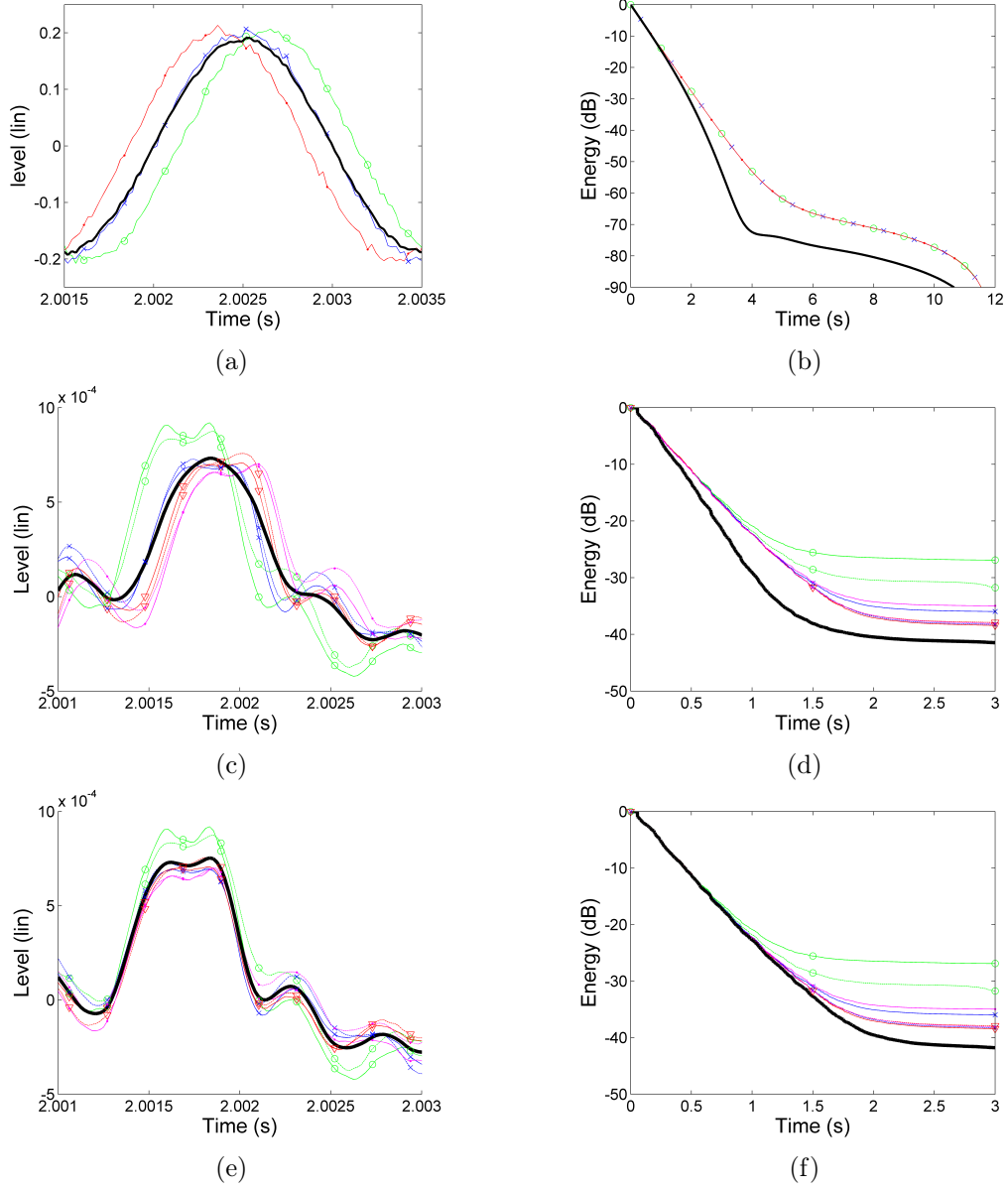


Figure 3.6: Analysis of synthesized, measured, and corrected measured RIRs. (a) Extract of the three synthesized RIRs and their average (thick line). (b) Schroeder decay curve of the synthesized RIRs and their average (thick line). (c) Extract of 8 measured RIRs of configuration **S1R8** (1<sup>st</sup> repetition ( $\circ$ ), 2<sup>nd</sup> ( $\times$ ), 3<sup>rd</sup> ( $\diamond$ ), and 4<sup>th</sup> ( $\cdot$ ); 1<sup>st</sup> sweep (—) and 2<sup>nd</sup> (— —)) and their average (thick line). (d) Schroeder decay curve of the measured RIRs and their average (for legend see Fig. 3.6c). (e) Extract of 8 corrected measured RIRs of configuration S01R8 and their average (for legend see Fig. 3.6c). (f) Schroeder decay curve of the measured RIRs and their average (for legend see Fig. 3.6c).

The proposed method reduced the slope of  $\tau$ 's linear-fit function to zero (see Fig. 3.5b). Consequently, the resulting eight RIRs and their average became more consistent (see Fig. 3.6e). The resulting decay curve in the 4000 Hz octave band can be observed to be straight, following the same decay rate as the eight separate RIRs (see Fig. 3.6f). This is also observable in the reverberation time estimations of the corrected RIRs (see Figs. 3.4a and 3.4b) which are significantly closer to the individual RIR numerical averages. Table 3.3 indicates an increased mean SNR across octave bands of ca. 7 dB as a result of averaging the corrected RIRs.

Finally, from  $\tau$ , it is possible to estimate the associated temperature change between measurements according to the formula  $c = 331 + 0.6T_c$  under the assumption of dry air [Pierce 1981]. Following the ISO 18233 equation previously mentioned, the maximum temperature change permitted for a time invariant system in the Notre-Dame is  $\Delta T = 0.02^\circ\text{C}$ . This value was exceeded during the measurement ( $\Delta T$  with set1sw(eep)1 (reference): set1sw2 =  $0.001^\circ\text{C}$ , set2sw1 =  $0.037^\circ\text{C}$ , set2sw2 =  $0.039^\circ\text{C}$ , set3sw1 =  $0.054^\circ\text{C}$ , set3sw2 =  $0.056^\circ\text{C}$ , set4sw1 =  $0.067^\circ\text{C}$ , set4sw2 =  $0.069^\circ\text{C}$ ). Differences between consecutive sweeps within each of the four sets vary minimally ( $0.001 - 0.002^\circ\text{C}$ ), while the temperature difference clearly increases over the four sets, separated by 10-20 minutes. However, it is noted that such variations in temperature are practically impossible to measure without high sensitivity specialized thermometers.

### 3.4.3.3 Discussion

Previous studies have shown the effect of a time-variant acoustic system on reverberation estimation of averaged RIRs. Previous compensations have been performed by time-stretching the separate RIRs to one reference temperature based on an iterative process. The newly presented method determined the time-stretching factor from the slope of the time lag, defined as the argument of the maximum cross-correlation using a 10 ms window with a reference RIR. After applying this method, mean EDT *error ratios* improved from ca. 9.4% to 1.0% with a mean SD of 0.8%. Mean T20 *error ratios* improved from ca. 15.8% to 2.3% with a mean SD of 0.7%. These values are well within one JND. Additionally, these *error ratios* were close to or within uncertainties as a consequence of measurement/analysis chains (ca. 2.9% in the 250–4000 Hz octave band) [James 2004] and the uncertainties due to the applied analysis tools (ca. 2% in the 1000 Hz octave band) [Katz 2004]. This validates the corrected averaged RIRs as being suitable for room acoustical parameter analysis and auralizations while providing a clearly improved SNR.

## 3.5 Summary

The first goal of this thesis is to study the 2<sup>nd</sup> aspect of *fully computed auralizations* i.e. sound propagation in 3D spaces. Therefore, this chapter presented (B)RIR measurements carried out in four rooms (Amphithéâtre, Théâtre de l'Athénée, Saint-Germain-des-Prés church, and Notre-Dame cathedral) which will serve as references

---

for subsequent calibrations and validations. It should be noted that the RIRs measured in the Notre-Dame cathedral were subject to a time-variant acoustic system. Therefore, a non-iterative correction method was proposed and validated. Chapter 4 will employ the measured RIRs to calibrate the room acoustic models of these four spaces. It was observed that average reverberation time estimations between measurements (1987 and 2015) in the Notre-Dame cathedral differed up to 3 JND. Therefore, additionally the calibrated room acoustic model of the Notre-Dame cathedral will be employed to study the potential cause of this difference. Additionally, the calibrated GA model of the Saint-Germain-des-Prés church will be employed to study the acoustic characteristic of a previous configuration. Chapter 5 will use the measured (B)RIRs to create auralizations and compare these to the simulated binaural counterpart employing listening tests.



# Model creation, calibration procedure, and objective validation<sup>1</sup>

---

## Contents

<b>4.1</b>	<b>Introduction</b>	<b>44</b>
<b>4.2</b>	<b>Calibration procedure</b>	<b>45</b>
<b>4.3</b>	<b>Creation of the geometrical models</b>	<b>47</b>
4.3.1	Amphithéâtre	48
4.3.2	Théâtre de l'Athénée	49
4.3.3	Notre-Dame cathedral	50
4.3.4	Saint-Germain-des-Prés church	50
<b>4.4</b>	<b>Calibration of the GA models</b>	<b>52</b>
4.4.1	Simple model study on repeatability and influence of parameters	52
4.4.2	Amphithéâtre	56
4.4.3	Théâtre de l'Athénée	59
4.4.4	Notre-Dame cathedral	62
4.4.5	Saint-Germain-des-Prés church	65
4.4.6	GA models adjustments	69
<b>4.5</b>	<b>Historical studies employing the calibrated GA models</b>	<b>69</b>
4.5.1	Saint-Germain-des-Prés model	70
4.5.2	Notre-Dame model	73
<b>4.6</b>	<b>Discussion</b>	<b>74</b>
<b>4.7</b>	<b>Summary</b>	<b>75</b>

---

<sup>1</sup>This work was partly presented in:

- B.N.J. Postma and B.F.G. Katz. *Creation and calibration method of virtual acoustic models for historic auralizations* Virtual Reality, vol. 19, no. SI: Spatial Sound, pp. 161-180, 2015.
- B.N.J. Postma and B.F.G. Katz. *Acoustics of Notre-Dame Cathedral de Paris* in Intl. Cong. on Acoustics (ICA), (Buenos Aires), 2016.
- B.N.J. Postma, A. Tallon, and B.F.G. Katz. *Calibrated auralization simulation of the abbey of Saint-Germain-des-Prés for historical study* In Intl. Conf. Auditorium Acoustics, Paris, 2015, pp. 190-197.



## 4.1 Introduction

In order to achieve realistic auralizations, a first step is to calibrate the GA model i.e. improve upon the 2<sup>nd</sup> aspect of *fully computed auralizations*: Sound propagation in a 3D space. Calibration of room acoustic models is necessary if one wishes to build scientific auralizations rather than simple audio novelties. In the selected test-cases the buildings still exist, and therefore physical measurements were carried out (see Chapter 3), allowing comparisons between the measured and simulated acoustic responses.

There have been few studies which examined calibration of GA simulations in order to create valid auralizations. [Vorländer 2010] proposed that a simulation is well calibrated when the difference between simulation and measurement is less than the JND per considered objective acoustical parameter. Comparisons of acoustic signals are also possible, such as those by [Katz 2005, Weinzierl 2015], who compared measured and simulated RIR envelopes.

These studies set out the final calibration goal, however, one should also consider the calibration process. An approach which went beyond comparison of envelopes was Pelzer and Vörlander [Pelzer 2013]. They presented a method that automatically matched absorption coefficients for every single wall in a GA model to measurements by applying an inverse room acoustics model. The inversion was performed numerically using a non-linear least-squares optimization process. The independent variables were all absorption coefficients and the goal was to minimize the error between measured and simulated impulse responses at all measured positions in the room.

However more typically, calibration procedures are based on bringing simulated average reverberation parameters within one JND of measurements [Alonso 2014, Pedrero 2014, Iannace 2015] by absorption and scattering coefficient adjustments. For instance, Galindo et al. [Galindo 2009] calibrated 6 GA models of Mudejar-Gothic churches employing the software CATT-Acoustic. The calibration method carried out employed an iterative process adjusting the absorption and scattering coefficients to bring the spatially average simulated reverberation times within one JND of the measurements. Katz and Wetherill [Katz 2005], calibrating a GA model of Sabine’s Fogg Art museum, regarded the simulated T15, T30, and Sabine’s reverberation formula bringing these within one JND for the single measurement position available. Martellotta [Martellotta 2009] adopted absorption coefficients from similar models which were previously calibrated. Additionally, final results were given as averages over five different predictions in order to take into account run-to-run variations.

Examples which go further in describing the calibration procedure are [Martellotta 2014, Alvarez-Morales 2015]. These studies state that an iterative adjustment process needs to be carried out starting from the absorption coefficients of the most unusual or uncertain surfaces, possibly covering large areas so that small variations from the originally assigned values could lead to a better agreement between simulated and measured parameters. As a tolerance range

for the absorption coefficient adjustments one can take variations between data sets [Beranek 1986, Beranek 1996, Vorländer 2008]. This iterative process also considered only reverberation times.

An alternative method was presented by Astolfi et al. [Astolfi 2008] when calibrating GA models of classrooms. The absorption coefficients of present materials were measured in a reverberation chamber. Then an iterative process was carried out adjusting the scattering coefficients of the linings in classrooms to match the reverberation time of GA models to measurements. The authors remarked that slight variation of the scattering coefficients of the different linings left the reverberation time estimations unaffected.

From these previous studies one can conclude that calibrations of GA models in general have been performed on the bases of bringing the simulated spatially averaged reverberation times within one JND of measured values. However, no calibration was performed regarding clarity parameters nor for individual positions. Additionally, the use of the scattering coefficients during calibrations was neglected or concluded not to affect parameters. Building upon these studies, a methodical calibration approach is proposed here.

Sec. 4.2 describes the general calibration procedure. Sec. 4.3 is devoted to the creation of the geometrical models of the considered rooms. In order to explore the calibration procedure, it was first applied to the Amphithéâtre which has a rather simple geometry and material selection in Sec. 4.4. Secondly, the procedure was applied to the more complex configurations of the Théâtre de l'Athénée, Notre-Dame cathedral, and the abbey church of Saint-Germain-des-Prés. Finally, Sec. 4.5 employs the calibrated Saint-Germain-des-Prés model to study a typical 17<sup>th</sup> century configuration and the calibrated Notre-Dame model to study the difference between the 1987 and 2015 measurement's room acoustical parameter estimations.

## 4.2 Calibration procedure

The general calibration procedure consists of 9 steps. While some details of the GA model construction may be specific to the selected software package, the general approach proposed should be applicable to any GA software that produces RIRs suitable for auralizations.

1. RIR measurements are carried out in the studied venue (see Chapter 3). The results of these measurements are used as a reference for the calibration.
2. The geometrical model is created and remains unchanged during the remainder of the calibration.
3. A search for the absorption coefficients in various databases of materials present in the model. The variation of the published data on absorption coefficients is used as the range within absorption coefficients are changed during the calibration process [Martellota 2014].

4. Preliminary acoustical properties are assigned to the geometrical model's surfaces, resulting in the geometrical acoustics model. Assigning coefficients for specific materials includes the need to take into account, in addition to the actual material:
  - The construction/installation variances of a given material relative to published measured data.
  - The adaptation of published normal incidence data to random incidence coefficients as required in GA software.
  - The assumptions on defining representative coefficients for geometrical model surfaces representing various materials.

Local geometry variations are represented by the *scattering coefficient* which is required for the definition of simpler geometrical surfaces and any associated acoustical texture. The direct impact of this variable on the simulation results is less obvious, and can vary as a function of 'mixing' properties of the specific geometry and absorption distribution.

5. There have been numerous studies on the variability and repeatability of acoustical RIR measurements [James 2003, James 2004] and RIR measurement analysis [Katz 2004]. Similarly, stochastic implementation of the Lambert function for scattering leads to run-to-run variations in the employed algorithm [Dalenbäck 2010]. Because of these stochastic variations, it is important to ascertain the degree of variance expected in the results for repeated "virtual measurements" in a given configuration. These variations can then be taken into account when comparing results and the degree of accordance achievable when attempting to obtain a calibrated model. For this purpose, ten simulation repetitions are run of the initial iteration and taking per acoustical parameter the average over the SD's per position, the run-to-run variation of the model is established.
6. Variations exist due to the various means by which different GA algorithms implement scattering models [Kulowski 1982]. Therefore, an initial step in the general calibration method is directed towards gauging the sensitivity of simulation results due to scattering coefficients for the given GA model and GA software's algorithms. For this purpose simulations are run of the initial model followed by two simulations with the same absorption coefficients, with the scattering coefficients set uniformly to 0% or 99%.
7. Absorption coefficients are adjusted to bring the reverberation parameters within 1 JND of the measured value. The first materials to be adjusted are those with the largest surface areas, since small variations lead to considerable effects.<sup>2</sup>

---

<sup>2</sup>To optimize computation time, this step was first carried out using the option *interactive estimate* in CATT-Acoustic, which runs a rapid global ray-tracing. Subsequently, the basic algorithm

8. With the aid of the knowledge gained in step 6, the scattering coefficients are used to bring the average clarity parameters within 1 JND.
9. Acoustical properties of local key surfaces are adjusted to minimize the standard deviation (SD) of the differences between measured and simulated results for reverberance and clarity parameters.

This procedure relies on several assumptions:

- The selected acoustic parameters for reverberation and clarity are sufficient metrics for the calibration procedure.
- Calibration according to a collection of objective acoustical parameters within 1 JND results in a valid simulation.
- Calibration of a GA model for a sufficient ensemble of discrete points (source and receiver positions) provides sufficient confidence in the quality of the simulated RIR at other positions.

### 4.3 Creation of the geometrical models

With the RIR measurements performed (step 1), the next step is the creation of the geometrical models. A primary consideration in the construction of a GA model is the level of detail. A model with too fine detail will result in high computational costs and at the same time will be inaccurate. Surface elements that are very small with respect to acoustic wavelength are poorly considered by GA algorithms, with small elements either reflecting more energy than their area permits or conversely being ignored as their surface area is too small. The result being that an overly detailed model will be more suited to only high frequency calculations. As such, one should avoid overly detailed models, and rather rely on somewhat larger surfaces and defined scattering properties to account for surface textures [Bork 2005c, Vorländer 2008].

To create the geometrical acoustics model from the geometrical model it is necessary to define the acoustic properties of each surface element. The sound absorption coefficient describes the fraction of sound that is absorbed at each reflection from a given surface, with this value being defined as a function of frequency. Numerous sources of data can be found for material properties [Beranek 1986, Beranek 1996, Vorländer 2008]. Coefficients in these databases correspond typically to random sound incidence conditions, as they were obtained from measurements in reverberation chambers and are often employed in reverberation time calculations according to the Sabine equation, or its derivatives. Other sources may provide normal incidence absorption condition coefficients,

---

for closed rooms was used (option: *short calculation, basic auralization*) to create RIRs and refine the calibration.

which are typically used in boundary element or finite element numerical simulations [Katz 2000a, Katz 2000b, Katz 2001] with the coefficient being modified for non-normal incidence.

Somewhat analogous to texture mapping and bump mapping in graphics, GA algorithms may incorporate diffraction- or scattering-type effects. These effects can be associated with two types of phenomena, those related to surface irregularities and those caused by the limited size of the surface as well as edge diffraction [Zeng 2006]. Both phenomena are a function of wavelength. Surface scattering can be estimated, based on wavelength of the incident sound compared to the surface depth variation. Scattering coefficients can also be measured or simulated using detailed numerical acoustical simulations, although few data sets are publicly available [ISO 17497 2007, Cox 2009, Vorländer 2008]. Algorithms for edge diffusion are typically based on the size of the object [Lokki 2002b, Pulkki 2003]. Inclusion of higher-order diffraction effects and especially true edge diffraction can significantly increase computational costs.

The frequency-dependent scattering coefficient ( $scatt_{coef}$ ) can be estimated as a function of a given characteristic depth ( $char_{depth}$ ) representative of the surface's depth variations or roughness. The estimation algorithm in Eq. 4.1 is included in CATT-Acoustic by the *estimate* function, though specific values can also be directly assigned as a function of frequency.

$$scatt_{coef}(f) \Bigg|_{\substack{\leq 0.99 \\ \geq 0.10}} = 0.5 \sqrt{\frac{char_{depth}}{\lambda}} \quad (4.1)$$

where  $\lambda$  is the wavelength. Values are clipped to the range ( $0.10 \leq scatt_{coef} \leq 0.99$ ). This clipping implies that values of  $char_{depth} \leq 0.004$  m result in a constant  $scatt_{coef} = 0.1$ , or 10% in the 4 kHz band, at which point scattering begins to increase. Values of  $char_{depth} \geq 0.11$  m are required to affect  $scatt_{coef}$  in the lowest 125 Hz band. This method of defining  $scatt_{coef}$  was selected as it provides a more intuitive and physically relevant control parameter and reduces the possibility of creating unrealistic frequency variations in scattering properties for general materials.

### 4.3.1 Amphithéâtre

The geometry of the Amphithéâtre was determined by on-site measurements (minimum modeled dimension = 0.34 m). Fig. 4.1a shows the interior of the room and Fig. 4.1b depicts the geometrical model. The plan of the room is rectangular (12 m  $\times$  14.9 m) with the exception of an intruding smaller rectangle (2.5 m  $\times$  1.9 m) to the right of the entrance (see door in Fig. 4.1a).

The materials and finishes of the amphitheater were determined by visual inspection. The floor is made of concrete. The wall, in which the entrance door is positioned, is unpainted plaster board (see Fig. 4.1b). The remaining walls are concrete for the lower part with profiled aluminum panels above. The ceiling is profiled

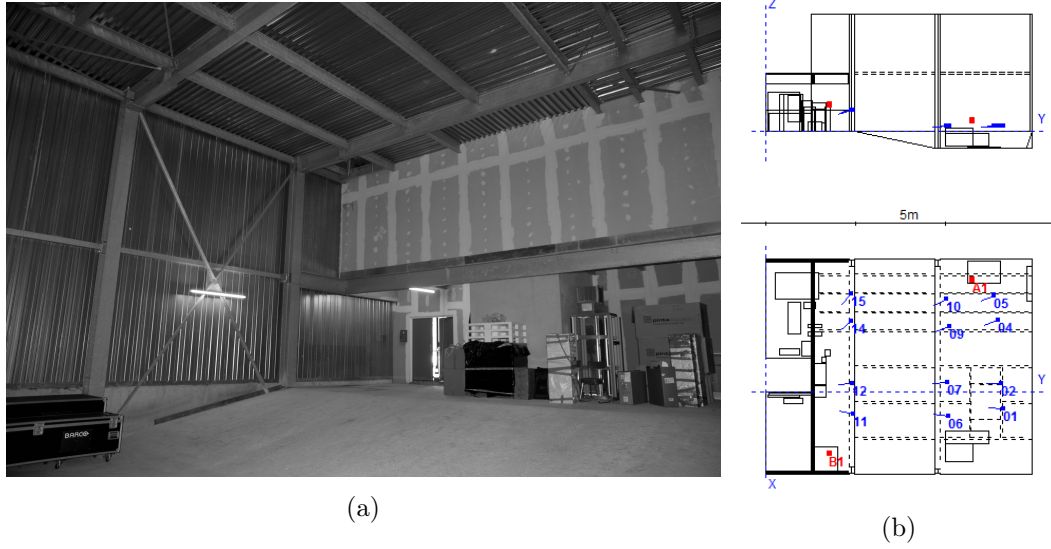


Figure 4.1: (a) Amphithéâtre (18-09-2014): Photo depicting the profiled aluminum panels, room construction, objects, and plaster board surfaces. (b) Geometrical model of the Amphithéâtre. Volume ca.  $1100 \text{ m}^3$ , ca. 270 planes. The lines depicted on the aluminum wall and ceiling panels are graphic indications that the surface is defined as a 1D diffuser and do not depict the actual relief lines.

aluminum panels. In front of these panels, metal construction elements are exposed. The profiled aluminum panels of the walls have a relief of 8 mm and those in the ceiling of 80 mm. As these undulating panels diffuse sound only along one dimension, they were modeled using the 1D scattering option, *L1*, in CATT-Acoustic, which includes the direction of scattering for the surface. To reproduce the protocol for the measured RIR, omnidirectional sources and receivers were employed in the simulations.

#### 4.3.2 Théâtre de l'Athénée

The geometry of the Théâtre de l'Athénée was determined from architectural plans and sections (minimum modeled dimension = 0.34 m). Fig. 4.2 compares the interior of the room to the geometrical model. The stage opening of this 4-floor theater is ca. 7.7 m, stage depth ca. 7.8 m, and height of the stage ca. 13.8 m. The audience area is 'horse shoe' shaped with a maximum width of ca. 11.0 m, length of ca. 14.3 m, and height of ca. 13.6 m. Box seats are present on the side and back walls on the first, second, and third floor.

Surface materials were determined from visual inspection. The Théâtre de l'Athénée has a wooden stage floor and concrete side and back stage walls. The audience area has a wooden floor, light velour walls, plastered balcony fronts, and plastered ceiling with a considerable chandelier. Upholstered chairs are positioned throughout the room.



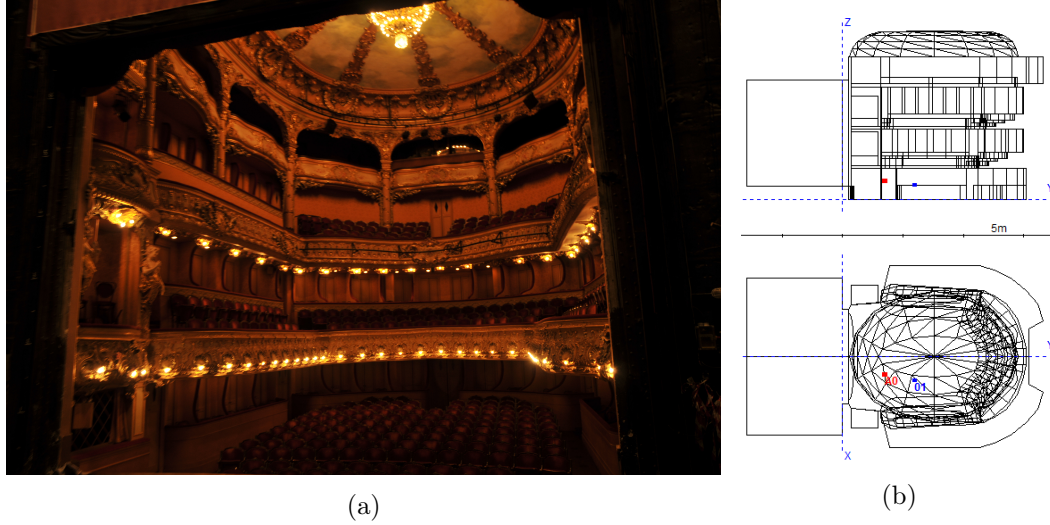


Figure 4.2: (a) Théâtre de l'Athénée: Photo depicting the audience area as seen from the stage (22-07-2014). (b) Geometrical model of the Théâtre de l'Athénée. Volume ca. 2,500 m<sup>3</sup>, planes ca. 1,300.

### 4.3.3 Notre-Dame cathedral

The geometry of the Notre-Dame cathedral was determined from a 3D laser scan point cloud<sup>3</sup> (minimum modeled dimension = 0.34 m). The cathedral (see Fig. 4.3) is ca. 130 m long, in the nave ca. 48 m wide, and ca. 35 m high. The nave consists of 7 bays with two side aisles at each side. The organ is positioned in the first bay above the entrance, the pulpit on the right side of the fourth bay, and the high altar behind the nave. The sanctuary, positioned behind the altar, consists of four bays with two side aisles and is rounded off.

A visual inspection determined the surface materials. The floor of the Notre-Dame cathedral is marble and the pulpit is wooden. Pews are positioned in the center part of the nave and first side aisles. The columns, walls, and ceiling are generally limestone. Confession booths and statues are positioned in the alcoves and large paintings hang on their walls.

Statues were modelled as simple bounding boxes with high scattering. The  $scatt_{coef}$  of the pews were specifically defined following guidelines in the CATT-Acoustic manual which recommends these to be modeled with  $scatt_{coef}$  of 30% to 80%, rising 10% per octave band [Dalenbäck 2009, p. 85].

### 4.3.4 Saint-Germain-des-Prés church

The geometry of the Saint-Germain-des-Prés church was determined from both a 3D laser scan point cloud<sup>4</sup> and architectural plans & sections<sup>5</sup> (minimum modeled

<sup>3</sup>Provided by Andrew Tallon of Vassar College.

<sup>4</sup>Provided by Andrew Tallon of Vassar College.

<sup>5</sup>Provided by Pierre Bloy Géometre-Expert D.P.L.G., architects.

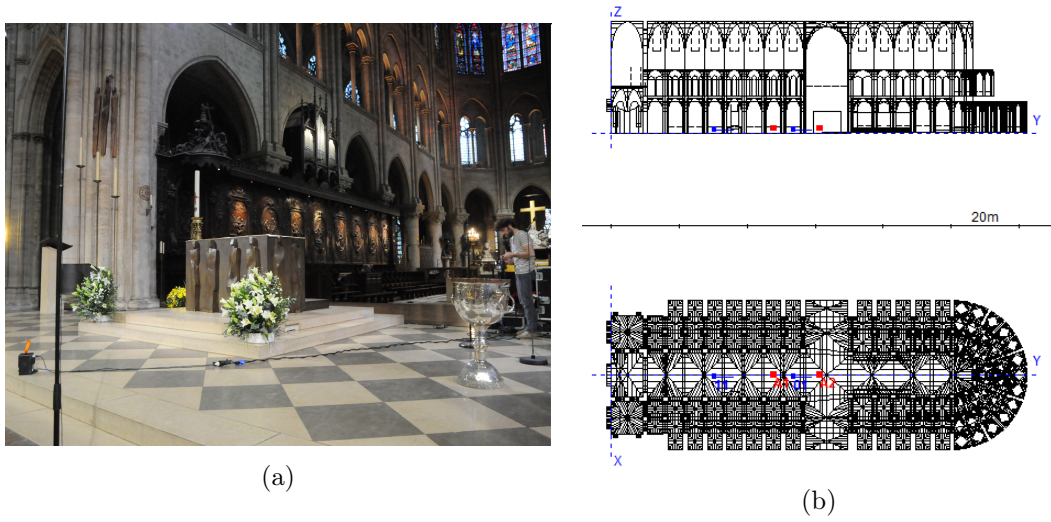


Figure 4.3: (a) Photo depicting the altar of the Notre-Dame cathedral (13-04-2015). (b) Geometrical model of the Notre-Dame. Volume ca.  $84,000 \text{ m}^3$ , planes ca. 14,700.

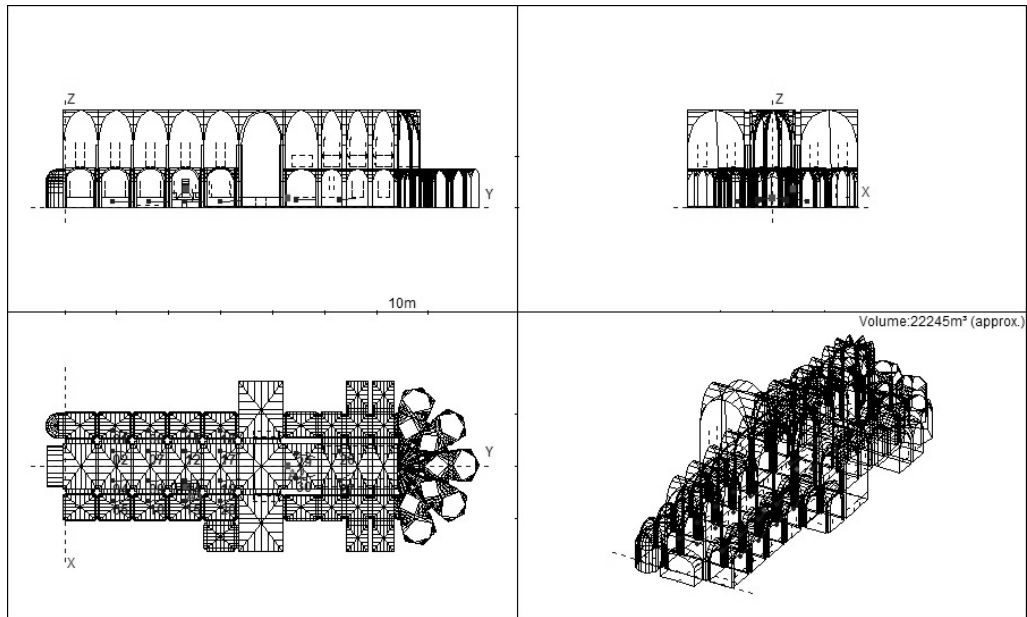


Figure 4.4: Geometrical model of the Saint-Germain-des-Prés church. Volume ca.  $22,200 \text{ m}^3$ .

dimension = 0.34 m). The abbey church is ca. 78 m long, in the nave ca. 21 m wide, and ca. 18 m high. The nave consists of 5 bays with one side aisle at both sides. The organ is positioned in the first bay above the entrance, the pulpit on the right side of the fourth bay, and the high altar behind the nave. The sanctuary, positioned behind the altar, consists of four bays with two side aisles and is rounded off at the





Figure 4.5: Photos (30-09-2014) of (a) altar position towards organ, (b) altar, and (c) pulpit in the Saint-Germain-des-Prés church.

Table 4.1: Summary of the Geometrical models of the Amphithéâtre, Théâtre de l’Athénée, Notre-Dame cathedral, and Saint-Germain-des-Prés church.

Room	Polygons	Volume (m <sup>3</sup> )	Floor plan area (m <sup>2</sup> )
Amphithéâtre	270	1,100	170
Théâtre de l’Athénée	1,300	2,500	300
Notre-Dame cathedral	14,700	84,000	4,800
Saint-Germain-des-Prés	2,200	22,200	1,800

end.

A visual inspection determined the surface materials. The floor of the Saint-Germain-des-Prés church and the pulpit are marble. Reed chairs are positioned in the center part of the nave and to the left and right of the high altar. The columns, walls, and ceiling are generally painted plaster, except in the sanctuary’s side aisles where these are exposed limestone. Large paintings hang in the left side aisles, except in the fourth nave where a large statue is present.

Regarding scattering, the plaster was modeled acoustically flat, with a  $char_{depth}$  of 1 mm, and the limestone was modeled with a  $char_{depth}$  of 10 mm. Statues and pews were similarly modeled to the Notre-Dame model. Table 4.1 presents a summary of all geometrical model details.

## 4.4 Calibration of the GA models

### 4.4.1 Simple model study on repeatability and influence of parameters

#### 4.4.1.1 Simulation repeatability

CATT-Acoustic employs a hybrid calculation algorithm. Crossover is based on the reflection order and reflection type, in the following way [James 2001]:

- *ISM* for 1<sup>st</sup> and 2<sup>nd</sup> order specular reflection so that all early reflections are included independent of the amount of rays used.
- Diffuse *radiosity* for 1<sup>st</sup> order diffuse reflection. Many small diffusely radiating surface sources are distributed over each diffusing surface. From the actual

sound source, vectors are drawn to each diffuse surface source and from each of those to the receivers (taking occlusion into account). To give the highest geometrical accuracy where it is needed, the number of these surface sources is increased for surfaces with low absorption coefficients and high scattering coefficients.

- Randomized cone-tracing for higher order reflections where ray directions are randomized like in ray-tracing so that, unlike with specular cone-tracing, diffuse reflection can be taken into account.

CATT-Acoustic provides parameter results based on energy echograms and RIRs as well as the possibility to generate RIRs suitable for auralization and analysis in external software. In order to study the effect of the chosen parameter analysis method, a small simple room model,  $12 \times 12 \times 10$  m, with one source and a  $6 \times 6$  receiver grid was employed. Material properties were defined homogeneously to obtain a T30 of ca. 3.0 s across all octave bands, comparable to the Amphithéâtre. Uniform scattering was defined at 10% over all frequency bands. While it is acknowledged that the use of such extremely simplistic rooms are somewhat at the limits of GA [Dalenbäck 2010], due to their highly symmetric nature, these effects should not affect the results here. A total of 10 simulation repetitions (Algorithm 1, no. of rays: ca. 50,000, Transition Order (TO) = 1) were carried out with results of T30 and C80 being tabulated for the two analysis methods within CATT-Acoustic, as well as the analysis of the exported RIR (employing the IRA toolkit) which is used for auralization.

Fig. 4.6 shows good agreement for the different methods and repetition across T30 results. While the SD over repetitions for the *Energy* based reported results do not vary with repetition the *IR* based reported results and analysis of the generated RIR using the *RIR analysis* show SD below 0.05 s for frequency octave bands of 500 Hz and above, with SD increasing for lower frequencies. These variations remain below the perceptual JND of 5%, or 0.15 s in this case.

Results for C80 show less agreement in both the mean of reported values and the SD over repetitions. With an accepted JND for C80 being 1 dB, the repeatability variance for the *IR* and *RIR analysis* are comparable to the perceptual threshold. As such, the degree of tolerance for C80 in calibrating the model to the measured data should realistically be extended beyond 1 JND.

These findings confirm [Katz 2004] that differences may occur in acoustic parameter results between different RIR analysis implementations, despite the existence of measurement standards. As the final application is the virtual reality auralization of the room, which uses the exported RIRs, analysis of these exported audio files with measured RIR values using the same analysis software was selected as the most reliable means of comparison. In order to avoid issues regarding automatic noise detection algorithms and the ideal noise-free RIRs from the simulation, low level white Gaussian background noise was added to the numerical RIRs.

Subsequently, the influence of the chosen algorithm and number of rays on run-to-run variation was studied. This was done running 10 repetitions employing the

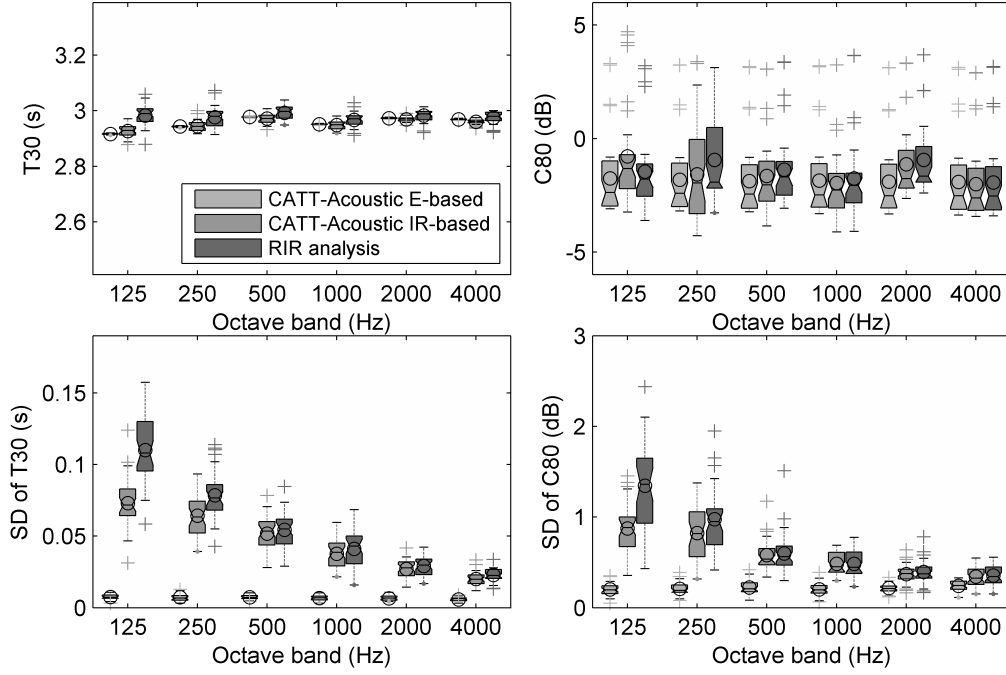


Figure 4.6: (upper) Distribution of the mean T30 and C80 results for all source-receiver combinations (36) over 10 repeated simulations with the internal CATT-Acoustic analysis tools and the exported RIR analysis. (lower) Distribution of the SD for all source-receiver combinations results over 10 repeated simulations (average number of rays: ca. 50,000). Notches display the confidence interval, box limits represent the 25% and 75% quartiles, outliers are indicated by (+), (-) indicates the median, (o) indicates the mean value.

same simple model, however with material properties defined homogeneously to obtain a T30 of ca. 1.0 s across all octave bands, with the default number of rays (ca. 50,000) and ten times that amount (500,000) using the three algorithms furnished within CATT-Acoustic:

- Algorithm 1 (Short calculation, basic auralization)
- Algorithm 2 (Longer calculation, detailed auralization)
- Algorithm 3 (Even longer calculation, detailed auralization)

According to the CATT-Acoustic manual [Dalenbäck 2009] algorithm 1 should be employed for closed geometrically mixing room and fairly even absorption distribution with limited risk for flutter echoes, algorithm 2 for cases with uneven absorption, open or very dry rooms, algorithm 3 for unusual open cases.

Fig. 4.7 shows that there is agreement between the run-to-run variation of the three algorithms. Additionally, the run-to-run variation remains similar when the amount of rays is raised. It should be noted that the run-to-run variation of algorithm 3 is higher in the 1000-4000 Hz octave bands, this might be due to the overly simple model simulated by a detailed algorithm. Therefore, the number of

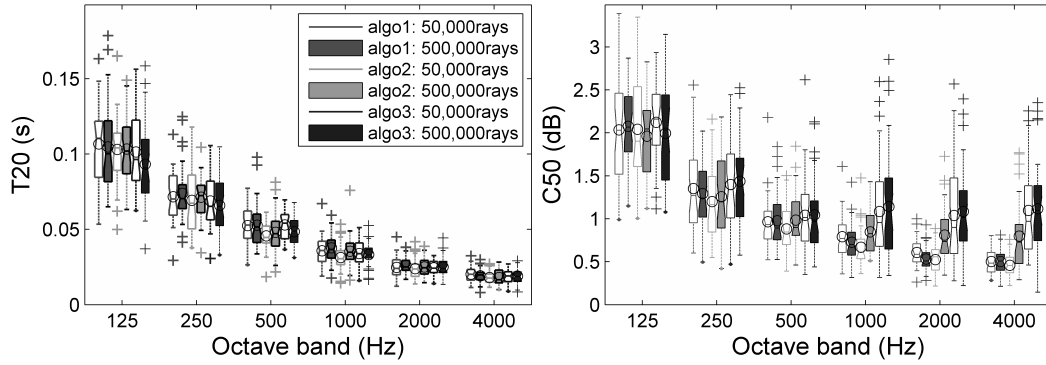


Figure 4.7: Distribution of T20 and C50 results for all source-receiver combinations. SD of results over 10 repeated simulations for all three algorithms with default number of rays (ca. 50,000) and ten times that amount (500,000). (See legend of Fig. 4.6 for boxplot notations).

rays and calculation algorithm needs to be selected based on the resulting energy echogram, complexity of the geometry, and distribution of the sound absorption. In order to reduce the effect of the remaining run-to-run variation, all subsequent calibration parameter results are the average of 5 simulation repetitions, comparable to [Martellotta 2009].

#### 4.4.1.2 Influence of input parameters

Before starting the calibration, regard is given to the three basic variables in room acoustics simulations as they are handled in CATT-Acoustic: geometry, absorption coefficients, and scattering properties. The *geometry* of a room is a relatively straightforward concept, with the main variable being the complexity of the model. Surface material properties such as *absorption coefficient* are relatively well understood. The effect of absorption in a geometrical acoustics algorithm is simple in concept, with incident rays diminishing in energy after a reflection from a surface by the absorption coefficient. Assigning coefficients for specific materials includes the need to take into account, in addition to the actual material, installation variances relative to published measured data, adaptation of published normal incidence data to random incidence coefficients as required in GA software, and assumptions on defining representative coefficients for surfaces representing various materials. Local geometry variations are represented by the *scattering coefficient* which allows for the defining of simpler geometrical surfaces and an associated acoustic texture. The direct impact of this variable on the model results is less obvious, and can vary due the various means by which GA software algorithms implement scattering. Therefore, step 6 in the general calibration method is dedicated to improve the understanding of the influence of scattering coefficients on the simulation results.

Table 4.2: Absorption coefficients,  $char_{depth}$  (mm), and surface area ( $m^2$ ) used in the initial and final iteration of the Amphithéâtre. Surfaces indicated with a  $\star$  are modeled with edge diffusion.  $char_{depth}$  indicated with a  $\ddagger$  have scattering coefficients in the 125 Hz octave band defined as [5%].

	125	250	500	1000	2000	4000	$char_{depth}$	surface
<b>initial material properties</b>								
concrete(floor & wall)	0.01	0.01	0.01	0.01	0.02	0.02	1	162
concrete(ramp)	0.02	0.03	0.03	0.03	0.04	0.07	1	55
aluminum(walls)	0.11	0.17	0.10	0.07	0.09	0.11	8	265
aluminum(ceiling)	0.11	0.17	0.10	0.07	0.09	0.11	80	132
plaster board	0.12	0.10	0.08	0.06	0.06	0.06	1	65
doors	0.11	0.12	0.12	0.12	0.10	0.10	1	24
objects $\star$	0.08	0.06	0.04	0.03	0.02	0.02	1	66
metal construction	0.01	0.01	0.02	0.06	0.03	0.03	1	100
<b>final material properties</b>								
concrete(floor & wall)	0.01	0.01	0.01	0.01	0.02	0.02	20 $\ddagger$	162
concrete(ramp)	0.02	0.03	0.03	0.03	0.04	0.07	20 $\ddagger$	55
aluminum(walls)	0.16	0.15	0.10	0.07	0.08	0.07	50 $\ddagger$	265
aluminum(ceiling)	0.16	0.15	0.10	0.07	0.08	0.07	80	132
plaster board	0.17	0.12	0.08	0.06	0.05	0.06	10 $\ddagger$	65
doors	0.11	0.12	0.12	0.12	0.10	0.10	1	24
objects	0.25	0.12	0.04	0.03	0.02	0.02	1	66
metal construction $\star$	0.01	0.01	0.02	0.06	0.03	0.03	1	100

#### 4.4.2 Amphithéâtre

The proposed calibration procedure was first applied to the Amphithéâtre<sup>6</sup>. Table 4.2 presents the absorption coefficients,  $char_{depth}$  used in the first iteration and the associated surface area. Curve *initial* in Fig. 4.9 shows the average T20 and C50 which resulted from the use of these material properties.

As proposed in step 5, ten simulation repetitions (Algorithm 1, no. of rays: ca. 23,000) were run. The upper graphs in Fig. 4.8 show that the SD of the T20 exceeds the JND in the 125 Hz octave band, the SD of the C50 exceeds the JND of 1 dB in the 125 and 250 Hz octave band. As such, the degree of tolerance for T20 and C50 in calibrating the model to the measured data should realistically be extended beyond 1 JND in these octave bands. Subsequently, simulations were run with all scattering set to 0% and all scattering set to 99%. The lower graphs in Fig. 4.8 depict that the T20 is stable when the scattering is varied, however the C50 increases when the scattering is lowered and vice versa.

According to step 7, the absorption coefficients of the *aluminum(walls)* and *aluminum(ceiling)*, the material most used in the Amphitheater, was iteratively

<sup>6</sup>Since the measured SNR in the 125 Hz octave band was lower than 45 dB the T30 is not considered during the calibration process, leaving the T20, EDT, C50, and C80 to be analyzed.

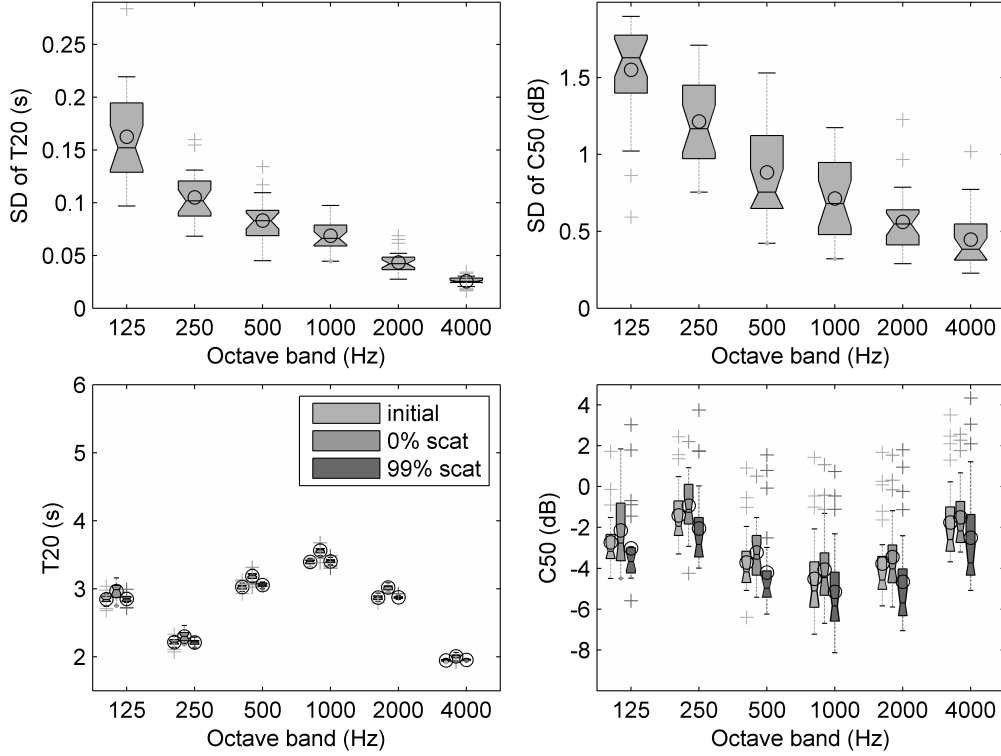


Figure 4.8: The distribution in the Amphithéâtre of SD of the T20 and C50 results for all source-receiver combinations (24) of the initial iteration (upper). Mean T20 and C50 distribution of the initial iteration with normal estimation of scattering, 0% scattering and 99% scattering (lower). (See legend of Fig. 4.6 for boxplot notations).

adjusted to arrive at  $[0.168, 0.15, 0.099, 0.072, 0.080, 0.072]$ , bringing EDT and T20 within 1 JND.

After step 7 the mean C50 was  $\Delta[-0.49 \text{ dB}, +0.21 \text{ dB}, +0.33 \text{ dB}, +0.32 \text{ dB}, +0.60 \text{ dB}, +0.50 \text{ dB}]$ <sup>7</sup>. Since Fig. 4.8 showed that increasing the scattering resulted in a lower clarity, the  $char_{depth}$  of the *aluminum(walls)* was increased to 0.055 m, the  $char_{depth}$  of the *plaster board* was increased to 0.01 m, the  $char_{depth}$  of the *concrete(floor & wall)* was increased to 0.02 m, and the  $char_{depth}$  of the *concrete(ramp)* was increased to 0.02 m. To raise the mean C50 in the 125 Hz octave band the scattering of these surfaces was set to 5%. Furthermore, the edge diffusion of the *objects* were omitted and the *metal construction* was modeled with edge diffusion. After these adjustments the mean difference between measured and simulated C50 changed to  $\Delta[-0.33 \text{ dB}, +0.06 \text{ dB}, +0.39 \text{ dB}, +0.20 \text{ dB}, +0.50 \text{ dB}, +0.35 \text{ dB}]$  which is an improvement across octave bands except for 1000 Hz.

At the start of step 9, the SD determined over the positions' difference between measured and simulated T20 [125 Hz, 250 Hz] was [0.29 s, 0.18 s] and EDT [125 Hz,

<sup>7</sup>The  $\Delta[\cdot]$  notation employed here denotes the difference in resulting values over the 6 octave bands relative to the measured values.

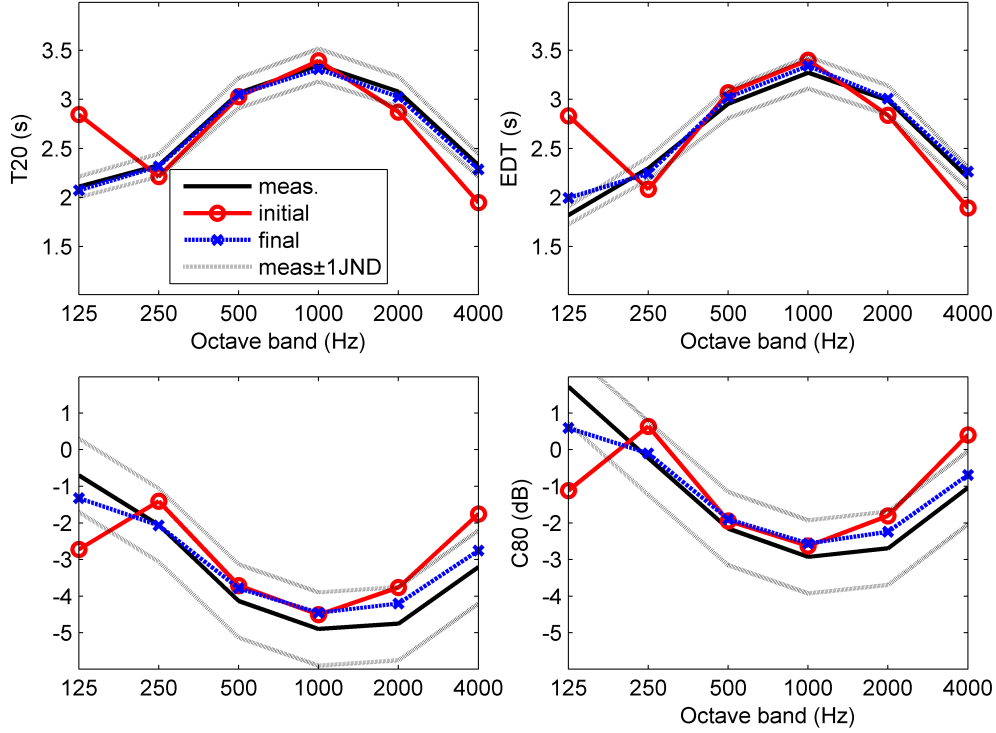


Figure 4.9: Comparison between measured T20, EDT, C50, and C80 ( $\pm 1$  JND, for reverberation parameters: 5%, and clarity parameters: 1 dB) in the Amphitheater and the simulation of the first and final iteration.

250 Hz] was [0.44 s, 0.25 s]. To lower these SDs a search was done for positional differences. It was found that the receivers on the upper platform (R11-15) overestimated the T20 [125 Hz, 250 Hz] with  $\Delta[+0.14$  s,  $+0.02$  s] while the receivers on the lower platform underestimated this with  $\Delta[-0.15$  s,  $-0.06$  s]. Since the receivers on the upper platform were closer to the *objects* and *plaster boards* their absorption coefficients were iteratively raised to arrive at respectively [0.25, 0.12] and [0.17, 0.12] while the absorption of the *aluminum(walls)* and *aluminum(ceiling)* was iteratively lowered to [0.156, 0.146] to retain the same mean reverberation and clarity. After these adjustments the receivers on the upper platform overestimated the T20 by  $\Delta[+0.11$  s,  $+0.02$  s], while the receivers on the lower platform underestimated the T20 by  $\Delta[-0.11$  s,  $-0.03$  s]. This resulted in a lower SD of the T20 [0.25 s, 0.15 s] and EDT [0.37 s, 0.25 s].

With the calibration concluded, the initial values are compared to the calibrated model. Fig. 4.9b shows the T20, EDT, C50, and C80 of the initial and final iteration and the measurement  $\pm 1$  JND. Where in the original configuration the T20 and EDT were overestimated with more than 1 JND in the 125 Hz octave band and underestimated with more than 1 JND in the 250 Hz, 2000 Hz, and 4000 Hz octave band, in the final configuration these were brought within 1 JND. The C50 and C80 were initially considerably overestimated in the 250 Hz and 4000 Hz octave bands



and these were a better estimation in the final iteration.

The main changes in the model's surface properties were the adjustment of the *aluminum* & *objects* absorption coefficient and the adjustments of scattering coefficient of the *concrete*, *aluminum*, and *plaster panels*. In the initial iteration the absorption coefficients of the *aluminum panels* were adopted from data concerning non-perforated aluminum panels on a 300 mm airspace measured in a reverberation chamber. However, in the Amphithéâtre insulation material is installed behind the aluminum panels. Therefore, the actual absorption properties of the aluminum panels differ from the initially simulated values. The absorption of the *objects* is difficult to estimate for the uncertainty in material and possible panel absorbing properties due to their hollow inside. There are additional reasons why absorption coefficients can differ from the databases: materials may have experienced transformations due to their age [Garcia 2014] or specific features were hidden or were difficult to be grasped from a visual inspection. As the *concrete*, *plaster board*, and *aluminum panels* are large flat surfaces and consequently the surface size is large in relation to the wavelength, it is justified to adopt a scattering of 5% in the 125-Hz octave band. Furthermore, raising the  $char_{depth}$  of the *aluminum (walls)* can be justified by the metal construction in front of the wall which makes the wall more diffusing in the higher octave bands than initially assumed, and increasing the  $char_{depth}$  of the concrete and plaster board is justified by the objects in front these which makes it more diffusing than originally assumed. Additionally, probably for similar reasons, we note that the general guidelines in the CATT-Acoustic manual recommend that it is better to overestimate scattering coefficients than to underestimate them [Dalenbäck 2009, p.111 Table 3].

#### 4.4.3 Théâtre de l'Athénée

Having established the proposed calibration protocol in Section 4.4.2, the procedure was then applied to a more complex case, the Théâtre de l'Athénée. Table 4.3 depicts the absorption coefficients,  $char_{depth}$ , and surface area, in the first iteration, assigned to these materials. Curve *initial* in Fig. 4.11 shows the average EDT and C80 of the first iteration.

After completing the material search of *step 3*, 10 simulation repetitions (Algorithm 2, no. of rays: 100,000) with the initial material properties were run to determine the variance of the model, according to *step 5*. The upper graphs in Fig. 4.10 show that the SD of T20 exceeds the JND in the 125 and 250 Hz octave band, the SD of C50 exceeds the JND of 1 dB in the 125 and 250 Hz octave band. As such, the degree of tolerance for T20 and C50 in calibrating the model to the measured data should realistically be extended beyond 1 JND in these octave bands. According to *step 6*, simulations were run with all scattering set to 0% and subsequently all scattering set to 99%. The lower graphs in Fig. 4.10 indicates that T20 is stable when the scattering is varied, but C50 increases when the scattering is reduced, and vice versa. This information is employed during the calibration procedure.



Table 4.3: Absorption coefficients,  $char_{depth}$  (mm), and surface area ( $m^2$ ) used in the initial and final iteration of the Théâtre de l'Athénée.  $char_{depth}$  indicated with a † are defined with scattering coefficients of [30%, 40%, 50%, 60%, 70%, 80%],  $char_{depth}$  indicated with a ★ are defined with one dimensional scattering, and  $char_{depth}$  indicated with a ‡ are defined with surface scattering.  $char_{depth}$  indicated with a ◇ have scattering coefficients across octave band defined as [5%].

	125	250	500	1000	2000	4000	$char_{depth}$	surface
<b>initial material properties</b>								
seating	0.56	0.68	0.79	0.83	0.86	0.86	†	378
light velour wall	0.03	0.04	0.11	0.17	0.24	0.35	10	329
stage floor	0.10	0.07	0.06	0.06	0.06	0.06	1	108
stage walls	0.06	0.05	0.05	0.04	0.04	0.04	30	504
stage ceiling	0.11	0.17	0.10	0.07	0.09	0.11	1000★	108
ceiling under balc.	0.11	0.17	0.10	0.07	0.09	0.11	30	108
loge partitions	0.03	0.04	0.11	0.17	0.24	0.35	10‡	121
floor audience	0.14	0.10	0.06	0.04	0.04	0.03	1	108
ceiling audience	0.14	0.10	0.06	0.04	0.04	0.03	10	108
logefront	0.14	0.10	0.06	0.04	0.04	0.03	1‡	100
<b>final material properties</b>								
seating	0.44	0.56	0.68	0.77	0.85	0.89	†	378
light velour wall	0.03	0.04	0.13	0.20	0.26	0.37	◇	329
stage floor	0.10	0.07	0.06	0.06	0.06	0.06	◇	108
stage walls	0.06	0.05	0.05	0.04	0.04	0.04	30	504
stage ceiling	0.11	0.17	0.10	0.07	0.09	0.11	1000★	108
ceiling under balc.	0.11	0.17	0.10	0.07	0.09	0.11	◇	108
loge partitions	0.03	0.04	0.13	0.20	0.26	0.37	10‡	121
floor audience	0.14	0.10	0.06	0.04	0.04	0.03	◇	108
ceiling audience	0.14	0.10	0.06	0.04	0.04	0.03	◇	108
logefront	0.10	0.09	0.12	0.14	0.12	0.13	10‡	100

Curve *initial* in Fig. 4.11 shows that the mean EDT was  $\Delta[-0.53$  s,  $-0.42$  s,  $-0.13$  s,  $+0.29$  s,  $+0.22$  s,  $+0.17$  s]. Therefore, the absorption coefficients of the *seating*, *light velour wall*, *loge partitions*, and *logefront* were adjusted within the range found during *step 3*, to arrive within one JND of the reverberation parameters.

The mean C50 at this stage was  $\Delta[-0.16$  dB,  $-0.45$  dB,  $-1.35$  dB,  $-1.50$  dB,  $-0.24$  dB,  $-0.62$  dB]. Since during *step 6* it was found that the clarity parameters increased when the scattering was diminished, the scattering coefficients of the surface with a large surface compared to the wavelength was diminished to 5%. These surfaces were the *light velour wall*, *stage floor*, *ceiling under the balconies*, *floor audience*, and *ceiling audience*. C50 at this stage was  $\Delta[+0.04$  dB,  $-0.23$  dB,  $-0.90$  dB,  $-1.01$  dB,  $+0.01$  dB,  $-0.25$  dB], an improvement across octave bands.

With the model being calibrated, the initial parameter values are compared to

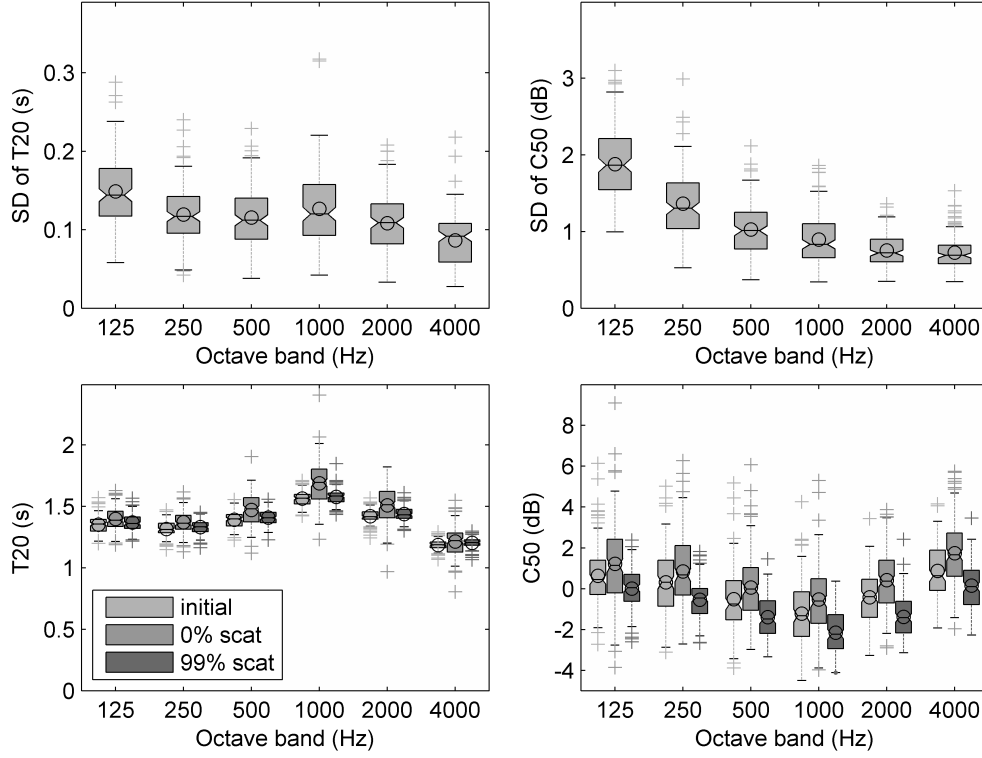


Figure 4.10: (upper) The distribution in the Théâtre de l’Athénée of SD (over ten simulations repetitions) of the T20 and C50 results for all source/receiver combinations (24) of the initial iteration. (Lower) Mean (over ten simulations repetitions) T20 and C50 distribution of the initial iteration with normal estimation of scattering, 0% scattering and 99% scattering. (See legend of Fig. 4.6 for boxplot notations).

those of the calibrated model. Fig. 4.11 shows the T20, EDT, C50, and C80 before and after the calibration compared to the measurement ( $\pm 1$  JND). In the initial configuration the reverberation parameters (T20 and EDT) were underestimated in the low octave bands (125-500 Hz) and overestimated in the high octave bands (1000-4000 Hz), and the clarity parameters (C50 and C80) were overestimated in the low octave bands (125-250 Hz) and underestimated for the high octave bands (1000-4000 Hz). In the final iteration the average parameter estimations were within 1 JND.

The initially chosen material properties are compared to those used in the final model. Various reasons can explain the main change in the model which was the *seating’s* absorption coefficient. Originally the values were adopted from *Seats, unoccupied- medium upholstered* [Beranek 1996], however as in the final iteration these values resembled the *Upholstered seats, unoccupied* found in [Barron 1993] this falls within the range found in *step 3*. Additionally, the absorption coefficient of *light velour wall*, *loge partitions*, and *logefront* were adjusted. The difference between initial and final absorption coefficients of *light velour wall* and *loge partitions*

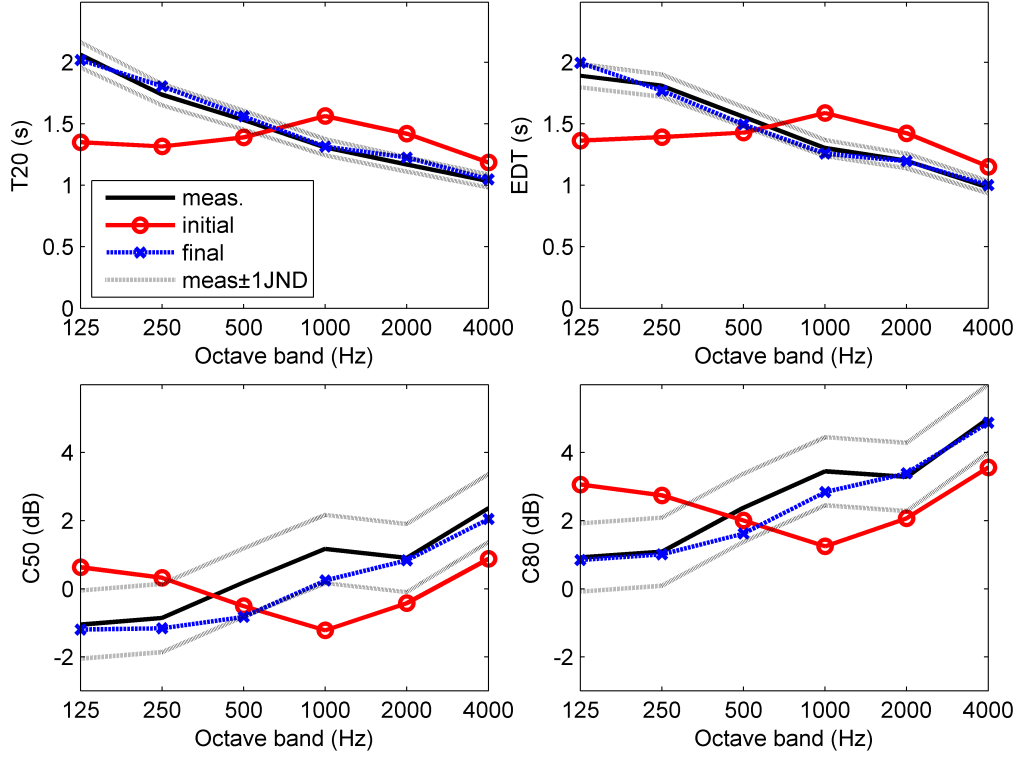


Figure 4.11: Comparison between measured T20, EDT, C50, and C80 ( $\pm 1$  JND, for reverberation parameters: 5%, and clarity parameters: 1 dB) in the Théâtre de l'Athénée and the simulation of the first and final iteration.

were  $\pm 3\%$  and thus fall within reasonable limits. Initial absorption coefficients of the *logefront* were adopted from the material *stage floor* [Vorländer 2008], the final values resembled more *pine wood* [Vorländer 2008]. Furthermore, lowering the scattering of the *light velour wall*, *stage floor*, *ceiling under balconies*, and *floor* can be justified by the relative large size in relation to the incident wavelength.

#### 4.4.4 Notre-Dame cathedral

To provide an acoustic contrast to the Théâtre de l'Athénée, the calibration procedure was also studied for the Notre-Dame cathedral. In a room-acoustical sense, this space is different, as it has generally well-distributed absorption and reverberation times which are not strongly dependent on scattering. Table 4.4 depicts the absorption coefficients,  $char_{depth}$ , and surface area, in first iteration, employed in the Notre-Dame model. Curve *initial* in Fig. 4.13 shows the average T20, EDT, C50, and C80 of the first iteration.

The calibration was initiated by running 10 simulation repetitions of the initial model configuration (Algorithm 1, no. of rays: 250,000). Fig. 4.12 shows that the mean SD of the T20 is across all octave bands lower than the JND, however the mean C50 across all octave bands is higher than the JND or approaching it. Subsequently,

simulations were run with the initial iteration and all scattering coefficients set to 0% and 99%. The lower graphs in Fig. 4.12 depict that the T20 is stable when the scattering is varied, the C50 increases when the scattering is lowered, and decreases when the scattering is raised.

After exploring the properties of the model regarding repeatability and influence of scattering, the average T20 and EDT were brought within 1 JND of the measured value by adjusting the absorption coefficient of the *limestone* to [0.028, 0.037, 0.044, 0.054, 0.076, 0.082], since this material represents the largest surface area and therefore has a principal influence on the acoustic response<sup>8</sup>.

<sup>8</sup>Receivers in the Notre-Dame cathedral were positioned in close proximity to **S2** and **S4**, as a result these two simulated source-receiver combinations exhibited variations relative to measured values EDT (single number frequency average (500-1000 Hz) -2.25 s) and C50 (single number

Table 4.4: Absorption coefficients,  $char_{depth}$  (mm), and surface area (m<sup>2</sup>) used in the initial and final iteration of the Notre-Dame cathedral.  $char_{depth}$  indicated with a † are defined with scattering coefficients of [30%, 40%, 50%, 60%, 70%, 80%] and  $char_{depth}$  indicated with a ★ are defined with scattering coefficients of [99%, 99%, 99%, 99%, 99%, 99%], and  $char_{depth}$  indicated with a ‡ are partly defined with scattering coefficients of [70%, 70%, 70%, 70%, 70%, 70%] (ca. 3448 m<sup>2</sup>).

	125	250	500	1000	2000	4000	$char_{depth}$	surface
<b>initial material properties</b>								
wall	0.02	0.03	0.04	0.05	0.06	0.07	30	13,889
ceiling	0.02	0.03	0.04	0.05	0.06	0.07	55	7,182
marble floor	0.01	0.01	0.01	0.02	0.02	0.02	1	4,676
columns	0.02	0.03	0.04	0.05	0.06	0.07	250	3,815
glass	0.30	0.20	0.14	0.10	0.05	0.05	300	2,342
pews	0.05	0.08	0.10	0.12	0.12	0.12	†	914
wood	0.14	0.10	0.06	0.08	0.10	0.10	50	589
arch	0.02	0.03	0.04	0.05	0.06	0.07	300	524
carpet	0.11	0.14	0.37	0.43	0.27	0.25	1	446
doors	0.14	0.10	0.06	0.04	0.04	0.03	20	427
<b>final material properties</b>								
wall	0.028	0.037	0.044	0.054	0.076	0.082	30‡	13,889
ceiling	0.028	0.037	0.044	0.054	0.076	0.082	55	7,182
marble floor	0.01	0.01	0.01	0.02	0.02	0.02	1	4,676
columns	0.028	0.037	0.044	0.054	0.076	0.082	★	3,815
glass	0.30	0.20	0.14	0.10	0.05	0.05	300	2,342
pews	0.05	0.08	0.10	0.12	0.12	0.12	†	914
wood	0.14	0.10	0.06	0.08	0.10	0.10	50	589
arch	0.028	0.037	0.044	0.054	0.076	0.082	300	524
carpet	0.11	0.14	0.37	0.43	0.27	0.25	1	446
doors	0.14	0.10	0.06	0.04	0.04	0.03	20	427

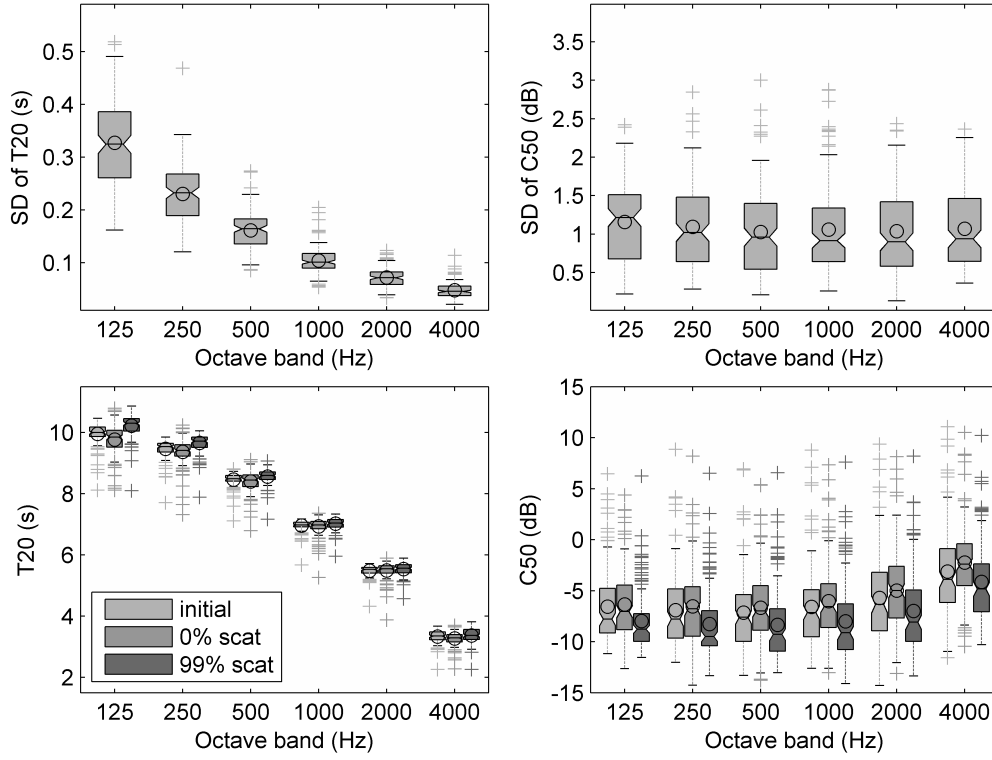


Figure 4.12: The distribution in the Notre-Dame cathedral of SD of the T20 and C50 results for all source-receiver combinations (128) of the initial iteration (upper). Mean T20 and C50 distribution of the initial iteration with normal estimation of scattering, 0% scattering and 99% scattering (lower). (See legend of Fig. 4.6 for boxplot notations).

Fig. 4.12 shows that by increasing the scattering the clarity parameters became lower and vice versa. At this stage the C80 differed  $\Delta[-1.46, +1.04, +1.57, +1.11, +1.88, +1.44]$  from the measurements. Since the clarity in the higher octave bands needed to decrease, the scattering coefficients of the columns were increased to 99% across frequency bands and the scattering of the walls in the alcoves was increased to 70%. The combination of these measures resulted in a C80  $\Delta[-0.80, +0.91, +1.30, +0.81, +1.87, +1.39]$  which better estimated the measurement across all octave bands.

With the calibration completed, the initial parameter values are compared to those of the calibrated model. Fig. 4.13 shows the T20, EDT, C50, and C80 before and after the calibration compared to the measurement ( $\pm 1$  JND). In the initial configuration the reverberation parameters (T20 and EDT) were overestimated, especially in the low octave bands, and the clarity parameters (C50 and C80) were

frequency average (500-1000 Hz)  $+7.6$  dB). Calibration of these source-receiver combination is difficult due to the perfect omni-directional directivity of the simulated sound source while the measured sound sources have slight variations in directivity.

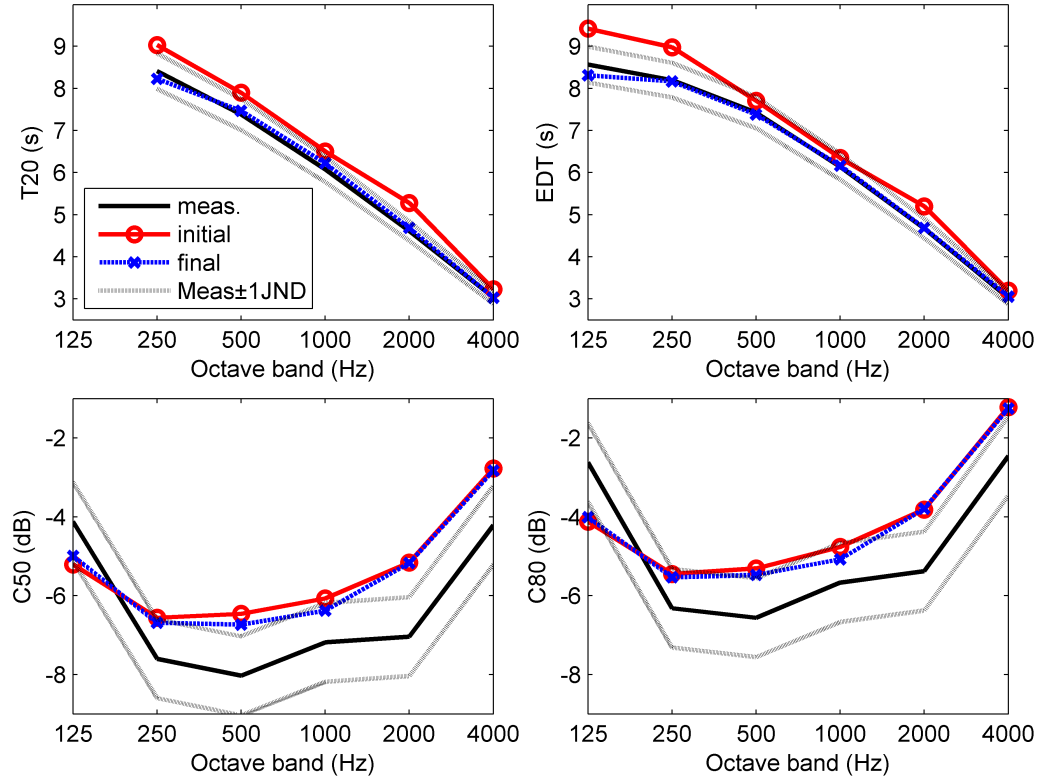


Figure 4.13: Comparison between measured T20, EDT, C50, and C80 ( $\pm 1$  JND, for reverberation parameters: 5%, and clarity parameters: 1 dB) in the Notre-Dame cathedral and the simulation of the first and final iteration.

overestimated. In the final iteration the average T20 and EDT were within 1 JND, and also the C50 and C80 were within 1 JND, except in the 500, and 2000-4000 Hz octave band. These differences could not be corrected for while staying within the material properties range established in step 3 of the calibration procedure.

The main adjustments in the model's surface properties were the adjustment of the limestone's absorption coefficient and the raising of the scattering coefficients of the walls in the alcoves and the columns. Various reasons can explain the difference between the initial and final modeled absorption coefficient of the limestone. The adjustments are within  $\pm 2\%$ , falling within the range found in *step 3*. Furthermore, raising the scattering coefficients of the wall in the alcoves can be justified by the numerous statues, paintings, and confession booths positioned here. Raising the scattering coefficients of the columns can be explained by their very rough surface finish.

#### 4.4.5 Saint-Germain-des-Prés church

Finally, the calibration procedure was applied to the GA model of the Saint-Germain-des-Prés church. In a room-acoustical sense, this site is comparable to

the Notre-Dame. Table 4.5 depicts the absorption coefficients,  $char_{depth}$ , and surface area, in first iteration, assigned to these materials. Curve *initial* in Fig. 4.15 shows the average T20, EDT, C50, and C80 of the first iteration<sup>9</sup>.

The calibration was initiated by running ten simulation repetitions (Algorithm 1, no. of rays: ca. 130,000). Fig. 4.14 shows that the mean SD of the T20 is across all octave bands lower than the JND, however the mean C50 across all octave bands is higher than the JND or approaching it. Subsequently, simulations were run of the

<sup>9</sup>In the Saint-Germain-des-Prés church, measurement positions which combine **S2** and the two closest receivers were omitted from analysis due to the very high direct-to-reverberant ratio, due to close proximity to the source, and the impact of this on the reliability of parameter assessment. The positions underestimated EDT (single number frequency average (500-1000 Hz)  $-0.68$  s) and overestimated C50 (single number frequency average (500-1000 Hz)  $+2.0$  dB) relative to measured results. As **S1** was positioned in the pulpit, receivers close to this source were (partially) shielded from the direct sound and included in the mean parameter analysis.

Table 4.5: Absorption coefficients,  $char_{depth}$  (mm), and surface area ( $m^2$ ) used in the final iteration of the Saint-Germain-des-Prés.  $char_{depth}$  indicated with † have defined scattering coefficients of [30%, 40%, 50%, 60%, 70%, 80%]. Surfaces indicated with ★ are modeled with edge diffusion.

	125	250	500	1000	2000	4000	$char_{depth}$	surface
<b>initial material properties</b>								
floor	0.01	0.01	0.02	0.02	0.03	0.03	1	1,201
stained glass	0.30	0.20	0.14	0.10	0.05	0.05	500	443
plaster wall	0.06	0.05	0.05	0.04	0.04	0.04	1	2,229
plaster column	0.06	0.05	0.05	0.04	0.04	0.04	150	1,020
plaster arch	0.06	0.05	0.05	0.04	0.04	0.04	150	678
plaster ceiling	0.06	0.05	0.05	0.04	0.04	0.04	1	1,578
limestone	0.03	0.03	0.03	0.04	0.05	0.07	10	1,406
reed chairs	0.06	0.10	0.10	0.20	0.30	0.20	†	500
pew	0.10	0.15	0.18	0.20	0.20	0.20	†	49
wood	0.14	0.10	0.06	0.08	0.10	0.10	50	352
paintings	0.11	0.21	0.29	0.33	0.40	0.50	80	76
<b>final material properties</b>								
floor	0.01	0.01	0.02	0.02	0.03	0.03	10	1,201
stained glass	0.30	0.20	0.14	0.10	0.05	0.05	100	443
plaster wall	0.013	0.022	0.032	0.032	0.035	0.037	55	2,229
plaster column ★	0.013	0.022	0.032	0.032	0.035	0.037	300	1,020
plaster arch	0.013	0.022	0.032	0.032	0.035	0.037	300	678
plaster ceiling	0.013	0.022	0.032	0.032	0.035	0.037	55	1,578
limestone	0.03	0.03	0.03	0.04	0.05	0.07	55	1,406
reed chairs	0.06	0.10	0.10	0.20	0.30	0.20	†	500
pew	0.10	0.15	0.18	0.20	0.20	0.20	†	49
wood	0.14	0.10	0.06	0.08	0.10	0.10	50	352
paintings	0.11	0.21	0.29	0.33	0.40	0.50	80	76

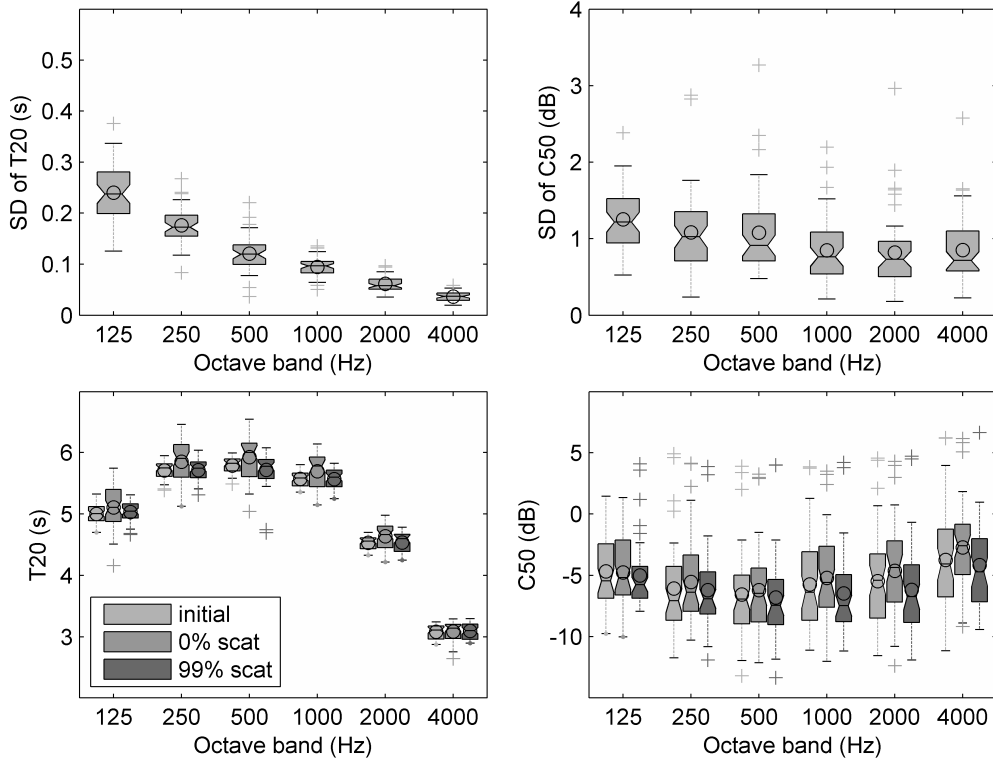


Figure 4.14: The distribution in the Saint-Germain-des-Prés church of SD of the T20 and C50 results for all source-receiver combinations (48) of the initial iteration (upper). Mean T20 and C50 distribution of the initial iteration with normal estimation of scattering, 0% scattering and 99% scattering (lower). (See legend of Fig. 4.6 for boxplot notations).

initial iteration and all scattering coefficients set to 0% and 99%. The lower graphs in Fig. 4.14 depict that the T20 is stable when the scattering is varied, the C50 is stable in the 125 Hz octave band, in the 250 Hz octave band the C50 increases when the scattering is lowered, and in the 500-4000 Hz octave band the C50 increases when the scattering is lowered and decreased when the scattering is raised.

After exploring the properties of the model regarding repeatability and influence of scattering, the average T20 and EDT were brought within 1 JND of the measured value by adjusting the absorption coefficient of the painted plaster to [0.01, 0.019, 0.028, 0.032, 0.038, 0.037], since this material represents the largest surface area and therefore has a principal influence on the acoustic response.

Fig. 4.14 shows that by increasing the scattering the clarity parameters in the octave bands 500-4000 Hz became lower and vice versa. At this stage the C80 differed  $\Delta[-1.35, +0.25, -0.26, +0.98, +1.60, +0.52]$  from the measurements. Since the clarity in the higher octave bands needed to decrease, the  $char_{depth}$  was increased for the columns to 0.30 m, for the floor to 0.01 m, for the walls to 0.055 m, and for the ceiling to 0.055 m. Additionally, the absorption coefficient of the painted



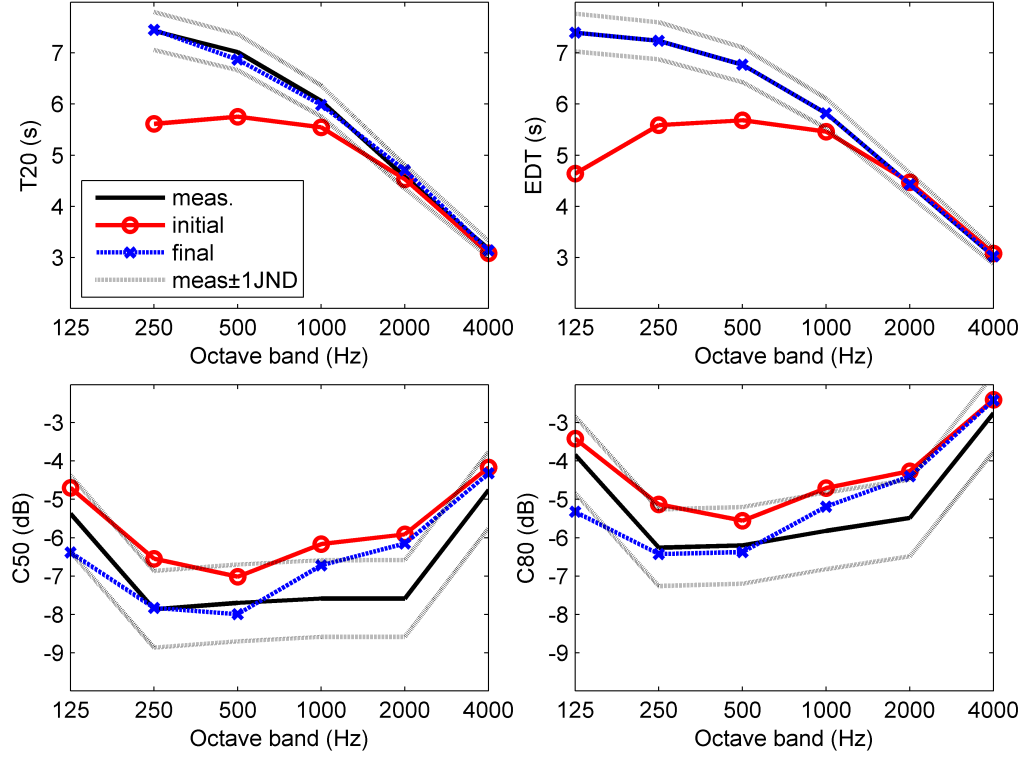


Figure 4.15: Comparison between measured T20, EDT, C50, and C80 ( $\pm 1$  JND, for reverberation parameters: 5%, and clarity parameters: 1 dB) in the Saint-Germain-des-Prés church and the simulation of the first and final iteration.

plaster in octave band 2000 Hz was lowered to 0.035. Furthermore, since Fig. 4.14 did not give an answer how to raise the C50 in the lower octave bands, iteratively adjusting various parameters resulted in finally adding edge diffusion to the columns and raising the absorption [125 Hz, 250 Hz, 500 Hz] to [0.013, 0.022, 0.032]. The combination of these measures resulted in a C50  $\Delta[-1.01, +0.03, -0.29, +0.87, +1.43, +0.43]$  which better estimated the measurement across all octave bands, except in the 500 Hz where it remained approximately stable.

With the calibration completed, the initial values are compared to those of the calibrated model. Fig. 4.15 shows the T20, EDT, C50, and C80 before and after the calibration compared to the measurement ( $\pm 1$  JND). In the original configuration the reverberation parameters (T20 and EDT) were underestimated, especially in the low octave bands, and the clarity parameters (C50 and C80) were overestimated. In the final iteration the average T20 and EDT were within 1 JND, and also the C50 and C80 were within 1 JND, except in the 2000 Hz octave band. This difference could not be corrected for while staying within the material properties range established in step 3 of the calibration procedure.

The main adjustments in the model's surface properties were the adjustment of the painted plaster's absorption coefficient and the raising of the  $char_{depth}$  of the

*marble floor, painted plaster & limestone walls, and painted plaster ceiling.* Various reasons can explain the difference between the initial and final modeled absorption coefficient of the painted plaster. In the initial iteration these values were adopted from the material *Concrete block, plastered*, in the calibrated model the absorption coefficient more resembles the material *Walls, hard surfaces average (brick walls, plaster, hard floors)* [0.02, 0.02, 0.03, 0.03, 0.04, 0.04] found in [Vorländer 2008]. Since these are both reasonable choices for the absorption coefficient, it is reasonable to state that these fall within the range found in *step 3* of the methodological calibration. Furthermore, raising the  $char_{depth}$  of the floor, walls, & ceiling and the addition of the edge diffusion to the columns follows the previous models and the need to overestimate the scattering coefficients.

#### 4.4.6 GA models adjustments

This section summarizes the degree and direction of changes made to material properties over the 4 models. Absorption coefficient changes in general were within the range found in *step 3* of the methodical calibration procedure. In some cases larger adjustments were necessary such as the *aluminum panels* in the Amphithéâtre. Possible explanations why absorption coefficients differed from the initial value are that materials may have experienced transformations due to their age or specific features were hidden or were difficult to be grasped from a visual inspection. More adjustments were carried out for the scattering coefficients. In the Amphithéâtre, Notre-Dame cathedral, and the Saint-Germain-des-Prés church the scattering coefficients were generally raised while in the Théâtre de l'Athénée they were lowered. A possible reason for this difference is the room-acoustical nature of the regarded space, in the Théâtre de l'Athénée the absorption is unevenly divided throughout the space while the other three rooms have generally well-distributed absorption and reverberation times which are not strongly dependent on scattering.

### 4.5 Historical studies employing the calibrated GA models

With the calibration of the GA models completed, modifications can be made in an effort to estimate the impact of architectural modifications. In the context of historical studies, modifications can be made to the GA model to revert rooms to previous states in their history, such as before important renovations [Katz 2005]. In the context of this thesis, this process was performed for two spaces: 1) the Saint-Germain-des-Prés model was employed to study the liturgical acoustical properties of the 17<sup>th</sup> century and 2) the Notre-Dame model was employed to study possible grounds for the different reverberation estimations between 1987 and 2015 RIR measurements.

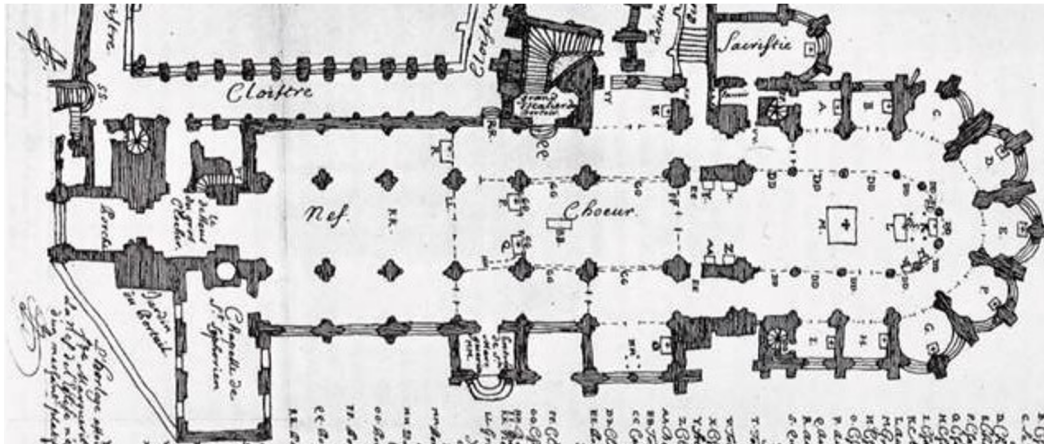


Figure 4.16: 17th century plan of the Saint-Germain-des-Prés church, from [Bibliothèque nationale de France 1898].

#### 4.5.1 Saint-Germain-des-Prés model

Historical studies of the Saint-Germain-des-Prés church have identified a 17<sup>th</sup> century plan of the building. Not only has the architecture of the abbey church changed over the centuries, but so have the customs concerning the application of adornments in churches. Therefore, the current acoustical conditions significantly differ from those in the 16<sup>th</sup> and 17<sup>th</sup> century.

##### 4.5.1.1 Former configuration of the Saint-Germain-des-Prés church

Fig. 4.16 depicts the former plan of the Saint-Germain-des-Prés church. The plan has changed at the 5<sup>th</sup> bay of the nave where next to the right side aisle a space was created. Furthermore, the sanctuary, the center of the high altar, and the center of the 5<sup>th</sup> bay of the nave were fully enclosed by screens, as these were the principal areas in the 16<sup>th</sup> and 17<sup>th</sup> century. A final screen, positioned to separate the sanctuary and choir, was probably acoustically and visually transparent.

Since there is little information about the adornments of the Saint-Germain-des-Prés during the 16<sup>th</sup> and 17<sup>th</sup> century, the Notre-Dame cathedral was examined as a reference in order to make an ‘educated guess’ regarding what kind of materials could have been used and where these could have been positioned [Wright 2008]. In the Notre-Dame cathedral, screens also enclosed the sanctuary and high altar. The screens, presumably extending halfway up the shaft of the sanctuary’s columns, were covered with draperies. Furthermore, Notre-Dame cathedral was furnished for festivities, especially in the area of the sanctuary, by tapestries. By the early 16<sup>th</sup> century, there was scarcely an area of the high altar and sanctuary that was not covered with some kind of fabric during principal festivities. Moreover, the floor immediately in front of the Notre-Dame cathedral’s altar was covered with rugs. Finally, it is assumed that the painted plaster was already present in the Saint-Germain-des-Prés church since this was a conventional finishing for churches of that

Table 4.6: Octave band absorption coefficients and assigned surface area ( $\text{m}^2$ ) used in the GA model of the 17<sup>th</sup> century liturgical configuration of the Saint-Germain-des-Prés. For these materials,  $\text{char}_{\text{depth}} = 1$  mm, with scattering coefficients determined by Eq. 4.1.

	125	250	500	1000	2000	4000	area
tapestries	0.07	0.11	0.16	0.21	0.31	0.44	288
draperies	0.20	0.25	0.30	0.35	0.40	0.50	406
rugs	0.09	0.08	0.21	0.26	0.27	0.37	85

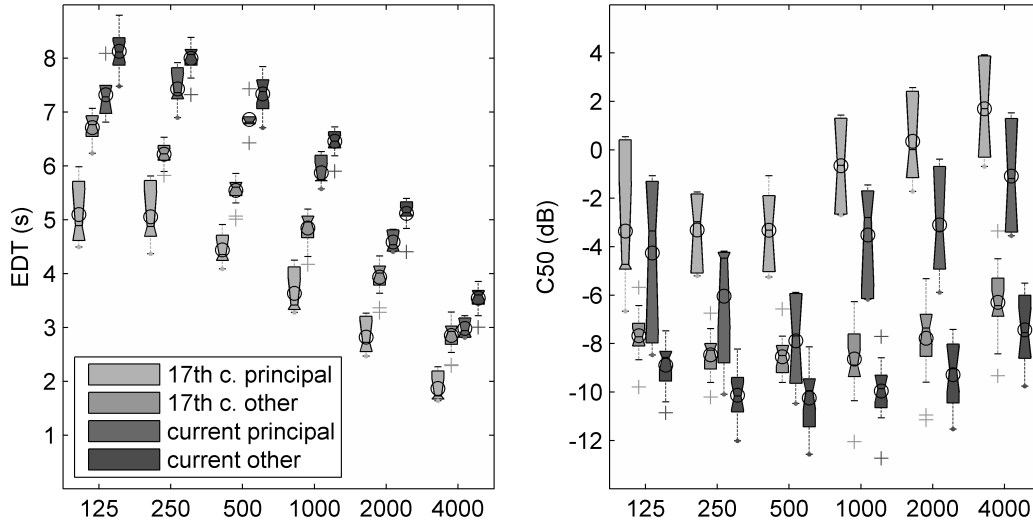


Figure 4.17: Summary of EDT and C50 results for the 17<sup>th</sup> century and the current configuration's calibrated GA model considering receivers inside the principal areas and those at the other positions. (See legend of Fig. 4.6 for boxplot notations).

time.

To acquire information about the absorption coefficients of materials described in the previous paragraph, papers which studied these materials and churches which still apply comparable materials were considered. The sound absorption of the tapestries was adopted from [Martelotta 2011], a study concerning 17<sup>th</sup> century tapestries. The absorption of the rugs was adopted from the database in [Vorländer 2008]. Additionally, in the cathedral of Seville, banners are positioned at the entrance [Alonso 2014]. It was assumed that the draperies in the 17<sup>th</sup> century configuration of the Saint-Germain-des-Prés could have had the same absorption coefficient. Table 4.6 shows the adopted absorption coefficients,  $\text{char}_{\text{depth}}$ , and their associated surface area in the GA model.

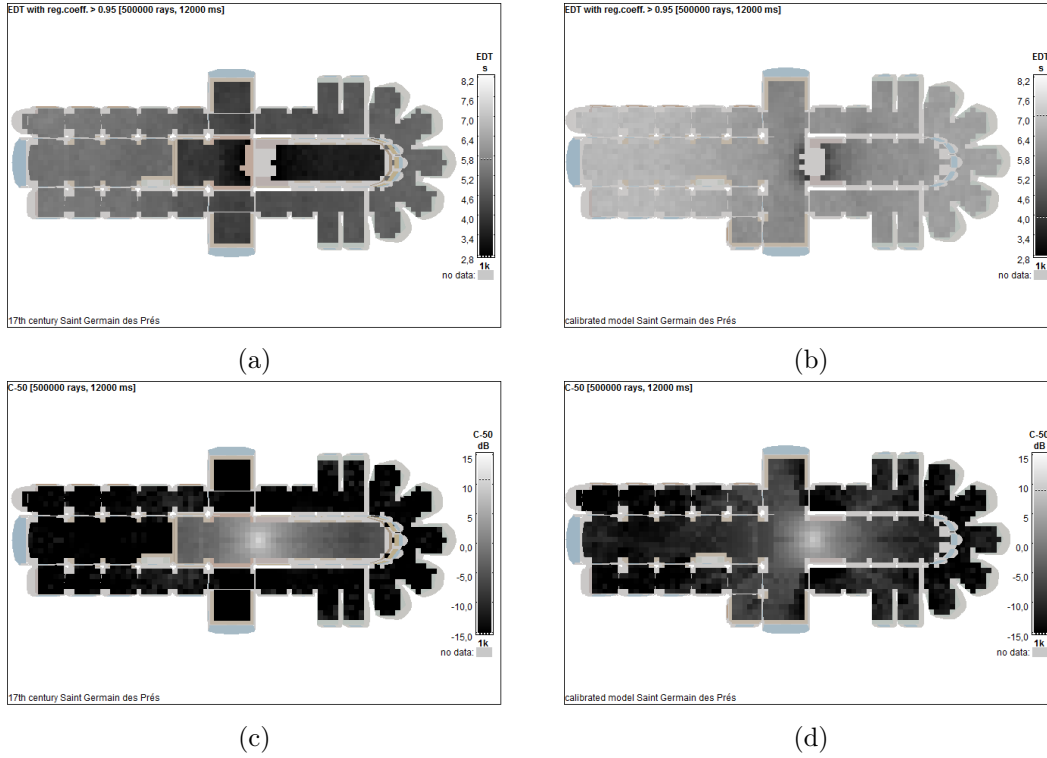


Figure 4.18: Simulation results for 1 kHz octave band. (a) EDT in the 17<sup>th</sup> century GA model. (b) EDT in the calibrated GA model. (c) C50 in the 17<sup>th</sup> century GA model. (d) C50 in the calibrated GA model. (500.000 rays used, ray truncation time 12000 ms).

#### 4.5.1.2 Acoustical conditions in a historical configuration of the Saint-Germain-des-Prés church

Unlike the nearby cathedral church of Notre-Dame, with its professional singers, the principle performers and auditors in the Saint-Germain-des-Prés church during the 17<sup>th</sup> century were monks, who were positioned in the principal areas of the high altar and sanctuary. Therefore, current and historical listening conditions are compared when the source is positioned in these areas and the receivers positioned inside and outside the principal areas.

Fig. 4.17 shows that for both the 17<sup>th</sup> century and current configuration the mean EDT is shorter and the mean C50 is greater across octave bands in the principal areas than in the other areas. In addition, this difference in acoustics between the two areas was greater during the 17<sup>th</sup> century.

Fig. 4.18 presents a mapping of EDT and C50 values in the 1000 Hz octave band for the calibrated reference GA model and the 17<sup>th</sup> century liturgical GA model<sup>10</sup>. The quantified results show, as could be expected with the addition of

<sup>10</sup>These plots are results from the internal CATT-Acoustic parameter mapping. This stands in contrast to the previously presented results which were based on RIR analysis.

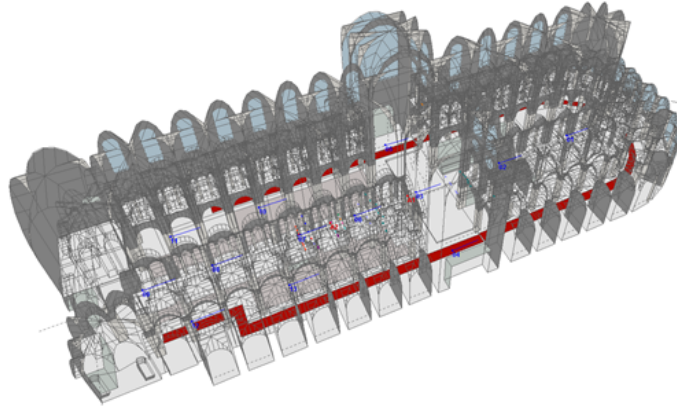


Figure 4.19: GA model of the Notre-Dame de Paris cathedral (ca. 14,700 surfaces). The red surfaces highlight the positions of the carpeting.

absorption, that the 17<sup>th</sup> century configuration had a perceptually shorter mean EDT and a perceptually higher mean C50, an effect which is more prominent in the principal areas. Shorter reverberation times and higher clarity are associated with higher speech intelligibility. It is probable, therefore, that the performance of the liturgy in the choir of Saint-Germain-des-Prés church changed in synchrony with the acoustics. For example, it would have been possible, with this increased clarity, to have performed the standard chant repertory at a higher tempo. The increased clarity of the performance space may have also encouraged the composition of new musical forms inspired by, and adapted to, the changed acoustical environment. Further research on the implications of these observations is the focus of ongoing studies.

#### 4.5.2 Notre-Dame model

Parameter results presented in Sec. 3.4 showed that the reverberation time of the Notre-Dame cathedral between the 1987 and 2015 has decreased. As data from [Hamayon 1996, Mercier 2002], (published in 1996 and 2002) was comparable to the 2015 measurements, it can be concluded that changes leading to the shorter reverberation time estimations were carried out between 1987 and 1996. As the volume of the Notre-Dame cathedral is rather large the reverberation time difference has to be the result of substantial changes. In a telephone conversation with the Notre-Dame cathedral it was confirmed that carpeting was installed in several areas and two confirmation booths were added in the two alcoves adjacent to the first two bays of the south naves during this time period. As the effect of the confirmation booths was considered marginal, simulations were performed only replacing the carpet by marble flooring. The possibility of the atmospheric conditions influencing the reverberation time results was considered. As temperature and relative humidity mainly effect reverberation estimations above 1000 Hz [Benedetto 1983], this can be excluded as the cause for the decrease in reverberation time.

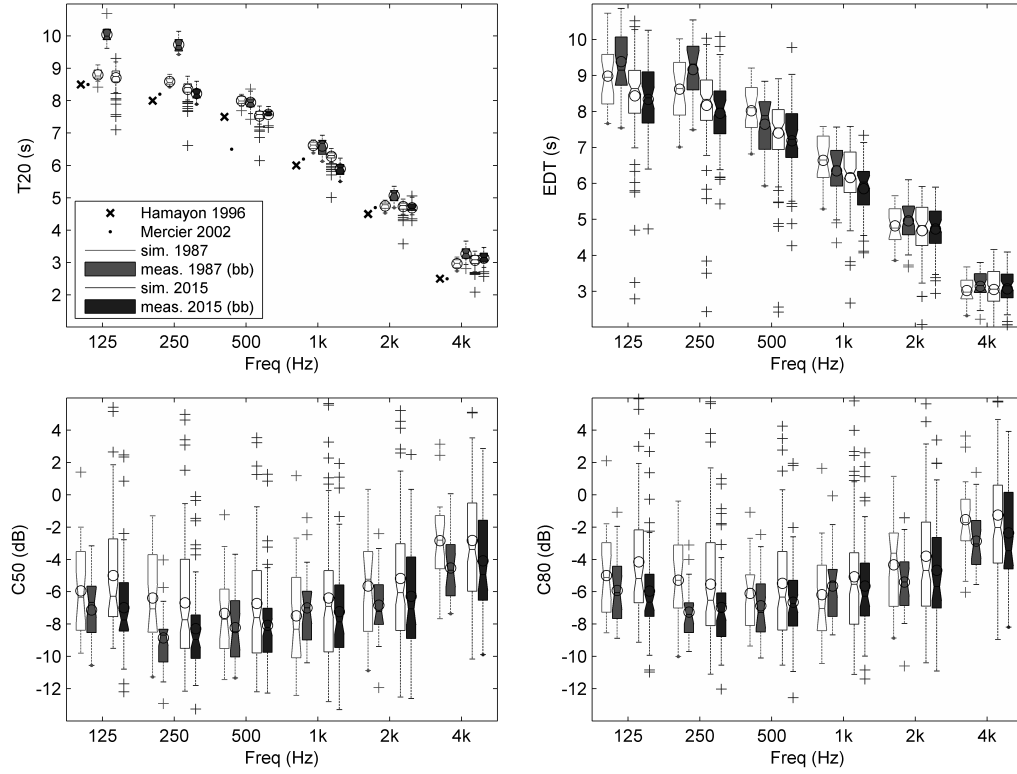


Figure 4.20: Distribution of T20, EDT, C50, and C80 results of the 1987 balloon burst and GA model, as well as 2015 measurements (combination of sine sweeps and balloon bursts), and its GA model. Additionally, results from previous Notre-Dame reverberation studies [Hamayon 1996, Mercier 2002] are included for T20. (See legend of Fig. 4.6 for boxplot notations).

Fig. 4.20 shows the distribution of the T20, EDT, C50, and C80 for the 1987 GA model and corresponding measurements. It shows that the reverberation changes can only be partly explained by the addition of the carpet. As no other refurbishments or renovations were carried out, one can only contemplate about other causes for the decreased reverberation time.

## 4.6 Discussion

The results from this chapter indicate that it is possible to calibrate GA models for reverberation and clarity parameters employing absorption and scattering coefficients. The presented calibration procedure is compared to previous methods [Martellota 2014, Alvarez-Morales 2015, Astolfi 2008]. It should be noted that these calibrations were carried out for parameter estimations, where the presented procedure aimed to calibrate for auralizations.

Therefore, the presented methodological calibration procedure employed pa-



parameter analysis based on the generated RIRs where the other studies used parameter estimations based on the energy echograms. As the implementation of Lambert scattering leads to run-to-run variations, especially for external RIR analysis, the final results were given as averages over five different predictions in order to take into account run-to-run variations, as was done in [Martellotta 2009]. The previous studies neglected the usage of scattering coefficients [Martellotta 2014, Alvarez-Morales 2015] or concluded them not to influence parameter estimations [Astolfi 2008], while the presented procedure devoted a step to gauging knowledge how changes in scattering effects influence parameter estimations and another step implementing this knowledge.

Parallels can be found in how previous and the presented methods calibrate reverberation parameters. [Martellotta 2014, Alvarez-Morales 2015] started with a search for variations in data sets to create a tolerance range for absorption adjustments. Subsequently, this range was employed to bring average reverberation parameters within one JND of the measurements. These steps were adopted into the proposed methodical calibration procedure.

## 4.7 Summary

This chapter presented a methodical calibration procedure for GA models. The procedure based on measured and simulated RIR comparison consists of 9 steps. One step is devoted to gauging knowledge regarding the effect of run-to-run variation caused by GA software's implementation of Lambert scattering. Another step gains knowledge of the effect of changing the scattering coefficients as it is not obvious. Results showed that applying this calibration method leads to GA models which sufficiently estimate reverberation and clarity parameters. Calibration according to parameter estimation can be considered a necessary step towards correctly estimating the sound propagation in 3D spaces, however auralizations with similar parameter estimation can still sound very different. Additionally, as the calibration procedure employed parameter estimations for reverberation and clarity estimations based on RIRs it is important to verify the perceptual equivalency of BRIRs as these also include spatial effects. Therefore, the subsequent chapter perceptually compares auralizations based on simulated (B)RIRs to measured auralizations employing listening tests.





# Perceptual evaluation of calibrated auralizations<sup>1</sup>

---

## Contents

<b>5.1 Introduction</b>	<b>77</b>
5.1.1 Verification in the 1980s and 1990s	78
5.1.2 Recent verifications	79
<b>5.2 Objective parameter results</b>	<b>81</b>
<b>5.3 Preliminary auralization comparison listening test</b>	<b>83</b>
5.3.1 Preparation of the measured (B)RIRs	84
5.3.2 Anechoic stimuli	84
5.3.3 Protocol	85
5.3.4 Results	86
<b>5.4 Binaural auralization comparison listening test</b>	<b>90</b>
5.4.1 Anechoic stimuli	90
5.4.2 Protocol	91
5.4.3 Results	92
<b>5.5 Discussion</b>	<b>98</b>
<b>5.6 Summary</b>	<b>99</b>

---

## 5.1 Introduction

Chapter 4 examined the ability to calibrate a GA model according to parameters to measured data of a real space, resulting in a methodical calibration procedure for GA models. This calibration procedure was objectively validated by means of

---

<sup>1</sup>This work was presented in part in:

- B.N.J. Postma and B.F.G Katz. *Perceptive and objective evaluation of calibrated room acoustic simulation auralizations* J. Acoust. Soc. Am. vol. 140, no. 6, pp. 4326-4337, 2016, doi:10.1121/1.4971422..
- B.N.J. Postma, A. Tallon, and B.F.G Katz. *Calibrated auralization simulation of the abbey of Saint-Germain-des-Prés for historical study* In Intl. Conf. Auditorium Acoustics, Paris, 2015, pp. 190-197.

the objective parameter reverberation and clarity comparisons for omni-directional sources and receivers. However, the binaural acoustical experience of an auralization can be very different than what descriptions of acoustic properties by abstract numerical quantities based on omni-directional receivers suggest. It is therefore important to examine how other perceptual parameters, especially spatial parameters, vary between measured and simulated binaural room auralizations.

Where calibration procedures of GA models are generally based on room acoustical parameters, validation of the calibration process can be performed employing listening tests where simulated auralizations are compared to actual recordings or measured auralizations [Lokki 2001]. To carry out auralization comparisons, besides ensuring the correct material property input, which can be based on acoustical parameter estimations, one needs to ascertain that other parts of the simulation are configured equal to the measurement as well. Specific attention is made here with regards to binaural auralizations which provide the most natural reconstruction of the listening experience. The use of a different source or receiver directivity or binaural HRTF can lead to coloration and other significant perceptual differences. On the other hand, preparation steps are required for the measured BRIR as well. Compensations must be carried out for the frequency response characteristics of the measurement system and for differences in SNR between measurement and simulation as the simulated BRIR is free of background-noise. Studies are discussed which compare measured and simulated auralizations.

### 5.1.1 Verification in the 1980s and 1990s

Since 1981, validations of simulated auralizations have been carried out by comparing measurements to simulations. Kleiner et al. [Kleiner 1981] studied speech intelligibility in a real and simulated theater. four situations were compared:

1. Direct listening in the theater
2. Indirect listening i.e. listening to dummyhead recordings made in the theater
3. Indirect listening to auralizations based on the energy echograms (series 1)
4. Indirect listening to auralizations based on the energy echograms (series 2)

The HRTF differed between direct listening and simulations on the one hand to recordings on the other. The background noise level difference nor the frequency response of the recording system were compensated for. On the other hand, source directivity was equal between recording and simulations. The results showed that it was hard to simulate the properties of the theater using calculated RIRs.

In a comparable study, Farina [Farina 1995a] compared measured and simulated responses of a church and sport arena. Source directivity was equal between simulation and measurement, however the employed HRTF differed. Additionally, the frequency response of the measurement system and background noise difference was not compensated for. Parameter comparison was presented by means of average RIR data [Farina 1995d]. Differences with more than one JND were observed for T30, EDT, and C80. Results of 14 participants who took the listening test indicated that the auralizations were perceived differently, however the most important

acoustical attributes were perceived similar, these being *Reverberance*, *Bass*, *Treble*, and *Spatial impression*.

### 5.1.2 Recent verifications

In 2001 Lokki et al. [Lokki 2001, Lokki 2002a] compared real head binaural recordings to simulated auralizations of a lecture room using listening tests. Simulations were performed with the DIVA real-time auralization system. The radiation characteristics of the sound source were measured and radiation filters were designed to fit the measured frequency response. The applied HRTFs were measured from the same person who did the real-head recordings. Recordings were also made to capture the background noise which was then added to the simulated auralizations. It does not appear that compensation for the frequency characteristics of the recording system was carried out. The recordings and simulated auralizations were compared according to the perceptual attributes: source location, externalization, sense of space, and timbre. They concluded that it was possible to create natural sounding virtual auditory environments with physics-based room acoustic modeling and advanced digital signal processing. However, some differences between recordings and auralizations were found, especially with a transient-like stimulus signal.

Choi et al. [Choi 2006] attempted to validate computer models of two concert halls by comparing recordings to simulated auralizations. The main aim was to determine whether computer generated auralizations were judged in the same way as recorded music in actual halls. BRIR measurements were carried out and anechoic stimuli were played and recorded in the considered rooms. In parallel, GA models of the spaces were created using the GA software . Measured and simulated BRIRs for five source-receiver combinations per concert hall showed several differences which exceeded the JND for room-acoustical parameters. Particularly, estimation of C80 seemed problematic with 7 out of 10 source-receiver combinations overestimating it by more than 1 JND and in some cases by more than 4 JND. In the subsequent listening test, recordings were compared to simulated auralizations. HRTFs differed between recording and simulation. Compensations for the frequency characteristics of the recording system were carried out, however no compensation was performed for differences in SNR. The listening test indicated that there was a significant difference in “preference” between recordings and simulated auralizations, indicating that measurement and simulation differed perceptually.

Yang et al. [Yang 2007] attempted to validate computed auralizations for use in speech-intelligibility studies. For this purpose, two classrooms (one with “extensive acoustical treatment” and the other with “no acoustical treatment”) were measured and also simulated using the GA software CATT-Acoustic. Simulated and measured RIRs showed several differences exceeding JNDs for parameter results. C50 was underestimated for one classroom and overestimated for the other by more than 1 JND for five of the six source-receiver combinations, with a maximum difference of 5.6 JND. Subsequently, speech-intelligibility tests were carried out in the classrooms in background noise free conditions and with added background noise. In

Table 5.1: Control factors between measurement and simulation in previous auralization comparison studies. A ‘✓’ indicates that the factor was controlled (i.e. equal) between measurement/recording and simulation, a ‘×’ indicates a difference. † this control factor was equal for the noisy conditions, however differed for the noise free conditions.

Factor	Lokki [Lokki 2001]	Choi [Choi 2006]	Yang [Yang 2007]
Source directivity	✓	✓	✓
Meas. system’s freq. response	×	✓	×
SNR	✓	×	†
HRTFs	✓	×	×

parallel the same speech-intelligibility tests were performed in a sound booth using the simulated auralizations in comparable background noise conditions. No compensation was performed for the frequency characteristics of the measurement system. The HRTF differed between the classrooms and simulated auralization tests. In the noisy conditions the SNR was equal, however no compensation was performed in the noise-free conditions. The simulated auralizations test and test in the classrooms gave comparable results for the “extensively acoustically treated” room in background noise free conditions. However, in the case of the “no acoustical treatment” classroom in the noise free condition and both cases with added background noise, results did not agree between actual classroom and simulated conditions. Table 5.1 provides an overview of which factors were controlled between measurement and simulation in these recent verifications.

Even though the main objectives of these previous studies differed from the current study, their results highlight potential problems in creating GA models to obtain perceptually similar auralizations. First, several control factors differed between measurement and simulation these being source/receiver directivity, HRTF, compensation for the frequency response characteristics of the measurement system, and differences in SNR between measurement and simulation. Second, measured and perceptible differences were identified between the measurement and simulation, clarity parameters seem especially problematic, a problem addressed in the previous chapter.

This chapter compares the perceptual quality of monaural (created from the omni-directional microphones) and binaural (created from the dummyhead microphones) simulated auralizations for the Théâtre de l’Athénée, Notre-Dame cathedral, and Saint-Germain-des-Prés church of several source-receiver combinations relative to measured data. Listening tests evaluated to what extent the calibration resulted in similar simulated to measured auralizations regarding important acoustical attributes.

To accomplish this study, this chapter presents first a comparison of the room acoustic parameter results for the binaural positions employed in the listening

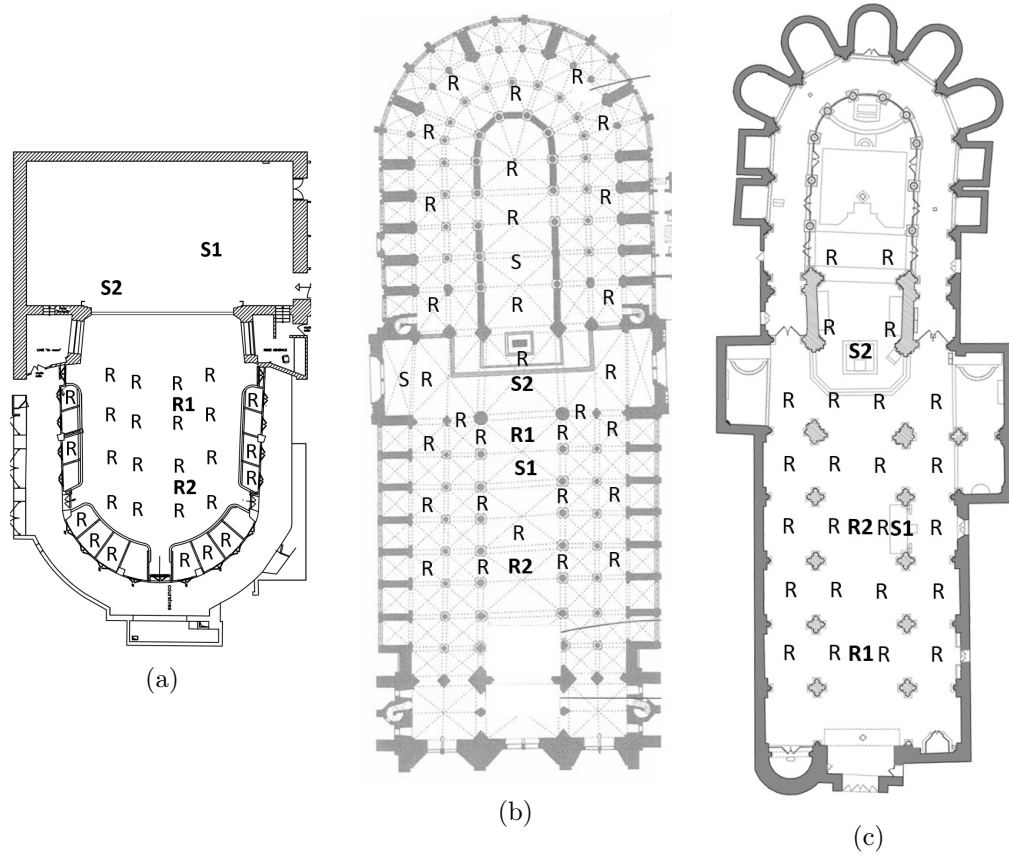


Figure 5.1: Measurement plans of (a) Théâtre de l'Athénée, (b) Notre-Dame cathedral, and (c) Saint-Germain-des-Prés church. **S** and **R** represent source and receiver positions (**S**# and **R**# were employed in the listening test).

test. Thereafter, monaural and binaural auralizations were compared for the Saint-Germain-des-Prés church in a preliminary listening test (see Sec. 5.3). As results indicated some perceptual differences regarding *Coloration* for the binaural auralizations, an additional preparation step was carried out. Finally, Sec. 5.4 describes the final listening test which compared binaural auralizations for the Théâtre de l'Athénée, Notre-Dame cathedral, and Saint-Germain-des-Prés church.

## 5.2 Objective parameter results

Chapter 4 showed that a model can be calibrated so that simulated and measured mean monaural receiver agree within a given tolerance. As binaural receiver positions were employed for the final listening test, their parameter results are compared here. The receiver positions are depicted in Fig. 5.1. The binaural GA-based simulation incorporated the previously measured HRTFs of the dummy head used during the measurement. Notre-Dame cathedral receiver position **S1R1** was omitted from

Table 5.2: Measured and simulated frequency averaged objective parameters (EDT, T20, C50, C80,  $IACC_e$ ,  $IACC_l$ ) per binaural receiver position in the Théâtre de l’Athénée (Athénée), Notre-Dame cathedral (ND), and Saint-Germain-des-Prés (SGdP) (Differences greater than 1 JND are *indicated*).

	Pos.	Reverberance			Clarity			Spaciousness		
		EDT(s) [JND=5%] (500–1000 Hz)			C50(dB) [JND=1.0] (500–1000 Hz)			1- $IACC_e$ [JND=0.075] (500–4000 Hz)		
		Meas.	Sim.	Diff.	Meas.	Sim.	Diff.	Meas.	Sim.	Diff.
Athénée	S1R1	1.68	1.31	+0.37	-0.3	-1.5	+1.2	0.52	0.51	+0.01
	S1R2	1.65	1.38	+0.27	-0.5	-0.7	+0.2	0.66	0.69	-0.03
	S2R1	1.37	1.33	+0.04	0.7	-1.5	+2.2	0.60	0.64	-0.04
	S2R2	1.35	1.29	+0.06	0.3	0.1	+0.2	0.74	0.72	+0.02
ND	S1R2	5.52	6.53	-1.01	-6.5	-6.5	-0.0	0.53	0.22	+0.31
	S2R1	6.76	6.59	+0.18	-2.4	-0.1	-2.3	0.21	0.16	+0.05
	S2R2	7.44	7.78	-0.34	-11.3	-10.9	-0.4	0.31	0.26	+0.05
SGdP	S1R1	5.65	5.56	+0.09	-7.1	-7.6	+0.5	0.65	0.69	-0.04
	S1R2	5.04	4.64	+0.40	-1.3	-0.4	-0.9	0.61	0.49	+0.12
	S2R1	7.11	7.23	-0.12	-11.6	-10.1	-1.5	0.54	0.27	+0.27
	S2R2	6.98	6.80	+0.18	-9.8	-8.0	-1.8	0.23	0.14	+0.09
		T20(s) [JND=5%] (500–1000 Hz)			C80(dB) [JND=1.0] (500–1000 Hz)			1- $IACC_l$ [JND=0.075] (500–4000 Hz)		
Athénée	S1R1	1.42	1.38	+0.04	1.6	1.7	-0.1	0.88	0.69	+0.19
	S1R2	1.44	1.42	+0.02	1.9	1.9	+0.0	0.86	0.63	+0.23
	S2R1	1.38	1.45	-0.07	3.0	1.6	+1.4	0.86	0.67	+0.19
	S2R2	1.40	1.49	-0.09	3.0	2.5	+0.5	0.81	0.61	+0.20
ND	S1R2	6.51	6.82	-0.31	-5.3	-5.3	-0.0	0.89	0.64	+0.25
	S2R1	6.19	6.14	+0.05	-1.8	0.7	-2.5	0.93	0.66	+0.27
	S2R2	6.57	6.69	-0.12	-10.1	-9.1	-1.0	0.89	0.60	+0.29
SGdP	S1R1	6.44	6.35	+0.09	-5.4	-5.7	+0.3	0.92	0.69	+0.23
	S1R2	6.46	6.37	+0.09	0.0	1.4	-1.4	0.91	0.69	+0.22
	S2R1	6.68	6.55	+0.13	-9.5	-8.3	-1.2	0.90	0.64	+0.26
	S2R2	6.12	6.30	-0.18	-7.8	-6.0	-1.8	0.91	0.62	+0.29

the listening test as the dummy head was orientated away from the sound source.

Table 5.2 provides a comparison of the measured and simulated single number frequency average objective parameters (EDT, T20, C50, C80,  $IACC_e$ ,  $IACC_l$ ) per position for the three considered rooms. T20 between measurements and simulations agreed well, with all simulated positions, except for one source-receiver combination in the Théâtre de l’Athénée, estimating the measurement within 1 JND. The clarity parameters show more positions with differences of more than 1 JND. Position **S2R1** in the Notre-Dame cathedral exhibited the maximum differences between measured and simulated C50 and C80, with 2.3 and 2.5 JND, respectively. It should be noted that the pews positioned in the nave and first side aisles were modeled as floor with high scattering coefficients. In the cathedral the pews have a height of  $\approx 0.9$  m,

Table 5.3: Measured and simulated single number frequency average parameters (EDT, T20, C50, and C80, according to ISO 3382 standard [ISO 2009]) per monaural position in the Saint-Germain-des-Prés church. (Differences greater than 1 JND are *indicated* ).

Position	Reverberance			Clarity		
	EDT(s)-[JND:0.31] (500-1000Hz)			C50(dB)-[JND:1.0] (500-1000Hz)		
	Meas.	Sim.	Diff.	Meas.	Sim.	Diff.
S1R1	5.76	5.56	+0.20	-8.7	-7.2	-1.5
S1R2	5.27	5.12	+0.15	-2.7	1.1	-3.8
S2R1	6.96	7.02	-0.06	-10.7	-10.1	-0.6
S2R2	6.76	6.73	+0.03	-10.1	-9.2	-0.9
	T20(s)-[JND:0.31] (500-1000Hz)			C80(dB)-[JND:1.0] (500-1000Hz)		
	Meas.	Sim.	Diff.	Meas.	Sim.	Diff.
	Meas.	Sim.	Diff.	Meas.	Sim.	Diff.
S1R1	6.31	6.42	-0.11	-6.9	-5.9	-1.0
S1R2	6.33	6.29	+0.04	-1.2	1.7	-2.9
S2R1	6.74	6.52	+0.20	-8.9	-8.6	-0.3
S2R2	6.35	6.05	+0.20	-8.1	-7.8	-0.3

this could have influenced the C50 and C80 for combinations with short source-receiver distance. Beside this maximum difference, more than half of the source-receiver combinations exhibit simulated clarity parameters within 1 JND. Finally, Table 5.2 shows that the differences between measured and simulated  $IACC_e$  and  $IACC_l$  exceed 1 JND for the majority of the positions, in some cases by more than 3 JND.

As in the preliminary listening test also monaural receiver positions were employed, Table 5.3 presents the single frequency averages for these positions in the Saint-Germain-des-Prés church. EDT and T20 between measurements and simulations agreed well, with all simulated positions estimating the measurement within 1 JND. The clarity parameters show more positions with differences of more than 1 JND. Position **S1R1** exhibits a C50 difference exceeding 1 JND and for position **S1R2** the C50 and C80 differ more than 1 JND.

### 5.3 Preliminary auralization comparison listening test

In order to establish the suitability of the proposed listening test and to establish the degree of similarity between measured and simulated auralizations, both for monaural and binaural receivers, a preliminary listening test was performed. Four monaural and four binaural measurement combinations were compared for two test stimuli for the Saint-Germain-des-Prés church.



### 5.3.1 Preparation of the measured (B)RIRs

Prior to commencing a listening tests, some additional processing is required for the measured (B)RIRs. First, the frequency response characteristics of the measurement system have been compensated for by creating an equalization filter. The measurement chain (one omnidirectional microphone only) was installed in an anechoic room (IRCAM, Paris) and the IR of the omnidirectional speaker was measured at  $5^\circ$  increments in the horizontal plane. The resulting IRs were time-windowed to 512 samples, to remove any reflection artifacts, from which the FFT was calculated and the mean over all directions of the magnitude determined. A filter was generated to match the inverse of this response using the recursive filter design yule-walk method. Non-linear frequency weighting followed a bark scale approximation, constraining the filter's level of detail to follow human hearing sensitivity.

Second, it is necessary to extend the signal beyond the noise floor since the dynamic range of the measured RIRs is lower than required for audio [Grillon 1995]. Therefore, differences in SNR between frequency bands were compensated for. The RIR was decomposed into  $1/3^{rd}$ -octave band components (spanning 100–16000 Hz). The noise floor was detected by determining the SNR for each  $1/3^{rd}$ -octave band. The signals were then windowed at the point 5 dB above the noise floor, eliminating the trailing noise. The decay rate (reverberation time) was then calculated over the entire window, and this decay rate was used to synthesize the continuation of a noise-free reverberant tail. This reverberant tail was created by multiplying a vector of random values by an exponential vector based on the decay rate, this synthesized RIR was filtered to the considered  $1/3^{rd}$  octave band. Since a reliable decay estimate reasonably requires at least 15 dB,  $1/3^{rd}$ -octave bands with SNR < 20 dB were discarded (muted in the final RIR). This typically resulted in omitting the lowest two  $1/3^{rd}$ -octave bands: 100 and 125 Hz (these were also omitted from the simulated RIRs). An equal power cross-fade between the measured and synthesized responses was applied over the last 10 dB decay of the measured and the first 10 dB decay of the synthesized response to provide a smooth transition between decay sessions, limiting audible artifacts. No effort was made to recreate the correlation between left and right ear synthesized tails. Processing left the perceivable correlation between left and right channel unaffected as the absolute differences between pre- and post-processed single number frequency average  $IACC_l$  (mean absolute difference: 0.004, SD: 0.003) were well within 1 JND.

### 5.3.2 Anechoic stimuli

The prepared measured and simulated (B)RIRs were convolved with two anechoic audio extracts appropriate to the acoustic function of the room: female soprano singing ‘*Abendempfindung*’, by W.A. Mozart, for details of the anechoic recording system see [Lokki 2008a]; male tenor performing ‘*A Chloris*’, by R. Hahn [Katz 2007]. Fig. 5.2 depicts the spectral composition of the chosen extracts. As the two lowest  $1/3^{rd}$ -octave bands were omitted from the RIRs, the extracts were

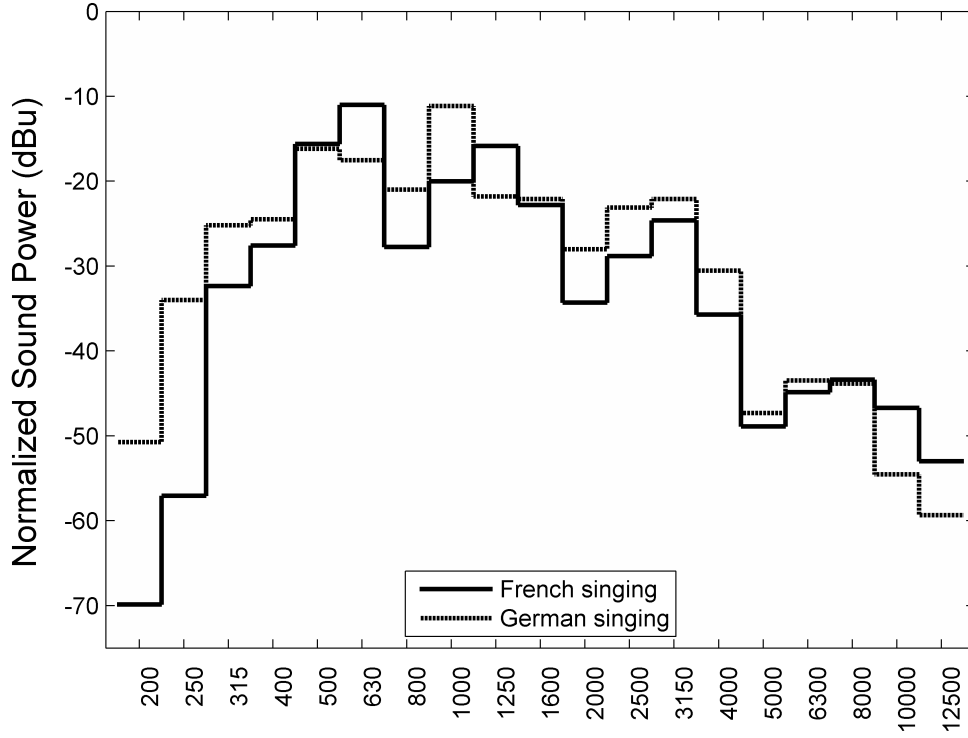


Figure 5.2: 1/3-octave RMS power spectrum for the anechoic stimuli.

chosen to have minimal energy in these bands. RMS of the measured and simulated convolutions was used for normalization. An alternative means of normalization could have employed A-weighted levels. As the auralizations employ the same anechoic recording and similar BRIRs these methods are roughly equivalent. This is confirmed by comparisons of the A-weighted levels of the normalized measured and simulated auralizations. For all auralization pairs, this difference was less than 1 dBA (mean absolute difference 0.48 dBA, SD: 0.27 dBA).

### 5.3.3 Protocol

The test was set up as an AB comparison. Stimuli were compared for both monaural and binaural receiver configurations for two source and receiver positions (Omni: **S1R1**, **S1R2**, **S2R1**, **S2R2**; Binaural: **S1R2**, **S1R2**, **S2R1**, **S2R2**<sup>2</sup>) resulting in 16 tested pairs. Four configurations were repeated to monitor the reliability of responses, resulting in 20 pairs. Additionally, participants were given three training pairs to ensure they understood the task. Results for these three training pairs were not analyzed.

Participants were given written instructions before commencing the listening

<sup>2</sup>It should be noted that the omnidirectional receivers employed were positioned  $\approx 2$  m to the left of the numbered binaural receiver.

test<sup>3</sup>. They were asked to rate the similarity of samples according to 5 perceptual attributes with the associated definitions:

- *Reverberance (reverb)* - The perception of the decay of sound. More reverberance is associated to a longer decay.
- *Clarity* - The degree to which discrete elements in the recorded musical performance stand apart from one another. If clarity is high, it is easy to spot individual notes in a musical piece, or individual phonemes in speech. If clarity is low, individual sounds merge, blend, or at the extreme can be confused and muddled.
- *Distance (dist.)* - The perceived distance to the sound source in the recording space.
- *Coloration (col.)* - Coloration represents changes in timbre, or frequency balance. It is qualified here as a comparison of the ratio of high to low frequency.
- *Plausibility (plaus.)* - Given the assumption that the recordings were made in a church, does the recording sound reasonable to you?

For binaural receiver pairs, participants were asked to additionally rate the similarity of samples according to 2 additional perceptual attributes with the associated definitions:

- *Apparent Source Width (ASW)* - Apparent source width describes the perceived width of a sound image. The source may sound ‘narrow’ (in the extreme case it is as if the sound is coming from a point). On the contrary, the source can also sound very ‘wide’.
- *Listener Envelopment (LEV)* - Listener envelopment describes the spatial distribution of the reverberant sound field. Higher “Listener envelopment” means a more uniform distribution, while less “Listener envelopment” means a more localized or directional reverberant sound.

Participants responded using a continuous graphic 100 pt scale, ranging from ‘A is much more ...’ to ‘B is much more ...’ corresponding respectively to values of  $-50$  and  $+50$ , with a center 0 response indicating no perceived difference. Presentation order and AB correspondence to simulation and measurement were randomized. Participants were able to listen to the compared pairs as many times as desired. Auralizations were presented via headphones (Sennheiser model HD 600) at an RMS level of 75 dBA. The experiment took place in a sound isolation booth at LIMSI, ambient noise level  $<30$  dBA. The 12 participants (mean age: 39.6 SD: 16.7) all reported normal hearing.

### 5.3.4 Results

Initial attention is given to the reliability of responses, determined from the mean absolute difference between the 4 repeated configurations and its SD (see Table 5.4

<sup>3</sup>Appendix C presents the instructions provided to the participants

Table 5.4: Mean absolute difference, corresponding SD and PCC between repeated pairs separated for monaural and binaural conditions.

		reverb.	clarity	dist.	Col.	Plaus.	ASW	LEV
Mon.	mean abs dif	12.5	11.4	9.6	9.0	9.0		
	SD	12.2	12.6	12.8	8.0	12.8		
	PCC	0.2	0.1	0.1	0.3	0.1		
Bin.	mean abs dif	10.7	9.7	8.9	7.3	8.2	11.0	8.6
	SD	12.2	13.2	9.2	8.2	7.3	13.5	10.1
	PCC	0.4	0.4	0.5	0.5	0.7	0.4	0.4

and Fig. 5.3a). The absolute mean difference ranged from 7.3 to 13.5 and corresponding standard deviation ranged from 7.3 to 13.2 on a 100 point scale. A metric for judging the repeatability of responses is Pearson’s Correlation Coefficient (PCC), defined as the covariance of the two sets of responses divided by the product of their standard deviations. It is interesting to note that the reliability of subject’s responses according to the PCC are uniformly better in the binaural condition than in the monaural condition.

Statistical analysis was performed employing confidence intervals (CIs). CI were calculated according to the Equation 5.1

$$z_{low} = \bar{x} - t \times \frac{s}{\sqrt{n}}, z_{up} = \bar{x} + t \times \frac{s}{\sqrt{n}} \quad (5.1)$$

where  $z$  are the limits of the CI interval,  $\bar{x}$  is the mean,  $t$  is the t-score,  $s$  is the SD, and  $n$  is the population size. The repetitions were excluded from further analysis. In order to compensate for multiple testing, the CI limits were extended, according to formula 5.2 [Dunn 1961].

$$Confidence\ level = 1 - \frac{\alpha}{m} \quad (5.2)$$

in which  $m$  is the number of tested conditions (‘Bonferroni’ correction) and  $\alpha$  is the significance level, in this manuscript taken as 0.05. As the monaural auralizations were tested for 5 attributes, with the typically employed  $\alpha$  of 0.05 this resulted in a 99% confidence interval. When the value 0 was included in the CI no significant difference between measurement and simulation was established.

Overall results were then compared employing 99% confidence intervals (CI) (see Fig 5.3b and Table 5.5). The simulated monaural auralizations were judged significantly less reverberant, less clear, and closer. The simulated binaural auralizations were rated significantly less reverberant, closer, less bright, less plausible, narrower, and less enveloping.

Subsequently the results were analyzed per position. Figs. 5.3c-d and Table 5.5 show that the degree of variance is smaller for the monaural receiver condition. The CI’s of the monaural auralizations included 0. The exception to this observation is **S1R2**, where the simulated auralization was rated significantly less reverberant, clearer, closer and brighter, which could be explained by the difference in C80 (meas–sim = –2.9 dB). The binaural condition responses exhibit more variation while

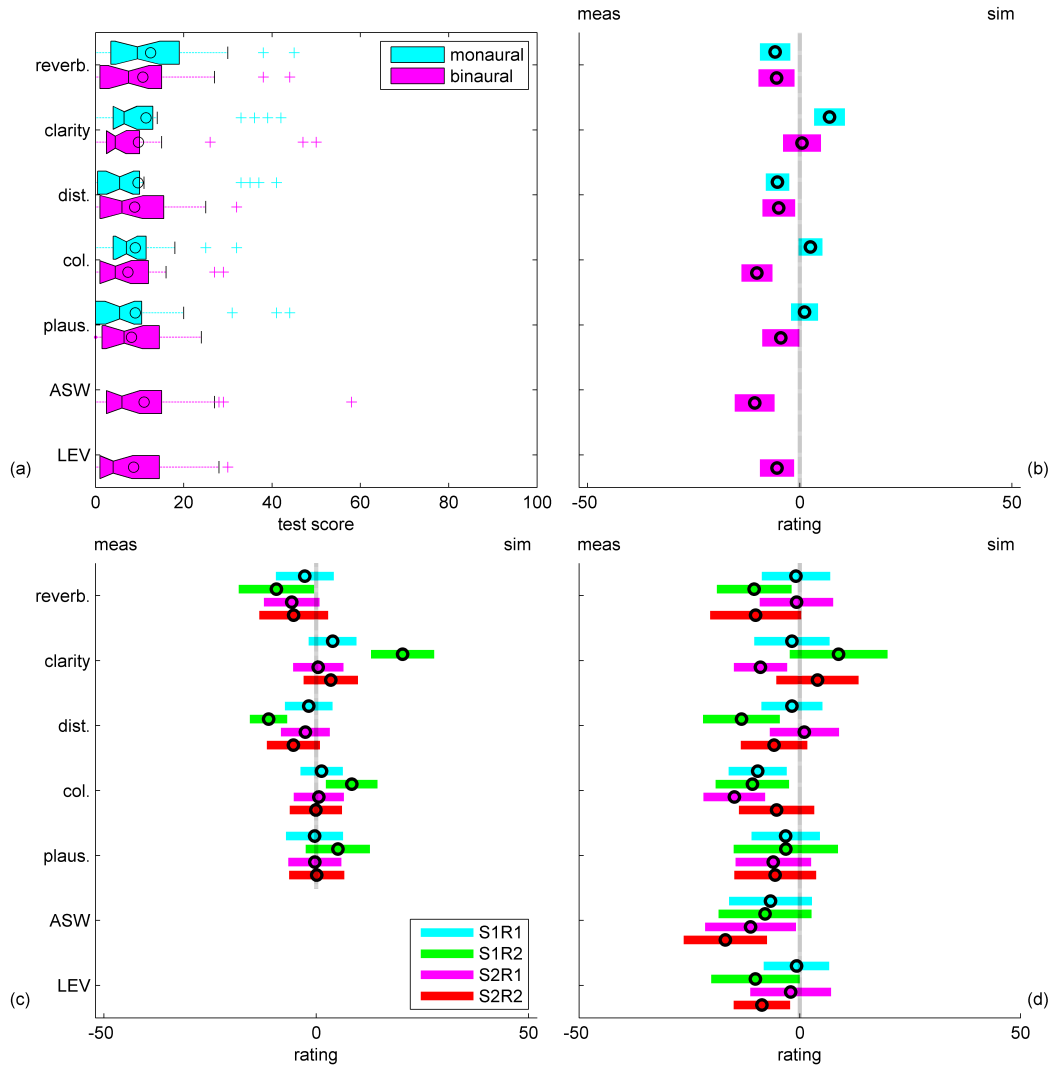


Figure 5.3: Measured on the left, Simulated on the right. (a) Notched boxplots of absolute differences for repeated pairs. Notches display the confidence interval, box limits represent the 25% and 75% quartiles, (—) indicates the median, (o) indicates the mean value. Remaining subplots: 99% CIs for all subjects on similarity of measured and simulated RIR auralizations. (b) Total results for the monaural and binaural auralizations. Results by position for (c) monaural auralizations and (d) binaural auralizations. Center dashed line represents a neutral response, (o) indicates the mean value.

the majority of the attribute ratings did not significantly differ. Specifically, the simulated BRIR at **S1R1** was perceived less bright, the simulated BRIR at **S1R2** was judged to be less reverberant, closer, and less bright, the simulated BRIR at **S2R1** was found to be less clear, less bright and narrower, and the simulated BRIR at **S2R2** was judged to be narrower and less enveloping than the measured BRIR

Table 5.5: Mean, lower and upper limits of the 99% CIs for the comparison between measured and simulated auralizations for the Monaural and Binaural conditions. (CI values which do not contain 0 are *indicated* ).

		pos.	CI values					pos.	CI values		
			low	mean	up				low	mean	up
Reverb.	Monaural auralizations	Total	-9.4	-5.8	-2.2	Binaural auralizations	Total	-9.7	-5.5	-1.3	
		S1R1	-9.5	-2.7	4.1		S1R1	-8.6	-0.9	4.1	
		S1R2	-18.2	-9.4	-0.5		S1R2	-18.8	-10.3	-1.9	
		S2R1	-12.3	-5.8	0.7		S2R1	-9.0	-0.8	7.5	
		S2R2	-13.4	-5.3	2.7		S2R2	-20.3	-10.0	0.3	
Clarity		Total	3.4	7.0	10.6		Total	-4.0	0.5	5.0	
		S1R1	-1.8	3.8	9.4		S1R1	-10.3	-1.8	6.7	
		S1R2	12.9	-20.3	27.8		S1R2	-2.3	8.8	19.8	
		S2R1	-5.5	0.5	6.4		S2R1	-14.9	-8.9	-2.9	
		S2R2	-3.0	3.4	9.8		S2R2	-5.3	4.0	13.3	
Dist.		Total	-8.0	-5.3	-2.5		Total	-8.8	-4.9	-1.1	
		S1R1	-7.4	-1.8	3.8		S1R1	-8.7	-1.8	5.1	
		S1R2	-15.6	-11.3	-6.9		S1R2	-21.9	-13.2	-4.5	
		S2R1	-8.3	-2.6	3.2		S2R1	-6.8	1.0	8.9	
		S2R2	-11.6	-5.4	0.8		S2R2	-13.3	5.8	1.7	
Col.		Total	-0.3	2.5	5.3		Total	-13.7	-10.1	-6.5	
		S1R1	-3.7	-1.3	6.2		S1R1	-16.1	-9.5	-3.0	
		S1R2	2.3	-8.3	14.4		S1R2	-19.0	-10.8	-2.5	
		S2R1	-5.3	0.6	6.5		S2R1	-21.8	-14.8	-7.9	
		S2R2	-6.2	-0.1	6.1		S2R2	-13.8	-5.3	3.3	
Plaus.		Total	-2.1	1.1	4.3		Total	-8.9	-4.5	-0.1	
		S1R1	-7.1	-0.4	6.3		S1R1	-10.9	-3.2	4.5	
		S1R2	-2.5	5.1	12.6		S1R2	-15.0	-3.2	8.6	
		S2R1	-6.6	-0.3	5.9		S2R1	-14.5	-6.0	2.5	
		S2R2	-6.4	0.1	6.6		S2R2	-14.8	-5.6	3.7	
ASW					Total	-15.3	-10.6	-5.9			
					S1R1	-16.0	-6.6	2.8			
					S1R2	-18.4	-7.9	2.6			
					S2R1	-21.4	-11.2	-0.9			
					S2R2	-26.3	-16.9	-7.5			
LEV					S1R1	-9.4	-5.4	-1.4			
					S1R1	-8.1	-0.8	6.6			
					S1R2	-20.1	-10.0	0.0			
					S2R1	-11.2	-2.1	7.1			
					S2R2	-15.0	-8.6	-2.2			

counterparts. It should be noted that the exact positions between monaural and binaural receiver conditions differed slightly ( $\approx 2$  m). Taking this into consideration, a two-way ANOVA which tested for auralization type (mon vs. bin) and position

found significant differences for *Reverberation* (Type:  $F = 0.1$ ,  $p = 0.88$ ; position:  $F = 3.2$ ,  $p = 0.02$ ; Type $\times$ position:  $F = 1.0$ ,  $p = 0.40$ ), *Clarity* (Type:  $F = 11.1$ ,  $p < 10^{-2}$ ; position:  $F = 16.5$ ,  $p < 10^{-2}$ ; Type $\times$ position:  $F = 1.9$ ,  $p = 0.14$ ), *Distance* (Type:  $F = 0.1$ ,  $p = 0.86$ ; position:  $F = 9.4$ ,  $p < 10^{-2}$ ; Type $\times$ position:  $F = 0.5$ ,  $p = 0.69$ ), *Coloration* (Type:  $F = 54.7$ ,  $p < 10^{-2}$ ; position:  $F = 2.2$ ,  $p < 10^{-2}$ ; Type $\times$ position:  $F = 3.1$ ,  $p = 0.03$ ), and *Plausibility* (Type:  $F = 7.3$ ,  $p < 10^{-2}$ ; position:  $F = 0.8$ ,  $p = 0.49$ ; Type $\times$ position:  $F = 0.3$ ,  $p = 0.83$ ). A subsequent  $t$ -test with ‘Bonferroni’ correction found significant differences between monaural and binaural receiver types for *Coloration* at positions **S1R2** and **S2R1**.

In general, binaural auralizations for *Clarity* agreed better than monaural auralizations. This is due to the outlier position **S1R2**. Source and receiver for this position are near the highly ornate pulpit and several columns. It is possible that the scattering properties for these elements were erroneous, leading to excessive early reflections and subsequently a perceivable higher *Clarity*. *Coloration* agreed better for monaural auralizations than binaural auralizations. Finally, binaural auralizations were judged slightly ‘wider’ and more ‘enveloping’ than their simulated counterparts.

## 5.4 Binaural auralization comparison listening test

The preliminary listening test showed that the use of the methodical calibration procedure resulted in a limited number of attributes for monaural auralizations which were rated significantly different for the GA model of the Saint-Germain-des-Prés church. Some significant differences between measured and simulated binaural auralizations were found, most prominently *Coloration* was rated more different in binaural conditions. These differences were most likely due to the mean spectral response of the binaural microphone setup which was not equalized for the measured response, contrary to the HRTF measurement protocol, which eliminates the frequency responses of the speaker and microphone, used for integrating the HRTF in the simulated responses. Therefore, an additional equalization filter was applied to the measured BRIRs. Example BRIRs from the three test rooms were analyzed by comparing the spectral differences of the late part of the reverberation. The average of these spectral differences across the three rooms was used to create an inverse filter, as previously described, which was then applied to the measured BRIRs. Additionally, as the monaural source-receiver combinations showed a good resemblance, for the final listening test only binaural source-receiver combinations were employed.

### 5.4.1 Anechoic stimuli

The measured and simulated BRIRs were convolved with three anechoic audio extracts appropriate to the acoustic function of the different rooms. For the Théâtre de l’Athénée: a French speaking male reciting a translated extract of ‘*Hamlet*’, by W. Shakespeare, and an Italian speaking male reciting an extract of ‘*Non recidere*,

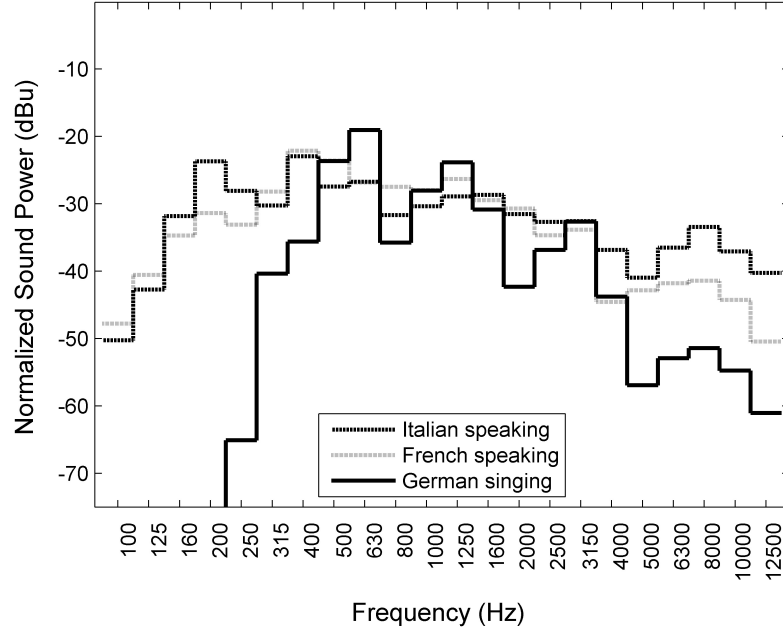


Figure 5.4: 1/3-octave RMS power spectrum for the anechoic stimuli.

*forbice, quel volto*’ by E. Montale. These anechoic stimuli were recorded in the anechoic room (IRCAM, Paris) using an omni-directional microphone (model 4006, DPA) at 4 m distance. For the Notre-Dame cathedral and Saint-Germain-des-Prés church: a female soprano singing ‘*Abendempfindung*’, by W.A. Mozart, for details of the anechoic recording system see [Lokki 2008a]. Fig. 5.4 depicts the spectral composition of the chosen extracts, each with a duration of  $\approx 13$  s. RMS of the measured and simulated convolutions was used for level normalization.

#### 5.4.2 Protocol

The resulting binaural auralizations were compared in a listening test. This test was set up as an AB comparison. In order to study the influence of the selected anechoic sound sample, two stimuli were used for the Théâtre de l’Athénée, resulting in eight auralization pairs ( $2 \text{ stimuli} \times 4 \text{ positions}$ : **S1R1**, **S1R2**, **S2R1**, **S2R2**). Only 1 stimuli was selected for the other rooms, resulting in 3 auralization pairs for the Notre-Dame cathedral (**S1R2**, **S2R1**, **S2R2**), and 4 for the Saint-Germain-des-Prés church (**S1R1**, **S1R2**, **S2R1**, **S2R2**), comprising 15 tested pairs. Additionally, 4 configurations were repeated to monitor the reliability of perceptual responses. 1 pseudo pair ( $A \equiv B$ ) was tested to determine the reliability of the participant. This resulted in a total of 20 pairs. Initially, participants were given 3 training pairs under supervision to ensure they understood the task. Results for these training pairs were not analyzed.

Participants were given written instructions before commencing the listening



test<sup>4</sup>. They were asked to rate the similarity of samples according to 8 perceptual attributes with the associated definitions: *Reverberance (reverb)*, *Clarity*, *Distance (dist.)*, *Tonal balance (ton. bal.)*, *Coloration (col.)*, *Plausibility (plaus.)*, *Apparent Source Width (ASW)*, *Listener Envelopment (LEV)*. In comparison to the preliminary listening test, definitions were adjusted for *Tonal balance (ton. bal.)* and *Coloration (col.)*:

- *Tonal balance (ton. bal.)* - Tonal balance represents changes in timbre, or frequency balance. It is qualified here as a comparison of the ratio of high to low frequency components so that more ‘tonal balance’ indicates more high frequency content.
- *Coloration (col.)* - Coloration represents modifications in the timbre of the sound source from its original timbre. With less coloration, the recording sounds more natural.

As in the preliminary listening test, participants responded using a continuous graphic slider scale (100 pt), with the end parts labeled ‘A is much more ...’ and ‘B is much more ...’ corresponding to end values of -50 and +50 respectively, with a center 0 response indicating no perceived difference. Presentation order and AB correspondence to simulation and measurement were randomized. Participants were able to listen to the compared pairs as many times as desired. Auralizations were presented via headphones (Sennheiser model HD 600) at an RMS level of 75 dBA.

In contrast to the preliminary listening test (see Sec. 5.3) where participants could only start listening at the beginning of the auralization, the test interface was improved so that participants were now able to select the starting play point during listening.

27 subjects participated, selected as having experience in either acoustics or vocal/instrumental performances. Before commencing the listening test, an audiogram was performed. One participant was immediately excluded due to poor hearing. Of the remaining 26 participants (mean age: 39.2 yrs. SD: 13.1) 21 were tested in an isolation booth located at LIMSI, ambient noise level < 30 dBA, and 5 were tested in a silent office located at the *Institut National d’Histoire de l’Art (THALIM)* (ambient noise level  $\sim$  31 dBA).

### 5.4.3 Results

Initial attention was given to the reliability of the subjects, determined from the results of the pseudo pair. Analysis of the absolute data for this pair showed that one participant gave outlier responses (exceeding  $\sim \pm 2.7\sigma$  and 99.3% coverage if the data is normally distributed) for all acoustics attributes indicating that the task was not well understood, and was therefore excluded from further analysis.

---

<sup>4</sup>Appendix C presents the instructions provided to the participants

Table 5.6: Mean absolute difference, corresponding SD and correlation coefficient between repeated pairs separated for monaural and binaural conditions.

	reverb.	clarity	dist.	Ton. Bal.	Col.	Plaus.	ASW	LEV
mean abs dif.	10.7	8.4	9.1	9.2	7.8	9.7	10.5	10.2
SD	12.4	8.9	10.8	10.5	8.8	11.4	11.8	10.1
PCC	0.1	0.4	0.3	0.4	0.1	0.1	0.1	0.4

#### 5.4.3.1 Reliability of responses

Analysis of the remaining 25 participants' results focused first on the reliability of responses, determined from the mean absolute difference between the 4 repeated configurations and its SD (see Table 5.6 and Fig. 5.5a). The absolute mean difference ranged from 7.8 to 10.7 and corresponding standard deviation ranged from 8.8 to 12.4 on a 100 point scale. A metric for judging the reliability of responses is Pearson's Correlation Coefficient (PCC), defined as the covariance of the two sets of responses divided by the product of their standard deviations (See Table 5.6 for results). It is interesting to note that for *Coloration* the absolute difference is low, in contrast the PCC is also low for this attribute. This raises questions about the use of PCC as a metric in this context.

#### 5.4.3.2 Perceptual attributes

In order to estimate whether acoustic attribute differed perceptually between measurement and simulation for the given test protocol and subjects according CIs were created according to Equation of which the extent was determined according to Equation 5.1 resulting in 99.4% CIs (see Fig. 5.6 and Table 5.7). When this CI excluded the rating 0 a significant difference was established.

Combined results were compared (see 'total' CIs in Fig. 5.6a). Simulated auralizations were rated significantly less reverberant, clearer, closer, and more plausible

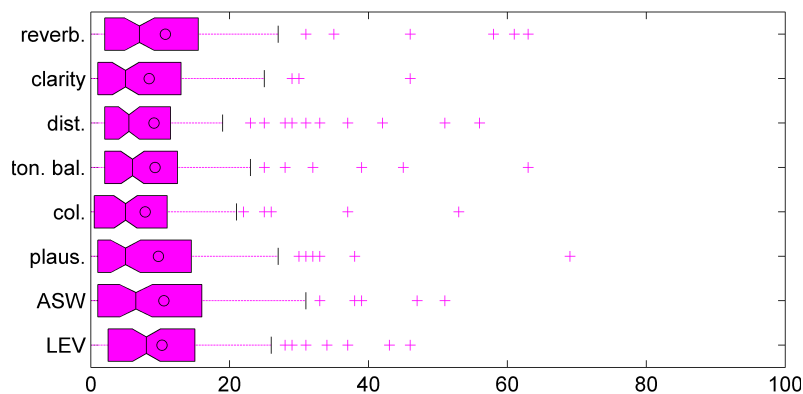


Figure 5.5: Absolute differences for repeated pair conditions. (See legend of Fig. 4.6 for boxplot notations).

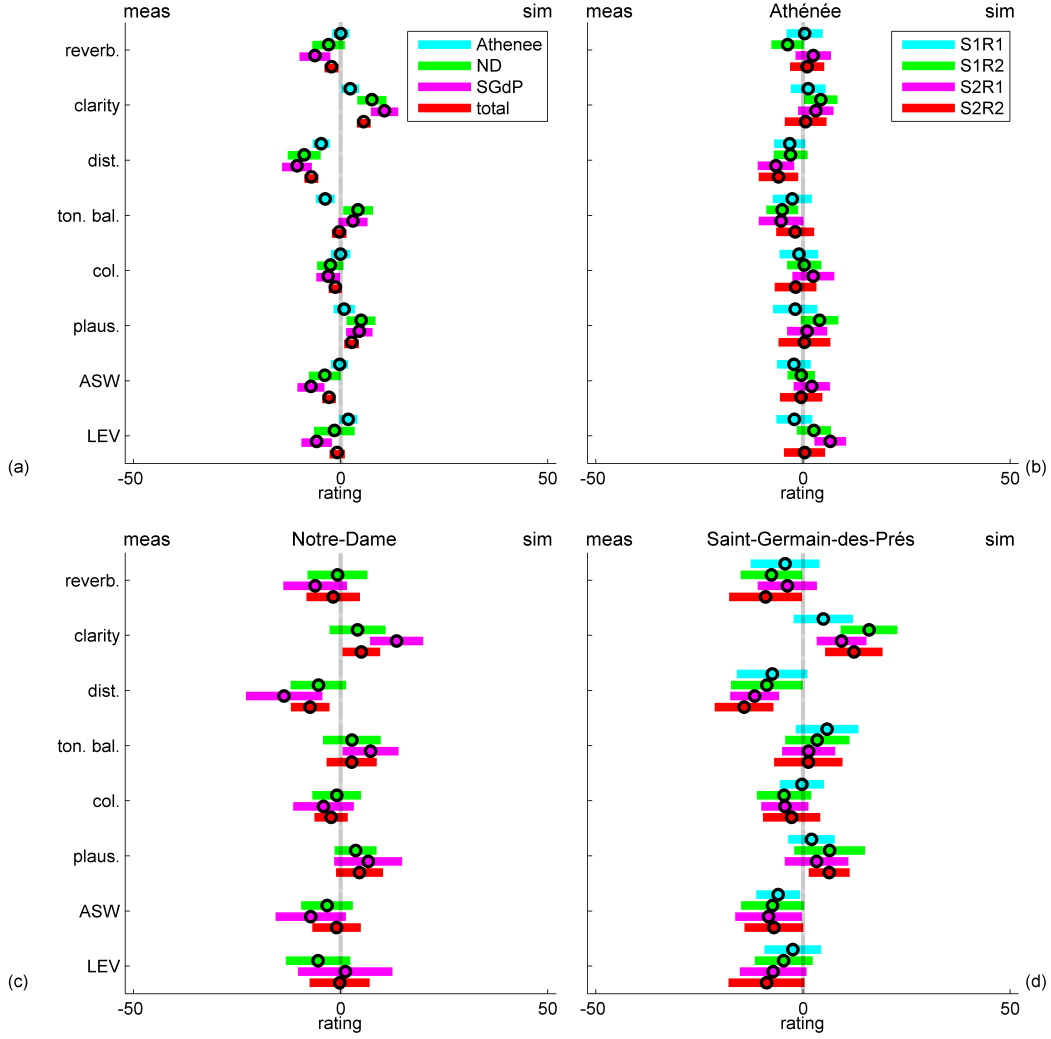


Figure 5.6: 99% CIs for all subjects on similarity of measured and simulated RIR auralizations. Measured on the left, Simulated on the right. (a) Combined results for Théâtre de l'Athénée, Notre-Dame, Saint-Germain-des-Prés, and overall results. Results by position for (b) Théâtre de l'Athénée, (c) Notre-Dame cathedral, and (d) Saint-Germain-des-Prés church. Center dashed line represents a neutral response, (o) indicates the mean value.

than their measured counterpart. Subsequently, results were analyzed for each individual space (see Fig. 5.6a). Simulated auralizations for the Théâtre de l'Athénée were judged clearer, closer, and less bright. Simulated auralizations for the Notre-Dame cathedral were perceived significantly clearer, closer, brighter, and more plausible. Simulated auralizations for the Saint-Germain-des-Prés church were judged significantly less reverberant, clearer, closer, more plausible, narrower, and less enveloping.

In order to study the grounds of these differences, results were examined by

Table 5.7: Mean, lower and upper limits of the 99% CIs for the comparison between measured and simulated auralizations for the Théâtre de l’Athénée, Notre-Dame, and Saint-Germain-des-Prés. (CI which do not contain 0 are *indicated* ).

		CI limits					CI limits					CI limits							
		pos.	low	mean	up		pos.	low	mean	up		pos.	low	mean	up				
Reverb.	Théâtre de l'Athénée	Total	-2.0	0.0	2.1	Notre-Dame cathedral	Total	-6.8	-2.9	1.0	Saint-Germain-des-Prés church	Total	-9.9	-6.2	-2.5				
		S1R1	-4.0	0.4	4.7		S1R1	-12.6	4.3	3.9		S1R1	-12.6	4.3	3.9				
		S1R2	-7.6	-3.7	0.2		S1R2	-8.0	-0.8	6.5		S1R2	-15.0	-7.6	-0.2	S1R2	-15.0	-7.6	-0.2
		S2R1	-1.8	2.5	6.7		S2R1	-13.8	-6.2	1.5		S2R1	-10.9	-3.8	3.3	S2R1	-10.9	-3.8	3.3
		S2R2	-3.1	1.0	5.1		S2R2	-8.2	-1.8	4.6		S2R2	-17.8	-9.0	-0.3	S2R2	-17.8	-9.0	-0.3
Clarity		Total	0.2	2.3	4.4		Total	4.1	7.5	11.0		Total	7.3	10.6	13.9				
		S1R1	-2.9	1.3	5.4		S1R1	-2.7	4.0	10.8		S1R1	-2.3	4.9	12.0				
		S1R2	0.2	4.3	8.3		S1R2	-2.7	4.0	10.8		S1R2	9.1	15.9	22.7				
		S2R1	-1.2	3.1	7.3		S2R1	7.1	13.6	19.9		S2R1	3.3	9.3	15.3				
		S2R2	-4.4	0.6	5.7		S2R2	0.5	7.4	9.5		S2R2	5.4	12.3	19.2				
Dist.		Total	-6.7	-4.7	-2.6		Total	-12.7	-8.8	-4.8		Total	-14.1	-10.5	-7.0				
		S1R1	-7.3	-3.2	0.5		S1R1	-12.0	-5.4	1.3		S1R1	-15.9	-7.4	1.1				
		S1R2	-7.1	-3.0	1.1		S1R2	-12.0	-5.4	1.3		S1R2	-17.4	-8.8	-0.1				
		S2R1	-10.9	-6.5	-2.2		S2R1	-22.8	-13.6	-4.4		S2R1	-17.5	-11.7	-5.8				
		S2R2	-10.7	-5.9	-1.2		S2R2	-12.0	-7.4	-2.7		S2R2	-21.3	-14.2	-7.1				
Ton. Bal.		Total	-6.0	-3.7	-1.5		Total	0.6	4.2	7.8		Total	-0.5	3.0	6.5				
		S1R1	-7.3	-2.6	2.1		S1R1	-4.2	2.7	9.7		S1R1	-1.8	5.8	13.4				
		S1R2	-8.9	-5.1	-1.3		S1R2	-4.2	2.7	9.7		S1R2	-4.3	3.4	11.2				
		S2R1	-10.7	-5.3	0.1		S2R1	0.5	7.2	14.0		S2R1	-5.0	1.4	7.7				
		S2R2	-6.5	-1.9	2.7		S2R2	-3.4	2.6	8.7		S2R2	-7.0	1.3	9.5				
Col.		Total	-2.3	0.0	2.3		Total	-5.6	-2.5	0.7		Total	-5.9	-3.0	-0.2				
	S1R1	-5.6	-1.0	3.6	S1R1	-6.8	-1.0	4.9	S1R1	-5.6	-0.3	5.0							
	S1R2	-3.8	0.3	4.4	S1R2	-6.8	-1.0	4.9	S1R2	-11.1	-4.6	1.9							
	S2R1	-2.6	2.5	7.5	S2R1	-11.4	-4.1	3.2	S2R1	-10.1	-4.4	1.3							
	S2R2	-6.8	-1.8	3.2	S2R2	-6.3	-2.3	1.7	S2R2	-9.6	-2.8	4.1							
Plaus.	Total	-1.7	0.9	3.4	Total	1.5	4.9	8.4	Total	1.3	4.5	7.7							
	S1R1	-7.2	-1.9	3.4	S1R1	-1.5	3.6	8.7	S1R1	-3.6	2.0	7.6							
	S1R2	-0.5	4.0	8.5	S1R2	-1.5	3.6	8.7	S1R2	-2.1	6.4	14.9							
	S2R1	-3.8	1.0	5.8	S2R1	-1.5	6.6	14.8	S2R1	-4.4	3.3	10.9							
	S2R2	-5.9	0.3	6.6	S2R2	-1.1	4.6	10.2	S2R2	1.4	6.3	11.2							
ASW	Total	-2.3	-0.3	1.8	Total	-7.6	-3.8	0.0	Total	-10.4	-7.2	-3.9							
	S1R1	-6.3	-2.2	1.8	S1R1	-9.5	-3.3	2.9	S1R1	-11.3	-6.0	-0.7							
	S1R2	-3.7	-0.4	2.9	S1R2	-9.5	-3.3	2.9	S1R2	-14.9	-7.4	0.2							
	S2R1	-2.2	2.1	6.5	S2R1	-15.7	-7.2	1.2	S2R1	-16.4	-8.4	-0.4							
	S2R2	-5.5	-0.5	4.6	S2R2	-6.8	-1.0	4.8	S2R2	-14.1	-7.0	0.0							
LEV	Total	-0.3	1.9	4.0	Total	-6.4	-1.5	3.4	Total	-9.4	-5.8	-2.1							
	S1R1	-6.4	-2.1	2.2	S1R1	-13.1	-5.4	2.3	S1R1	-9.3	-2.5	4.3							
	S1R2	-1.5	2.6	6.7	S1R2	-13.1	-5.4	2.3	S1R2	-11.6	-4.7	2.3							
	S2R1	2.8	6.6	10.4	S2R1	-10.2	1.1	12.5	S2R1	-15.2	-7.2	0.8							
	S2R2	-4.6	0.4	5.3	S2R2	-7.5	-0.2	7.0	S2R2	-17.9	-8.8	0.3							

room in detail. For the Théâtre de l’Athénée, simulated auralization **S1R2** was rated significantly clearer and brighter, simulated auralization **S2R1** was perceived closer and less enveloping, and simulated auralization **S2R2** was judged to be closer. Concerning the Notre-Dame cathedral, the simulated **S2R1** was rated clearer, closer, and brighter and the simulated auralization at **S2R1** was judged clearer and closer. Most differences were observed for the Saint-Germain-des-Prés. First, simulated auralization at **S1R1** was rated narrower. Second, simulated auralization at **S1R2** was perceived less reverberant, clearer, and closer. Additionally, simulated auralization at **S2R1** was judged clearer, closer, and narrower. Finally, simulated auralization

at **S2R2** was considered significantly less reverberant, clearer, closer, and more plausible. The remainder of the attributes were not rated significantly different.

In comparison to the preliminary listening test, the results of the Saint-Germain-des-Prés church have significantly improved for *Tonal Balance*<sup>5</sup>, due to the binaural dummyhead recording equalization. In contrast to the preliminary test results, simulated auralizations in the current test were found to be clearer and closer. These differences could be due to the added ability of participants to start auralization playback at different points in the extract, allowing them to focus on different elements, instead of having to start listening at the beginning of the auralizations each time. The remaining attribute results were comparable to the preliminary test results.

#### 5.4.3.3 Objective attributes

Listening test results were compared to the measured parameters presented in Sec. 5.2. Between measurement and simulation, only a single T20 instance was found to slightly exceed 1 JND (Théâtre de l'Athénée: **S2R2**), where three tested auralization pairs (Théâtre de l'Athénée: **S1R2**, Saint-Germain-des-Prés: **S1R2** and **S2R2**) *reverberance* were rated significantly different.

As auralizations with spoken voice were employed for the Théâtre de l'Athénée, C50 results are compared to the attributes *clarity* and *distance* [Thiele 1953]. C50 results indicated that simulated auralizations for positions **S1R1** and **S2R1** would be expected to be perceived less clear and further away. However, listening test results showed that the simulated auralization **S1R2** was perceived clearer and simulated auralizations **S2R1** and **S2R2** were perceived closer. It should be noted that the direct correspondence between perceived *Clarity* and the parameters C50 and C80 is not well defined, and that examining these parameters individually as a function of frequency may be overemphasizing their quality as a metric.

As singing stimuli were employed for the Notre-Dame cathedral and Saint-Germain-des-Prés church, C80 was compared to *clarity* and *distance* in these rooms [Reichardt 1975]. For the Notre-Dame results the simulated auralizations **S2R1** and **S2R2** were judged clearer and closer and with the simulated C80 being respectively 2.5 and 1.0 JND higher. Correspondence between listening test results and C80 for the Saint-Germain-des-Prés is good, with the simulated auralizations for the positions **S1R2**, **S2R1**, and **S2R2** having a perceptually higher C80 while also being judged clearer and closer in the listening test.

The correspondence between  $ASW$  and  $IACC_e$  is good. Simulated positions **S1R2** in the Notre-Dame and **S1R2**, **S2R1**, and **S2R2** in the Saint-Germain-des-Prés underestimated the measured  $IACC_l$ 's with more than 1 JND while **S1R2** and **S2R1** in the Saint-Germain-des-Prés were rated wider in the listening test. The correspondence between  $LEV$  and  $IACC_l$  is rather poor. Simulations underestimated the measured  $IACC_l$ 's with more than 1 JND for all positions in the three

<sup>5</sup>In the preliminary test defined as *Coloration*

rooms, however only the measured auralization of positions **S2R1** in the Théâtre de l’Athénée was rated more enveloping.

#### 5.4.3.4 Stimuli effect

Listening test results as a function of stimuli were analyzed employing the different anechoic stimuli in the Théâtre de l’Athénée. Fig. 5.7 shows that the majority of attributes were judged equally between the *French speaking* and *Italian speaking* stimuli, except *Clarity* and *Distance*. In order to establish the grounds for these differences a group-wise two-way ANOVA tested for anechoic type ( $F = 3.0$ ,  $p = 0.03$ ), position ( $F = 10.3$ ,  $p < 10^{-2}$ ), and Type $\times$ position ( $F = 0.8$ ,  $p = 0.47$ ). A subsequent  $t$ -test with ‘Bonferroni’ correction in order to compensate for multiple testing found significant for position **S2R2** regarding *Clarity* and *Distance*.

As *Clarity* and *Distance* relate to C50, this parameter’s results at position **S2R2** were compared between measurement and simulation in the  $1/3^{rd}$  octave bands where the two anechoic stimuli recordings differed more than 5 dB in spectral content, namely 200, 4000, 6300, 8000, 10000, and 12500 Hz (see Fig. 5.4). It was observed that C50 was overestimated in the simulation for these  $1/3^{rd}$  octave bands by an average of 1.4 dB with a SD of 0.9 dB. In these  $1/3^{rd}$  octave bands, the *Italian speaking* stimuli contained more energy with the early-to-late energy being overestimated in the simulation. It can be hypothesized that the combination of these factors resulted in the observed differences in *Clarity* and *Distance* judgments as a function of stimuli.

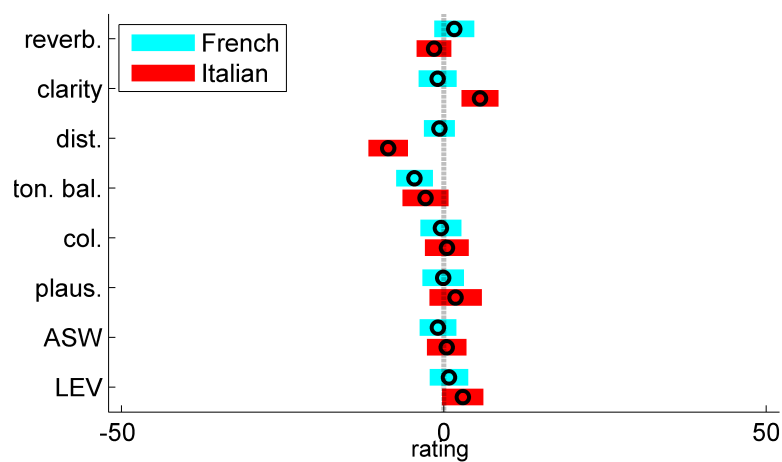


Figure 5.7: CIs for the Théâtre de l’Athénée separated by stimuli. Center dashed line represents a neutral response.

## 5.5 Discussion

This chapter presented the results of listening tests which compared simulated and measured binaural auralizations. These binaural auralizations were compared according to eight perceptual acoustic attributes commonly used in room acoustic evaluations. It should be noted that the tested positions in the Théâtre de l'Athénée and Saint-Germain-des-Prés church differed from the calibrated omni-directional combinations. The use of binaural auralization presents an additional extrapolation as the data used for calibration employed omni-directional microphones, not the measured BRIRs. Perceptual test results indicated small and mostly statistically insignificant differences between measurement and simulations for the attributes *Reverberance*, *Tonal Balance*, *Coloration*, *Plausibility*, *ASW*, and *LEV*. Larger differences were found for *Clarity* and *Distance* in particular for the Notre-Dame and Saint-Germain-des-Prés models. These differences could be associated to small variations in clarity parameter results between measurement and simulation RIRs which were unable to achieve the calibration goal of  $<1$  JND error for clarity while still adhering to other calibration procedure conditions. The geometrical approximations used for the model construction, specifically the means of simplified modelling of the seating, could be a source of common error in these GA models.

However, these significant differences were not observed in the preliminary listening test, where participants were required to listen to the entire extract from beginning when judging. We assume that the observed differences are due to the ability to carry out more detailed listening on specific sections of the extracts, rather than global ratings as was the case in the preliminary test. It is therefore recommended that for critical auralization listening tests, participants be able to select which elements of the extract they listen to, so that they can focus on key points where differences may be more evident.

Additionally, reverberation, clarity, and  $IACC_e$  parameters corresponded rather well to the perceptual attributes *Reverberance*, *Clarity*, and *ASW*. However, results showed limited significant differences for *LEV* while differences between measurement and GA simulations exceeded JND reference values for all  $IACC_l$ .

Furthermore, the influence of anechoic stimuli selection was examined in auralizations of the Théâtre de l'Athénée using two different talkers. Results indicated that significant differences in some attribute judgments existed between the anechoic stimuli, linked to fine differences in spectral content. Therefore, in order to achieve more generalizable results, it is recommended that listening tests employ multiple stimuli with varying frequency content.

Finally, parameter estimations and listening test results are compared to previous studies [Lokki 2001, Lokki 2002a, Choi 2006, Yang 2007]. Sec. 5.1 showed that using this calibration method, objective parameter estimation for the 3 considered GA models was better than previous studies. The binaural parameter estimations showed fewer positions exhibiting differences  $> 1$  JND and a smaller maximum difference than those reported in previous studies. In addition, these studies failed to control for several factors between measurement and simulation where the em-

ployed auralizations in the presented listening test did. These being source/receiver directivity, HRTF, compensation for the frequency response characteristics of the measurement system, and differences in SNR between measurement and simulation. Even though the main objective of these studies differed from creating perceptually equal auralizations, a combination of these aspects resulted in simulated auralizations which varied from the measured auralizations. This study found for the majority of the tested attributes no significant differences between the measured and simulated auralizations.

## 5.6 Summary

One can regard this chapter and the previous two as this thesis' study regarding the second aspect of *fully computed auralizations*, i.e. study of the sound propagation in a 3D space. This chapter presented two listening tests which compared measured to calibrated simulated auralizations. The first listening test which employed the Saint-Germain-des-Prés GA model found limited significant differences between monaural auralizations, however perceptual differences regarding *Coloration* and *ASW* differences were found between binaural auralizations. Therefore, an additional compensation regarding the mean spectral response of the binaural microphones was performed. The second listening test found only small differences between measured and simulated binaural auralizations of the Théâtre de l'Athénée, Notre-Dame, and Saint-Germain-des-Prés. With the second aspect of *fully computed auralizations* studied, the following chapter will study the first step, i.e. the 3D directivity of sound sources.





# Voice-directivity: incorporation into auralizations and evaluation<sup>1</sup>

---

## Contents

---

<b>6.1</b>	<b>Introduction</b>	<b>101</b>
<b>6.2</b>	<b>Source decomposition according to overlapping beam forming approach</b>	<b>104</b>
6.2.1	Employed auralizations	104
6.2.2	Protocol	105
6.2.3	Results	107
<b>6.3</b>	<b>Evaluation of dynamic voice directivity in auralizations</b>	<b>109</b>
6.3.1	Employed auralizations	111
6.3.2	Protocol	112
6.3.3	Results	113
<b>6.4</b>	<b>Discussion</b>	<b>115</b>
<b>6.5</b>	<b>Summary</b>	<b>116</b>

---

## 6.1 Introduction

With the 2<sup>nd</sup> aspect of *fully computed auralizations* (Sound propagation in a 3D space) sufficiently addressed in the previous chapters, attention was focused on the source excitation in the calibrated GA model. The aim of this chapter is to improve upon the 1<sup>st</sup> aspect of *fully computed auralizations*: 3D directivity of sound sources. The calibrated of the Théâtre de l'Athénée is employed for the following study.

It should be noted that sources in the previous chapter were of an omnidirectional nature. However, the source directivity of the human voice and instruments has been shown to be more complex [Meyer 1978, Meyer 1993, Katz 2006].

---

<sup>1</sup>This work was partly presented in:

- B.N.J. Postma, H. Demontis, and B.F.G Katz. *Subjective evaluation of dynamic voice directivity for auralizations* Acta Acustica united with Acustica, Vol. 103, pp. 181-184, Jan. 2017.
- B.N.J. Postma and B.F.G Katz. *Dynamic voice directivity in room acoustic auralizations* in German Annual Conf. on Acoustics (DAGA), pp. 352-355, Mar 2016.

These studies presented measurements taken in a spherical pattern around the source in the octave bands 125-4000 Hz.

Source directivity has been shown to impact measured RIRs and the consequent parameter estimations [Prince 1994, San Martin 2007, Knüttel 2013]. Additionally, studies [Dalenbäck 1993, Savioja 1999] have shown it to be an important factor to take into consideration when creating more realistic auralizations. This was confirmed by Wang et al. [Wang 2008] who found an influence of source directivity on simulated parameters and various listening tests [Giron 1996, Otondo 2004, Wang 2008] which found differences between source directivity types.

These studies were based on sources with a static nature. Otondo et al. [Otondo 2004] concluded that a static representation of sources was inadequate for describing the variations in source ordination due to varying radiation patterns and dynamic source orientation. For these reasons, studies [Rindel 2004, Otondo 2005] proposed to achieve the inclusion of dynamic directivity through the usage of multi-channel source directivity auralization<sup>2</sup>. This method employs anechoic multi-channel recordings. The radiation sphere source is divided into segments representing each microphone position. The RIR is then calculated for each segment and convolved with the corresponding microphone channel of the anechoic recording. Convolutions of each channel are then down-mixed to create a multi-channel source directivity auralization. This source representation follows changes in direction, movement, asymmetry, and orientation of the recorded source, unlike simple single channel source representations. Multi-channel source directivity auralizations were compared by listening tests to a static directivity source type. The GA software Odeon was employed to create auralizations of an anechoic clarinet recordings convolved with 2, 5, and 10 channels and a single channel with a static clarinet directivity. Fig. 6.1 depicts the multi-channel sources which were combined without overlap to represent the spherical recording area around the musician. A listening test compared these auralizations in terms of *perceived spaciousness of sound in the room* and *perceived naturalness of timbre of the clarinet*. Results of that study indicated that the 10-channel representation was judged significantly less spacious than the three other source representations. Additionally, the test subjects significantly preferred the 10-channel auralization over the other in terms of *perceived naturalness*.

Vigeant et al. [Vigeant 2011] compared 1-, 4-, and 13-channel source directivity auralizations by means of a listening test. The multi-channel source directivity representations and employed GA software were the same as the previously mentioned studies. The first phase of the test compared the different source representations for a violin, trombone, and flute in terms of *realism* and *source size*. Subjects rated the 13-channel auralization significantly more realistic than the other two. No significant trend was found regarding *source size*. In the second phase, the effect of orientation (facing the audience and facing 180° from the audience) of the 4-channel

<sup>2</sup>The original referenced paper termed this application multi-channel auralization. In order to avoid confusion with distributed sources multi-channel auralizations, this article will employ the term multi-channel source directivity auralization.









Type	1 Channel	2 Channels	5 Channels	10 Channels
Horizontal Plane				
Vertical Plane				

Figure 6.1: partial sources used for multi-channel source directivity auralizations (from [Rindel 2004]).

and 13-channel auralization on *Clarity* were studied. The results indicated that the 13-channel auralization was perceived clearer when the source faced the audience. No significant difference regarding *Clarity* was observed when the sources faced  $180^\circ$  from the audience.

In contrast to previous studies, the final goal of this chapter is to employ multi-channel source directivity for the inclusion of dynamic source directivity using a single channel anechoic recording. Advantages are a better representation of source directivity, simulations need to be run only once even when the selected instrument is adjusted, and source directivity can be adjusted post-simulation in real-time. A first step towards this goal is taken in Sec. 6.2, by perceptually examining the usage of a newly established source decomposition. Where previous studies employed segmented directivity approaches, here a multi-channel source decomposition using an overlapping beamforming approach was investigated. This source was placed in the calibrated model of the Théâtre de l'Athénée and resulting auralizations were compared to auralizations with static source directivities. As results of this preliminary listening test indicated that the inclusion of phoneme dependent directivity and dynamic orientation leads to perceptibly different auralizations, Sec 6.3 explores the final implementation. A listening test compared the resulting auralization with dynamic voice directivity to auralizations with static voice directivity and omnidirectional source ordination. All of these auralizations employed single channel anechoic recordings. Sec. 6.4 compares the listening test's results to previous studies.

## 6.2 Source decomposition according to overlapping beam forming approach

### 6.2.1 Employed auralizations

First, vocal recordings were performed in an anechoic chamber using 20 microphones geometrically positioned at the vertices of a dodecahedron [Lokki 2008a, Pätynen 2011]. The singer's mouth was situated at the center of the array. The selected extract for this study was a female soprano singing '*Abendempfindung*', by W.A. Mozart.

Based on the microphone configuration of this anechoic recording a beam pattern was established with slightly overlapping segments. The beams were designed to have minimal overlap while having an equal gain sum for all sections in order to approximate an omni-directional pattern. The following control points were employed:

- $0^\circ$  - No attenuation
- $21^\circ$  - (center of the rib) to sum 2 beams to 0 dB
- $42^\circ$  - (center of the pentagon) to sum 5 beams to 0 dB
- $90^\circ$  - Maximum attenuation

The 2D beam was produced (see Fig. 6.2) using a spline interpolation in  $5^\circ$  steps. It was rotated around its symmetry axis to create the 3D beam. The 20 instances of the beam pattern were aimed at one of the 20 microphone positions of the anechoic recordings. The result of the summation produces an omni-directional sphere, with a variation in the directivity pattern of  $\pm 0.2$  dB.

The established 20 sources were positioned in a GA model of the Théâtre de l'Athénée. Simulations were run with 400,000 rays using Algorithm 2. As a baseline

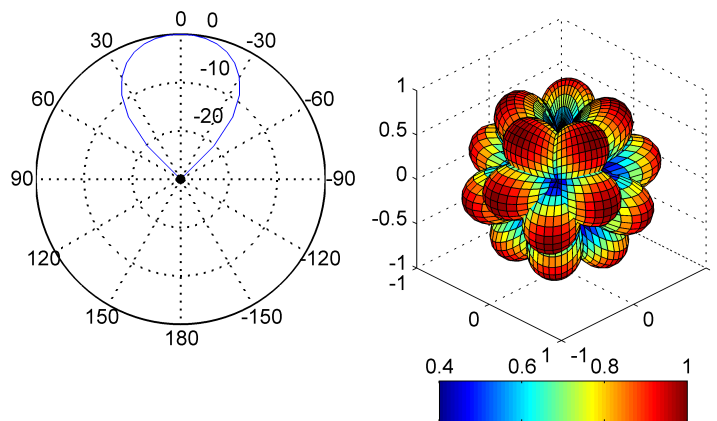


Figure 6.2: (left): 2D polar plot of a single beam (dB scale), and (right): superposition of the 20 3D beam patterns (linear scale) to show orientations.

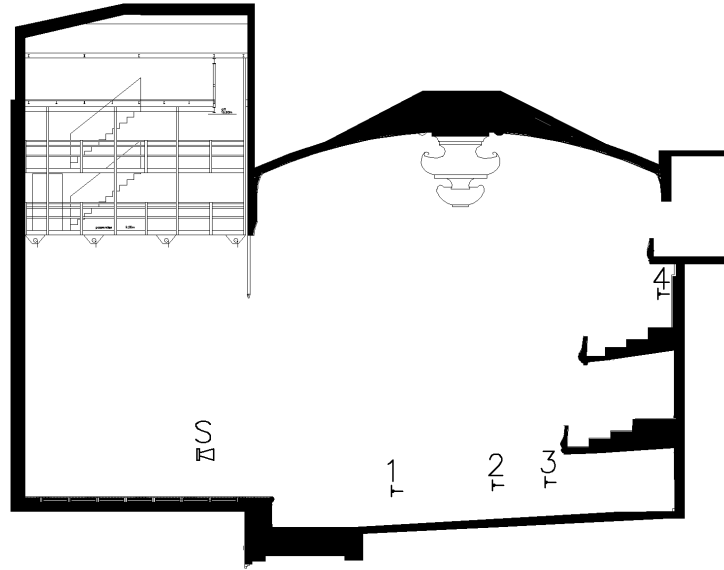


Figure 6.3: Section of the Théâtre de l'Athénée depicting the source position on stage and the 4 receiver positions (1, 2, 3 first floor, and 4: third floor).

validation of the multi-channel source directivity auralization, the mix of the channels RIRs (reconstructed RIR) should be perceptibly equal to the RIR resulting from a simulation with an omni-directional source (omni-directional RIR). Therefore, in addition to the designed multi-channel source, simulations were performed with an omni-directional source. In order to be able to compare the multi-channel source directivity auralization with static sources, simulations were also carried out using sources with static singer [Marshall 1985], and static loudspeaker [Choueiri 2010] directivities.

All sources were positioned on the center of the stage. Four binaural receiver positions were simulated on the center axis of the theater at various positions (see Fig. 6.3). Post-simulation, the reconstructed, omni-directional, static singer, and static loudspeaker RIRs were convolved with a single-channel recording of the selected extract. Finally, the 20 channel beam RIRs were convolved with the corresponding 20 channels of the anechoic recording and summed. This resulted in five binaural auralizations per receiver position. RMS of the binaural auralizations was used for normalization.

### 6.2.2 Protocol

The resulting auralizations were compared by means of a listening test. The test was setup as a randomized experiment with five variants corresponding to the source directivity-types. Binaural auralizations were compared per receiver position 1, 2,

3, and 4. Additionally, one iteration was repeated in order to monitor the reliability of the test's responses (receiver position 3). Participants were initially presented one training iteration with the test administrator present in order to ensure they understood the task (receiver position 2), resulting in six iterations. The training session results were not tabulated in the presented results.

Participants were given written instructions before commencing the listening test<sup>3</sup>. For each iteration, participants compared and rated the five auralizations in terms of *Plausibility*, *Clarity*, *Distance*, *ASW*, and *LEV*<sup>4</sup> on a discrete scale ranging from 1 ('least ...') to 7 ('most ...'). A discrete scale was used to reduce the test's duration. Participants were forced to use the 2 extreme scale values at least once per attribute. They were allowed to give auralizations the same rating. Presentation order of the receiver position and correspondence to source directivity-type were randomized. This protocol is similar to [Vigeant 2011], which employed similar acoustic attributes and a 7 point scale. However, the current study employed two additional acoustical attributes, the auralizations were compared during the same iteration, and participants were forced to use the extreme scale values.

28 participants (mean age: 35.3 SD: 12.6) who all reported normal hearing in an audiogram took part in the study. They were selected to have experience with either room acoustics or vocal/instrumental performances as it was hypothesized that experienced listeners would perform significantly better than untrained listeners [Olive 2003]. 15 participants took the test in a sound isolation booth at LIMSI (ambient noise level < 30 dBA), 10 in an isolation booth at the *Institut Jean le Rond d'Alembert* (LAM) (ambient noise level < 30 dBA), and 3 participants in a quiet office at the *Institut National d'Histoire de l'Art* (THALIM) (ambient noise level = 31 dBA). Participants were given written instructions before commencing the test which explained the task, described the attribute definitions, and illustrated the software usage. Participants were able to select the starting play point in the auralizations and to listen to the auralizations as many times as desired. Auralizations were presented via headphones (Sennheiser model HD 650) at an RMS level of 80 dBA.

Table 6.1: Mean absolute difference, corresponding SD and PCC between repeated conditions.

	Plausibility	Clarity	Distance	ASW	LEV
mean abs dif.	1.9	2.1	1.4	1.7	1.5
SD	2.7	2.9	2.2	2.5	2.3
PCC	0.4	0.1	0.2	0.3	0.2

### 6.2.3 Results

Initial attention is given to the reliability of the responses, determined from the mean absolute difference between the repeated configurations and its SD (see Table 6.1 and Fig. 6.4). The absolute mean difference ranged from 1.4 to 2.1 and corresponding standard deviation ranged from 2.2 to 2.9 on a 7 point scale. A metric for judging the reliability of responses is Pearson’s Correlation Coefficient (PCC), defined as the covariance of the two sets of responses divided by the product of their standard deviations. It is interesting to note that for *Distance* the absolute difference is lowest with lowest corresponding SD, in contrast the PCC is also low for this attribute. As observed in Chapter 5, these results raise the question of the suitability of the PCC in the current context.

As 5 attributes were tested according to Equations 5.1 and 5.2, Fig. 6.5 presents the 99% CIs for the combined results per acoustic attribute per directivity type. First, a group-wise two-way ANOVA ( $p$  – value = 0.05) testing for the effects of directivity type ( $F = 9.6$ ,  $p < 10^2$ ), position ( $F = 0.4$ ,  $p = 0.73$ ), and directivity type  $\times$  position ( $F = 2.0$ ,  $p = 0.02$ ) was employed. As these results indicated significant results a pair-wise  $t$ -test with ‘Bonferroni’ correction in order to compensate for multiple testing was employed to estimate whether an acoustic attribute was rated significantly different between auralization types (see Table 6.2).

In order to validate the multi-channel auralization the omni-directional and reconstructed omni-directional auralization are compared first. The omni-directional auralization at **SR1** was perceived closer and at **SR2** narrower than the reconstructed omni-directional auralizations. The limited significant differences indicate that auralizations with omni-directional and reconstructed omni-directional source directivity were perceived equal.

Subsequently, the multi-channel directivity source auralizations are compared to the static loudspeaker and singer auralizations. These comparisons resulted in more differences. For *SR1* the static singer was perceived closer, narrower, and less en-

<sup>3</sup>Appendix C presents the instructions provided to the participants

<sup>4</sup>See Sec. 5.3 for descriptions provided to the participants.

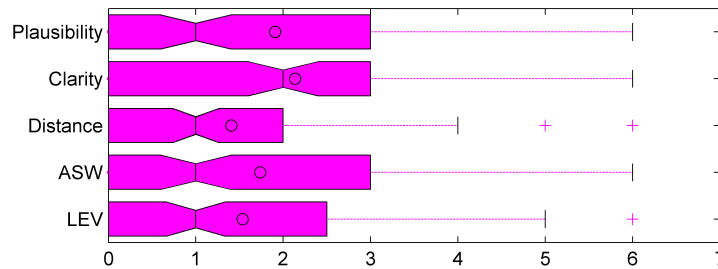


Figure 6.4: Absolute differences for repeated conditions. (See legend of Fig. 4.6 for boxplot notations).



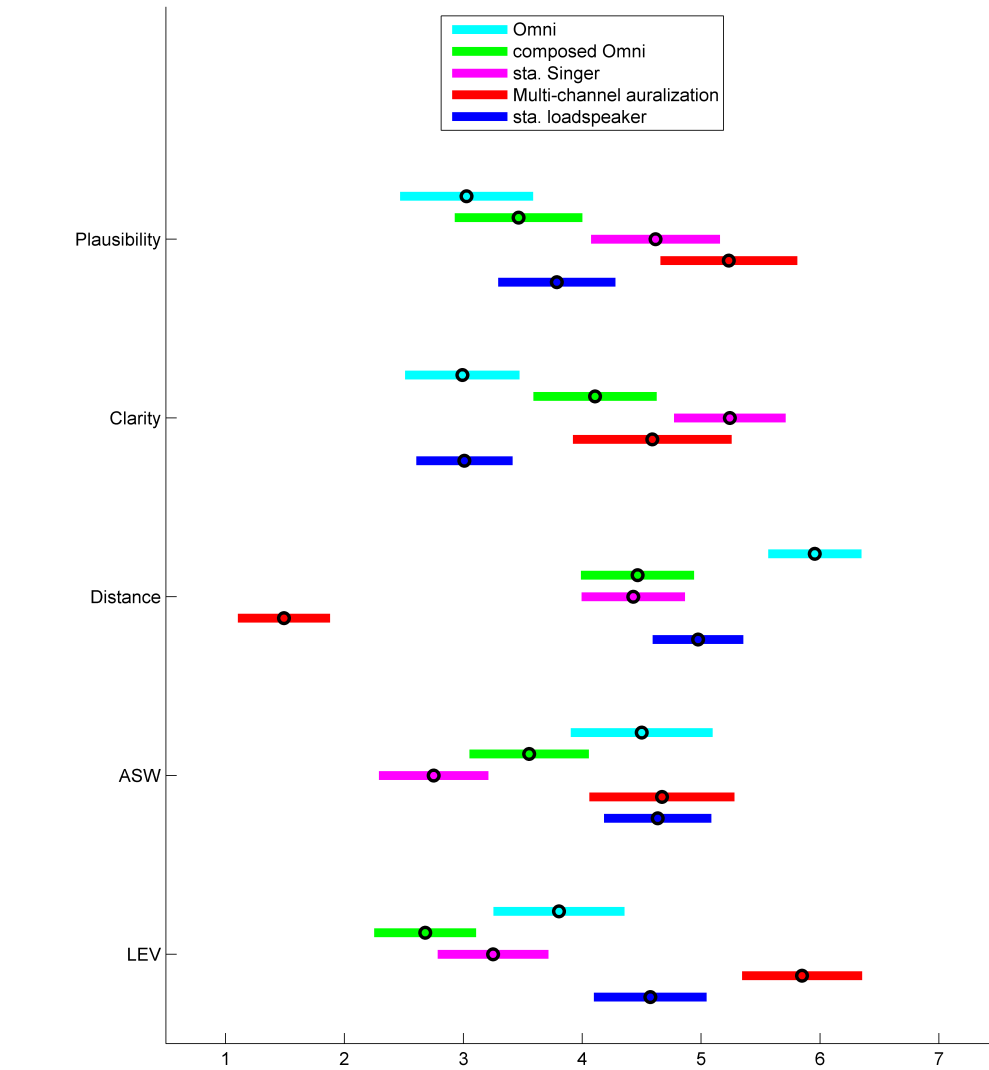


Figure 6.5: 99% confidence intervals of **Total** results for the (Cyan) omni-directional, (Green) recomposed omni-directional, (Magenta) static singer directivity, (Red) multi-channel singer directivity, and (Blue) static loudspeaker directivity.

veloping and the loudspeaker further. At *SR2* the static singer and loudspeaker were perceived further away. Additionally, at *SR3* loudspeaker directivity auralization was perceived less plausible, further away, and less enveloping and the loudspeaker further away and less enveloping. Finally, for *SR3* the static singer was rated further away, narrower, and less enveloping and the loudspeaker was considered further away.

Table 6.2: *T*-test results comparing the omni-directional to the reconstructed omni-directional auralizations (rec. omni) as well as the static loudspeaker (sta. loud.) or static singer (sta. sing.) to the multi-channel auralization (multi-chan.) (a ✓ indicates a significant difference).

	Acoustic Attributes	Omni- rec. Omni	sta. loud – multi-chan.	sta. sing – multi-chan.
SR1	Plaus.			
	Clarity			
	Dist.	✓	✓	✓
	ASW			✓
	LEV			✓
SR2	Plaus.			
	Clarity			
	Dist.		✓	✓
	ASW			
	LEV	✓		
SR3	Plaus.		✓	
	Clarity			
	Dist.		✓	✓
	ASW			
	LEV		✓	✓
SR4	Plaus.			
	Clarity			
	Dist.		✓	✓
	ASW			✓
	LEV			✓

### 6.3 Evaluation of dynamic voice directivity in auralizations

The results of the previous listening test showed perceptual differences between the presented multi-channel source directivity application and static singer source auralizations in terms of perceived *Distance*, *ASW*, and *LEV*. Therefore, it was concluded that the inclusion of phoneme dependent directivity and dynamic orientation leads to perceptibly different auralizations than those based on static directivity source types. This notion justified further endeavors towards creating auralizations with dynamic source directivity employing single channel anechoic recordings.

Initial attention was focused on studying the variation in phoneme dependent radiation. For this purpose, recordings were carried out in an anechoic chamber using 20 microphones geometrically positioned at the vertices of a dodecahedron (same measurement setup as presented in [Lokki 2008a]). A speaker's mouth was positioned at the center of the array. This speaker uttered the 15 most commonly

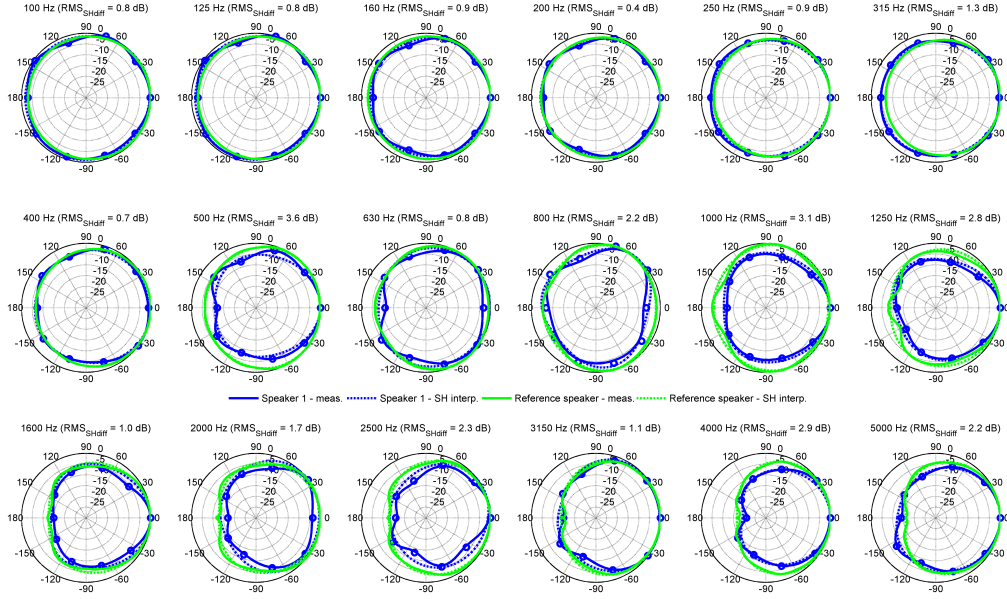


Figure 6.6: Measured and spherical harmonic decomposed  $1/3^{rd}$  octave band filtered directivity pattern of a normally speaking person [Chu 2002] and phoneme *a*.

used phonemes in the French language. Each phoneme was filtered in  $1/3^{rd}$  octave bands. A decomposition was made into the  $1/3^{rd}$  order spherical harmonics domain. Fig. 6.6 compares the directivity pattern of a typical phoneme to the directivity pattern of a speaking person [Chu 2002]. As differences were rather minimal, it was decided to omit the phoneme dependent radiation from further analysis.

Subsequent study focused on the inclusion of dynamic voice directivity into auralizations. A baseline requirement for this implementation was that the convolutions could be performed in real-time on a single computer. Therefore, the number of spatially decomposed source beams had to be reduced. Instead of the beams focusing at the vertices of the dodecahedron, they were aimed at the faces, resulting in 12 beams (instead of 20). In correspondence with the previous beam-forming pattern, the beams were designed to have minimal overlap while also resulting in an equal gain sum for all directions in order to approximate an omni-directional pattern. The following control points were employed:

- $0^\circ$  No attenuation
- $42^\circ$  (dodecahedron vertices) 3 beams sum to 0 dB
- $90^\circ$  Maximum attenuation

The 2D beam was generated using a spline interpolation ( $5^\circ$  steps, see Fig. 6.7) rotated around its symmetry axis to create the 3D beam. The result of the equal-weighted summation accurately reproduced an omni-directional sphere ( $\pm 0.3$  dB, see Fig. 6.7).

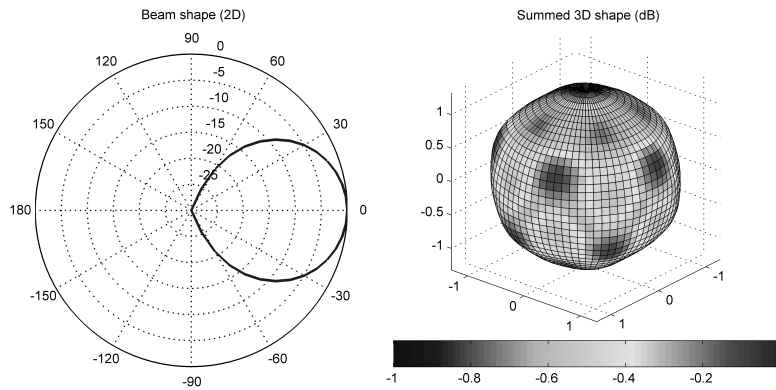


Figure 6.7: (Left) 2D polar plot of a single beam (dB scale). (Right) Summed directivity response (dB) of the 12 beams for a reconstructed omni-directional source (spatial dodecahedron configuration of beams).

The final implementation required the Théâtre de l’Athénée GA model in performance conditions rather than the empty stage used for the calibration measurements. For this reason, curtains were modeled in front of the back and side walls of the stage and the chairs were simulated as occupied. As a reference for the new calibration T20 values from measurements by [Polack 2012] were used. Subsequently, the established 12 beam source was positioned on the center of the stage and 3 receiver positions in the audience on the first floor were simulated (see Fig. 6.8a). Simulations were run with 200,000 rays using Algorithm 3.

### 6.3.1 Employed auralizations

To provide an ‘anechoic’ sound file with natural corresponding orientation information, a performance with two actors (“Ubu Roi” by Alfred Jarry) was recorded with both head-worn mics and a Kinect 2 RGB/Depth sensor in a small theater (Théâtre de la Reine Blanche, Paris). The close-mic recordings were employed in analogy with anechoic recordings. The coupled pair of RGB/Depth videos were used to generate a 3D point-cloud rendering of the actors. A 17 s extract of this performance was selected which highlighted variations in actors’ head orientation and movement during the scene. Actors’ head orientation within the point-cloud rendering was tracked with a specifically designed *Blender* module.

Voice directivity variations were controlled via gain modulations of the 12-beam components. Frequency dependent gain weighting factors were determined based on measured voice directivity data [Chu 2002]. This data was converted to octave band values and then encoded using 3<sup>rd</sup> order spherical harmonic decomposition to facilitate interpolation. A real-time *Max/MSP* patch was designed which modulated the gains of the 12-beam RIR contributions to create the desired directivity. First the ‘anechoic’ recording was convolved with the 12-beam source RIRs. These convolved channels were then filtered in octave bands 125–4000 Hz. Gain weighting factors

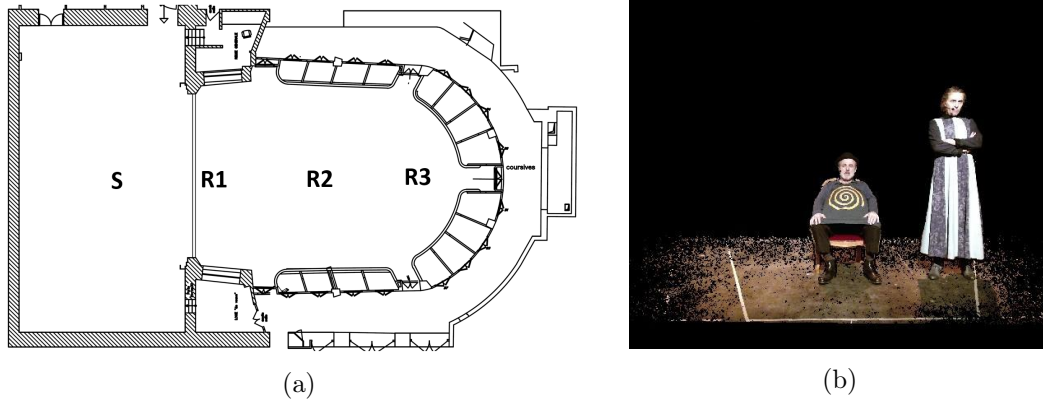


Figure 6.8: (a) Plan of the Théâtre de l'Athénée depicting the source position on stage and the 3 receiver positions. (b) Picture of the RGB video shown to serve as a reference for the acoustical attribute *Plausibility*.

were calculated by first applying any source rotation in the spherical harmonic domain. The octave-band rotated directivity data was then decoded to the 12-beam directions, resulting in a set of weighting factors which were applied to the convolved filtered channels which were subsequently summed. For an *omni-directional* source, all weightings factors were set to unity. For the *static voice* directivity auralization, the source orientation was fixed towards the audience. For *dynamic voice* directivity, source rotation used the actors' head orientation. These convolutions were decoded from 2<sup>nd</sup> order Ambisonic to binaural, employing *Spat*'s<sup>5</sup> virtual speaker array approach (18 speakers uniformly distributed defined) (see e.g. [Noisternig 2003b]) and the provided KEMAR dummy head HRTF, to produce the final binaural auralizations. The different auralizations were then recorded and normalized in RMS level.

### 6.3.2 Protocol

The resulting auralizations were compared by means of a listening test which resembled the preliminary test. The test had 3 variants corresponding to the source directivity-types. Binaural auralizations were compared for 3 receiver positions (see Fig. 6.8a). Every trial was repeated 3 times, resulting in (3×3) 9 trials. Participants were initially presented one training trial (**SR2**) with the test administrator present in order to ensure the task was understood. Results for this trial were not tabulated in the presented results.

Participants were given written instructions before commencing the listening test<sup>6</sup>. For each trial, participants compared and rated the 3 auralizations in terms of *Plausibility* (*Plaus.*), *Distance* (*Dist.*), *ASW*, and *LEV*<sup>7</sup> on a discrete scale ranging

<sup>5</sup><http://forumnet.ircam.fr/product/spat-en/>

<sup>6</sup>Appendix C presents the instructions provided to the participants

<sup>7</sup>See Sec. 5.3 for descriptions provided to the participants.

from 1 ('least ...') to 7 ('most ...'). Participants were forced to use the 2 extreme scale values at least once per attribute, but were allowed to give 2 auralizations the same rating. Presentation order of the receiver position and correspondence to source directivity-type were randomized.

21 participants (mean age: 36.2 SD: 11.5) took part in the study. Before commencing the listening test, an audiogram was performed in order to confirm the quality of the participant's hearing. All participants had experience with either room acoustics or vocal/instrumental performances. The test was carried out in a sound isolation booth at *l'Institut Jean le Rond d'Alembert* (ambient noise level  $\approx 32$  dBA). Participants were given written instructions before commencing the test which explained the task, described the attribute definitions, and illustrated the test interface. Additionally, before the test commenced, participants were shown the silent RGB video in order to provide them with a contextual reference for the attribute *Plausibility*. A video was selected over a still image in order to put the subjects in mind of a real play, with actor dynamics included. Participants were able to listen to the auralizations as many times as desired and were able to select the starting play point in the auralizations. Auralizations were presented via headphones (Sennheiser model HD 650) at an RMS level of 80 dBA.

### 6.3.3 Results

Initial attention is given to the reliability of the responses, determined from the mean absolute difference between the repeated configurations and its SD (see Table 6.1). The absolute mean difference ranged from 1.7 to 2.2 and corresponding standard deviation ranged from 2.1 to 2.3 on a 7 point scale. A metric for judging the reliability of responses is Pearson's Correlation Coefficient (PCC), defined as the covariance of the two sets of responses divided by the product of their standard deviations. It is interesting to note that for *LEV* the absolute difference is low with lowest corresponding SD, in contrast its PCC which is also low for this attribute.

Fig. 6.9 presents the combined results and results per position. As 4 attributes were tested, according to equations 5.1 and 5.2 98.8% CIs were established in order to compensate for multiple testing. A group-wise two-way repeated measures ANOVA ( $\alpha = 0.05$  level) was conducted regarding position ( $F = 1.1$ ,  $p = 0.34$ ), directivity type ( $F = 13.2$ ,  $p < 10^{-2}$ ), and position $\times$ directivity type ( $F = 2.0$ ,  $p = 0.11$ ). As these results indicated significant differences a subsequent *t*-test with 'Bonferroni' correction was employed in order to estimate whether an acoustic attribute was

Table 6.3: Mean absolute difference, corresponding SD and PCC between repeated conditions.

	Plausibility	Distance	ASW	LEV
mean abs dif.	2.2	1.7	2.0	1.8
SD	2.3	2.1	2.2	2.1
PCC	0.3	0.5	0.2	0.2

rated significantly different (see Table 6.4). Combined results of the dynamic voice auralizations were compared to the static voice and omni-directional auralizations. Dynamic voice auralizations were perceived significantly more plausible, more enveloping, and exhibiting a wider source width than the omni-directional and static voice auralizations. Additionally, dynamic voice auralizations were perceived closer than the omni-directional auralizations.

Results of the dynamic voice auralizations were compared to the static voice and omni-directional auralizations per position (see Table 6.4 and Fig. 6.9b-d). For all 3 positions, the dynamic voice sources were considered significantly closer, wider, and the auralizations more enveloping than the omni-directional auralizations. Additionally, for **SR2** and **SR3** the dynamic voice auralizations were perceived as being more plausible than the omni-directional auralizations. The dynamic voice at **SR2** was judged to more enveloping than the static voice. For **SR3**, the dynamic voice was considered wider and the auralization more plausible and more enveloping than

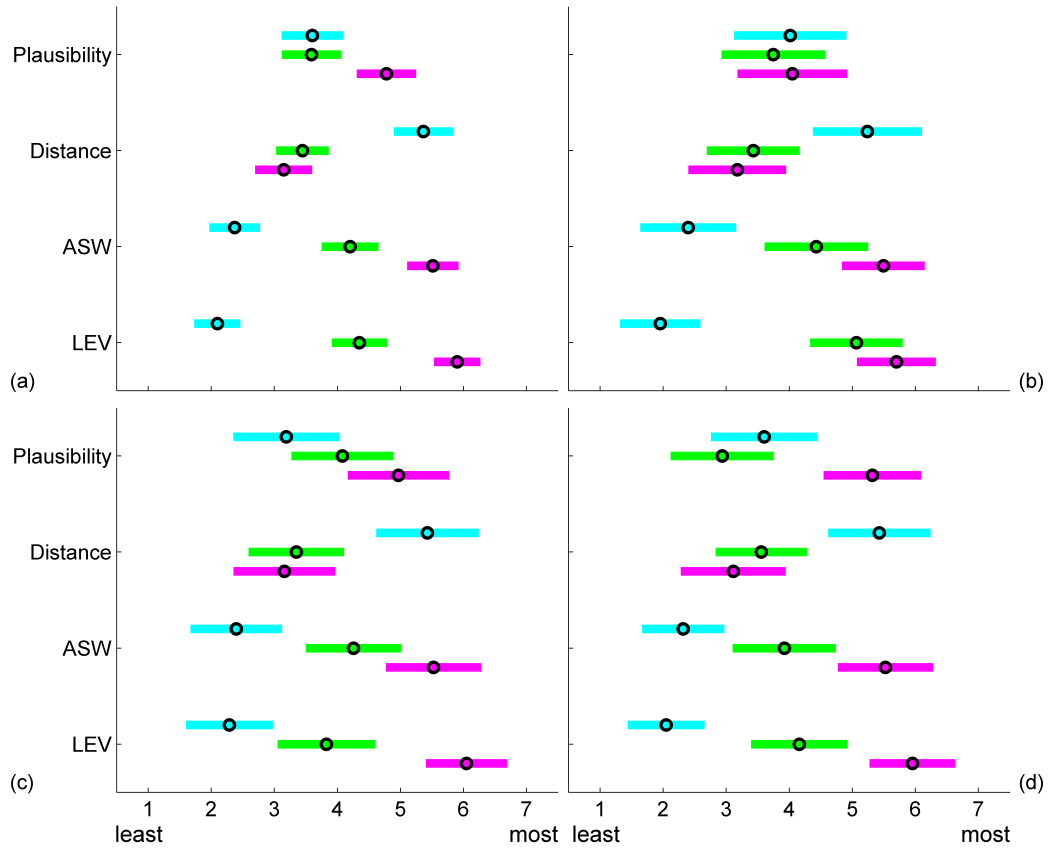


Figure 6.9: 98.8% confidence intervals of (a) **Total** results and individual positions (b) **SR1**, (c) **SR2**, and (d) **SR3**. (Upper) omni-directional, (Middle) static voice, and (Lower) dynamic voice directivity. (Cyan) omni-directional, (Green) static voice, and (Magenta) dynamic voice directivity.

Table 6.4:  $T$ -test results comparing the dynamic voice (dyn.) to the omni-directional (omni.) and static voice (sta.) directivity (a ✓ indicates a significant difference).

	Acoustic Attributes	$t$ -test	
		omni- dyn.	sta.- dyn.
Total	Plaus.	✓	✓
	Dist.	✓	
	ASW	✓	✓
	LEV	✓	✓
SR1	Plaus.		
	Dist.	✓	
	ASW	✓	
	LEV	✓	
SR2	Plaus.	✓	
	Dist.	✓	
	ASW	✓	
	LEV	✓	✓
SR3	Plaus.	✓	✓
	Dist.	✓	
	ASW	✓	✓
	LEV	✓	✓

the static voice auralization.

## 6.4 Discussion

These results indicate that auralizations with dynamic voice directivity orientation are perceptually more plausible, wider, and more enveloping than auralizations with an omni-directional source or static voice directivity. Comparing these results to previous studies [Rindel 2004, Otondo 2005, Vigeant 2011], it should be noted that those studies employed multi-channel source directivity auralization whereas this study used a single channel approach. In [Rindel 2004, Otondo 2005] dynamic source auralizations were judged more *natural* and in [Vigeant 2011] more *realistic*; this study confirms these findings where these were judged more *plausible*. In [Rindel 2004, Otondo 2005] dynamic source auralizations were judged less *spacious* and [Vigeant 2011] did not observe a trend in terms of *source size*; in contrast, this study found that the dynamic voice directivity was considered wider and the auralizations more enveloping. Possible explanations for these differences are slight differences between attribute definitions or test protocol and variation between anechoic recordings. This study employed a scene which highlighted the variation in the actors' head orientation, while no additional information was given about source orientation in anechoic recordings employed in other studies.



## 6.5 Summary

The first aspect of *fully computed auralizations* is the 3D directivity of sound sources. This chapter presented a manner to include dynamic voice directivity into auralizations. As it was concluded that including phoneme dependent radiation did not perceptually affect architectural auralizations, only dynamic voice directivity was simulated and not phoneme dependent radiation. The resulting dynamic voice directivity auralizations were compared to auralizations with static source ordinations in a listening test. Employing the calibrated GA model of the Théâtre de l'Athénée, these listening tests were performed with realistic acoustic conditions. It was found that auralizations which included dynamic voice directivity were perceived more plausible and more enveloping as well as exhibiting a wider source width than auralizations with static source ordinations. The next chapter studies the influence of including visuals with the acoustical experience of auralizations. Having established the quality sound propagation in the 3D space and shown that the inclusion of dynamic voice directivity leads to more plausible auralizations, this can be carried out having confidence that the results also apply to real-life conditions.

# Multi-modal auralizations and subjective evaluation<sup>1 2</sup>

---

## Contents

---

<b>7.1 Introduction</b>	<b>117</b>
<b>7.2 Framework for multi-modal auralization</b>	<b>120</b>
7.2.1 Room acoustic rendering	120
7.2.2 Visual room rendering	120
7.2.3 Dynamic performance recording, point-cloud rendering in the VR world	123
<b>7.3 Coherent visual-aural experiment</b>	<b>123</b>
7.3.1 Protocol	123
7.3.2 Results	125
<b>7.4 Incoherent visual-aural experiment</b>	<b>129</b>
7.4.1 Protocol	129
7.4.2 Results	131
<b>7.5 Discussion</b>	<b>137</b>
<b>7.6 Summary</b>	<b>138</b>

---

## 7.1 Introduction

The increased plausibility of the sound source and the proven ecologically valid sound propagation in the 3D space achieved in the previous chapters have improved the perceptually validity of auralizations. With this increased validity it is possible to study room acoustic experience employing virtual reality, having confidence that

---

<sup>1</sup>This work was partly presented in:

- D. Poirier-Quinot, B.N.J. Postma, and B.F.G. Katz, *Augmented auralization: Complimenting auralizations with immersive virtual reality technologies*, in Intl. Sym on Music and Room Acoustics (ISMRA), (La Plata), pp. 1-10, Sept 2016.

<sup>2</sup>Appendix B details a related study which investigated the current CPU/GPU limitations regarding multi-modal auralizations with an ambitious project which auralized and visualized a complicated scene.

the results can also be applied to real-life situations. These real-life situations can entail concerts, opera performances, and theater plays.

Experiments in this thesis so far have been carried out in solely aural conditions. However, a basic assumption of multi-modal perception is that one perceives complex perceptual information in everyday-life [Gibson 1979]. According to this vision, complicated perception relies on convoluted environmental information rather than uni-modal sensations. As in theaters, opera halls, and concert halls both visual and acoustical information is conveyed to the audience, this chapter studies the impact of visuals on acoustical experience. Multi-sensory research has shown influence of visuals on acoustic perception [Howard 1966, McGurk 1976]. However, there are few studies which have studied the influence of visuals on room acoustical experience.

Work by Barron [Barron 1988] on British concert halls indicated visual influences on the acoustical experience. He analyzed the combination of objective acoustic measurements with the results of questionnaires after live concert conditions in the same rooms. *Intimacy* and *Loudness* ratings were found to remain constant while moving away from the source, even when the actual sound level decreased. Barron hypothesized that listeners may be compensating for the distance and making loudness judgements relative to expectations.

Larsson et al. [Larsson 2001] studied the influence of visuals on the acoustic experience by means of VR. They tested the hypothesis that varying degrees of visual realism would affect judgments of aural room qualities. The visual stimuli were virtual (photographs or VR-models) or real concert halls, theaters, and practice rooms in Musikhögskolan in Gothenburg, Sweden. The auditory stimuli were auralizations of these rooms created with CATT-Acoustic. In a listening test, 80 participants were assigned to one of four conditions:

1. Auralizations only (Sound condition)
2. Auralizations while viewing still pictures of the room (Picture condition)
3. Auralizations while participants navigated in a virtual model of the room (VR condition)
4. Auralizations while participants were in the actual rooms (Real condition)

Participants rated the sound file on ASW, aurally perceived room size, and aurally perceived distance to sound source. Results indicated that both the VR and real conditions were considered significantly wider than the auralization only and auralization combined with picture conditions. Additionally, perceived room size and distance to source were significantly smaller and shorter for the real condition.

In a comparable study, Maempel et al. [Maempel 2013a, Maempel 2013b] studied to what extent the perceived room size and egocentric source distance were experimentally influenced by acoustic and visual stimuli. In order to acquire auralizations with corresponding visualizations, binaural auralizations simulations were performed and stereoscopic images were taken of four rooms. In a listening test, 35 participants were asked to assess among other attributes, source distance and room size in acoustic, visual, and visual-acoustic conditions. Auditory perceived distance was perceived furthest, visually perceived distance closest, and visual-acoustic

Table 7.1: Studied acoustical attribute and visual stimuli per study. A ‘✓’ indicates that the acoustical attribute was studied and the identified visual stimuli was employed in the study.

Acoustical Attribute	Barron	Larson	Maempel	Menzel	Jeon
Source distance		✓	✓		
ASW		✓			
Room size		✓	✓		
Loudness	✓			✓	
Seat preference					✓
Visual stimulus					
Still photo		✓	✓	✓	✓
Visual VR model		✓			
Real Hall	✓	✓			
Visible source?	✓			✓	

conditions in between. Additionally, it was concluded that optical room size had a greater perceptual weight than acoustic room size.

That visual cues can also affect the loudness judgement of auralizations was shown by Menzel et al. [Menzel 2008]. As in previous studies, they performed a listening test which combined auralizations with a visual stimuli. The visual stimuli were still photographs of the same sports car in four different colors. 4 s long auralizations of an accelerating sports car at four different levels of loudness was employed as the aural cue. 16 participants judged auralizations while the red sports car was shown to be perceived louder than when other colored cars were visible.

Studies by Jeon et al. [Jeon 2005, Jeon 2008] also performed listening tests which combined auralizations with visual stimuli. The visual stimuli were still photographs of an opera hall’s stage view from nine different positions. The auralizations were binaural auralizations of the same theater at corresponding positions. 50 participants pair-wisely compared auditory only and subsequently visually only stimuli on *preference*. Finally, two experiments on audio-visual cross modality were conducted. The first was a repetition of the preference judgement and the second was a comparable pair-wise comparison with cross-matched visual and acoustical stimuli. In the cross-modal conditions, both auditory and visual preferences were shown to contribute to overall impression, however auditory cues were more influential than the static visual cues.

Table 7.1 provides an overview of which acoustical attributes were studied and which visual stimuli were employed in the discussed studies. Summarizing, results from the discussed studies indicated that the inclusion of visual cues influenced source distance, ASW, room size, and loudness judgements. However, in the majority of these studies, the visual cues were represented by photographs and the sound source was not visible in the photos.

In the conduct of the current study, a framework was developed which combines

animated avatars of the sound sources (actors, performers, etc.), positioned in a virtual visual model with the calibrated dynamic voice auralizations (see Sec. 7.2). As this framework enables studies of visuals' influence of the space and sources on the room-acoustical experience, two multi-modal listening tests were carried out: 1) the final test of the previous chapter was repeated employing this framework (see Sec. 7.3) and 2) a listening test which compared dynamic voice auralizations coherently matched with visual positions to incoherently matched audio-visual pairs (see Sec. 7.4).

## 7.2 Framework for multi-modal auralization

Fig. 7.1 illustrates the augmented auralization framework, from the creation of a GA model to the real-time auralization of the VR scene. Section 7.2.1 details the auralization part of the multi-modal application. Section 7.2.2 discusses how the visual model of the theater was added to the virtual scene. Initially designed in 3ds Max, the model was then exported to Blender to be rendered in real-time on a VR architecture using the BlenderVR extension [Katz 2015a]. Finally, the virtual avatar creation method is detailed in Sec. 7.2.3. Actors were recorded with both microphones and a Kinect 2 sensor, the latter producing a coupled pair of RGB/Depth videos. These videos were used to generate a point-cloud of the actors in the rendered scene<sup>3</sup>.

### 7.2.1 Room acoustic rendering

A Max/MSP patch was designed to handle the auralizations (see Figure 7.2) using the 2<sup>nd</sup> order Ambisonic (9 channels) audio stream presented in Sec. 6.3. The current auralization was selected according to the user position. Decoding was based on the virtual speaker array for binaural rendering<sup>4</sup>. Real-time head-tracking of user orientation was applied as a rotation to the resulting Ambisonic stream prior to binaural decoding. Current user position and orientation in the virtual environment, or more precisely virtual camera position and orientation, were determined in BlenderVR (see Sec. 7.2.2) and sent via OSC (Open Sound Control) protocol to the Max/MSP patch.

### 7.2.2 Visual room rendering

The initial mesh creation and texturing of the visual Théâtre de l'Athénée model was performed in 3ds Max, based on the materials collected for the creation of the GA model in CATT-Acoustic. It was then imported in Blender for real-time rendering in the Blender Game Engine. The whole scene was ported to BlenderVR [Katz 2015a]

<sup>3</sup>see <http://www.youtube.com/watch?v=arFU8yFe73Q> for a video of the final result of the virtual Théâtre de l'Athénée case-study (single screen version)

<sup>4</sup>If one has a loudspeaker array it is possible to play it over an ambisonic system (for *Computed multiple-loudspeaker auralization*)

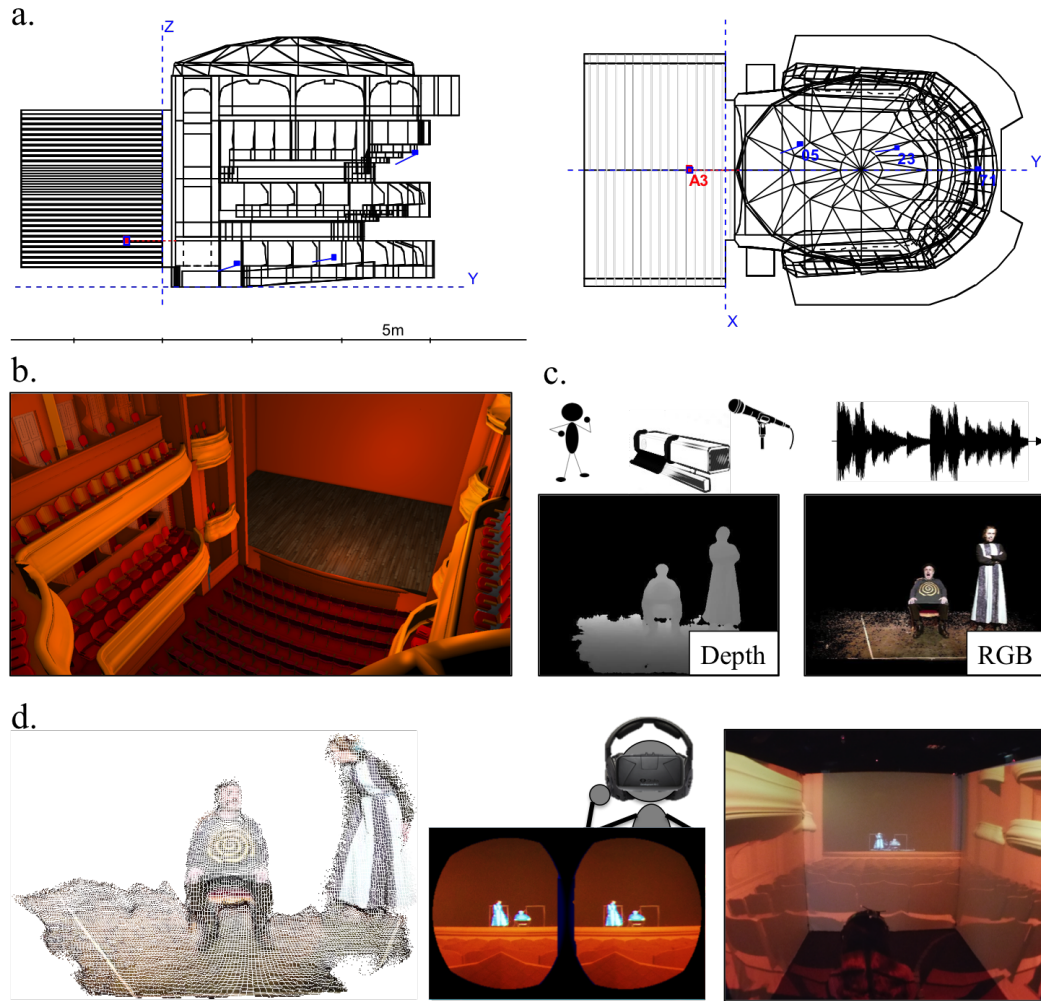


Figure 7.1: Conceptual overview of the Augmented Auralization framework. (a) Creation of the Théâtre de l'Athénée GA model and RIRs simulation for source-receiver positions. (b) Creation of the visual model. (c) Audio (dry) and Visual (RGB and Depth) recording of the performance. (d) Rendering the performers' avatar as a point-cloud, created from RGB and Depth recordings, which is integrated in the virtual environment for real-time augmented auralizations.

to be rendered on a 3-screen video-wall architecture<sup>5</sup>. Lighting was adapted to the targeted VR architecture, based on aesthetic considerations.

A lightweight system architecture was conceived to allow for the creation of a 3-wall CAVE-like system as illustrated in Figure 7.3. The objective was to project the virtual scene on 3 screens using a single wide-angle lens projector, based on a technique similar to standard homography. A set of 3 virtual cameras was positioned in the theater scene, each camera rendering its image on a virtual screen in an intermediate layer “projection” scene. The relative positions and dimensions of the virtual screens in this scene matched those of the physical screens of the

<sup>5</sup>The framework also enables rendering on an Oculus Rift DK2 HMD

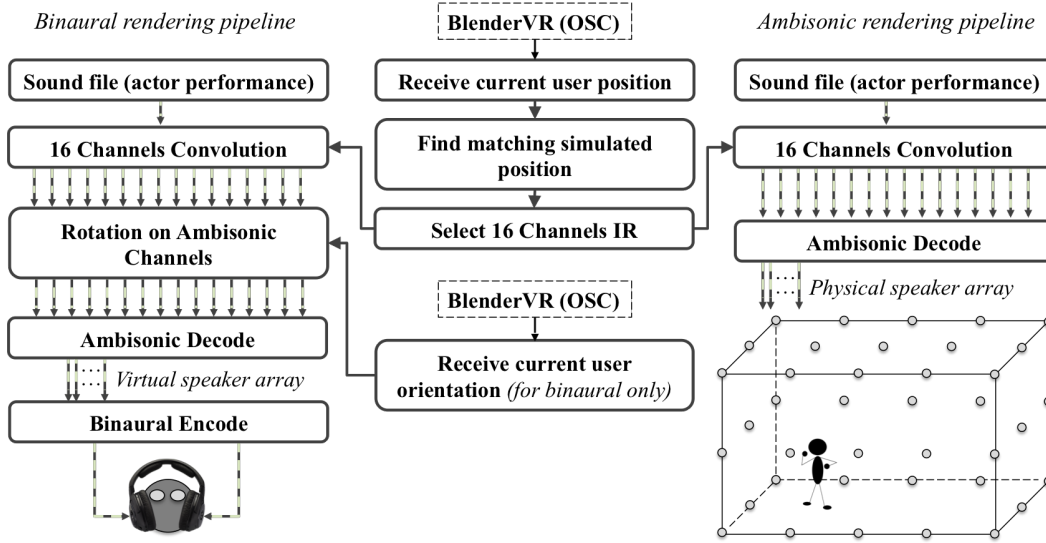


Figure 7.2: Conceptual schematic detail of the real-time auralization implemented in *Max/MSP*. Left and right hands of the figure respectively illustrate binaural and Ambisonic rendering pipelines.

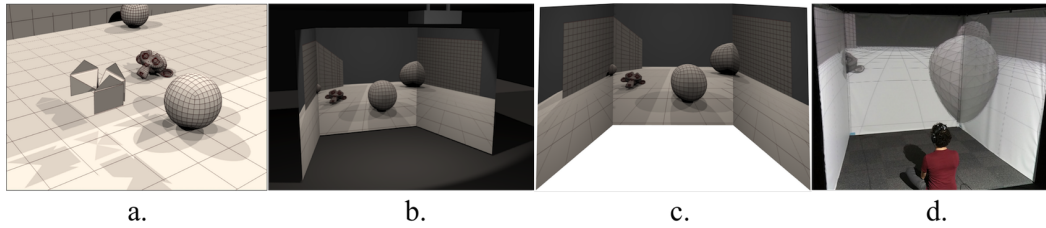


Figure 7.3: Illustration of the single-projector based rendering on the 3 screen CAVE. (a) Virtual scene with the set of 3 virtual cameras. (b) “Projection” scene with the 3 virtual screens and the virtual camera (cube-like shape on the top of the image), whose position, extrinsic, and intrinsic parameters match those of the screens and projector of the real world. (c) Pre-distorted image of the 3 screen CAVE architecture, as seen from the virtual camera of the “projection” scene, to be projected by the projector. Adaptive rendering is achieved by synchronizing the position of the 3 virtual cameras with the position of the user tracked in the CAVE. (d) Final projection on the 3 screen CAVE.

CAVE. A virtual camera sharing the extrinsic and intrinsic parameters of the physical projector rendered the image that the latter projected on the actual CAVE. Finally, head-tracking was accomplished using a set of OptiTrack infrared cameras to adapt the current rendered viewpoint to the user’s actual position, providing a stable virtual environment (with respect to the real world).



### 7.2.3 Dynamic performance recording, point-cloud rendering in the VR world

The performance was recorded in the Théâtre de la Reine Blanche, a 140-seat theater in Paris, using two headset microphones and a Kinect 2 sensor. As described in Sec. 6.3, the direct-to-reverberant ratio is high for close mic recordings, therefore these were employed as approaching anechoic recordings. The video stream of the Kinect 2 sensor was handled by a script based on the libfreenect2 library<sup>6</sup>, recording current time stamp and both RGB and Depth images to disk. RGB and Depth videos were created from these images with a MatLab script checking for frame-per-second (fps) regularity of the image recording. Both videos were then combined during the real-time rendering in BlenderVR to produce a  $512 \times 424$  pixel point-cloud of the actors. The term point-cloud here refers to a GLSL texture rather than a 2D deformable mesh to reduce CPU consumption, projected in the VR world from a point in the virtual environment corresponding to the Kinect camera's position. The Depth video was used to define the spatial position of the point-cloud pixels, the RGB to define their color. The work of Pagliari et al. [Pagliari 2015] was used to define the mapping between the HUE of the Depth video gray-scale and the pixels depth position, along with the X/Y scaling coefficients of the 3D volumetric pyramid projection. The global scale of the point-cloud was defined to produce life-sized avatars in the VR scene's. Noise in the captured Depth video was removed frame-by-frame using filters for pepper-noise removal, forcing consistency amongst neighboring pixel values (medfilt2, MatLab Image Processing Toolbox).

## 7.3 Coherent visual-aural experiment

Using the presented multi-modal framework the effect of the inclusion of visuals on the acoustic experience of auralizations was studied. For this purpose, the listening test presented in the previous chapter was repeated in the established framework. The same auralizations were employed, however, tracking of the participant was included creating a more ecological experience.

### 7.3.1 Protocol

The test had 3 variants corresponding to the source directivity-types.  $2^{nd}$  order ambisonic auralizations rendered on headphones were compared for 3 receiver positions (see Fig. 7.4). Every trial was repeated 3 times, resulting in  $(3 \times 3)$  9 trials. Participants were initially presented one training trial with the test administrator present in order to ensure they understood the task (SR2). The training session results were not included in the results.

Participants were given written instructions before commencing the listening

---

<sup>6</sup>Libfreenect2: Open source drivers for the Kinect for Windows v2 device, <https://github.com/OpenKinect/libfreenect2>



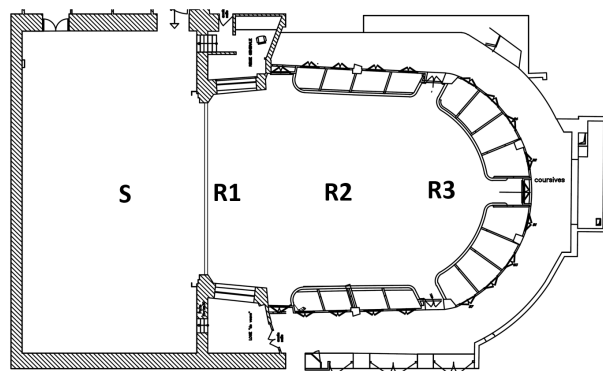


Figure 7.4: Plan of the Théâtre de l'Athénée depicting the source position on stage and the 3 receiver positions.



Figure 7.5: Test setup (a) without and (b) with the visual virtual theater rendering.

test<sup>7</sup>. For each trial, participants compared and rated the 3 auralizations in terms of *Plausibility* (*Plaus.*), *Distance* (*Dist.*), *ASW*, and *LEV*<sup>8</sup> on a discrete scale ranging from 1 ('least ...') to 7 ('most ...'). Participants were forced to use the 2 extreme scale values at least once per attribute, but were allowed to give 2 auralizations the same rating. Presentation order of the receiver position and correspondence to source directivity-type were randomized.

20 participants (mean age: 34.8 SD: 11.4) took part in the study. First, an audiogram was carried out ensuring the quality of the participant's hearing. All participants had experience with either room acoustics or vocal/instrumental performances. The test was carried out in the ambisonic room at LIMSI (ambient noise level  $\approx 37$  dBA,  $T_{20}$  (500-1000 Hz) = 0.2 s (see Fig. 7.5)). Participants were given written instructions before commencing the test which explained the task, described the attribute definitions, and illustrated the test interface. Participants were able to listen to the auralizations as many times as desired and were able to select the starting play point in the auralizations. Auralizations were presented via

<sup>7</sup>Appendix C presents the instructions provided to the participants

<sup>8</sup>See Sec. 5.3 For descriptions provided to the participants.

Table 7.2: Absolute mean difference, corresponding SD and PCC between repeated conditions.

	Plausibility	Distance	ASW	LEV
mean abs dif.	1.9	1.5	1.7	1.8
SD	2.3	2.1	2.2	2.2
PCC	0.3	0.2	0.4	0.1

headphones (Sennheiser model HD 650) at an RMS level of 80 dBA.

### 7.3.2 Results

Initial attention is given to the reliability of the responses, determined from the mean absolute difference between the repeated configurations and its SD (see Table 6.1).

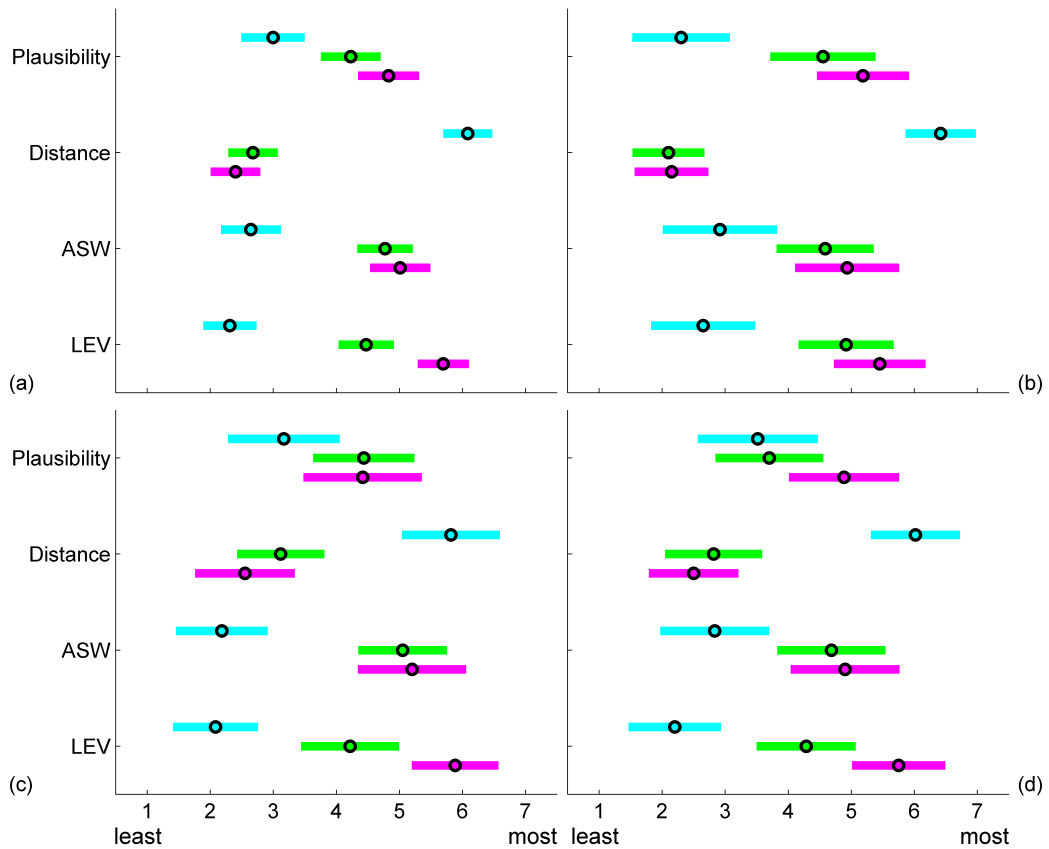


Figure 7.6: 98.8% confidence intervals of (a) **Total** results and individual positions (b) **SR1**, (c) **SR2**, and (d) **SR3**. (Upper) omni-directional, (Middle) static voice, and (Lower) dynamic voice directivity. (Cyan) omni-directional, (Green) static voice, and (Magenta) dynamic voice directivity.

The absolute mean difference ranged from 1.5 to 1.9 and corresponding standard deviation ranged from 2.1 to 2.3 on a 7 point scale. A metric for judging the reliability of responses is Pearson's Correlation Coefficient (PCC), defined as the covariance of the two sets of responses divided by the product of their standard deviations. It is interesting to note that the attribute with the lowest mean absolute difference (*Distance*) has a rather low PCC.

Fig. 7.6 presents the combined results and results per position. As 4 attributes were tested, according to equations 5.1 and 5.2 98.8% CIs were created in order to compensate for multiple testing. Subsequently, a group-wise two-way repeated measures ANOVA ( $\alpha = 0.05$  level) was conducted regarding position ( $F > 0.1$ ,  $p = 1.00$ ), directivity type ( $F = 4.7$ ,  $p = 0.02$ ), and position $\times$ directivity type ( $F = 1.1$ ,  $p = 0.37$ ). As these results indicated significant differences, a *t*-test with 'Bonferroni' correction in order to compensate for multiple testing was employed to estimate whether an acoustic attribute differed perceptually between auralization types (see Table 7.3).

Combined results of the dynamic voice auralizations were compared to the static voice and omni-directional auralizations. Dynamic voice auralizations were perceived significantly more enveloping than the omni-directional and static voice auralizations. Additionally, the dynamic voice auralization was perceived more plausible and closer as well as exhibiting a wider source width than the omni-directional

Table 7.3: *T*-test results comparing the dynamic voice (dyn.) to the omni-directional (omni.) and static voice (sta.) directivity (a ✓ indicates a significant difference).

	Acoustic Attributes	<i>t</i> -test	
		omni– dyn.	sta.– dyn.
Total	Plaus.	✓	
	Dist.	✓	
	ASW	✓	
	LEV	✓	✓
SR1	Plaus.	✓	
	Dist.	✓	
	ASW	✓	
	LEV	✓	
SR2	Plaus.		
	Dist.	✓	
	ASW	✓	
	LEV	✓	✓
SR3	Plaus.		
	Dist.	✓	
	ASW	✓	
	LEV	✓	✓

auralizations.

Results of the dynamic voice auralizations were compared to the static voice and omni-directional auralizations per position (see table 7.3 and Fig. 7.6b-d). For all 3 positions, the dynamic voice sources were considered significantly closer, wider, and the auralizations more enveloping than the omni-directional auralizations. Additionally, the **SR1** dynamic voice source was rated more plausible. Finally, the dynamic voice auralization at **SR2** and **SR3** was judged to be significantly more enveloping than the static voice.

Finally, these results (*coherent multi-modal experiment*) are compared to the previous listening test (*voice directivity test*) in order to establish the effect of the addition of visuals. Initial attention is focused on the significant differences between source directivity types. The results between omni-directional and dynamic voice auralizations show good agreement between both experiments estimating the dy-

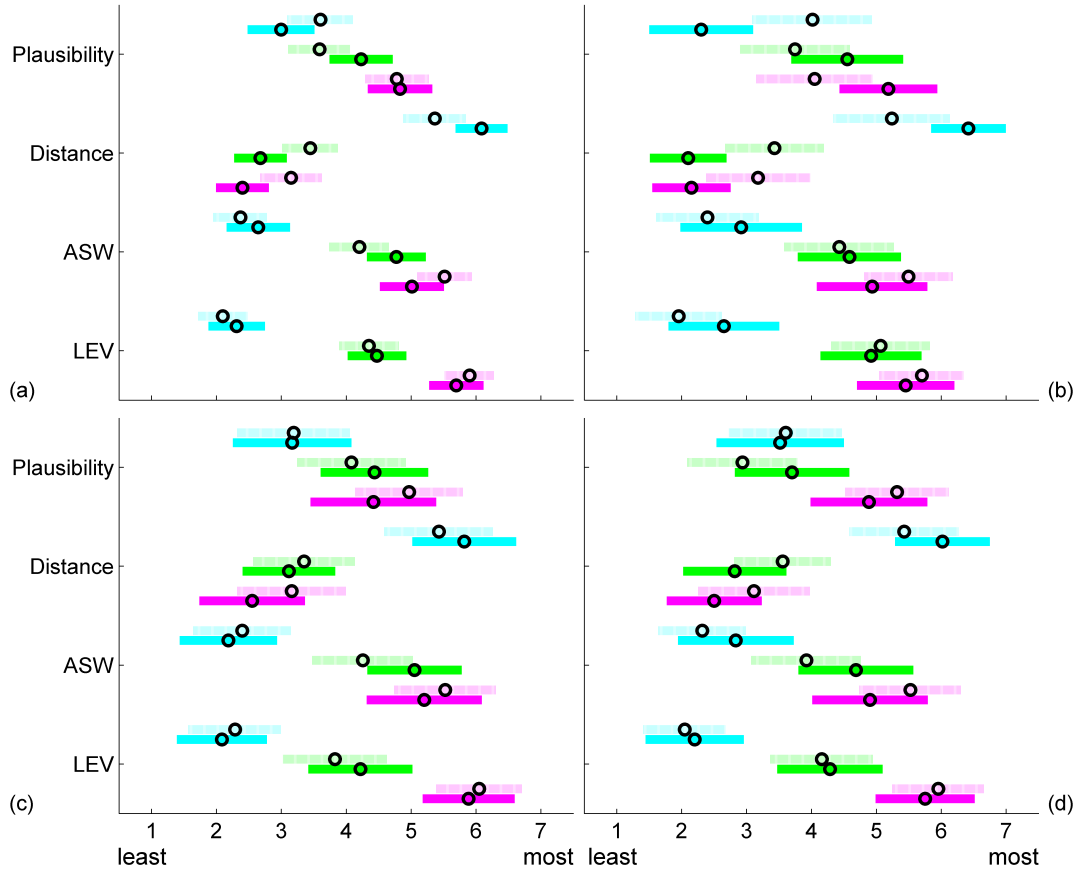


Figure 7.7: Comparison of 98.8% confidence intervals of (- -) *voice directivity test* and (—) *coherent multi-modal experiment* for (a) **Total** results and individual positions (b) **SR1**, (c) **SR2**, and (d) **SR3**. (Cyan) omni-directional, (Green) static voice, and (Magenta) dynamic voice directivity.

dynamic voice directivity closer, wider, and more enveloping for all positions. The only difference was that for the *voice directivity test* no significant difference was found between omni-directional and dynamic voice auralizations for *Plausibility* for **SR1** while the *coherent multi-modal experiment* found no significant difference for positions **SR2** and **SR3**. More differences can be observed for the comparisons between static and dynamic voice directivity auralizations. Combined auralizations of the dynamic voice directivity of the *voice directivity test* were judged more plausible while also exhibiting a wider apparent source width than the static voice directivity while the *coherent multi-modal experiment* found no significant differences. For position **SR3** the dynamic voice directivity of the *voice directivity test* were judged more plausible and exhibiting a wider apparent source width than the static voice directivity while the *coherent multi-modal experiment* found no significant differences for these attributes. The significance judgment for the remainder of the attributes was equal between experiments.

Finally, the results of both tests were compared with a group-wise two-way repeated measures ANOVA ( $\alpha = 0.05$ ) which tested per position for directivity and test (see Table 7.4). As these results indicated significant differences, subsequently a pair-wise *t*-test with ‘Bonferroni’ correction was conducted between the two tests for one directivity type for one position to establish whether attributes were rated significantly different (see Table 7.4). For the combined auralizations, the omni-

Table 7.4: Two-way repeated measures ANOVA results (F-value = variation between sample means / variation within the samples) testing for directivity type (dir.), test, and directivity type  $\times$  test (*p*-values less than 0.05 are *indicated*) and *t*-test results comparing the *voice directivity test* and *coherent multi-modal experiment*, A ‘✓’ indicates a significant difference.

Pos.	two-way ANOVA			acoustic attribute	omni-directional	static voice directivity	dynamic voice directivity
		<i>F</i>	<i>p</i> – value				
Total	dir.	152.6	$10^{-2}$	Plaus.			
	test	0.3	0.64	Dist.	✓	✓	✓
	dir. $\times$ test	5.3	0.08	ASW			
				LEV			
SR1	dir.	22.7	0.01	Plaus.	✓		
	test	0.1	0.74	Dist.	✓	✓	
	dir. $\times$ test	0.6	0.53	ASW			
				LEV			
SR2	dir.	77.1	$10^{-2}$	Plaus.			
	test	0.4	0.60	Dist.			
	dir. $\times$ test	7.6	0.04	ASW			
				LEV			
SR3	dir.	37.4	$10^{-2}$	Plaus.			
	test	0.4	0.60	Dist.			
	dir. $\times$ test	2.1	0.24	ASW			
				LEV			

directional auralization was judged further while the static and dynamic voice directivity were rated further. This was mainly due to the results of auralizations **SR1** where the omni-directional was judged further while the static voice directivity was rated further. Additionally, for the auralizations at **SR1** the omni-directional auralization was rated significantly less plausible. Finally, the  $t$ -test found no significant differences for *ASW* and *LEV*.

## 7.4 Incoherent visual-aural experiment

Results from the previous experiment indicate that the inclusion of the visual model as presented in Sec. 7.2 influenced the perceptual acoustic attributes *Plausibility* and *Distance*. Another means of studying the influence of visuals on auralization judgment is comparison of coherently matched auralizations with visualizations to incoherently matched audio-visual simulations. As previous studies [Barron 1988, Menzel 2008] have also found an influence of visuals on *Loudness* judgment, this acoustic attribute was also included in the *Incoherent visual-aural experiment*.

### 7.4.1 Protocol

The stimuli in the test were 3 visual positions combined with three 2<sup>nd</sup> order ambisonic auralizations rendered over headphones, with both matched and mismatched auditory-visual cues (see Fig. 7.8<sup>9</sup>). It should be noted that **SR1** is the same as in the previous experiment, however receiver position **SR2** has been moved to the side of the audience area and **SR3** is in the current experiment on the first balcony. Every trial was repeated 3 times, resulting in  $(3 \times 3 \times 3)$  27 iterations. Participants were initially presented with 3 training iterations with the test administrator present in order to ensure they understood the task. Results for these training iterations were not analyzed.

<sup>9</sup>It should be noted that the visual and aural orientation for **SR2** is towards the Source

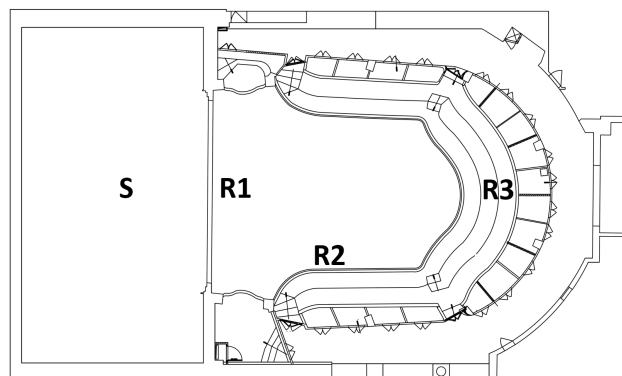


Figure 7.8: Plan of the Théâtre de l'Athénée depicting the source position on stage and the 3 receiver positions.

Participants were given written instructions before commencing the listening test<sup>10</sup>. For each iteration, participants rated the auralization on *Plausibility*, *Distance*, *Loudness*, *ASW*, and *LEV* on a discrete scale ranging from 1 (‘not ...’) to 7 (‘very ...’). See Sec. 5.3 for descriptions of *Distance*, *ASW*, and *LEV* provided to the participants. The provided descriptions for *Plausibility* and *Loudness* were

- *Plausibility* - Does the example sound plausible/realistic in relation to your seating position in the theater?
- *Loudness* - Overall impression of the intensity of sound. Loudness should be assessed relative to what you consider acceptable for the theatrical performance involved.

Presentation order of visual receiver position and correspondence to aural receiver position were randomized. This protocol is similar to the second auditory-visual test in [Jeon 2008], which also compared coherently matched audio-visual stimuli to incoherently matched. However, this protocol includes VR models with visual sources and it compares acoustical attributes instead of seat preference.

22 participants (mean age: 37.6 SD: 11.6) took part in the study. An audiogram was performed in order to ensure that the participants hearing was sufficient. All test participants had experience with room acoustics or vocal/instrumental performances. The test was performed in the same room as the coherent audio-visual experiment. Participants were given written instructions before commencing the test which explained the task and attribute definitions. Participants were able to listen to the auralizations as many times as desired and were able to select the starting play point in the auralizations.

As no audio calibration data was available for the audio recording, playback level was calibrated based on sound pressure level measurements of speech [Olsen 1998]. As these measurements were carried out in an anechoic room with a source-receiver distance of 1 m, a similar room with the same source-receiver distance was simulated in CATT-Acoustic. Sound pressure of the source was set to the same level as for the Théâtre de l’Athénée model. A voice directivity auralization of the regarded scene (“Ubu roi”) was created for the anechoic room model. Playback RMS level of the framework was set to 82 dBA (female standard shouting level) for this auralization. With the calibration of the source level in anechoic conditions at 1 m, the resulting RMS levels of the tested auralizations were **SR1**=75 dBA, **SR2**=74 dBA, and **SR3**=73 dBA when the head of the listener was still, orientated forward. Auralizations were presented via closed headphones (Sennheiser model HD 380) for improved sound isolation.

---

<sup>10</sup>Appendix C presents the instructions provided to the participants

Table 7.5: Absolute mean difference, corresponding SD and PCC between repeated conditions.

	Plausibility	Distance	Loudness	ASW	LEV
mean abs dif.	0.8	0.7	0.7	0.8	0.8
SD	0.8	0.7	0.8	0.8	0.8
PCC	0.2	0.3	0.2	0.2	0.2

## 7.4.2 Results

### 7.4.2.1 Statistical analysis approach, data normalization, and subject reliability

For this test one statistical analysis procedure is employed to estimate whether ratings differed perceptually between conditions. Analysis was carried out with the software package MatLab: first a group-wise one-way repeated measures ANOVA ( $\alpha = 0.05$ ) was applied regarding visual position/auralization in order to establish whether acoustic attribute ratings comprised significant differences. Then, a pair-wise t-test with ‘Bonferroni’ correction in order to compensate for multiple testing was employed to compare conditions.

Initial attention was focused on the normalization of the responses in order to create comparable ratings between participants. Normalization was performed according to Equation 7.1

$$z = \frac{x - \mu}{\sigma} \quad (7.1)$$

in which  $z$  is the normalized response (distance between the response and the participant’s mean response for the same acoustic attribute in units of the standard deviation (SD)),  $x$  is the response,  $\mu$  is the mean of the participant’s responses for the acoustic attribute over all trials, and  $\sigma$  is the SD for the same acoustic attribute.

Initial attention is given to the reliability of the responses, determined from the mean absolute difference between the repeated configurations and its SD (see Table 7.5). The absolute mean difference ranged from 0.7 to 0.8 and corresponding standard deviation ranged from 0.7 to 0.8 on a 7 point scale. A metric for judging the reliability of responses is Pearson’s Correlation Coefficient (PCC), defined as the covariance of the two sets of responses divided by the product of their standard deviations. It is interesting to note that the mean absolute differences are significantly lower than for the previous test while the PCC are comparable.

#### 7.4.2.2 Coherent audio-visual position conditions

As 5 attributes were tested, according to equations 5.1 and 5.2 99% CIs were created in order to compensate for multiple testing. Fig. 7.9a and Table 7.7 presents results for the coherently matched audio-visual combinations. Regarding *Distance*, the audio-visual condition at **R3**<sup>11</sup> was considered acoustically further away than **R1**

<sup>11</sup>It should be noted that visual positions are denoted with R# and auralizations with SR#



and **R2**. The *ASW* of audio-visual condition at **R1** and **R2** was rated significantly wider than at **R3**. The audio-visual condition at **R1** was perceived louder than at **R2** and **R3**. Finally, the audio-visual condition at **R1** and **R2** were perceived more enveloping than at **R3**.

#### 7.4.2.3 Incoherent audio-visual position conditions

Results were compared between coherently and incoherently matched audio-visual conditions. Figure 7.9b presents results for all the separate audio-visual conditions. The lower rows of Table 7.7 present the results for all audio-visual combinations. As *Plausibility* was rated according to seating position, statistical comparisons between various coherently matched audio-visual conditions was deemed redundant. Results for *Plausibility* are therefore provided separately (see upper graph of Figure 7.9 and Table 7.6). First, ratings were compared per visual position. Participants positioned at visual positions **R1** and **R2** rated auralizations **SR1** and **SR2** significantly more plausible than **SR3**. No significant results were observed for visual position **R3**. Second, ratings were compared per auralization. Auralizations **SR1** and **SR2** were found most plausible for visual position **R2**. Auralization **SR3** was rated most plausible for visual position **R3**.

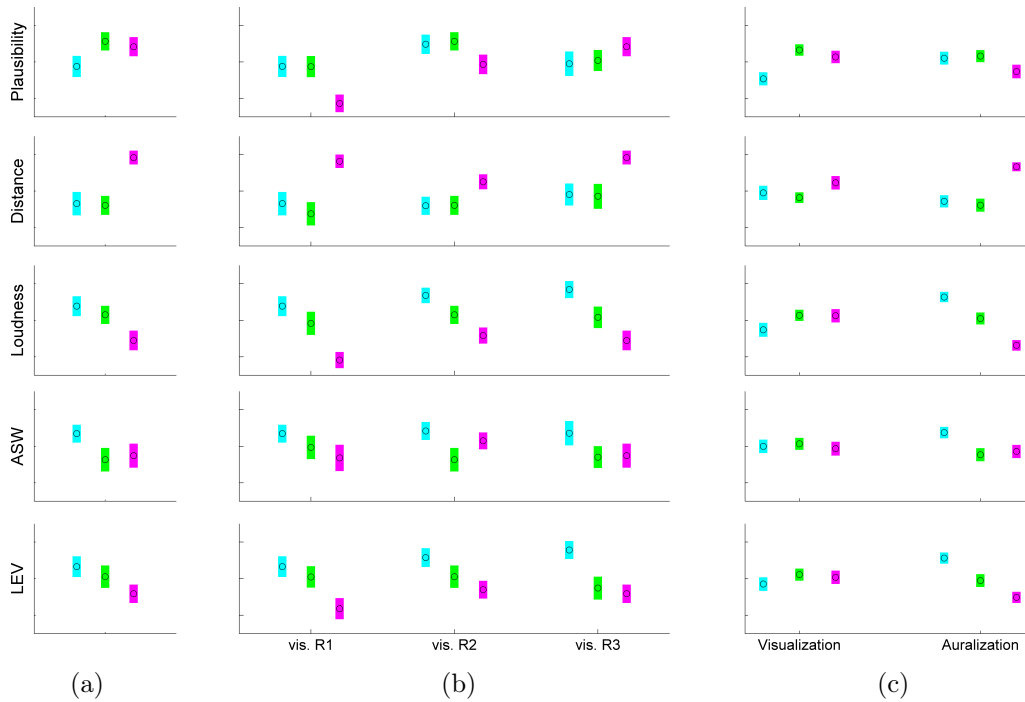


Figure 7.9: 99% confidence intervals for (a) coherently matched audio-visual combinations, (b) results separated per audio-visual combination, (c) combined results for visual position **SR1**, **SR2**, and **SR3** as well as auralizations **SR1**, **SR2**, and **SR3**. (Cyan) (S)R1, (Green) (S)R2, and (Magenta) (S)R3.

Table 7.6: One-way repeated measures ANOVA results ( $p$ -values  $< 0.05$  are indicated) and cross-modal t-test results comparing *Plausibility* ratings as a function of (upper) visual position and (lower) auralization (a ‘✓’ indicates a significant difference).

	Pos.	ANOVA		t-test		
		$F$	$p$ -value	SR1–SR2	SR1–SR3	SR2–SR3
Visual	R1	22.9	$10^{-2}$		✓	✓
	R2	9.3	$10^{-2}$		✓	✓
	R3	2.9	0.06			
				R1–R2	R1–R3	R2–R3
Aural	SR1	5.8	0.01	✓		✓
	SR2	8.7	$10^{-2}$	✓		✓
	SR3	39.9	$10^{-2}$	✓	✓	✓

The influence of auralizations was examined comparing the ratings of different auralizations at the same visual positions for the remaining parameters (see Table 7.7):

- For visual position **R1**, auralizations **SR1** and **SR2** were rated closer than **SR3** and auralizations were perceived louder with decreasing source-receiver distance. Additionally, auralization **SR1** exhibited a wider source width than **SR3**. Auralizations **SR1** and **SR2** were rated more enveloping than **SR3**.
- For visual position **R2**, auralizations **SR1** and **SR2** were rated closer than **SR3**. Auralization **SR1** and **SR3** were perceived to exhibit a wider source width than **SR2**. Furthermore, auralizations were experienced louder and more enveloping with decreasing source-receiver distance.
- For visual position **R3**, auralizations **SR1** and **SR2** were rated closer than **SR3**. Auralizations were experienced louder with decreasing source-receiver distance. Additionally, at this position auralization **SR1** exhibited a wider source width and was rated more enveloping than **SR2** and **SR3**.

The influence of visual position was studied comparing ratings of the auralizations at different visual positions:

- Auralization **SR1** was experienced louder and more enveloping at visual position **R3** than **R1**.
- No significant differences were observed for auralization **SR2**.
- Auralization **SR3** was rated closer at **R2** than at **R1** and **R3**. Additionally, auralization **SR3** was rated louder and more enveloping at **R2** and **R3** than at **R1**.

Table 7.7: One-way repeated measures ANOVA results ( $p$ -values  $< 0.05$  are *indicated*). (upper row) Comparing coherent audio-visual conditions. (lower three rows) Results across all cross-modal positions as a function of (left) visual positions (**R1**, **R2**, and **R3**) and (right) auralizations (**SR1**, **SR2**, and **SR3**) (a ‘✓’ indicates a significant difference).

	ANOVA		Acoustic Attribute	t-test		
	$F$	$p$ -value		<b>SR1–SR2</b>	<b>SR1–SR3</b>	<b>SR2–SR3</b>
coherent	32.9	$10^{-2}$	Distance		✓	✓
	12.4	$10^{-2}$	Loudness		✓	✓
	5.8	0.01	ASW	✓	✓	
	7.9	$10^{-2}$	LEV		✓	✓
Vis. R1	32.3	$10^{-2}$	Distance		✓	✓
	41.3	$10^{-2}$	Loudness	✓	✓	✓
	4.1	0.02	ASW		✓	
	15.5	$10^{-2}$	LEV		✓	✓
Vis. R2	10.7	$10^{-2}$	Distance		✓	✓
	33.4	$10^{-2}$	Loudness	✓	✓	✓
	11.8	$10^{-2}$	ASW	✓		✓
	14.0	$10^{-2}$	LEV	✓	✓	✓
Vis. R3	24.0	$10^{-2}$	Distance		✓	✓
	41.8	$10^{-2}$	Loudness	✓	✓	✓
	4.3	0.02	ASW	✓	✓	
	27.3	$10^{-2}$	LEV	✓	✓	
	ANOVA			t-test		
	$F$	$p$ -value		<b>R1–R2</b>	<b>R1–R3</b>	<b>R2–R3</b>
Aur. SR1	1.6	0.22	Distance			
	5.0	0.01	Loudness		✓	
	0.1	0.89	ASW			
	5.1	0.01	LEV		✓	
Aur. SR2	2.7	0.08	Distance			
	0.7	0.48	Loudness			
	2.1	0.14	ASW			
	2.8	0.07	LEV			
Aur. SR3	14.4	$10^{-2}$	Distance	✓		✓
	10.1	$10^{-2}$	Loudness	✓	✓	
	3.1	0.06	ASW			
	6.0	$10^{-2}$	LEV	✓	✓	

#### 7.4.2.4 Combined visual and aural conditions

Subsequently, combined results per visual position are compared. Figure 7.9c presents combined results for the visual positions and the auralizations. Such analysis allows for investigation of the contribution of the position in one modality irre-

Table 7.8: One-way repeated measures ANOVA results ( $p$ -values  $< 0.05$  are indicated) and cross-modality position t-test results of pairwise comparisons of combined results at (upper) visual positions **R1**, **R2**, and **R3** or (lower) auralizations **SR1**, **SR2**, and **SR3** ('S' omitted for brevity, a '✓' indicates a significant difference).

	ANOVA		Acoustic Attribute	t-test		
	$F$	$p$ -value		<b>R1-R2</b>	<b>R1-R3</b>	<b>R2-R3</b>
<b>Visual</b>	5.3	0.01	Distance		✓	✓
	6.5	$10^{-2}$	Loudness	✓	✓	
	0.7	0.49	ASW			
	3.9	0.03	LEV	✓		
<b>Aural</b>	42.4	$10^{-2}$	Distance		✓	✓
	91.8	$10^{-2}$	Loudness	✓	✓	✓
	7.9	$10^{-2}$	ASW	✓	✓	
	27.1	$10^{-2}$	LEV	✓	✓	✓

spective of the position of the other modality's rendering. The statistical analysis results are presented in Table 7.8.

Auralizations at visual position **R3** were judged significantly further away than at **R2**. Auralizations at visual position **R2** and **R3** were perceived significantly louder than at **R1**. No significant trends were observed regarding *ASW* and *LEV*.

Finally, the combined auralization ratings, combined across all visual positions, are compared (see Table 7.8). Auralization **SR3** was perceived overall significantly further away than **SR1** and **SR2**. The *Loudness* and *LEV* ratings decreased with increasing source-receiver distance. Auralization **SR1** exhibited a wider source width than **SR2** and **SR3**.

#### 7.4.2.5 Test population grouping analysis

Observations of the individual result distributions indicated that there were potentially several categories of responses, as a function of participant. Similar analysis was performed by Andre et al. [André 2012], in the context of evaluating spatial audio quality in 3D-cinema. An experiment compared three different audio conditions (one coherent and two incoherently matched) for a single video. They found that the sound condition did not affect the entire test population. However, participants could be classified according to their *Presence* ratings independently of the sound condition. In the sub-group with the highest score, the coherent soundtrack lead to a decreased sense of *Presence*.

In order to subdivide the test population t-test compared combined results at visual positions **R1** and **R3** per participant. Visual position **R2** was omitted from this analysis in order to evaluate the visual influence between the two extreme positions. Based on these results, participants could be divided into three sub-groups:

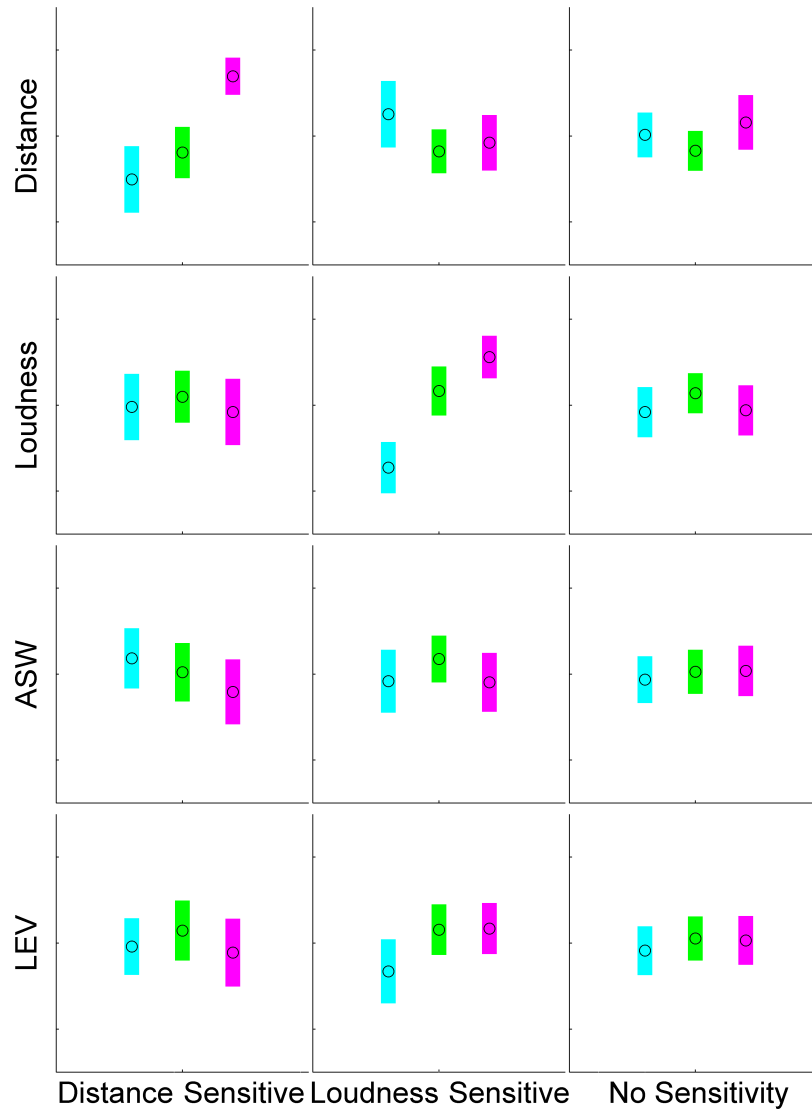


Figure 7.10: 99% confidence intervals of the combined normalized ratings for separate visual positions separated by participants on who increased visual distance influenced *Distance* or *Loudness* perception or participants who were uninfluenced. (Cyan) R1, (Green) R2, and (Magenta) R3.

- *Distance Sensitive*: participants who significantly judged auralizations at visual position **R3** further away than **R1** (6 participants)
- *Loudness Sensitive*: participants who significantly judged auralizations at visual position **R3** louder than **R1** (7 participants)
- *No Sensitivity*: remaining participants (10 participants).

One participant significantly judged auralizations at visual position **R3** louder and further away than **R1**. This participant was classified into the first sub-group as the  $p$ -value was lower for *Distance* ratings.

Table 7.9: One-way repeated measures ANOVA results ( $p$ -values  $< 0.05$  are indicated) and t-test results comparing combined results at visual position **R1**, **R2**, and **R3** (a ‘✓’ indicates a significant difference) classified by participant sub-group cross-modal attribute sensitivity analysis.

ANOVA		Acoustic	t-test		
<i>F</i>	<i>p</i> -value	Attribute	<b>R1–R2</b>	<b>R1–R3</b>	<b>R2–R3</b>
<i>Distance Sensitive</i> sub-group					
34.8	< 10 <sup>−2</sup>	Distance		✓	✓
0.2	0.82	Loudness			
1.0	0.40	ASW			
2.3	0.14	LEV			
<i>Loudness Sensitive</i> sub-group					
2.7	0.11	Distance			
43.6	< 10 <sup>−2</sup>	Loudness	✓	✓	✓
1.0	0.38	ASW			
6.4	0.01	LEV	✓	✓	
<i>No Sensitivity</i> sub-group					
1.2	0.32	Distance			
2.3	0.12	Loudness			
0.1	0.88	ASW			
0.2	0.80	LEV			

Figure 7.10 and Table 7.9 present combined results at the visual positions **R1**, **R2**, and **R3**, separated per sub-group. The *Distance Sensitive* sub-group rated the combined auralizations at visual position **R3** further away than **R1** and **R2**. More significant trends were observed for the *Loudness Sensitive* sub-group. *Loudness* judgment for this sub-group increased with a increasing visual source-receiver distance. Additionally, this sub-group rated auralizations at visual position **R2** and **R3** more enveloping than at **R1**. No significant trends were observed for the *No Sensitivity* sub-group.

## 7.5 Discussion

This chapter studied the influence of visuals on the room acoustic experience of auralizations. The results are discussed per studied room acoustic attribute *Plausibility*, *Distance*, *Loudness*, *ASW*, and *LEV*. Additionally, the results are compared to previous studies [Barron 1988, Larsson 2001, Maempel 2013a, Menzel 2008]<sup>12</sup>. It should be noted that [Barron 1988] studied real concert hall experience and objective parameters, [Larsson 2001, Maempel 2013a] employed still photos or visual models but the sound sources were not visible, and [Menzel 2008] studied the loudness of car

<sup>12</sup>[Jeon 2008] is not included as it regarded seat preference.

images while this study employed the presented VR multi-modal framework with a visible source.

In this chapter's first experiment (Sec. 7.3) the dynamic voice auralization was rated more plausible than static voice though not significantly, while in the uni-modal (Sec. 6.3) listening test this resulted in a significant difference. A possible explanation is that in the auralization the source remained at the same position while in the visualization the actors changed position. For the second experiment, the *Plausibility* rating for auralization **SR1** was judged more plausible for visual position **SR2** and auralization **SR1** and **SR2** were judged to be equal distant. It is possible that beside the static source position, reverberance related to the close proximity of **SR1** to the stage house in the Théâtre de l'Athénée has affected these ratings.

Additionally, the results from this chapter indicated that acoustical *Distance* and *Loudness* were dependent on visual distance. A subdivision of the test population indicated that the participants could be subdivided into three groups: 1) participants who rated auralization at a further visual position acoustically further away, 2) participants who rated auralizations louder with an increased visual source-receiver distance, 3) participants whose acoustic ratings were uninfluenced by the visuals. This corresponds to previous studies [Larsson 2001, Maempel 2013a] which found an influence of visuals on acoustical *Distance* perception. [Barron 1988] found in correspondence with these experiments that loudness judgment increased with increasing visual source-receiver distance.

Furthermore, results of this chapter indicate that the influence of visuals on *ASW* ratings is of a different nature. Comparison between the final listening test of Chapter 6 and this chapter's first experiment showed no significant difference in *ASW* rating between aural only and audio-visual conditions. The second experiment also found no significant trends in *ASW* ratings between auralizations combined with varying source-receiver distances. This contradicts [Larsson 2001] who compared uni-modal to multi-modal conditions and found the addition of a visual led to wider *ASW* judgment.

In both experiments no significant trends were observed which indicated an effect of visuals on the acoustical attribute *LEV*. In order to limit the time usage for test participants, only a limited number of acoustical attributes was studied. In subsequent studies, the influence of visuals on reverberance, tonal color, spatial image, and intimacy could be studied.

## 7.6 Summary

Chapters 3-5 studied the perceptual validity of sound propagation in a 3D space and chapter 6 increased the *Plausibility* of the 3D directivity of the source. With this increased perceptual validity, audio-visual experiments were performed having confidence that the results can also apply to real-life situations. The results of multi-modal experiments indicated that by means of individual statistical analysis

---

the test population could be divided into three sub-groups: 1) participants on which increased visual source-receiver distance influenced acoustical distance perception, 2) participants who rated auralizations with greater visual source-receiver distance louder, and 3) participants who rated all tested acoustical attributes similar under different visual conditions. The next chapter discusses conclusions from the calibrations, validations, listening tests, and audio-visual experiments and presents suggestions for subsequent research.





# Conclusion and future work

---

## Contents

---

<b>8.1 Conclusion</b>	<b>141</b>
<b>8.2 Recommendations</b>	<b>143</b>
8.2.1 Dynamic voice auralization framework	143
8.2.2 Multi-modal auralization framework	143

---

## 8.1 Conclusion

Over recent decades, auralizations have become more prevalent in architectural acoustics and virtual reality. Despite their numerous use-cases, they have rarely been part of scientific studies. Therefore, this thesis presented advances regarding the calibration of room acoustical simulations, the inclusion of dynamic voice directivity into auralizations, and studies of the influence of visuals on the acoustic experience of auralizations. The major contributions are:

- A methodical calibration approach for geometrical acoustics models is presented and the extent to which it was successful was studied by means of comparing room-acoustical parameters between measurements and simulations and listening tests.
- A framework is presented which enables the inclusion of dynamic voice directivity into auralizations.
- A framework is presented which couples real-time convolution based auralization and 3D visualization with the inclusion of a 3D audio-visual (3D-AV) recorded performance.

The calibration of room acoustical simulations is necessary if one desires to have confidence in the results or to construct a scientific tool. Chapter 4 presented a methodical approach to calibrate geometrical acoustics models. The method assumed the existence of physical room acoustical measurements which were presented in Chapter 3. The measurement results were used as a reference for the simulation. Objective acoustic parameters representing subjective acoustical perceptions were employed. Acceptable tolerances in the calibration were based on the generally accepted JND metric for the various parameters, taking into account run-to-run variations for GA algorithms incorporating stochastic components. Measured and

simulated RIRs were analyzed using a common impulse response analysis tool in order to minimize the effect of analysis implementation variations.

The proposed calibration method started with the selection of absorption coefficients for the various materials in the geometrical model from available databases. This range of values for each given material is explored in order to adjust the mean reverberation parameters to within 1 JND of the measured value. It should be noted that the selected JND values vary as a function of room usage, but they were chosen as a base model tolerance reference value for the purpose of this study. In a specific instance, a more suitable threshold should be used which is specifically appropriate to the room's function. Adjustments begin with the materials with the largest surface area in the room. Scattering coefficients were adjusted to bring the clarity parameters to within 1 JND of the measured results. The sensitivity to scattering modifications can be studied by running the simulation with the GA model for two extreme cases of scattering definitions. These variances are quantified using the standard deviation of differences between measured reference parameters and those of the simulated data.

Two perceptual tests were carried out in order to evaluate to what extent the calibration resulted in similar simulated to measured auralizations. These tests compared both monaural and binaural measured auralizations to simulated auralizations on the selected parameters for the calibration. Additionally, 6 other perceptual attributes were tested. The combined results showed that with a current commercially available GA software (specifically CATT-Acoustic (v.9.0.c:3, TUCT v1.1a:4) in these studies), employing the proposed calibration approach based on few parameters evaluated for omni-directional source-receiver pairs, one is capable of producing simulated auralization which show limited number of significant differences in relation to their measured counterpart.

It should be noted that auralizations in these listening tests employed an omni-directional source directivity definition. As real sources have more complicated directivity patterns subsequent study explored the effect of dynamic voice directivity with regards to dynamic direction. Therefore, a preliminary listening test validated a specifically overlapping beam pattern which correctly reproduced an omni-directional pattern. Subsequently, a manner to include dynamic voice directivity into auralizations in post-processing in real-time employing a single channel anechoic recording was presented. As the results of the listening tests indicated perceptual differences between dynamic voice directivity and regular static source orientations in 3 acoustical attributes (including a more plausible rendering in reference to a corresponding RGB video), it highlighted the importance of including dynamic orientation into auralizations.

With the calibrated room acoustic model and increased plausibility of the directivity of the sound source, one can study the influence of visuals on the acoustic experience with a reasonable degree of confidence that this is also valid for real-life situations. For this purpose, a framework was developed which combined a 3D visual rendering and 3D auralizations. Two experiments were employed to study the effect of visuals on the room acoustical experience of the auralizations. Results

indicated that the room-acoustic experience of the test population was influenced by increased source-receiver distance in three separate manners: 1) participants who rated auralizations more acoustically distant with and increased visual source-receiver distance, 2) participants who rated auralizations louder with an increased visual source-receiver distance, and 3) participants who rated auralizations similar despite increased visual source-receiver distance.

## 8.2 Recommendations

As the sound propagation in the 3D space part seems to be sufficiently estimated, the recommendations concern the dynamic voice directivity and the multi-modal auralization frameworks.

### 8.2.1 Dynamic voice auralization framework

There are several subsequent studies which could further improve the dynamic voice auralization framework. In the current framework the source remains at the same position while real-life sources can move. Subsequent studies should create frameworks which enable the inclusion of source movements and study the effect on acoustical experience. However, it should be noted that the current framework employs (9 ambisonic channels ( $2^{nd}$  order)  $\times$  12 beam sources) 108 real-time convolutions and is consequently demanding on the CPU usage. A possible means of including moving sources is interpolating between source positions which would mean a doubling or tripling of the convolutions which is a significant strain on CPU usage. Therefore, a better option is probably to record  $2^{nd}$  order ambisonic dynamic voice directivity auralizations for multiple source positions on stage. Post-recording a panning can be performed between these auralizations based on the positions of the actors.

### 8.2.2 Multi-modal auralization framework

To support future research on augmented auralization, further developments of the framework are planned, focusing primarily on real-time point-cloud capture. The point cloud rendering is included to give the listener an idea of the position and orientation of the source. Currently, this rendering is grainy because the point cloud is recorded with a single Kinect depth sensor. Preliminary testing with multiple Kinect cameras showed promise for the capture of more complex scenes and improvement of the point-cloud rendering.

A possible means of usage for the framework is direct interaction between performer and virtual theater. For this purpose, the Kinect 2 Depth and RGB images need to be streamed directly into BlenderVR to generate the point-cloud in real-time which will allow further diversification of applications of the presented framework. The live audio stream of the performers would likewise be used directly for the auralization. With the final framework, performers should be able to interact with

the room space and acoustics in real-time, interacting with virtual avatars or other performers in different physical spaces, or to record themselves for latter assessment.

Additionally, suggestions are provided for further studies which are enabled by this framework. Studies presented in this thesis have been based on binaural presentations. However, currently the 3-screen system is positioned in a room with a loudspeaker array. Therefore, studies can compare the acoustical experience as a result of binaural and ambisonic presentations (*computed multiple-loudspeaker auralizations*). This will create more generalizable conclusions.

Another possible study concerns the visual rendering. In the current framework, the rendering is performed on a 3-screen rendering (cave-like application). The framework also allows for a visual rendering over an HMD or actual VR cave system. Therefore, future studies could also compare the acoustical experience between a cave-like rendering on the one hand and an HMD or VR cave system on the other, also resulting in more generalizable results.

In order to limit the time usage for participants for the final experiment, only the influence of visuals on the acoustical attributes *Plausibility*, *Distance*, *Loudness*, *ASW*, and *LEV* was studied. In subsequent studies, it should be possible to investigate the effect of visuals on for instance *Reverberance*, *Tonal Color*, and *Intimacy*.

A final study could investigate the addition of an audience to the Théâtre de l'Athénée VR experience both in terms of visuals and acoustics. The visual audience can be modeled by still avatars. However, in real-life visitors make slight movements, which should therefore be included. The acoustical audience could be added by simulating 20 omni-directional sources positioned throughout the audience. The resulting RIRs will be convolved with anechoic recordings of typical audience noise. With these additions the VR application would more closely approximate real-life conditions. Subsequently, studies can compare the current to the future VR experience in order to establish the effect on the acoustical experience.

# A History of the Use of Reflections Arrival Time in Pre-Sabinian Concert Hall Design<sup>1</sup>

---

## A.1 Introduction

Before Sabine’s work on reverberation and its use as a design element, basic ideas of “echo theory” attempted to quantify the perception threshold between direct and reflected sounds. Various methods to determine this threshold resulted in different outcomes. A number of conclusions from these theories on echoes were used in the architectural design of buildings intended for both speech and music. This “echo theory” most resembles recent studies on early reflections.

In a previously published paper this quantified acoustic guideline which influenced the design of three concert halls was discussed [Postma 2013]. It developed from theories on echoes which were proposed during the 17<sup>th</sup>, 18<sup>th</sup>, and 19<sup>th</sup> centuries. In the current article additional literature on “echo theory” is included, new information regarding the use of “echo theory” in the three previously considered concert halls is given, and three new venues are described in which this theory on echoes was used in their design.

Sec. A.2 presents a brief overview of a number of studies carried out concerning the relation between path length differences and the detection of echoes. Sec. A.3 then presents examples of several constructions which used these results as an acoustic and architectural guideline. This work concludes with a discussion of modern criteria in relation to these historic works in Sec. A.4.

## A.2 Echo theory before the discovery of Sabine’s reverberation formula

In 1620, Guiseppe Biancani performed experiments on the occurrence of echoes [Biancani 1620]. He stated that the reflection consisting of the sound of exactly

---

<sup>1</sup>This work was presented in:

- B.N.J. Postma and B.F.G. Katz, *A history of the use of reflections arrival time in pre-Sabinian concert hall design* in Forum Acusticum, (Krakow), pp. 1-6, Sept. 2014.

one syllable was perceived when the observer was located  $\approx 18$  m from a reflecting surface.

In 1636, Marin Mersenne carried out experiments regarding echoes as well [Mersenne 1636]. He found that it took exactly one second to reflect the sound of seven syllables. Because he calculated with a sound speed of 157 m/s, he concluded that a reflection consisting of exactly one syllable was perceivable when the reflected sound traveled  $\approx 22$  m more than the direct sound.

In 1751, Jean le Rond d'Alembert stated that when a violinist played more than ten tones per second, the tones were not separately distinguishable anymore [d'Alembert 1751]. Therefore, he concluded that the human ear could not distinguish tones which followed within a very short amount of time after each other. He concluded that not all reflections were echoes because there was a minimum extra path length reflections had to travel to become separately audible from the direct sound.

In 1759, Joseph-Louis Lagrange described that the human ear can distinguish two tones when they are separated by a time period of at least 63 ms [de Serret 1867].

In 1778, Jean-Etienne Montucla concluded from an experiment with a violinist that the human ear could distinguish 12 tones/s. He therefore stated that an echo was audible when the reflected sound arrived 83 ms later than the direct sound [Ozanam 1778]. Using a sound speed of  $\approx 360$  m/s, he concluded an echo was observable with a path difference of more than  $\approx 30$  m.

In a comparable way to Montucla, Johann Gehler in 1787 determined that the human ear could distinguish 9 tones/s [Gehler 1787]. Using a sound speed of  $\approx 350$  m/s, he concluded that an echo was observable when the reflected sound had an extra path length of  $\approx 39$  m.

In 1800, this theory led to a guideline for architects. Johann Rhode described that through careful observations it had been established that echoes occurred when a wall was more than  $\approx 19$  m away from the observer [Rhode 1800]. He concluded that when all listeners were within 19 m of the sound source in an auditorium the shape did not influence the acoustics. When listeners were  $> 19$  m from the sound source the architect needed to take extra measures for the acoustics.

In 1848, Gustave Schilling stated that the human ear could distinguish ten sounds of speech and song per second [Schilling 1848]. Therefore, he concluded that the difference in path length between direct and reflected sounds should not exceed  $\approx 34$  m.

In 1865, Emil Winkler stated that the average single sound lasts for approximately 125 ms [Winkler 1865]. In this time interval sound travels  $\approx 43$  m. Therefore he argued that a perfect echo occurred when the path difference between direct and reflected sound is 43 m.

In 1869, Rudolphe Radau stated that if the reflection of the first syllable reached the speaker again during the first 100 ms after it was produced it was found to enhance the sound of this syllable [Radau 1869]. Because Radau built in a safety margin, he stated that to create an echo the reflection should travel an extra path length of  $\approx 32$  m instead of Schilling's 34 m.

Table A.1 presents a summary of the different studies on echoes, highlighting the type of sound stimuli used. Some studies resulted in direct distances of these measures, while others were calculated from time measurements and an assumed speed of sound. Due to the large historical variations in the value of this constant, revised conversions have been included to allow for a more coherent comparison.

### A.3 Venues in which these theories were used

#### Iffland Theater, Berlin (1800-1817)

In 1800 Carl Gotthard Langhans wrote about the acoustics of the Iffland Theater, of which he was the architect, referring to Gehler's theory regarding echoes [Langhans 1800]. He stated that in the design of the Iffland theater the maximal distance between direct and reflected sound would be 4 m because of the elliptical shape of the theater. Langhans argued that this would not cause an echo.

Despite the room-acoustical study in advance of the construction, the acoustic properties turned out to be poor. Ludwig Catel argued that echoes were perceivable and hence proposed to install cloth on the inside of the theater [Weinbrenner 1809]. Carl Ferdinand Langhans, son of Carl Gotthard Langhans, tried to explain the poor acoustics of the theater [Langhans 1810]. As superintendent during the construction of the Iffland theater, Langhans Jr. also used Gehler's theory which stated that an echo could be perceived when the reflected sound arrived 111 ms after the direct. He added that an unpleasant prolongation of sound was experienced when the reflected sound arrived between 56 ms and 111 ms later than the direct, which he called reverberation.

Langhans Sr. did not seem to take into account the reflections from surfaces other than the walls. Therefore, Langhans Jr. studied the curved ceiling of the theater which according to him reflected sound which had an extra path length of 16 m or a delayed arrival time of 46 ms. He concluded therefore that echoes or reverberation were not the cause of the disappointing acoustics. Subsequently he turned his eye to the elliptical shape of the theater in which, according to Langhans Jr., focal points would arise. In regard to music he argued that each instrument of an orchestra, while playing near the first focal point of an elliptical auditorium, would have its own acoustical focal point near the second focal point of the elliptical-shaped theater. This would disturb ensemble playing because the sound strengths of the instruments would vary.

Furthermore, regarding the spoken word he stated that speech consists of the connection between vowels and consonants. It is important that the vowels and consonants are not perceived simultaneously. The duration in speech between a vowel and a consonant is shorter than the 56 ms, making reflections, which arrive before 56 ms, important to study. Langhans Jr. argued that due to the sound concentrations in the elliptical theater hall, the reflected sound in the focal points is almost as strong as the direct sound, making at that position the reflections of the previously spoken vowel as loud as the direct sound of the following spoken



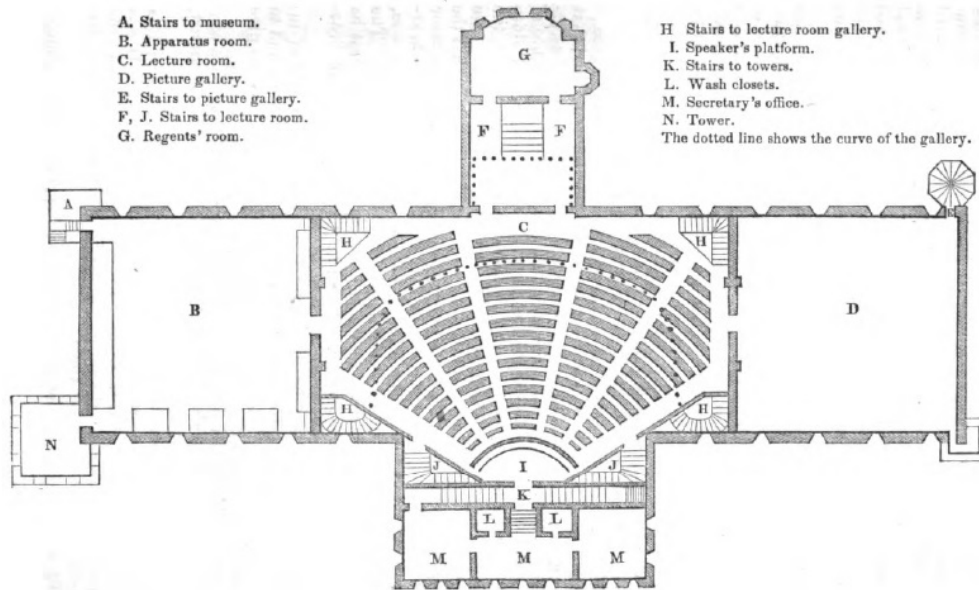


Figure A.1: Map of the lecture hall in the Smithsonian Building. ([Henry 1857] p.232)

consonant, causing an unnatural merging of the two.

### Lecture hall in the Smithsonian Building, Washington D.C. (1854-1865)

In 1857 Joseph Henry described his experiences during the period he was the acoustical consultant of a lecture hall in the Smithsonian Building (see Figure A.1) [Henry 1857]. Unlike Langhans, Henry did his own experiments on the occurrence of echoes. By clapping his hands in front of a wall of the west facade of the Smithsonian Building, he determined the distance at which the direct sound and reflected sound were not separately audible anymore. He found that the reflection was distinctly heard from the direct sound when the distance was  $\approx 9$  m. He called this the “limit of perceptibility”. Using a speed of sound of 343 m/s this would equate to 53 ms to distinguish between two sounds. He however concluded that it took 63 ms. Henry designed the lecture hall to have a height of 8 m which ensured that reflections from this surface reached the audience within the “limit of perceptibility”. Furthermore, he stated that the speaker’s position was between two oblique walls which were positioned within the “limit of perceptibility” of the center of the platform.

### The theaters at the Place du Châtelet, Paris (1862)

The Théâtre du Châtelet and the Théâtre-Lyrique, today called the Théâtre de la Ville, both designed by Gabriel Davioud opened in 1862 [Daly 1865]. In a book published about the design of both theaters the architect discussed the occurrence

of echoes. According to him, the lowest frequency perceivable by the human ear was 32 Hz. He argued therefore that the shortest perceivable interval between two separate tones was  $1/32$  of a second or 31 ms. Using a sound speed of 327 m/s, he argued that the reflected sound had to travel an extra path of  $\approx 10$  m to be heard as an echo, but then stated that this was an extreme case and in practice this limit could be raised to 15 or 20 m.

For this reason, in both designs the acoustic concept was based on the consideration that the back wall of an auditorium, and the ceiling when too high, should never be used to reflect sound back into the hall. Furthermore, the architect stated that only the wall behind the stage in an auditorium could be used for reflections because this surface is close enough to the emission point not to cause an echo.

### Palais du Trocadero, Paris (1878-1934)

Even though Davioud was, together with Jules Bourdais, the architect of the Palais du Trocadero they did not use his earlier statement that echoes occurred when the path difference between direct and reflected sound was more than  $\approx 10$  m. To come to the acoustic concept of this concert hall they used the assumption that echoes arose when the extra path length of the reflections was 34 m, possibly adopted from Schilling's theory [Davioud 1878]. The concert hall was "horse shoe" shaped, 62 m in length and 55 m in height. Therefore, numerous surfaces were  $> 17$  m from the center of the stage and additional measures had to be taken. The architects covered all surfaces  $< 17$  m from the center of the stage with plaster and surfaces located  $> 17$  m from the center of the stage were covered with painted silk.

Davioud and Bourdais realized that this measure alone was not enough to generate satisfying acoustics. Therefore, they placed 100 reflectors on the back wall of the stage. Each reflector directed towards a different  $1/100^{\text{th}}$  of the audience. This was tested with an optical acoustic scale model. The outcome of the first test was that the sound would be evenly distributed over the concert hall. These result however did not satisfy the architects as they thought the most remote parts in the hall needed more reflections. Therefore, the reflectors of the back wall were realigned and the test was repeated. Both architects were satisfied with the results of a second scale model test. Despite this acoustic study, the acoustic properties of the Palais du Trocadero turned out to be poor.

Between 1903 and 1911, under the leadership of Gustave Lyon, an attempt was made to improve the acoustics. He suspected that echoes were the basis of the acoustical difficulties. He based his search for reflecting surfaces causing echoes on the difference of path lengths between the direct and reflected sound [Fournier 1909]. When this differed by:

- a) 0–8.50 m: The sound was warm, colored, and enveloping. These reflections were necessary reinforcement for good listening conditions. Sounds arrive with a maximum delay of 25 ms.
- b) 8.50–11.33 m: These reflections are useful reinforcement. The maximum delay is 33 ms.

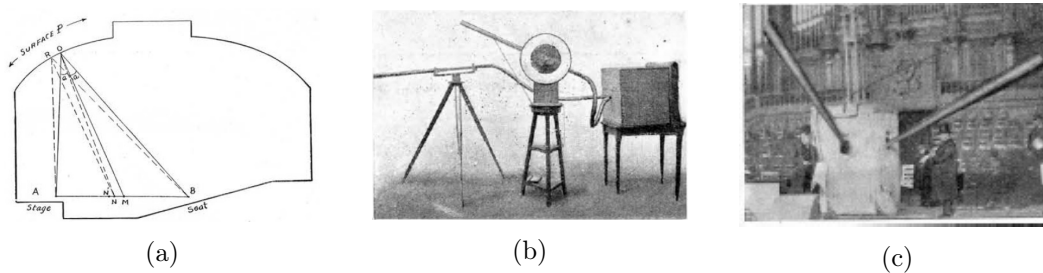


Figure A.2: Palais du Trocadero: (a) Mathematical determination of the surface which caused the echoes ([Anon. 1909] p.462), (b) the smaller sound device to test which surfaces were responsible for the echoes ([Anon. 1909] p.462), and (c) the target apparatus to test which surfaces were responsible for the echoes ([Anon. 1909] p.462).

- c) 11.33–17 m: These reflections are tolerable. The maximum delay is 50 ms.
- d) 17–23 m: These reflections are the cause of resonance and this is the area where it becomes troublesome. The maximum delay is 63 ms.
- e) 23–34 m: These reflections caused echoes and made listening impossible. The maximum delay is 100 ms.

With these starting points, Lyon looked for the surfaces which caused the echoes. He divided the stage into  $2 \times 2$  m sectors from which he made a sound by clapping two pieces of wood together [Anon. 1909]. When auditors, positioned in various seats of the concert hall, heard an echo they raised their hand. Lyon determined, from the results of the experiments and the above five statements about the perception of echoes, in a mathematical way the reflecting elements which caused the echoes, as depicted in Figure A.2a.

To check his theoretical findings, Lyon constructed a device comprising a sound box with two long metal tubes connected to the box via rubber hoses as depicted in Figure A.2b and A.2c. A sound was created in the box which was “aimed” via one of the metal tubes at a surface which supposedly caused the echo and the other metal tube was directly pointed at the position where the echo was supposed to be audible. When the sound was heard double at the test position Lyon concluded that the considered surface was responsible for the echo. To demonstrate this in a more striking manner Lyon devised a second larger apparatus. From his calculations and experiments Lyon concluded that 90% of the echoes were caused by the concave ceiling above the organ, surface *P* in Figure A.2aa. To suppress the reflections Lyon covered this surface with a double layer of cloth with a few inches air space in between.

### Neue Gewandhaus, Leipzig (1884-1944)

Architects Martin Gropius and Heino Schmieden of the Neue Gewandhaus paid special attention to the acoustics of the concert hall. “Echo theory” was adopted

from the Palais du Trocadero. In an article published after construction, Schmieden stated that an echo was perceived when reflected sound arrived 83 ms later than the direct sound [Schmieden 1886]. The architects realized that to avoid echoes in the rectangular form of the Neue Gewandhaus different measures had to be taken than in the round Palais du Trocadero. They decided to breach the wall opposite the stage where possible with columns and cover it with rough cloth. Furthermore, at this wall the openings of the lodges were equipped with thick curtains. To avoid flutter echoes between the parallel side walls, these were breached with columns which divided these walls into seven zones. These zones were covered with tapestry-like paintings.

### Salle Pleyel, Paris (1927)

Gustave Lyon was involved in the acoustic design of another Parisian concert hall, the Salle Pleyel [Calfas 1927]. Before he began, Lyon undertook several acoustical experiments. He tested how far his voice reached in a wide open field covered with snow arriving at a distance of  $\approx 11$  m. Because the same voice carried for 2400 m over the lake of Annecy, Lyon regarded the action of reflectors as essential. To establish when echoes occur, Lyon undertook a second experiment. High in the mountains, two assistants simultaneously hit the ice with their ice axes at various distances. Lyon perceived the sound together unless the distance between both assistants was  $> 22$  m. Therefore, he concluded that reflected sound arriving  $> 63$  ms after the direct sound is perceived as an echo. Furthermore, he stated that this time difference was not the same for every sound, being longer for speech and music.

Lyon chose for the Salle Pleyel a shape and proportions to avoid echoes by ensuring that at every position the difference in path lengths between direct and reflected sound was  $\leq 22$  m. The concert hall was fan shaped with a maximum height of 20 m. The width of the stage was 20 m at the back and 21.5 m at the front. Furthermore, Lyon divided the hall into three stacked horizontal zones. For these zones the side walls were inwardly inclined to act as reflectors. The wall behind the orchestra possessed three different inclinations for the same purpose, as shown in Figure A.3.

### Summary

The relevant studies in this section are also included in the summary Table A.1. It shows that except for Biancani, acousticians who performed experiments using a speech syllable or musical impulses measured a time interval and then calculated the distance for which an echo would occur, while acousticians performing experiments using impulse stimulus determined the distance and then calculated the corresponding time interval. The average minimum time interval (sound speed corrected) for a perceived echo was 112 ms for a syllable, 98 ms for a musical tone, and 60 ms for an impulse stimulus.

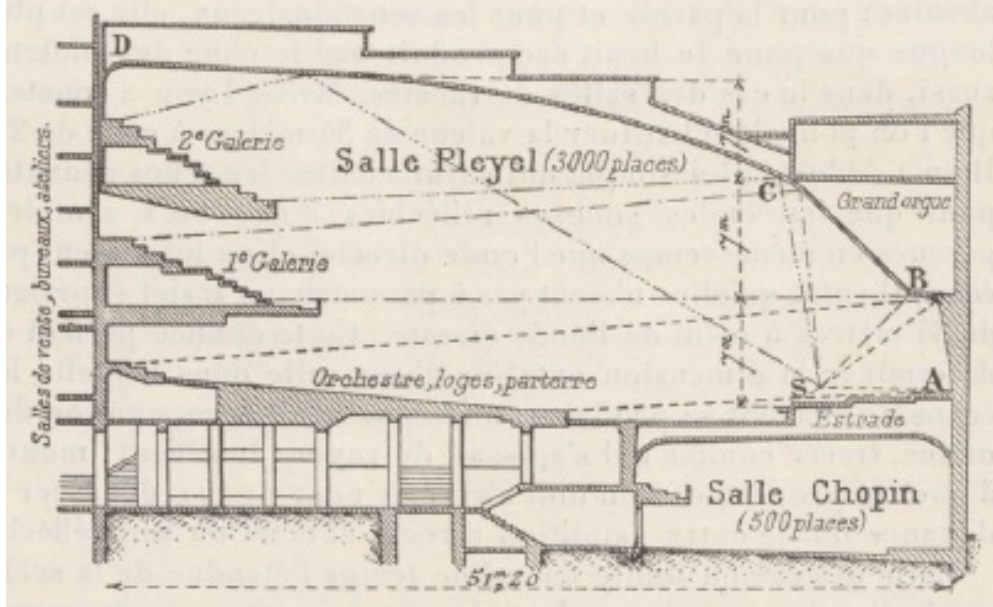


Figure A.3: Longitudinal section of the Salle Pleyel. ([Calfas 1927] p.424)

#### A.4 Current findings regarding early reflections

The described works can be compared to the contemporary use of early reflections. The development of the parameters Definition, Clarity, Initial Time Delay Gap, and Center Time in the second half of the 20<sup>th</sup> century indicate the importance of designing with early reflections. Furthermore, it has been found that the time interval when early reflections are considered beneficial for the acoustics differs between speech and music. In this respect, Thiele in 1953 concluded that reflected sound was detrimental for speech when it arrived  $> 50$  ms later than the direct sound [Thiele 1953]. In 1975, Reichardt et al. found that beneficial reflections for music arrive within 80 ms of the direct sound [Reichardt 1975].

More recently, Jurkiewicz et al. [Jurkiewicz 2012] in 2012 have stated that the acoustic success of a concert hall is possibly more dependent on early reflections than on reverberation [Jurkiewicz 2012]. This study proposed a formula to calculate early energy termed mean early strength ( $G_{em}$ ) which is influenced by the surface area of the audience ( $s_{aud}$ ), the total efficient solid angle ( $\Omega_{eff}$ ), and the average angle of incidence of early reflections on the audience ( $\Omega_m$ ). According to this formula a larger audience area results in a lower  $G_{em}$ . The  $\Omega_{eff}$ , which is the solid angle of all efficient surfaces measured from the omni-directional sound source, can only be increased to a certain degree. The authors stress the importance of  $\Omega_m$ . To create higher  $G_{em}$  in very large rooms, early reflections should arrive from surfaces positioned low in the room. Therefore, the authors conclude that reflectors with more visibility from the stage have a larger positive effect on  $G_{em}$ . However, they realize that increasing the solid angle to very large values increases the risk of excessive loudness. Furthermore, they state that the design should allow enough

Table A.1: Time interval and distance of first reflection resulting in the perception of an echo in historical studies. Values in [-] determined from measured values in associated publications. Values indicated by † determined from statements on how many tones/s were separately audible. Revised values are provided using a common sound speed for comparison.

stimulus	Author	reported interval (ms)	reported distance (m)	assumed speed of sound (m/s)	Recalculated values ( $v = 343$ m/s)	
					interval (ms)	distance (m)
Syllable	Biancani		36		105	36
	Mersenne	142†	[22]	157	142	49
	Schilling	100	[34]	340	100	34
	Radau	100	[32]	340	100	34
Musical	d'Alembert	100†			100	34
	Montucla	83	[30]	360	83	28
	Gehler	111	[39]	350	111	38
Impulse	Henry	[63]	18	343	53	18
	Lyon	[67]	22	340	67	22
no stated stimulus	Lagrange	67†			67	23
	Davioud	31	[10]	327	31	11
	Winkler	125	[43]	343	125	43

energy to remain for the late reverberant field.

## A.5 Discussion & Conclusion

During the 17<sup>th</sup>, 18<sup>th</sup>, and 19<sup>th</sup> centuries a number of studies were carried out to determine the threshold when an echo was heard. These studies either directly determined the distance at which an echo was audible, used a certain number of tones per second which were separately audible, or determined the time it took to reflect the sound consisting of exactly one syllable. The selected method influenced the determined time interval or propagation path distance difference for which an echo was audible. In contrast to our current understanding, early acousticians determined a shorter time interval for musical stimuli than for a speech syllable stimulus. However, modern values are related to impressions of clarity and intelligibility, and not directly to single echo perception. As expected, acousticians who used an impulsive stimuli obtained shorter values.

These theories on echoes were used as quantified acoustical and architectural guidelines in at least seven venues with both speaking and musical purposes. The architects of these venues were correct in assuming that reflections that arrived shortly after the direct sound are beneficial for the perception of music and speech. However, the architects probably did not base the decision of the employed time

interval on the stimulus which was used during the experiment. For the venues with speaking purposes, “echo theories” were used which were based on a musical stimulus, impulse stimulus, or on a theory without reference to a stimulus (Iffland Theater = a fast playing violin, the lecture hall in the Smithsonian building = clapping of hands, and the theaters at the Place du Châtelet = theory without a stated stimulus), while the venues with musical purposes were based on an impulse or syllable stimulus (The Palace du Trocadero and the Neue Gewandhaus = syllable stimulus, and the Salle Pleyel = hitting axes on ice).

Nevertheless, it should be noted that the time intervals adopted for venues with musical purposes (Palais du Trocadero = 100 ms, Neue Gewandhaus = 83 ms, and Salle Pleyel = 67 ms) were significantly longer than those used in venues with speech purposes (Iffland Theater = 46 ms, lecture hall in the Smithsonian Building = 53 ms, and the theaters at the Place du Châtelet = 31 ms), which is in agreement with current knowledge. It can be observed that average time intervals used in venues with musical purposes (83 ms) and venues with speech purposes (40 ms) greatly resemble the currently used delimiting values of 80 ms for music and 50 ms for speech in the Clarity and Definition parameters. This is notable considering the limited measuring equipment available to 19<sup>th</sup> century acousticians.

When these historical findings are compared to parameters such as Clarity, Definition, Initial Time Delay Gap, and Center Time, the early applications of echo theory only take into account first order reflections while current parameters take the early acoustical energy into account. Furthermore, recent studies, such as Jurkiewicz, propose to take into account the arrival direction of the reflections and the energy distribution over the room.



# Virtual Reality Performance Auralization in a Calibrated Model of Notre-Dame Cathedral<sup>1</sup>

---

## B.1 Introduction

The use of Virtual Reality (VR) technologies has increased over the last decennia due to the improvement of available computing power and the quality of numerical modelling software. This study explored the current potential of VR technologies which combine auralizations and 3D graphics. The global concept of this project was to present a complex VR scene, with numerous acoustical sources, in which the listener could move around having a realistic experience throughout the regarded venue.

Several studies have reconstructed historical sites in terms of audio and visuals. The ERATO project [Magnenat-Thalmann 2006] constructed acoustical and visual models of archaeological open-air and roofed theatres. Acoustical simulations used the GA software ODEON. Visual reconstructions were created with the *3ds Max* software based on architectural drawings, photos, and videos. The visitor was able to navigate within the visual scene. Auralizations were linked to interactive area triggers, allowing the visitor to perceive and experience the simulated voices from specific positions.

Game engines are a useful platform for combining visuals and audio in VR applications [Moloney 2004]. They offer interactive rendering of visual environments while also enabling the integration of audio and visuals. Lindebrink et al. [Lindebrink 2015] employed a software platform combining the game engine TyrEngine and the room acoustical software BIM/CAD. RIRs were calculated and convolved with anechoic recordings offline. When progressing through the visual scene, the audio rendering was performed by playing the sound file of the nearest neighbor.

Another VR application [Taylor 2010a, Taylor 2010b] created audio-visual scenes employing the game engine *Gamebryo* and rendered the room response in real-time.

---

<sup>1</sup>This work was presented in:

- B.N.J. Postma, D. Poirier-Quinot, J. Meyer, and B.F.G. Katz, *Virtual reality performance auralization in a calibrated model of Notre-Dame Cathedral* in Euroregio, (Porto), pp. 1-10, June 2016.



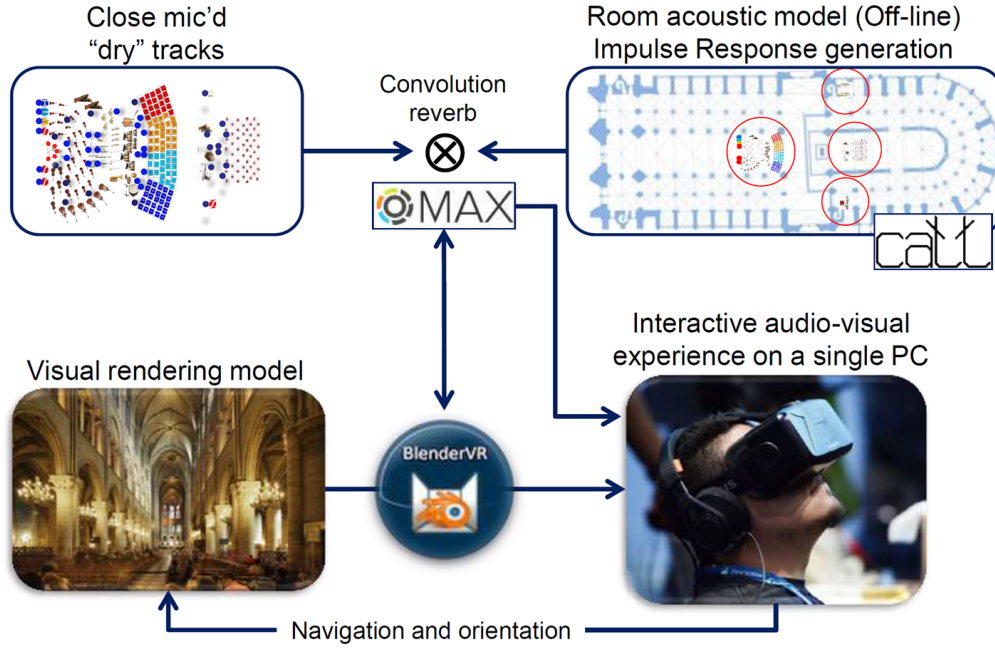


Figure B.1: Schematic diagram of the architecture of the VR experience

In order to enable real-time convolution, the RIR was divided into an early and late part. An underlying GA based algorithm computed specular reflections, diffuse reflections, and edge diffraction on a multi-core system. The late reverberation time was simulated by a statistical estimation technique. Physical restrictions were imposed on the motion of source and receiver to generate an artifact-free rendering. In 2010, this application was used to present the visual rendering of the Sibenik cathedral at 20-30 fps in combination with a binaural audio representation of 12 instruments, taking into account the listener's position and orientation.

As with these discussed studies, the current project employed a game engine platform to create an audio-visual reconstruction of an orchestral performance of '*La Vierge*' in the Notre-Dame cathedral. The cathedral's complicated geometry and considerable dimensions (length:  $\sim 130$  m, width:  $\sim 48$  m, height:  $\sim 33$  m, volume:  $\sim 84,000$  m<sup>3</sup>) as well as the number of musicians results in a complex scenario. In contrast to previous studies, the audio-visual rendering was designed to be suitable for a tracked Head Mounted Display (HMD), requiring a higher frame rate than ordinary desktop screens. The combination of a complex scene with the high technical requirements rendered the audio-visual reconstruction suitable to explore the contemporary potential of VR platforms. As the study proposed an exploration of technology, emphasis was placed on identifying technological limitations and perceptual aberrations.

## B.2 Project overview

The first step was to conceive the global project architecture. A recording was made of the ‘*La Vierge*’ concert in the Notre-Dame cathedral. These recordings were convolved with 3rd order Ambisonic RIRs obtained from the calibrated Notre-Dame GA model. In parallel, a visual model was created of the Notre-Dame cathedral in *3ds Max*<sup>2</sup>, subsequently ported to the *Blender Game Engine*<sup>3</sup>. The visual and acoustical models were integrated using a platform which combined the interactive VR environments of BlenderVR and the audio software *Max/MSP*. Fig. B.1 depicts the conceptual architecture of the presented audio-visual VR application.

This paper presents an overview of the different essential elements necessary to achieve the proposed immersive VR experience according to the proposed global architecture.

## B.3 Recordings

On 24-April-2013, a grand concert was organized in the Notre-Dame cathedral, to celebrate its 850<sup>th</sup> anniversary. A symphonic orchestra, 2 choirs, and 7 soloists performed ‘*La Vierge*’, composed by Jules Massenet in 1880. Fig. B.2a depicts the placement of the instruments and section microphones during the concert. The event was recorded by the Conservatoire de Paris and made accessible to this study thanks to the BiLi project. Each instrument section and soloist were recorded using a total of 44 microphones in close proximity. As the direct-to-reverberant ratio is high for close mic recordings, these were employed as approaching anechoic recordings for the purpose of auralization.

## B.4 Room-acoustic model

Source positions were defined in the calibrated Notre-Dame cathedral GA model according to the 44 microphones used during the recording. In order to approximate the directivity properties of instruments, directivity patterns were defined in the GA model, based on [Olson 1967, Le Carrou 2010, Marshall 1985, Dalenbäck]. Additional details can be found in [Meyer 2015]. All sources were aimed at the conductor position. As no directivity data for a harmonium or glockenspiel was found, these were defined as omnidirectional sources. To reduce the number of RIRs necessary to compute, and to limit GPU/CPU load at run-time, 1D linear, as opposed to 2D planar, interpolating was selected for the interactive navigation component of the VR application. A single trajectory path was defined along which the visitor was free to move (see Fig. B.2). Receiver positions were defined along this trajectory at  $\sim 3$  m intervals, resulting in 88 receiver positions. The use of 3rd order

<sup>2</sup>3ds Max: 3D modelization software. <http://www.autodesk.fr/products/3ds-max>

<sup>3</sup>Blender: 3D creation software. <https://www.blender.org>

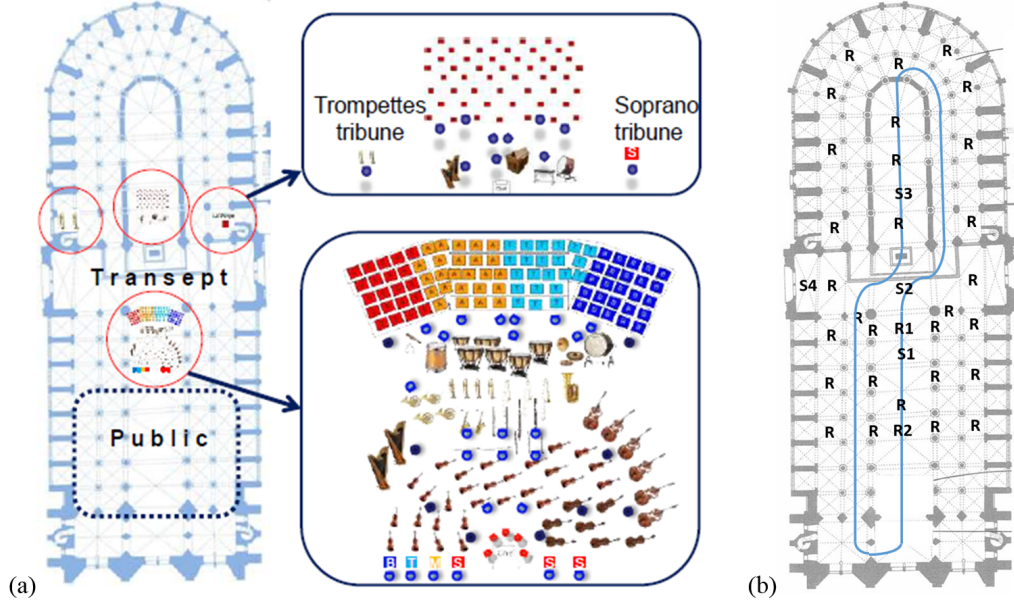


Figure B.2: (a) Orchestra and microphone ( $\bar{o}$ ) layout for the concert in the Notre-Dame cathedral. (b) Measurement plan in the Notre-Dame cathedral. S and R represent source and receiver positions. The blue line depicts the VR trajectory experience.

Ambisonic microphone RIRs allowed for real-time binaural conversion of the HOA audio stream, taking into account the HMD's head orientation at run-time.

## B.5 Visual model

### Visual model of the Notre-Dame cathedral

To accompany the virtual acoustic reconstruction, a 3D visual model of Notre-Dame cathedral was created. The visual model was created in 3ds Max and subsequently ported to the Blender Game Engine. The model geometry was based on a 3D laser scan point cloud, plans and sections, as well as visual inspection. The final model consisted of  $\sim 500,000$  triangles. The textures employed in the model were based on photos taken during on-site visits. Fig. B.3 shows a photo of the cathedral and the Blender visual model.

### Visual animations

In order to create a more engaging immersive VR experience, animations were added to the static cathedral *Blender* model. First, instruments were represented in the 3D environment and positioned in the virtual cathedral to visualize the different components of the orchestra. These instruments included an animated shadow, changing shape as a function of the associated audio track's amplitude in real-time.



Figure B.3: (a): Picture of the Notre-Dame de Paris cathedral from the altar towards the organ (from [Murray a]) . (b): Similar view in the Blender VR model.

Second, a ‘magic carpet’ was added, upon which the visitor sat while exploring the venue. The carpet provided a visual anchor for the ‘plausibility’ of flying while also avoiding the visual sensation of being suspended several meters in the air with no support, as the HMD allows one to freely look in all directions. On the magic carpet, visitors were free to progress along the predefined trajectory path at a user defined speed and direction using a simple forward-backward joystick controller.

## B.6 Integration of acoustics and visuals to create the VR experience

### Integration architecture

Visual models and room-acoustic simulations were integrated using *BlenderVR* and the audio software *Max/MSP*. The visitor was able to navigate along the trajectory path in real-time within the visual model in the *Blender Game Engine*. Visitor’s position (magic carpet) and head orientation (HMD) were communicated to *Max/MSP* which used the position information to perform an amplitude interpolating between the two nearest HOA receiver positions. Subsequently, head orientation was employed as an HOA rotation prior to decoding the panned Ambisonic stream auralization to the final binaural rendering (see [Noisternig 2003a, Picinali 2014]).

### Technological limitations

In order to create a smooth audio-visual VR application, it was required that (1) the auralizations were perceived without audible crackles and (2) the visuals were rendered with a sufficiently high frame rate as well as no visual pixelization. In the design process several technological limitations regarding the VR experience were encountered.

The audio was originally conceived to be rendered via real-time convolutions of the 3rd order Ambisonic RIRs with the recordings. In order to prevent audible artifacts due to buffer updating and fast speeds, 5 audio buffers for nearest-neighbor receiver positions needed to be loaded and processed for the 1D linear interpolating. This resulted in the real-time convolution of 3,520 channels (5 receiver positions  $\times$  16 channel HOA RIR  $\times$  44 instrument tracks). As this put too great a demand on the CPU for a single PC application, it was decided to perform the convolutions offline. Creating preconvolved HOA audio for each received position leads to a drastic increase in the data storage for the application (88 receiver positions  $\times$  24 bit audio wav  $\times$  16 channel audio =  $\sim$ 12 GB/min). Consequently, the present VR example demonstration application was limited to a 6-minute audio extract, instead of the entire concert performance.

Visually, it was intended that the visitor could see the entire length of cathedral with the animated orchestra, giving the dynamic option to select the time of the day for the visit, comprising an adjustable lighting scene. As a selectable light scene put too much strain on the GPU/CPU, it was decided to ‘bake’ the shadows according to a single lighting scheme: night-time conditions. Additionally, the depiction of the entire cathedral from one end to the other leads to pixelization issues for distant elements with the HMD’s resolution. Therefore, visuals beyond a distance of  $\sim$ 40 m were ‘clipped’. These decisions resulted in a VR experience which can only present night-time conditions, using a dynamic illumination which followed the visitor’s progression along the predefined trajectory, lighting only the nearby sections.

### Rendering

The created VR experience demonstration was made available on two platforms with different performance requirements. The first option allowed exploration of the rendering interactively along the path with an HMD (*Oculus Rift DK2*). The visitor was able to control their speed as well as the direction of the ‘magic carpet’ and the entire field of view was available, providing a highly immersive solution. A second lighter option enabled exploration via a tablet. As the *Blender Game Engine* (the foundation of *BlenderVR*) was unable to run complex environments in real-time on a standard tablet, it was necessary to pre-render the visual part of the scene using *Blender Cycles*. This required predefining the visitor’s progression along the path and rendering a high definition 360° video and associated single HOA mixed audio track. This approach allowed for exploration of the cathedral through a tablet equipped with orientation sensors, behaving like an orientable window to the

360° virtual world. The tablet's orientation is used to rotate both the visual 360° rendering and the HOA stream which is then converted to binaural in real-time.

### Observed perceptual artifacts

Several artifacts were observed regarding both the acoustical and visual rendering. Two issues were caused by limitations of the 'dry track' assumption of the recording which actually contained a non-negligible level of cross-talk between audio tracks (i.e. other instrument sections). First, when close to a given instrument's position, the sound from different sections seemed to spatially blur instead of being able to distinguish the positions of separate instrument groups. Second, as the visual avatar animations were based on the audio track levels, during loud passages the visual avatars all appeared active instead of only the active instrumental sections (e.g. kettle drum instances). Finally, when the trajectory passed behind pillars, the expected acoustical variations were absent. For these positions the acoustics vary considerably for relatively small displacements. As such, the chosen RIR calculation/interpolating step size (3 m) was probably too large for the rate of architectural variations, resulting in an amplitude interpolating result that did not correctly represent the expected acoustical details.

## B.7 Conclusion

The current potential of VR technologies which combine realistic auralizations and 3D graphics in complex geometries was explored. For this purpose, an ambitious project attempted to reconstruct a large concert in the Notre-Dame cathedral in Paris. Visitors were able to experience this immersive interactive audio-visual VR application on high performance system as well as a portable platform. The presented application enabled realistic audio-visual visits to the complex scene of an extract of the '*La Vierge*' concert.

There remains room for improvements regarding the immersion and accuracy of this and comparable VR scenes. For non-cluster based renderings, today's available GPUs/CPU's limit the presented application in several ways. With increased available computational power, the inclusion of real-time convolution and lighting scene of choice will become possible and consequently comparable VR applications will become more immersive and interactive.

In order to provide a more realistic representation of the reconstructed sound scene, it is recommended that one carefully considers the receiver positions in the room-acoustic model with regards to expected spatial variations of the sound field. A denser receiver grid could be used at locations where acoustical variations are expected to vary considerably with relative small displacements, though the use of a denser grid may impose limitations on movement speed in order to avoid audio buffer switching artifacts. Alternatively, such highly varying areas could be excluded from visitor accessibility.

Additional information regarding this work and YouTube videos of the final rendering can be found at [groupeaa.limsi.fr/projets:ghostorch](https://groupeaa.limsi.fr/projets:ghostorch).

# Instructions perceptual tests

---



## Comparison 1 Study

### Instructions

Bart Postma

postma@limsi.fr

### Introduction

Thank you for agreeing to participate in this study as part of the ECHO project.

In this test you will be asked to compare pairs of audio samples regarding subjective attributes relating to room acoustics. Should you wish to pause the experiment at any time, you may do so.

### Starting the test

The test supervisor will start the test by typing your name in the box on the top of the page and choosing the list you will listen to. Press the “Confirm” button to go to the test screen, depicted in the Fig. below.

2/23  
iteration / max iterations

A Off On B

Submit

Reverberance	Duration of the reverberant decay. Well audible at the end of signals.	A is much longer	Neutral	B is much longer
Clarity	Clarity/clearness with respect to any characteristic of elements of a sound scene.	A is much clearer	Neutral	B is much clearer
Distance	Source distance from this location	A is much further away	Neutral	B is much further away
Tonal Balance	Timbral impression which is determined by the ratio of high to low frequency components.	A is much brighter	Neutral	B is much brighter
Coloration	modifications in the timbre of the sound source from its original timbre	A is much more colored	Neutral	B is much more colored
Plausability	Plausability of the sound file	A is much more plausible	Neutral	B is much more plausible
Apparent Source Width	Perceived extent of a sound source in horizontal direction.	A is much wider	Neutral	B is much wider
Listener envelopment	Sensation of being spatially surrounded by the reverberation.	A is much more enveloping	Neutral	B is much more enveloping

In the test you will compare 23 pairs of sound files, which can be omnidirectional or binaural. If the sound files are omnidirectional the sound files are compared according to the following attributes: *reverberance*, *clarity*, *distance*, *coloration*, and *plausability* (definitions are given below). If the sound files are binaural, additionally the attributes *apparent source width* and *listener envelopment* (definitions are given below) are compared.

To listen to sound file A press button A; to listen to B press button B. You can listen to the sound files as many times as needed. **Listen first at least once or twice to both complete sounds files before you start rating them!**

The response slider is a continuous scale going from ‘A is much more...’ to ‘B is much more ...’. A response in the middle means that there is no difference between the samples for that attribute. The scale is comparable for all attributes.

Click on the scale to rate the corresponding attribute.

If you have additional comments which cannot be explained by the given attributes, or if you want to give more details, please note them down in the text box right next to the sliders.

Press submit to go to the next comparison once you have evaluated all attributes. Moving to the next page will not be possible unless you have listened to both audio samples and rated all attributes by clicking on each of them at least once.

During the first three pairs of audio files the test supervisor will remain with you to answer possible questions.

### **Acoustical attributes**

#### Reverberance

'Reverberance' is the perception of the decay of sound.

#### Clarity

'Clarity' is the degree to which discrete sounds in a musical performance stand apart from one another. If clarity is high, it is easy to spot individual notes in a musical piece, or individual phonemes in speech. If clarity is low, individual sounds merge, blend, or at the extreme can be confused and muddled.

#### Distance

Distance is the perceived distance to the sound sources.

#### Coloration

Coloration represents changes in timbre, or frequency balance. It is qualified here as a comparison of the ratio of high to low frequency components so that more coloration indicates more high frequency content.

#### Plausibility

Given the assumption that the recording was made in a church (with the soprano singing) or in a theatre (with the woman talking): Does the recording sound reasonable to you?

#### Apparent Source Width

Apparent source width describes the perceived width of a sound image. The source may sound 'narrow' (in the extreme case it is as if the sound is coming from a point). On the contrary the source can also sound very 'wide'.

#### Listener Envelopment

Listener envelopment describes the impression of the strength and directions from which the reverberant sound seems to arrive. "Listener envelopment" is judged highest when the reverberant sound seems to arrive at equally from all directions, to be surrounded by the sound.

**Thank you for your participation!**

## Comparison 2 Study

### Instructions

Bart Postma

postma@limsi.fr

### Introduction

Thank you for agreeing to participate in this study as part of the ECHO project.


In this test you will be asked to compare pairs of audio samples regarding subjective attributes related to room acoustics. You may pause the experiment at any time.

Audio examples were made using a dummy head microphone.

### Starting the test

The test supervisor will start the test by typing your name in the box on the top of the page and choosing the list you will listen to. Press the “**Confirm**” button to go to the test screen, depicted in the Fig. below.

2/23  
iteration / max iterations




A

Off

On

Submit



B

Reverberance	<div style="text-align: center; font-size: 0.8em;">Duration of the reverberant decay. Well audible at the end of signals.</div> <div style="display: flex; justify-content: space-between; font-size: 0.8em;"> <span>A is much longer</span> <span>Neutral</span> <span>B is much longer</span> </div> <div style="height: 20px; background: linear-gradient(to right, #ccc 49%, #ccc 49% 51%, #ccc 51%);"></div>
Clarity	<div style="text-align: center; font-size: 0.8em;">Clarity/clearness with respect to any characteristic of elements of a sound scene.</div> <div style="display: flex; justify-content: space-between; font-size: 0.8em;"> <span>A is much clearer</span> <span>Neutral</span> <span>B is much clearer</span> </div> <div style="height: 20px; background: linear-gradient(to right, #ccc 49%, #ccc 49% 51%, #ccc 51%);"></div>
Distance	<div style="text-align: center; font-size: 0.8em;">Source distance from this location</div> <div style="display: flex; justify-content: space-between; font-size: 0.8em;"> <span>A is much further away</span> <span>Neutral</span> <span>B is much further away</span> </div> <div style="height: 20px; background: linear-gradient(to right, #ccc 49%, #ccc 49% 51%, #ccc 51%);"></div>
Tonal Balance	<div style="text-align: center; font-size: 0.8em;">Timbral impression which is determined by the ratio of high to low frequency components.</div> <div style="display: flex; justify-content: space-between; font-size: 0.8em;"> <span>A is much brighter</span> <span>Neutral</span> <span>B is much brighter</span> </div> <div style="height: 20px; background: linear-gradient(to right, #ccc 49%, #ccc 49% 51%, #ccc 51%);"></div>
Coloration	<div style="text-align: center; font-size: 0.8em;">modifications in the timbre of the sound source from its original timbre</div> <div style="display: flex; justify-content: space-between; font-size: 0.8em;"> <span>A is much more colored</span> <span>Neutral</span> <span>B is much more colored</span> </div> <div style="height: 20px; background: linear-gradient(to right, #ccc 49%, #ccc 49% 51%, #ccc 51%);"></div>
Plausability	<div style="text-align: center; font-size: 0.8em;">Plausability of the sound file</div> <div style="display: flex; justify-content: space-between; font-size: 0.8em;"> <span>A is much more plausible</span> <span>Neutral</span> <span>B is much more plausible</span> </div> <div style="height: 20px; background: linear-gradient(to right, #ccc 49%, #ccc 49% 51%, #ccc 51%);"></div>
Apparent Source Width	<div style="text-align: center; font-size: 0.8em;">Perceived extent of a sound source in horizontal direction.</div> <div style="display: flex; justify-content: space-between; font-size: 0.8em;"> <span>A is much wider</span> <span>Neutral</span> <span>B is much wider</span> </div> <div style="height: 20px; background: linear-gradient(to right, #ccc 49%, #ccc 49% 51%, #ccc 51%);"></div>
Listener envelopment	<div style="text-align: center; font-size: 0.8em;">Sensation of being spatially surrounded by the reverberation.</div> <div style="display: flex; justify-content: space-between; font-size: 0.8em;"> <span>A is much more enveloping</span> <span>Neutral</span> <span>B is much more enveloping</span> </div> <div style="height: 20px; background: linear-gradient(to right, #ccc 49%, #ccc 49% 51%, #ccc 51%);"></div>

In the test you will compare 23 pairs of sound examples.

The sound examples are to be compared for the following attributes (definitions are given below):

- *Reverberance*
- *Clarity*
- *Distance*
- *Tonal balance*
- *Coloration*
- *Plausability*
- *Apparent Source Width*
- *Listener envelopment*

To listen to sound example A press button **A**; to listen to B press button **B**.

You can listen to the sound files as many times as needed. We suggest that you listen to both complete examples once or twice before you start rating them.

The response slider is a continuous scale going from ‘**A is much more...**’ to ‘**B is much more ...**’. A response in the middle means that you perceive no difference between the samples for that attribute. The scale size is comparable for all attributes. Click on the scale to rate the

corresponding attribute.

If you have additional observations which are not represented by these attributes, or if you want to give more detailed comments, please add them in the text box next to the sliders.

Once you have evaluated all attributes, press “**Submit**” to go to the next pair. You will not be able to proceed until you have listened to both sound samples and rated all attributes by clicking on each of them at least once.

The first three pairs are for you to get familiar with the test. During these three pairs the test supervisor will remain with you to answer any questions you may have.

### **Acoustical attributes**

#### Reverberance

The perception of the decay of sound. More reverberance is associated to a longer decay.

#### Clarity

The degree to which discrete elements in the recorded musical performance stand apart from one another. If clarity is high, it is easy to spot individual notes in a musical piece, or individual phonemes in speech. If clarity is low, individual sounds merge, blend, or at the extreme can be confused and muddled.

#### Distance

The perceived distance to the sound source in the recording space.

#### Tonal balance

Tonal balance represents changes in timbre, or frequency balance. It is qualified here as a comparison of the ratio of high to low frequency components so that more ‘tonal balance’ indicates more high frequency content.

#### Coloration

Coloration represents modifications in the timbre of the sound source from its original timbre. With less coloration a file sounds more natural.

#### Plausibility

Given the assumption that the recordings were made in a church (singing voice) or in a theatre (spoken voice) with the specified recording technology: Does the recording sound reasonable to you?

#### Apparent Source Width

Apparent source width describes the perceived width of a sound image. The source may sound ‘narrow’ (in the extreme case it is as if the sound is coming from a point). On the contrary the source can also sound very ‘wide’.

#### Listener Envelopment

Listener envelopment describes the spatial distribution of the reverberant sound field. Higher “Listener envelopment” means a more uniform distribution, while less “Listener envelopment” means a more localized or directional reverberant sound.

**Thank you for your participation!**

## Comparison Study 3

### Instructions

Bart Postma

postma@limsi.fr

### Introduction

Thank you for agreeing to participate in this study as part of the ECHO project. In this test you will be asked to rate 5 audio samples regarding subjective attributes related to room acoustics. You may pause the experiment at any time.

The audio examples have been recorded in the same small wooden and fabric-covered theater using a dummy head with microphone placed in the public, while a singer on stage was recorded. You should therefore imagine yourself listening to a \*live\* performance within a theatre. You will hear examples from different seats within the theatre. Please keep your head relatively still, looking forward, during listening to avoid perceptual conflicts of your head orientation relative to that of the recording dummy head.

### Starting the test

The test supervisor will start the test by typing your name in the box on the top of the page and choosing the list you will listen to. Press the “**Confirm**” button to go to the test screen, depicted in the Fig. below.

The interface shows five audio samples labeled A, B, C, D, and E. Above them is a 'Submit' button and a progress indicator '1/7 iteration / max iterations'. Below the samples are five rating scales for different attributes:

- Plausibility/Realism for listening to a live operatic vocal performance in a theatre.** Scale from 'least plausible' to 'most plausible'.
- Clarity/clearness with respect to any characteristic of elements of a sound scene.** Scale from 'least clear' to 'most clear'.
- Distance of the singer from your seating position.** Scale from 'closest' to 'furthest'.
- Horizontal extent of the acoustic image of the singer.** Scale from 'least wide source' to 'most wide source'.
- Sensation of being surrounded by the sound of the singer and room.** Scale from 'least enveloping' to 'most enveloping'.

On the left, there are labels for each attribute: Plausibility, Clarity, Distance, Apparent Source Width, and Listener Envelopment. On the right, there is a text box for additional comments with the prompt: 'If you have additional comments which cannot be explained by the given attributes, or if you want to give more details, please note them down here.' and a 'no comment' button.

In the test you will rate 7 groups of sound examples. The top right corner will tell which iteration you are at. The sound samples are to be rated for the following attributes (definitions are given below): *Plausibility, Clarity, Distance, Apparent Source Width, and Listener envelopment.*

To listen to sound example A press button **A**; to listen to B press button **B** ... etc.  
You can listen to the sound files as many times as needed. We suggest that you listen to all complete examples once or twice before you start rating them.

The task is to compare and rate the five sound samples per acoustical attribute on a discrete scale from 1 (**'least ...'**) to 7 (**'most ...'**). You must use the 2 extreme scale values (1 and 7) at least once per attribute to be able to submit the rating. Multiple sound samples may have the same rating.

If you have additional observations which are not represented by these attributes, or if you want to give more detailed comments, please add them in the text box next to the sliders.

Once you have evaluated all attributes, press **"Submit"** to go to the next iteration.

The first iteration is for you to get familiar with the test. During this iteration the test supervisor will remain with you to answer any questions you may have.

#### **Acoustical attributes**

##### Plausibility/Naturalness

Given the assumption that you are listening to a live operatic performance of a singing soprano in the same small wooden and fabric-covered theater. Which recording sounds most natural to you?

##### Clarity/ Intelligibility of the singing voice

The degree to which discrete elements in the recorded musical performance stand apart from one another. If clarity is high, it is easy to spot individual notes in a musical piece, or individual phonemes in speech. If clarity is low, individual sounds merge, blend, or at the extreme can be confused and muddled.

##### Distance

Distance of the singer from your seating position.

##### Apparent Source Width

Apparent source width describes the perceived horizontal extent of the acoustic image of the singer. The source may sound 'narrow' (in the extreme case it is as if the sound is coming from a point). On the contrary the source can also sound very 'wide'.

##### Listener Envelopment

Listener envelopment describes the sensation of being surrounded by the sound of the singer and room. Higher "Listener envelopment" means a more uniform distribution, while less "Listener envelopment" means a more localized or directional reverberant sound.

**Thank you for your participation!**

## Comparison Study 4

### Instructions

Bart Postma

postma@limsi.fr

### Introduction

Thank you for agreeing to participate in this study as part of the ECHO project. In this test you will be asked to rate 3 audio samples regarding subjective attributes related to room acoustics. You may pause the experiment at any time.

The audio examples have been recorded in a small theater (see Figure below) using a binaural microphone in the audience, while a play with two actors was performed on stage. You should therefore imagine yourself listening to a \*live\* performance within this theatre. You will hear examples from different seats within the theatre. You are asked during listening to keep your head relatively still, looking forward as there is no head-tracking.

### Starting the test

The test supervisor will start the test by typing your name in the box on the top of the page and choosing the list you will listen to. Additionally on this page a video is presented which will provide you with a reference for the attribute *Plausibility*. Watch this video paying special attention to the movement of the actors and the orientation of their heads. Once you have memorized these two aspects, press the “**Confirm**” button to go to the test screen, depicted in the Fig. below.

The screenshot shows a test interface with three audio samples labeled A, B, and C. Above the samples is a 'Submit' button and a progress indicator '1/7 iteration / max iterations'. Below the samples are five attributes to be rated, each with a discrete scale from 1 to 7. The attributes are:

- Plausibility/Realism for listening to a live performance in a theater**: Scale from 'most plausible' to 'least plausible'.
- Distance of the singer from your seating position**: Scale from 'furthest' to 'closest'.
- Horizontal extent of the acoustic image of the singer**: Scale from 'widest source' to 'narrowest source'.
- Sensation of being surrounded by the sound of the singer and room**: Scale from 'most enveloping' to 'least enveloping'.

At the bottom left, there are two checkboxes: 'exit' and 'escape fullscreen'.



In the test you will rate 10 groups of sound examples. The top right corner will tell which iteration you are at. The sound samples are to be rated for the following attributes (definitions are given below): *Plausibility*, *Distance*, *Apparent Source Width*, and *Listener envelopment*.

To listen to sound example A press button **A**; to listen to B press button **B** ... etc.

You can listen to the sound files as many times as needed. We suggest that you listen to all complete examples once or twice before you start rating them.

The task is to compare and rate the 3 sound samples per acoustical attribute on a discrete scale from 1 ('least ...') to 7 ('most ...'). You must use the 2 extreme scale values (1 and 7) at least once per attribute to be able to submit the rating. Multiple sound samples may have the same rating.

If you have additional observations which are not represented by these attributes, or if you want to give more detailed comments, please add them in the text box next to the sliders.

Once you have evaluated all attributes, press “**Submit**” to go to the next iteration.

The first iteration is for you to get familiar with the test. During this iteration the test supervisor will remain with you to answer any questions you may have. If you understand the task & acoustical attributes definitions and you have no further questions for the test supervisor, you can proceed to the next iteration without giving detailed responses.

### **Acoustical attributes**

#### Plausibility/Realism

Given the assumption that you are listening to a live theater performance of two actors (see video) in a small theater (see Figure). Which recording sounds most natural to you?

#### Distance

Distance of the actors from your seating position.

#### Apparent Source Width

Apparent source width describes the perceived horizontal extent of the acoustic image of the actors. The sources may sound ‘narrow’ (in the extreme case it is as if the sound is coming from a point). On the contrary the source can also sound very ‘wide’.

#### Listener Envelopment

Listener envelopment describes the sensation of being surrounded by the sound of the actors and room. Higher “Listener envelopment” means a more uniform distribution, while less “Listener envelopment” means a more localized or directional reverberant sound.

**Thank you for your participation!**



## Comparison Study 5

### Instructions

Bart Postma

postma@limsi.fr

### Introduction

Thank you for agreeing to participate in this study as part of the ECHO project. In this test you will be asked to rate 3 audio samples regarding subjective attributes related to room acoustics. You may pause the experiment at any time.

The audio examples have been recorded in a small theater using ambisonic microphones in the audience, while a play with two actors was performed on stage. You should therefore imagine yourself listening to a *\*live\** performance within this theater. You will hear examples from different seats within the theater. You are able to move your head during listening, as there is *head-tracking*.

### Starting the test

The test supervisor will start the test by typing your name in the box on the top of the page and choosing the list you will listen to. Press the **“Confirm”** button to go to the test screen, depicted in the Fig. below.

The screenshot shows a test interface with a dark background. At the top left, there are two small icons: a circle with a dot and the text 'quit', and a circle with a cross and the text 'escape fullscreen'. In the top center is a large grey button labeled 'Submit'. To the right of the button is a progress indicator '1/10' with 'iteration / max iterations' below it. Below these are three columns of red circles with white letters 'A', 'B', and 'C' inside, each with a small play button icon below it. Below the columns are five rows of rating scales, each with a label on the left and a title above the scales. The scales consist of seven circles with a dot in the center, and the first circle is filled. The attributes and their scales are: 1. 'Plausibility' with title 'Plausibility/Realism for listening to a live performance in a theater', scales for A, B, and C, and labels 'most plausible' and 'least plausible'. 2. 'Distance' with title 'Acoustical distance of the actors from your seating position.', scales for A, B, and C, and labels 'furthest' and 'closest'. 3. 'Apparent Source Width' with title 'Horizontal extent of the acoustic image of the singer.', scales for A, B, and C, and labels 'widest source' and 'narrowest source'. 4. 'Listener Envelopment' with title 'Sensation of being surrounded by the sound of the singer and room.', scales for A, B, and C, and labels 'most enveloping' and 'least enveloping'. 5. A fifth row of scales without a label, but with the same structure.

In the test you will rate 10 groups of sound examples. The top right corner will tell which iteration you are at. The sound samples are to be rated for the following attributes (definitions are given below): *Plausibility*, *Distance*, *Apparent Source Width*, and *Listener envelopment*.

To listen to sound example A press button **A**; to listen to B press button **B**; to listen to C press button **C**. You can listen to the sound files as many times as needed. We suggest that you listen to all complete examples once or twice before you start rating them.

The task is to compare and rate the 3 sound samples per acoustical attribute on a discrete scale from 1 ('least ...') to 7 ('most ...'). You must use the 2 extreme scale values (1 and 7) at least once per attribute to be able to submit the rating. Multiple sound samples may have the same rating.

Once you have evaluated all attributes, press "**Submit**" to go to the next iteration.

The first iteration is for you to get familiar with the test. During this iteration the test supervisor will remain with you to answer any questions you may have. If you understand the task & acoustical attributes definitions and you have no further questions for the test supervisor, you can proceed to the next iteration without giving detailed responses.

#### **Acoustical attributes**

##### Plausibility/Realism

Given the assumption that you are listening to a live theater performance of two actors in a small theater. Which recording sounds most natural to you?

##### Distance

Distance of the actors from your seating position.

##### Apparent Source Width

Apparent source width describes the perceived horizontal extent of the acoustic image of the actors. The sources may sound 'narrow' (in the extreme case it is as if the sound is coming from a point). On the contrary the source can also sound very 'wide'.

##### Listener Envelopment

Listener envelopment describes the sensation of being surrounded by the sound of the actors and room. Higher "Listener envelopment" means a more uniform distribution, while less "Listener envelopment" means a more localized or directional reverberant sound.

**Thank you for your participation!**

## Comparison Study 6

### Instructions

Bart Postma  
postma@limsi.fr

#### Introduction

Thank you for agreeing to participate in this study as part of the ECHO project. In this test you will be asked to rate a series of extracts regarding subjective attributes related to room acoustics. You may pause the experiment at any time.

The extracts are reproductions of recordings made in several theaters of a play with two actors on stage. You should therefore imagine yourself attending a rehearsal of a \*live\* performance within a theater. You will have examples from different seats within the theaters. You are free to move/rotate your head during the extract, as there is *head-tracking*, though please remain in your assigned seat.

#### Starting the test

The test supervisor will start the test by typing your name in the box on the top of the page and choosing the list of presented extracts. Press the “**Confirm**” button to go to the test screen, depicted in the Figure below.

The screenshot shows a test interface with a dark background. At the top left, there are two small icons: a circle with a dot and the text 'quit', and a circle with a dot and the text 'escape fullscreen'. In the top center is a large grey button labeled 'Submit'. In the top right corner, it says '1 / 30' and 'iteration / max iterations'. Below the 'Submit' button is a red circle with the word 'play' in white. Below the 'play' button is a small progress bar. The interface is divided into five rating sections, each with a title and a vertical scale of radio buttons:

- Plausibility/Realism for listening to a live performance in a theater:** A vertical scale with three radio buttons labeled 'very plausible', 'neutral', and 'implausible'.
- Acoustical distance of the actors from your seating position:** A vertical scale with five radio buttons labeled 'very far', 'far', 'neutral', 'close', and 'very close'.
- Perceived loudness of the sound source:** A vertical scale with three radio buttons labeled 'loud', 'neutral', and 'quiet'.
- Horizontal extent of the acoustic image of the actors:** A vertical scale with three radio buttons labeled 'very wide', 'neutral', and 'very narrow'.
- Sensation of being surrounded by the sound of the actors and room:** A vertical scale with three radio buttons labeled 'very enveloping', 'neutral', and 'not enveloping'.

In the test you will rate separately 30 extracts. The top right corner will tell the current iteration. The extracts are to be rated for the following acoustical attributes (definitions given below):

*Plausibility, Distance, Loudness, Apparent Source Width, and Listener envelopment.*

To start each extract press play. You can replay the extract as many times as you like. We suggest that you watch the complete extract once or twice before you start rating it.

The task is to rate the extract for each acoustical attribute on a discrete scale. Scale extremes are provided in the test template. Try to use consistent criteria for the scales during the test.

Once you have evaluated all attributes, press “**Submit**” to go to the next iteration.

The first 3 iterations are for you to get familiar with the test. The test supervisor will remain with you during this familiarization phase to answer any questions you may have. If you understand the task & acoustical attributes definitions and you have no further questions for the test supervisor, you may proceed to the actual test without giving detailed responses.

**ACOUSTICAL attributes for judging the performance in the current seat/theatre**

Plausibility/Realism

Does the example sound plausible/realistic in relation to your seating position in the theater?

Distance

Acoustical distance of the actors from your seating position.

Loudness

Overall impression of the intensity of sound. Loudness should be assessed relative to what you consider acceptable for the theatrical performance involved.

Apparent Source Width

Apparent source width describes the perceived horizontal extent of the acoustic image of the actors. The sources may sound ‘narrow’ (in the extreme case it is as if the sound is coming from a point). On the contrary the source can also sound very ‘wide’.

Listener Envelopment

Listener envelopment describes the sensation of being surrounded by the sound of the actors and room. Higher “Listener envelopment” means a more uniform distribution, while less “Listener envelopment” means a more localized or directional reverberant sound.

**Thank you for your participation!**



# Bibliography

- [Ahearn 2008] M. Ahearn, M. Schaeffler, R. Celmer and M. Vigeant. *Investigation of the just noticeable difference of the clarity index for music, C80*. J. Acoust. Soc. Am., vol. 126, no. 1, page 2288, 2008. (Cited on page 7.)
- [Allen 1979] J.B. Allen and D.A. Berkley. *Image method for efficiently simulating small-room acoustics*. J. Acoust. Soc. Am., vol. 65, pages 943–950, 1979. (Cited on page 14.)
- [Allred 1958a] J.C. Allred and A. Newhouse. *Applications of the Monte Carlo Method to Architectural Acoustics*. J. Acoust. Soc. Am., vol. 30, no. 1, pages 1–3, 1958. (Cited on page 15.)
- [Allred 1958b] J.C. Allred and A. Newhouse. *Applications of the Monte Carlo Method to Architectural Acoustics. II*. J. Acoust. Soc. Am., vol. 30, no. 10, pages 903–904, 1958. (Cited on page 15.)
- [Alonso 2014] A. Alonso, J.J. Sendra and R. Suárez. *Sound Space Reconstruction in the Cathedral of Seville for major feasts celebrated around the main chancel*. In Proceedings of the Forum Acusticum, 2014. (Cited on pages 44 and 71.)
- [Alvarez-Morales 2015] L. Alvarez-Morales and F. Martellotta. *A geometrical acoustic simulation of the effect of occupancy and source position in historical churches*. Applied Acoustics, vol. 91, pages 47–58, 2015. (Cited on pages 44, 74 and 75.)
- [André 2012] C.R. André, Rébillat M., J.-J. Embrechts, J.G. Verly and B.F.G. Katz. *Sound for 3D cinema and the sense of presence*. In Proc. Intl. Conf. on Auditory Display (ICAD), pages 14–21, Atlanta, USA, 2012. (Cited on page 135.)
- [Anon. 1909] Anon. *Method of Correcting Faulty Acoustic Properties of Public Halls*. American scientist, page 462, 19-06-1909. (Cited on page 150.)
- [Aoshima 1981] N. Aoshima. *Computer-generated pulse signal applied for sound measurement*. J. Acoust. Soc. Am., vol. 69, no. 5, pages 1484–1488, 1981. (Cited on page 21.)
- [Asensio 2005] J.C. Asensio. *Liturgia y música en la Hispania de la Alta Edad Media; el canto visigótico, hispánico o mozárabe*. In Proceedings of X Jornadas de Canto Gregoriano: De nuevo con los mozárabes. Zaragoza, pages 135–155, 2005. (Cited on page 24.)
- [Astolfi 2008] A. Astolfi, V. Corrado and A. Griginis. *Comparison between measured and calculated parameters for the acoustical characterization of small classrooms*. Applied Acoustics, vol. 69, pages 966–976, 2008. (Cited on pages 45, 74 and 75.)

- [Azevedo 2014] M. Azevedo and J. Sacks. *Auralization as an architectural design tool*. In Proc. of the EAA Joint Symposium on Auralization and Ambisonics, Berlin, Germany, pages 162–168, 2014. (Cited on page 25.)
- [Barron 1979] M Barron and C.B. Chinoy. *1:50 Scale acoustic models for objective testing of auditoria*. Applied Acoustics, vol. 12, no. 5, pages 361–375, 1979. (Cited on page 9.)
- [Barron 1988] M. Barron. *Subjective Study of British Symphony Concert Halls*. Acustica, vol. 66, pages 1–14, 1988. (Cited on pages 118, 129, 137 and 138.)
- [Barron 1993] M. Barron. Auditorium acoustics and architectural design. London: Taylor & Francis, 1993. (Cited on page 61.)
- [Becker 1993] E. Becker J. und Mommertz. *Die Modellierung der Wandreflexionen in der raumakustischen Simulation*. In Eng.: “The modelling of wall reflections in room acoustic simulations”. In Proceeding of Fortschritte des Akustik, DAGA, pages 211–215, 1993. (Cited on page 20.)
- [Benedetto 1983] G. Benedetto and R. Spagnolo. *A method for correcting the reverberation times of enclosures as a function of humidity and temperature*. Applied Acoustics, vol. 16, no. 6, pages 463–470, 1983. (Cited on page 73.)
- [Beranek 1986] L.L. Beranek. Acoustics. Acoustical Society of America, 1986. (Cited on pages 45 and 47.)
- [Beranek 1996] L.L. Beranek. How they sound concert and opera halls. Acoustical Society of America, 1996. (Cited on pages 1, 6, 45, 47 and 61.)
- [Berkhout 1980] A.J. Berkhout, D. De Vries and M.M. Boone. *A new method to acquire impulse responses in concert halls*. J. Acoust. Soc. Am., vol. 68, no. 1, pages 179–183, 1980. (Cited on page 21.)
- [Biancani 1620] G. Biancani. Spheara mundi. Bologna: Sebastiani Bonomij, 1620. (Cited on page 145.)
- [Bibliothèque national de France 1898] Bibliothèque national de France. *Plans manuscrits de l’église de Saint-Germain-des-Prés, en 1644, “l’année de son rétablissement” [Plan documents of the abbey of Saint-Germain-des-Prés in 1644 “year of reinstatement”]*. In Annales de l’abbaye de Saint-Germain-des-Prés (555-1743), et pièces annexes jusqu’en 1753. MS. FR. 18816, FOL. 72, volume 265. Fondation de L’Abbey, 1898. (Cited on page 70.)
- [Bilbao 2001] S. Bilbao. *Wave and scattering methods for the numerical integration of partial differential equations*. PhD thesis, Stanford University, 2001. (Cited on page 10.)

- [Borish 1984] J. Borish. *Extension of the image model to arbitrary polyhedra*. J. Acoust. Soc. Am., vol. 6, no. 75, pages 1827–1836, 1984. (Cited on pages 13 and 14.)
- [Bork 2000] I. Bork. *A Comparison of Room Simulation Software - The 2nd Round Robin on Room Acoustical Computer Simulation*. Acustica, vol. 86, pages 943–956, 2000. (Cited on page 18.)
- [Bork 2005a] I. Bork. *Report on the 3rd round robin on room acoustical computer simulation. Part I: Measurements*. Acta Acustica united with Acustica, vol. 91, pages 740–752, 2005. (Cited on page 18.)
- [Bork 2005b] I. Bork. *Report on the 3rd Round Robin on Room Acoustical Computer Simulation, Part II: Calculations*. Acta Acustica united with Acustica, vol. 91, pages 753–763, 2005. (Cited on page 18.)
- [Bork 2005c] I. Bork. *Report on the 3rd Round Robin on Room Acoustical Computer Simulation Part II: Calculations*. Acustica, vol. 91, pages 753–763, 2005. (Cited on page 47.)
- [Botteldoorn 1995] D. Botteldoorn. *Finite-difference time-domain simulation of low-frequency room acoustic problems*. J. Acoust. Soc. Am., vol. 96, no. 6, pages 3302–3308, 1995. (Cited on page 10.)
- [Bradley 1995a] B.S. Bradley and G.A. Soulodre. *The influence of late arriving energy on spatial impression*. J. Acoust. Soc. Am., vol. 97, no. 4, pages 2263–2271, 1995. (Cited on page 6.)
- [Bradley 1995b] B.S. Bradley and G.A. Soulodre. *Objective measures of listener envelopment*. J. Acoust. Soc. Am., vol. 98, no. 5, pages 2590–2597, 1995. (Cited on page 6.)
- [Bradley 1999] J. Bradley, R. Reich and S. Norcross. *A just noticeable difference in C50 for speech*. Applied Acoustics, vol. 58, no. 1, pages 99–108, 1999. (Cited on page 7.)
- [Calamia 2009] P.T. Calamia. *Advances in Edge-Diffraction Modeling for Virtual-acoustic Simulations*. PhD thesis, Princeton University, 2009. (Cited on pages 9, 10, 13 and 19.)
- [Calfas 1927] P. Calfas. *La Nouvelle Salle de Concert Pleyel*. Le Génie Civil, pages 421–427, 29-10-1927. (Cited on pages 151 and 152.)
- [Castellengo 1987] M. Castellengo, B. Fabre and E. Vivie. *Etude acoustique pour la situation d'un nouvel orgue à Notre-Dame de Paris*. Rapport technique, Laboratoire d'acoustique musicale de l'Université Paris VI, Paris, 1987. (Cited on page 32.)



- [Chen 2006] S.H. Chen. Finite difference method. Taipei: High-Field Physics and Ultrafast Technology Laboratory, 2006. (Cited on page 10.)
- [Choi 2006] Y.-J. Choi and F.R. Fricke. *A Comparison of Subjective Assessments of Recorded Music and Computer Simulated Auralizations in Two Auditoria*. Acta Acustica united with Acustica, vol. 92, pages 604–611, 2006. (Cited on pages 3, 79, 80 and 98.)
- [Choueiri 2010] E. Choueiri. *Genelec 8351A directivity*. <https://www.princeton.edu/3D3A/Directivity/Genelec%208351A/images/Plots/>, note = accessed: 2016-01-19, 2010. (Cited on page 105.)
- [Chu 2002] W.T. Chu and A.C.C. Warnock. *Detailed Directivity of Sound Fields Around Human Talkers*. NRC-CNRC, NRC Publications Archive Archives des publications du CNRC, 2002. (Cited on pages 110 and 111.)
- [Costabel 1987] M. Costabel. *Principles of boundary element methods*. Comp. Phys. Reports, vol. 6, pages 243–274, 1987. (Cited on page 12.)
- [Cox 1993] T.J. Cox, W.J. Davies and Y.W. Lam. *The Sensitivity of Listeners to Early Sound Field Changes in Auditoria*. Acustica, vol. 79, no. 1, pages 280–284, 1993. (Cited on page 7.)
- [Cox 2009] T.J. Cox and P. D’Antonio. Acoustic absorbers and diffusers: Theory, design and application. CRC Press, 2009. (Cited on page 48.)
- [Cremer 1982] L. Cremer and H.A. Müller. Principles and applications of room acoustics, vol. 1. London: Applied Science Publishers, 1982. (Cited on page 17.)
- [d’Alembert 1751] J. L. R. d’Alembert. L’encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers. Paris: Briasson, David, Le Breton and Durand, 1751. (Cited on page 146.)
- [Dalenbäck ] B.I. Dalenbäck. *Directivity of instruments included in CATT-Acoustic: Dalenbäck, B.I. Instrument directivity*. <http://www.catt.se/udisplay.htm>, note = accessed: 2015-06-23, 2015. Based on measurements performed by PTB, Braunschweig, Germany. (Cited on page 157.)
- [Dalenbäck 1992] B.I. Dalenbäck, P. Svensson and M. Kleiner. *Room acoustic prediction and auralization based on an extended image source model*. J. Acoust. Soc. Am., vol. 92, no. 4, page 2346, 1992. (Cited on page 20.)
- [Dalenbäck 1993] B.I. Dalenbäck, M. Kleiner and P. Svensson. *Audibility of changes in geometric shape, source directivity, and absorptive treatment-experiments in auralization*. J. Audio Eng. Soc., vol. 41, pages 905–913, 1993. (Cited on page 102.)

- [Dalenbäck 1995] B.I. Dalenbäck. *A new model for room acoustic prediction and auralization*. PhD thesis, Chalmers University of Technology, Gothenburg, Sweden, 12 1995. (Cited on page 20.)
- [Dalenbäck 1996] B.I. Dalenbäck. *Room acoustic prediction based on a unified treatment of diffuse and specular reflection*. J. Acoust. Soc. Am., vol. 100, no. 2, pages 899–909, 1996. (Cited on pages 16, 17, 19 and 21.)
- [Dalenbäck 2009] B.I. Dalenbäck. *CATT-Acoustic v9 powered by TUCT user manuals*. Computer Aided Theatre Technique, Gothenburg (Sweden):, 2009. (Cited on pages 50, 54 and 59.)
- [Dalenbäck 2010] B.I. Dalenbäck. *Engineering principles and techniques in room acoustics prediction*. In Baltic-Nordic. Acoustics Meeting, 2010. (Cited on pages 46 and 53.)
- [Daly 1865] C. Daly and G. Davioud. *Les théâtres de la place du châtelet*. Paris: Librairie générale de l’architecture et des travaux publics, 1865. (Cited on page 148.)
- [Davioud 1878] G. Davioud and J. Boudais. *Le palais du trocadero*. Paris: A. Morel et Cie, librairies-editeurs, 1878. (Cited on pages 8 and 149.)
- [Davis 1927] A.H. Davis and G.W.C. Kaye. *The Acoustics of Buildings*. G. Bell and sons, ltd., 1927. (Cited on page 9.)
- [de Serret 1867] M. de Serret. *Oeuvres lagrange*. Paris: Guathier Villars, 1867. (Cited on page 146.)
- [Desai 2001] C.D. Desai and T. Kundu. *Introductory finite element method*. New York: CRC press, 2001. (Cited on page 11.)
- [Drumm 2000] I.A. Drumm and Y.W. Lam. *The adaptive beam-tracing algorithm*. J. Acoust. Soc. Am., vol. 107, no. 3, pages 1405–1412, 2000. (Cited on page 18.)
- [Dunn 1961] O.J. Dunn. *Multiple Comparisons Among Means*. Journal of the American Statistical Association, vol. 56, no. 293, pages 52–64, 1961. (Cited on page 87.)
- [Elorza 2005] D.O. Elorza. *Room acoustics modeling using the ray-tracing method: implementation and evaluation*. PhD thesis, University of Turku, 2005. (Cited on page 10.)
- [Eyring 1930] C.F. Eyring. *Reverberation time in “dead” rooms*. J. Acoust. Soc. Am., vol. 1, no. -, pages 217–241, 1930. (Cited on page 14.)
- [Farina 1995a] A. Farina. *Auralization software for the evaluation of the results obtained by a pyramid tracing code: Results of subjective listening tests*. In Proc. of the Wallace C. Sabine Centennial Symposium, pages 81–84, June 1995. (Cited on page 78.)

- [Farina 1995b] A. Farina. *Pyramid tracing vs. ray tracing for the simulation of sound propagation in large rooms*. In Proc. of Int. Conf. on Comput. Acoustics and its Environmental Applications (COMACO95), Southampton, pages 1–8, 1995. (Cited on page 16.)
- [Farina 1995c] A. Farina. *RAMSETE - A new pyramid tracer for medium and large scale acoustic problems*. In Proc. Euro-Noise 95, Lyon, pages 1–6, 1995. (Cited on page 16.)
- [Farina 1995d] A. Farina. *Verification of the accuracy of the pyramid tracing algorithm by comparison with experimental measurements of objective acoustic parameters*. In Proc. 15th Intl. Congress on Acoustics (ICA), Trondheim, pages 1–4, 1995. (Cited on pages 16 and 78.)
- [Farina 1997] A. Farina and F. Righini. *Software implementation of an MLS analyzer, with tools for convolution, auralization and inverse filtering*. In Preprints of the 103rd AES Convention, New York, pages 1–24, 1997. (Cited on page 21.)
- [Farina 2000a] A. Farina. *Introducing the surface diffusion and edge scattering in a pyramidtracing numerical model for room acoustics*. In Proc. 108th Aud. Engr. Conv., Paris, pages 19–22, 2000. (Cited on page 16.)
- [Farina 2000b] A. Farina. *Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique*. In Proceedings of the 108th AES Convention, London, UK, pages 1–24, 2000. (Cited on pages 21 and 22.)
- [Fournier 1909] L. Fournier. *Acoustique: La Suppression de l’Echo dans la Salle du Trocadero*. La Nature Revue des Sciences, pages 326–331, 19-06-1909. (Cited on page 149.)
- [Funkhouser 2014] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J.E. West, G. Pingali, P. Min and A. Ngan. *A beam tracing method for interactive architectural acoustics*. J. Acoust. Soc. Am., vol. 115, no. 2, pages 739–756, 2014. (Cited on pages 13, 14, 15 and 16.)
- [Gade 2004] A.C. Gade, M.L. Lisa, C.L Christensen and J.-H. Rindel. *Roman Theatre Acoustics: Comparison of Acoustic Measurement and Simulation results from the Aspendos Theatre, Turkey*. In 18th International Congress on Acoustics, Kyoto, Japan, 2004. (Cited on page 24.)
- [Galindo 2009] M. Galindo, T. Zamarreño and Girón. *Acoustic simulations of Mudejar-Gothic churches*. J. Acoust. Soc. Am., vol. 126, no. 3, pages 1207–1218, 2009. (Cited on page 44.)
- [Garcia 2014] F. Garcia, A Planells, S. Cerdá, R. Montell and A. Giménez. *Archaeological acoustics of the venue of the Misteri d’Elx, Oral and Intangible Cultural Heritage (UNESCO): Basilica de Santa Maria de Elche*. In Proceedings of Forum Acusticum, pages 1–6, 2014. (Cited on pages 24 and 59.)

- [Gehler 1787] J. Gehler. *Physikalisches Wörterbuch, oder Versuch einer Erklärung der vornehmsten Begriffe und kunstwörter der Naturlehre, mit kurzen Nachrichten von der Geschichte der Erfindungen und Beschreibungen der Werkzeuge begleitet.* Leipzig: Schwikerschen Verlag, 1787. (Cited on page 146.)
- [Gibbs 1972] B.M. Gibbs and D.K. Jones. *A simple image method for calculating the distribution of sound pressure levels within an enclosure.* *Acustica*, vol. 26, no. 1, pages 24–36, 1972. (Cited on page 14.)
- [Gibson 1979] J. Gibson. *The ecological approach to visual perception.* Lawrence Erlbaum Associates, Hillsdale, NJ, 1979. (Cited on page 118.)
- [Giron 1996] F. Giron. *Investigations about the directivity of sound sources.* PhD thesis, Aachen Technical University, 1996. (Cited on page 102.)
- [Gournay 1985] I. Gournay. *Le nouveau trocadéro = the new trocadero.* Brussel: Mardaga, 1985. (Cited on page 8.)
- [Grillon 1995] V. Grillon, X. Meynial and J.-D. Polack. *Auralization in Small-Scale Models: Extending the Frequency Bandwidth.* In Audio Engineering Society Convention 98, Feb 1995. (Cited on page 84.)
- [Hamayon 1996] L. Hamayon. *L’acoustique des bâtiments.* Paris: Le Moniteur, 1996. (Cited on pages 32, 34, 73 and 74.)
- [Hamilton 2016] B. Hamilton. *Finite Difference and Finite Volume Methods for Wave-based Modelling of Room Acoustics.* PhD thesis, University of Edinburgh, 2016. (Cited on pages 9, 10, 11 and 12.)
- [Hargreaves 2005] J.A. Hargreaves and T.J. Cox. *A Transient Boundary Element Method for Room Acoustics.* In Presented at BuHu’s 5th International Post-grad Research Conference, pages 869–879, 2005. (Cited on pages 10 and 12.)
- [Henry 1857] J. Henry. *Annual report of the board of regents of the smithsonian institution.* Washington: A.G.F. Nicholson, Printer, 1857. (Cited on page 148.)
- [Hodgson 1991] M. Hodgson. *Evidence of diffuse surface reflections in rooms.* *J. Acoust. Soc. Am.*, vol. 89, no. 2, pages 765–771, 1991. (Cited on page 15.)
- [Hornickx 2015] M.C.J. Hornickx, M. Kaltenbacher and S Marburg. *A platform for benchmark cases in computational acoustics.* *Acta Acustica united with Acustica*, vol. 101, no. 4, pages 811–820, 2015. (Cited on page 19.)
- [Howard 1966] I.P. Howard and W.B. Templeton. *Human spatial orientation.* London: Wiley, 1966. (Cited on pages 3 and 118.)
- [Iannace 2015] G.A. Iannace, A. Berardi and C. Ianniello. *Study of a historical church based on acoustic measurements and computer simulation.* In 22nd international congress on sound and vibration, July 2015. (Cited on page 44.)

- [ISO 17497 2007] ISO 17497. *Acoustics – Sound-scattering properties of surfaces – Part 1: Measurement of the random-incidence scattering coefficient in a reverberation room (ISO 17497-1:2004); Part 2: Measurement of the directional diffusion coefficient in a free field (DIS 2007)*. Int. Organization for Standardization, 2007. (Cited on page 48.)
- [ISO 2009] ISO, Geneva, Switzerland: International Organization for Standardization. *ISO 3382-1:2009(E). Measurement of the reverberation time of rooms with reference to other acoustical parameters*, 2009. (Cited on pages 7, 31 and 83.)
- [James 2001] A. James, A. Naqvi and B.I. Dalenbäck. *Computer modelling with CATT-Acoustic - Theory and practice of diffuse reflection and array modelling*. In Proc. of Inst. of Acoustics, pages 1–8, 2001. (Cited on pages 18 and 52.)
- [James 2003] A. James. *Results of the NPL study into comparative room acoustic measurement techniques Part 1, Reverberation time in large rooms*. In Proc. Institute of Acoustics, volume 25 p4, 2003. (Cited on page 46.)
- [James 2004] A. James. *Results of the workshop on room acoustics measurements*. Rapport technique 9558/2, NPL, September 2004. (Cited on pages 40 and 46.)
- [Jeon 2005] J.Y. Jeon, Y.H. Kim, S.Y. Kim, D. Cabrera and J. Bassett. *The Effects of Visual Input on the Evaluation of the Acoustics in an Opera House*. In Forum Acusticum, pages 2285–2289, 2005. (Cited on page 119.)
- [Jeon 2008] J.Y. Jeon, Y.H. Kim, S.Y. Kim, D. Cabrera and J. Bassett. *The effect of visual and auditory cues on seat preference in an opera theater*. J. Acoust. Soc. Am., vol. 123, no. 6, pages 4272–4282, 2008. (Cited on pages 3, 119, 130 and 137.)
- [Jordan 1941] V.L. Jordan. *Elektroakustiske undersøgelser af materialer og modeller*. PhD thesis, Reitzels Forlag, 1941. (Cited on page 9.)
- [Junius 1959] J.W. Junius. *Raumakustische Untersuchungen mit neueren Meßverfahren in der Liederhalle Stuttgart*. In Eng.: "Room acoustic investigations with new measuring methods in the Liederhall Stuttgart". Acustica, vol. 9, no. 4, pages 289–303, 1959. (Cited on page 17.)
- [Jurkiewicz 2012] Y. Jurkiewicz, T. Wulfrank and E. Kahle. *Architectural shape and early acoustic efficiency in concert halls*. J. Acoust. Soc. Am., vol. 132, no. 3, pages 1253–1256, 2012. (Cited on page 152.)
- [Karjalainen 2005] M. Karjalainen. *Digital waveguides networks for room modeling and auralization*. In Proceedings of Forum Acusticum, pages 1–6, 2005. (Cited on pages 10 and 11.)

- [Katz 2000a] B.F.G. Katz. *Acoustic absorption measurement of human hair and skin within the audible frequency range*. J. Acoust. Soc. Am., vol. 108(5), pages 2238–2242, 2000. (Cited on page 48.)
- [Katz 2000b] B.F.G. Katz. *Method to resolve microphone and sample location errors in the two-microphone duct measurement method*. J. Acoust. Soc. Am., vol. 108(5), pages 2231–2237, 2000. (Cited on page 48.)
- [Katz 2001] B.F.G. Katz. *Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements*. J. Acoust. Soc. Am., vol. 110, no. 5, pages 2449–2455, 2001. (Cited on page 48.)
- [Katz 2004] B.F.G. Katz. *International Round Robin on Room Acoustical Response Analysis Software 2004*. J. Acoust. Soc. Am., vol. 4, pages 158–164, 2004. (Cited on pages 40, 46 and 53.)
- [Katz 2005] B.F.G. Katz and E.A. Wetherill. *Fogg Art Museum Lecture Room, A Calibrated Recreation of the Birthplace of Room Acoustics*. In Proceedings of Forum Acusticum, 2005. (Cited on pages 24, 44 and 69.)
- [Katz 2006] B.F.G. Katz, F. Prezati and C. d’Alessandro. *Human voice phoneme directivity pattern measurements*. In 4th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, page 3359, Honolulu (Hawaiï), 2006. (Cited on pages 3 and 101.)
- [Katz 2007] B.F.G. Katz and C. d’Alessandro. *Directivity measurements of the singing voice*. In Intl. Cong. on Acoustics 19, pages 1–6, Madrid (Spain, 2007. (Cited on page 84.)
- [Katz 2015a] B.F.G. Katz, D.Q. Felinto, D. Touraine, D. Poirier-Quinot and P. Bourdot. *BlenderVR: Open-source framework for interactive and immersive VR*. In Proc of IEEE VR, pages 203–204, 2015. (Cited on page 120.)
- [Katz 2015b] B.F.G. Katz, Y. Jurkiewicz, T. Wulfrank, G. Parseihian, T. Scélo and H. Marshall. *La Philharmonie de Paris - Acoustic scale model study*. In Intl. Conf. on Auditorium Acoustics, volume 37, pages 431–438, Paris, October 2015. Institute of Acoustics. (Cited on page 9.)
- [Kirszenstein 1984] J. Kirszenstein. *An Image Source Computer Model for Room Acoustics Analysis and Electroacoustic Simulation*. Applied Acoustics, vol. 17, no. 4, pages 275–290, 1984. (Cited on page 14.)
- [Kleiner 1981] M. Kleiner. *Speech intelligibility in real and simulated sound fields*. Acustica, vol. 47, no. 2, pages 55–71, 1981. (Cited on page 78.)
- [Kleiner 1993] M. Kleiner, B.-I. Dalenbäck and P. Svensson. *Auralization-an overview*. J. Audio Engineering Society, vol. 41, no. 11, pages 861–875, 1993. (Cited on page 2.)



- [Knüttel 2013] T. Knüttel, I.B. Witew and M. Vorländer. *Influence of “omnidirectional” loudspeaker directivity on measured room impulse responses*. J. Acoust. Soc. Am., vol. 134, no. 5, pages 3654–3662, 2013. (Cited on page 102.)
- [Koutsouris 2013] G.I. Koutsouris, J. Brunskog, C.-H. Jeong and F. Jacobsen. *Combination of acoustical radiosity and the image source method*. J. Acoust. Soc. Am., vol. 133, no. 6, pages 3963–3974, 2013. (Cited on page 17.)
- [Krokstad 1968] A. Krokstad, S. Strøm and S. Sørsdal. *Calculating the Acoustical Room Response by the Use of a Ray Tracing Technique*. J. Sound Vib, vol. 8, no. 1, pages 118–125, 1968. (Cited on page 15.)
- [Kulowski 1982] A. Kulowski. *Error investigation for the ray tracing technique*. Applied Acoustics, vol. 15, no. 4, pages 263–274, 1982. (Cited on pages 16 and 46.)
- [Kuttruff 1993] H. Kuttruff. *Auralization of impulse responses modelled on the basis of ray-tracing results*. J. Audio Eng. Soc., vol. 41, no. 11, pages 876–880, 1993. (Cited on pages 14, 16 and 20.)
- [Kuttruff 2000] H. Kuttruff. Room acoustics, 4th ed. Applied Science Publishers LTD, London, 2000. (Cited on page 17.)
- [Langhans 1800] K.G. Langhans. Vergleichung des Neuen Schauspielhauses zu Berlin mit verschiedenen ältern und neuen Schauspielhäusern in Rücksicht auf akustische und optische Grundfäse. Berlin: Johann Friedrich Unger, 1800. (Cited on page 147.)
- [Langhans 1810] C.F. Langhans. über Theater oder Bemerkungen über Katakustik. Berlin: Gottfr. Hayn, 1810. (Cited on pages 14 and 147.)
- [Larsson 2001] P. Larsson, D. Västfjäll and M. Kleiner. *Ecological acoustics and the multi-modal perception of rooms: real and unreal experiences of auditory-visual virtual environments*. In Proceedings of the 2001 Intl. Conf. on Auditory Display, pages 245–249, July 29-August 2001. (Cited on pages 118, 137 and 138.)
- [Le Carrou 2010] J.-L. Le Carrou, Q. Leclere and F. Gautier. *Some characteristics of the concert harpâs acoustic radiation*. J. Acoust. Soc. Am, vol. 127, no. 5, pages 3203–3211, 2010. (Cited on page 157.)
- [Lee 1988] H. Lee and B.-H. Lee. *An efficient algorithm for the image model technique*. Applied Acoustics, vol. 24, no. 2, pages 87–115, 1988. (Cited on page 14.)
- [Lehnert 1993] H. Lehnert. *Systematic errors of the ray-tracing algorithm*. Applied Acoustics, vol. 38, no. -, pages 207–221, 1993. (Cited on page 16.)

- [Lewers 1993] T. Lewers. *A combined beam tracing and radiant exchange computer model of room acoustics*. Applied Acoustics, vol. 38, no. -, pages 161–178, 1993. (Cited on page 16.)
- [Lindebrink 2015] J. Lindebrink and J. Nätterlund. *An engine for real-time audiovisual rendering in the building design process*. In Proc. Acoustics 2015, pages 1–8, 2015. (Cited on pages 22 and 155.)
- [Lisa 2004] M.L. Lisa, C.L. Christensen and J.-H. Rindel. *Predicting the acoustics of ancient open-air theatres: The importance of calculation methods and geometrical details*. In Proceedings of the BNAM, 2004. (Cited on page 24.)
- [Lisa 2006] M.L. Lisa, C.L. Christensen and J.-H. Rindel. *Predicting the acoustics of ancient open-air theaters: the importance of calculation methods and geometrical details*. In Proceedings of the Institute of Acoustics, 2006. (Cited on page 24.)
- [Lokki 2001] T. Lokki and H. Järveläinen. *Subjective Evaluation of Auralization of Physics-Based Room Acoustics Modeling*. In Intl. Conf. on Auditory Display, pages 1–6, 2001. (Cited on pages 3, 78, 79, 80 and 98.)
- [Lokki 2002a] T. Lokki and V. Pulkki. *Evaluation of geometry-based parametric auralization*. In AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio, pages 367–376, Jun 2002. (Cited on pages 3, 79 and 98.)
- [Lokki 2002b] T. Lokki, P. Svensson and L. Savioja. *An Efficient Auralization of Edge Diffraction*. In AES 21st Intl. Conf.: Architectural Acoustics and Sound Reinforcement, Jun 2002. (Cited on page 48.)
- [Lokki 2008a] T. Lokki, J. Pätynen and V. Pulkki. *Recording of Anechoic Symphony Music*. In Proc. Acoustics, pages 6431–6436, June-July 2008. (Cited on pages 84, 91, 104 and 109.)
- [Lokki 2008b] T. Lokki and L. Savioja. *State of the in auralization of concert halls models – what is still missing?* In BNAM, pages 1–7, 2008. (Cited on page 2.)
- [Luizard 2013] P. Luizard, M. Otani, J. Botts, L. Savioja and B.F.G. Katz. *Comparison of sound field measurements and predictions in coupled volumes between numerical methods and scale model measurements*. Proc of Meetings on Acoustics, vol. 19, no. 1, pages 1–9, 2013. (Cited on pages 11, 12 and 19.)
- [Maempel 2013a] H.-J. Maempel and M. Jentsch. *Audio-visual interaction of size and distance perception in concert halls - a preliminary study*. In Intl. Symp. on Room Acoustics, pages 1–16, 2013. (Cited on pages 118, 137 and 138.)
- [Maempel 2013b] H.-J. Maempel and M. Jentsch. *Auditory and visual contribution to egocentric distance and room size perception*. Building Acoustics, vol. 20, no. 4, pages 383–401, 2013. (Cited on page 118.)



- [Maercke 1993] D. v. Maercke and J. Martin. *The Prediction of Echograms and Impulse Responses within the Epidaure Software*. Applied Acoustics, vol. 38, no. -, pages 93–114, 1993. (Cited on pages 16 and 19.)
- [Magenat-Thalmann 2006] N. Magrenat-Thalmann, A.E. Foni and N. Cadi-Yazli. *Real-time animation of ancient roman sites*. In Proc. 4th Intl. Conf. Computer graphics and interactive techniques in Australasia and Southeast Asia (GRAPHITE), pages 19–30, 2006. (Cited on page 155.)
- [Mak 2015] C.M. Mak and Z. Wang. *Recent advances in building acoustics: An overview of prediction methods and their applications*. Building and environment, vol. 91, pages 118–126, 2015. (Cited on page 19.)
- [Marshall 1985] A.H. Marshall and J. Meyer. *Directivity and Auditory impression of Singers*. Acustica, vol. 58, pages 130–140, 1985. (Cited on pages 105 and 157.)
- [Martellota 2014] F. Martellota and L.A. Morales. *Virtual Acoustic Reconstruction of the Church of Gesù in Rome: a Comparison between Different Design Options*. In Proceedings of Forum Acusticum, 2014. (Cited on pages 44, 45, 74 and 75.)
- [Martellotta 2009] F. Martellotta. *Identifying acoustical coupling by measurements and prediction-models for St. Peter’s Basilica in Rome*. J. Acoust. Soc. Am., vol. 126, no. 3, pages 1175–1186, 2009. (Cited on pages 44, 55 and 75.)
- [Martelotta 2011] F. Martelotta and M.L. Castiglione. *On the Use of Paintings and Tapestries as Sound Absorbing Materials*. In Proc. of Forum Acusticum, 2011. (Cited on page 71.)
- [McGurk 1976] H. McGurk and J. MacDonald. *Hearing lips and seeing voices*. Nature, vol. 264, pages 746–748, 1976. (Cited on pages 3 and 118.)
- [Menzel 2008] D. Menzel, H. Fastl, R. Graf and J. Hellbrück. *Influence of vehicle color on loudness judgments*. J. Acoust. Soc. Am., vol. 123, pages 2477–2499, 2008. (Cited on pages 119, 129 and 137.)
- [Mercier 2002] D. Mercier. *Le livre des techniques du son*. Paris: D. Dunod, 2002. (Cited on pages 32, 34, 73 and 74.)
- [Mersenne 1636] M. Mersenne. *Traité de l’harmonie universelle*. Paris: Sébastien Cramoisy, 1636. (Cited on page 146.)
- [Meyer 1978] J. Meyer. *Acoustics and the performance of music*. Frankfurt/Main: Verlag das instrument, 1978. (Cited on page 101.)
- [Meyer 1993] J. Meyer. *The sound of the orchestra*. J. Audio Eng. Soc., vol. 41, no. 4, pages 203–213, 1993. (Cited on page 101.)

- [Meyer 2015] J. Meyer. *Auralisation d'une simulation acoustique calibrée de la Cathédrale Notre-Dame de Paris*. PhD thesis, Université Pierre-et-Marie-Curie, 2015. (Cited on page 157.)
- [Miles 1984] R.N. Miles. *Sound field in a rectangular enclosure with diffusely reflecting boundaries*. J. Sound Vib., vol. 92, no. -, pages 203–226, 1984. (Cited on page 17.)
- [Moloney 2004] J. Moloney and L Harvey. *Visualization and 'Auralization' of architectural design in a game engine based collaborative virtual environments*. In Proc 8th Intl. Conf. Information Visualisation (IV), pages 827–832, 2004. (Cited on page 155.)
- [Morgan 1914] M.H. Morgan. The ten books on architecture. vitruvius. Cambridge: Harvard University Press. London: Humphrey Milford. Oxford University Press, 1914. (Cited on page 1.)
- [Müller 2001] S. Müller and P. Massarani. *Transfer-Function Measurement with Sweeps*. J. Audio Eng. Soc., vol. 49, no. 6, pages 443–471, 2001. (Cited on page 21.)
- [Murphy 2000] D.T. Murphy. *Digital Waveguide Mesh Topologies in Room Acoustics Modelling*. PhD thesis, The University of York, 2000. (Cited on page 10.)
- [Murphy 2016] D. Murphy, S. Shelley, A. Foteinou, J. Brereton and H. Daffern. *Acoustic Heritage and Audio Creativity: the Creative Application of Sound in the Representation, Understanding and Experience of Past Environments*. Internet Archaeology, vol. Special issue, pages 1–23, 2016. (Cited on page 24.)
- [Murray a] S. Murray, A. Tallon and R. O'Neill. *Paris, Cathédrale Notre-Dame*. <http://mappinggothic.org/building/1164>, accessed 2016-03-27, 2016. (Cited on pages 28 and 159.)
- [Murray b] S. Murray, A. Tallon and R. O'Neill. *Paris, Église Saint-Germain-des-Prés*. <http://mappinggothic.org/building/1165>, accessed 2016-03-27, 2016. (Cited on page 28.)
- [Niemi 2014] H. Niemi, M. Kylliäinen, J. Jäppinen and M. Lindqvist. *Acoustics of vanished concert halls of Helsinki: preliminary results*. In Proceedings of Forum Acusticum, pages 1–6, 2014. (Cited on page 23.)
- [Noisternig 2003a] M. Noisternig, T. Musil, A. Sontacchi and R. Höldrich. *A 3D Ambisonic based binaural sound reproduction system*. In AES 24th Intl. Conf. Multichannel Audio, pages 174–178, Alberta (Canada), 2003. (Cited on page 159.)
- [Noisternig 2003b] M. Noisternig, T. Musil, A. Sontacchi and R. Holdrich. *A 3D real time Rendering Engine for binaural Sound Reproduction*. In Proc. of the

- 9th Intl. Conf. on Auditory Display (ICAD), pages 107–110, 2003. (Cited on page [112](#).)
- [Nosal 2004] E.-A. Nosal, M. Hodgson and I. Ashdown. *Improved algorithms and methods for room sound-field prediction by acoustical radiosity in arbitrary polyhedral rooms*. J. Acoust. Soc. Am., vol. 116, no. 2, pages 970–980, 2004. (Cited on page [17](#).)
- [Olive 2003] S. Olive. *Differences in performance and preference of trained versus untrained listeners in loudspeaker tests: a case study*. J Audio Eng Soc, vol. 51, no. 9, pages 806–825, 2003. (Cited on page [106](#).)
- [Olsen 1998] W.O. Olsen. *Average Speech Levels and Spectra in Various Speaking/Listening Conditions: A Summary of the Pearson, Bennett, & Fidell (1977) Report*. American Journal of Audiology, vol. 7, pages 1–5, 1998. (Cited on page [130](#).)
- [Olson 1967] H.F Olson. Music, physics and engineering. New York (USA): Dover publications, 1967. (Cited on page [157](#).)
- [Otondo 2004] F. Otondo and J.H. Rindel. *The influence of the directivity of musical instruments in a room*. Acta Acustica united with Acustica, vol. 90, pages 1178–1184, 2004. (Cited on pages [3](#) and [102](#).)
- [Otondo 2005] F. Otondo and J.H. Rindel. *A new method for the radiation representation of musical instruments in auralizations*. Acta acustica united with acustica, vol. 91, no. 5, pages 902–906, 2005. (Cited on pages [3](#), [102](#) and [115](#).)
- [Ozanam 1778] J. Ozanam. *Récréations mathématiques et physiques*. Paris: Cl. Ant. Jombert, Libraire du roi pour le Génie & l’Artillerie, 1778. (Cited on page [146](#).)
- [Pagliari 2015] D. Pagliari and L. Pinto. *Calibration of Kinect for Xbox One and comparison between the two generations of Microsoft sensors*. Sensors, vol. 15, no. 11, pages 27569–27589, 2015. (Cited on page [123](#).)
- [Pätynen 2011] J. Pätynen, B.F.G. Katz and T. Lokki. *Investigations on the balloon as an impulse source*. J Acoust Soc Am, vol. 129, no. 1, pages EL27–EL33, 2011. (Cited on pages [21](#), [34](#) and [104](#).)
- [Pedrero 2014] A. Pedrero, A. Díaz-Chyla, C. Díaz, S. Pelzer and M. vorländer. *Virtual Restoration of the Sound of the Hispanic Rite*. In Proceedings of the Forum Acusticum, 2014. (Cited on pages [24](#) and [44](#).)
- [Pelzer 2013] S. Pelzer and M. Vörlander. *Inversion of a room acoustics model for the determination of acoustical surface properties in enclosed spaces*. In Proc. of Meetings on Acoustics, pages 1–9, 2013. (Cited on page [44](#).)

- [Picinali 2014] L. Picinali, A. Afonso, M. Denis and B.F.G. Katz. *Exploration of architectural spaces by the blind using virtual auditory reality for the construction of spatial knowledge*. Intl. J. Human-Computer Studies, vol. 72, no. 4, pages 393–407, 2014. (Cited on page 159.)
- [Pierce 1981] A.D. Pierce. *Acoustics: An introduction to its physical principles and applications*. Acoustical Society of America, 1981. (Cited on page 40.)
- [Polack 2012] J.-D. Polack, F. Leao Figueiredo and S. Liu. *Statistical analysis of a set of Parisian Concert Halls and Theatres*. In Proc. of the Acoustics 2012 Nantes Conference, pages 911–916, 2012. (Cited on page 111.)
- [Postma 2013] B.N.J. Postma. *A History of the Use of Time Intervals After the Direct Sound in Concert Hall Design Before the Reverberation Formula of Sabine Became Generally Accepted*. Building Acoustics, vol. 20(2), pages 157–176, 2013. (Cited on pages 1 and 145.)
- [Prince 1994] D. Prince and R. Talaske. *Variation of room acoustic measurements as a function of source location and directivity*. In W. Clement Sabine cent. symp., pages 211–214, 1994. (Cited on page 102.)
- [Pulkki 2003] V. Pulkki and T. Lokki. *Visualization of edge diffraction*. Acoustics Research Letters Online, vol. 4, no. 4, pages 118–123, 2003. (Cited on page 48.)
- [Radau 1869] R. Radau. *Die lehre vom schall*. Munich: Verlag von R.A. Oldenbourg, 1869. (Cited on page 146.)
- [Rébillat 2011] M. Rébillat, R. Hennequin, E. Corteel and B.F.G. Katz. *Identification of cascade of Hammerstein models for the description of nonlinearities in vibrating devices*. J. Sound and Vibration, vol. 330, pages 1018–1038, 2011. (Cited on page 21.)
- [Reichardt 1975] W. Reichardt, O. Abdel Alim and W. Schmidt. *Definition und Messgründe eines objektiven Masses zur Ermittlung der Grenze zwischen brauchbarer und unbrauchbarer Durchsichtigkeit bei Musikaarbietung.* In English: “Definition and Basis of Making an Objective Evaluation to Distinguish Between Useful and Useless Clarity Defining Musical Performances. Acustica, vol. 32, no. 3, pages 126–137, 1975. (Cited on pages 96 and 152.)
- [Rhode 1800] J. Rhode. *Theorie der verbreitung des schalles für baukünstler*. Berlin: Heinrich Frölich, 1800. (Cited on page 146.)
- [Rindel 2000] J.H. Rindel. *The Use of Computer Modeling in Room Acoustics*. Journal of Vibroengineering, vol. 4, no. 3, pages 219–224, 2000. (Cited on page 10.)

- [Rindel 2003] J.H. Rindel and C.L. Christensen. *Room Acoustic Simulation and Auralization: How Close Can We Get to the Real Room?* In Proceeding of the Western Pacific Acoustics Conference, 2003. (Cited on page 25.)
- [Rindel 2004] J.H. Rindel, F. Otondo and C.L. Christensen. *Sound source representation for auralization*. In Proc. Int Symp on Rm Acoust: design and science, pages 1–8, Hyogo (Japan), 2004. (Cited on pages 3, 102, 103 and 115.)
- [Rindel 2006] J.-H. Rindel and M.L. Lisa. *The ERATO project and its contribution to our understanding of the acoustics of ancient Greek and Roman theaters*. In ERATO Project Symposium, 2006. (Cited on page 24.)
- [Sabine 1913] W.C. Sabine. *Theatre Acoustics*. American Architect, Incorporated, page 104:257, 1913. (Cited on page 9.)
- [Sabine 1922] W.C. Sabine. *Collected papers on acoustics*. Cambridge: Harvard university press, 1922. (Cited on page 1.)
- [Sakuma 2002a] T. Sakuma. *Benchmark platform on computational methods for architectural/environmental acoustics*. <http://news-sv.aij.or.jp/kankyo/s26/AIJ-BPCA/index.html>, note = accessed: 2017-01-04, 2002. (Cited on page 19.)
- [Sakuma 2002b] T. Sakuma, P. Svensson, A. Franck and S. Sakamoto. *A round-robin test program on wave based computational methods for room-acoustic analysis*. In Proceedings of Forum Acusticum Sevilla, pages 1–6, 2002. (Cited on page 19.)
- [San Martin 2007] R. San Martin, I.B. Witew, M. Arena and M. Vorlander. *Influence of the source orientation on the measurement of acoustic parameters*. Acta Acustica united with Acustica, vol. 93, pages 387–397, 2007. (Cited on page 102.)
- [Satoh 2002] F. Satoh, M. Nagayama and H. Tachibana. *Influence of Time-Variance in Auditorium on Impulse Response Measurement*. In Proc. Forum Acusticum, Sevilla, 2002. (Cited on page 36.)
- [Satoh 2007] F. Satoh, M. Sano, Y. Hayashi, S. Sakamoto and H. Tachibana. *Acceptable Temperature Changes during Averaging for Reverberation Time measuring by Swept-Sine Method*. In Proc. 19th Intl. Cong. on Acoustics, Madrid, volume 19, 2-7 September 2007. (Cited on page 36.)
- [Savioja 1999] L. Savioja, J. Huopaniemi, T. Lokki and Väänänen. *Creating Interactive Virtual Acoustic Environment*. J. Audio Eng. Soc., vol. 47, no. 9, pages 675–705, 1999. (Cited on page 102.)
- [Savioja 2015] L. Savioja and U.P. Svensson. *Overview of geometrical room acoustic modeling techniques*. J. Acoust. Soc. Am., vol. 138, no. 2, pages 708–730, 2015. (Cited on pages 9, 13 and 20.)

- [Schiettecatte 2003] B. Schiettecatte, A. Nackaerts and B.D. Moor. *Real-time acoustics simulation using mesh-tracing*. In Proc. Intl. Comp. Music Conf. (ICMC), Singapore, pages 1–4, 2003. (Cited on page 18.)
- [Schilling 1848] G. Schilling. *Akustik oder die lehre vom klange*. Stuttgart: Verlags Bureau, 1848. (Cited on page 146.)
- [Schmieden 1886] H. Schmieden. *Das neue Gewandhaus in Leipzig*. Zeitschrift fur Bauwesen, pages 1–14, 1886. (Cited on page 151.)
- [Schroeder 1961] M.R. Schroeder. *Novel Uses of Digital Computers in Room Acoustics*. In J. Acoust. Soc. Am, 1961. (Cited on page 9.)
- [Schroeder 1970] M.R. Schroeder. *Digital simulation of sound transmission in reverberant spaces*. J. Acoust. Soc. Am., vol. 47, no. 2, pages 424–431, 1970. (Cited on page 15.)
- [Seraphim 1958] H. Seraphim. *Untersuchungen uber die unterschiedsschwelle exponentiellen abklingens von rauschbandimpulsen. [Investigations about the difference threshold in exponential decay of noise band pulses]*. Acustica, vol. 5, no. 1, pages 280–284, 1958. (Cited on page 7.)
- [Siltanen 2007] S. Siltanen, T. Lokki, S. Kiminki and L. Savioja. *The room acoustic rendering equation*. J. Acoust. Soc. Am., vol. 122, no. 3, pages 1624–1635, 2007. (Cited on pages 14 and 17.)
- [Siltanen 2010] S. Siltanen, T. Lokki and L. Savioja. *Rays or Waves? Understanding the Strengths and Weaknesses of Computational Room Acoustics Modeling Techniques*. In Intl. Symp. on Room Acoustics (ISRA), pages 1–6, 29-31 August 2010. (Cited on pages 9 and 12.)
- [Southern 2013] A. Southern, S. Siltanen, D. T. Murphy and L. Savioja. *Room Impulse Response Synthesis and Validation Using a Hybrid Acoustic Model*. IEEE Transactions, vol. 21, no. 9, pages 1940–1952, 2013. (Cited on page 18.)
- [Spandöck 1934] F Spandöck. *Akustische modellversuche*. Annalen der Physik, vol. 412, no. 4, pages 345–360, 1934. (Cited on pages 1 and 9.)
- [Stan 2002] G.-B. Stan, J.-J. Embrechts and D. Archambeau. *Comparison of different impulse response measurement techniques*. J. Audio Eng. Soc., vol. 50, no. 4, pages 249–262, 2002. (Cited on pages 21 and 22.)
- [Stephenson 1996] U.M. Stephenson. *Quantized Pyramidal beam tracing - a new algorithm for room acoustics and noise immision prognosis*. Acta Acustica, vol. 82, no. 3, pages 517–525, 1996. (Cited on page 16.)
- [Suh 1999] J.S. Suh and P.A. Nelson. *Measurement of transient response of rooms and comparison with geometrical acoustic models*. J. Acoust. Soc. Am., vol. 105, no. 4, pages 2304–2317, 1999. (Cited on page 14.)



- [Summers 2004] J.E. Summers, K. Takahashi, Y. Shimizu and T. Yamakawa. *Assessing the accuracy of auralizations computed using a hybrid geometrical-acoustics and wave-acoustics method*. J. Acoust. Soc. Am., vol. 115, no. 5, page 2514, 2004. (Cited on page 18.)
- [Suzuki 1995] Y. Suzuki, F. Asano, H.-Y. Kim and T. Sone. *An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses*. J. Acoust. Soc. Am., vol. 97, no. 2, pages 1119–1123, 1995. (Cited on page 21.)
- [Takala 2014] J. Takala and M. Kylliäinen. *Comparison of modelled performance of a vanished building with historical information on its acoustics*. In Proceedings of Forum Acusticum, pages 1–6, 2014. (Cited on page 23.)
- [Tallon 2016] A. Tallon. “l’espace acoustique de l’abbatiale de saint-germain-des-prés” [the acoustic space of the abbey of saint-germain-des-prés] in saint-germain-des-prés. mille ans d’une abbaye à paris. Paris (France): Académie des Inscriptions et Belles-Lettres, 2016. (Cited on page 29.)
- [Taylor 2010a] M. Taylor, A. Chandak, L. Antani and D. Manocha. *Interactive Geometric Sound Propagation and Rendering*. In Intel Academic Spotlight, pages 1–16, 2010. (Cited on pages 22 and 155.)
- [Taylor 2010b] M. Taylor, A. Chandak, Q. Mo, C. Lauterbach, C. Schissler and D. Manocha. *iSound: Interactive GPU-based Sound Auralization in Dynamic Scenes*. In Tech. Report TR10-006, Computer Science. University of North Carolina, pages 1–10, 2010. (Cited on pages 22 and 155.)
- [Thiele 1953] R. Thiele. *Richtungsverteilung und Zeitfolge der Schallruckwürfe in Räumen*. In English: “Directional distribution and sequence of the sound reflections in rooms. Acustica, vol. 3, no. 12, pages 291–302, 1953. (Cited on pages 96 and 152.)
- [Tsingos 1997] N. Tsingos and J.-D. Gascuel. *A general model for simulation of room acoustics based on hierarchical radiosity*. In Technical Sketches, SIGGRAPH 97 Visual Proceedings, pages 1–2, 1997. (Cited on page 17.)
- [Van Dorp Schuitman 2011] J. Van Dorp Schuitman. *Auditory modelling for assessing room acoustics*. PhD thesis, Delft University of technology, 2011. (Cited on page 22.)
- [Van Duyne 1993] S. Van Duyne and J. Smith. *Physical modeling with the 2-D digital waveguide mesh*. In Proc. Intl. Comp. Music Conf. (ICMC), pages 40–47, 1993. (Cited on page 10.)
- [Van Mourik 2014] J. Van Mourik, S. Oxnard, A. Foteinou and D.T. Murphy. *Hybrid Acoustic Modelling of Historic Spaces Using Blender*. In Proceedings of Forum Acusticum, pages 1–8, 2014. (Cited on page 18.)

- [Vermeulen 1936] R. Vermeulen and J. de Boer. -. Philips Techn. Review, vol. 1, page 46, 1936. (Cited on page 9.)
- [Vigeant 2011] M.C. Vigeant, L.M. Wang and J.H. Rindel. *Objective and subjective evaluations of the multichannel auralization technique as applied to solo instruments*. Applied Acoustics, vol. 72, no. 6, pages 311–323, 2011. (Cited on pages 3, 102, 106 and 115.)
- [Vissilantonopoulos 2001] S.L. Vissilantonopoulos and J.M. Mourjopoulos. *Virtual acoustic reconstruction of ritual and public spaces of ancient Greece*. Acta Acustica united with Acustica, vol. 87, pages 604–609, 2001. (Cited on page 23.)
- [Vorländer 1989] M. Vorländer. *Simulations of the Transient and Steady-State Sound Propagation in Rooms Using a New Combined Ray-Tracing/Image-Source Algorithm*. J. Acoust. Soc. Am., vol. 86, no. 1, pages 172–178, 1989. (Cited on page 18.)
- [Vorländer 1995] M. Vorländer. *International Round Robin on Room Acoustical Computer Simulations*. In Proc. of the Intl. Cong. on Acoustics, 1995. (Cited on page 18.)
- [Vorländer 2008] M. Vorländer. *Auralizations fundamentals of acoustics, modeling, simulation, algorithms and acoustic virtual reality*. Springer-Verlag, 2008. (Cited on pages 2, 45, 47, 48, 62, 69 and 71.)
- [Vorländer 2010] M. Vorländer. *Prediction tools in Acoustics - Can We Trust the PC?* In BNAM, 2010. (Cited on page 44.)
- [Vorländer 2011] M. Vorländer. *Models and algorithms for computer simulations in room acoustics*. In ISVA, pages 72–82, 2011. (Cited on page 2.)
- [Wang 2008] L.M. Wang and M.C. Vigeant. *Evaluations of output from room acoustic computer modelling and auralization due to different sound source directionalities*. Applied Acoustics, vol. 69, pages 1281–1293, 2008. (Cited on page 102.)
- [Wang 2014] X. Wang. *Model Based Signal Enhancement for Impulse Response Measurement*. PhD thesis, Aachen Technical University, Comeniushof, Gubener Str. 47, D-10243 Berlin, 12 2014. (Cited on pages 37 and 38.)
- [Wayman 1977] J.L. Wayman and J.P. Vanyo. *Three-dimensional computer simulation of reverberation in an enclosure*. J. Acoust. Soc. Am., vol. 62, no. 1, pages 213–215, 1977. (Cited on page 15.)
- [Weinbrenner 1809] F. Weinbrenner. *über theater in architektonischer hinsicht mit beziehung auf plan und ausführung des neuen hoftheater zu carlsruhe*. Tübingen: J.G. Cottaschen Buchhandlung, 1809. (Cited on page 147.)



- [Weinzierl 2010] S. Weinzierl, H. Rosenheinrich, J. Blickensdorff, M. Horn and A. Lindau. *Die Akustik der Konzertsäle im Leipziger Gewandhaus. Geschichte, Rekonstruktion und Auralisation*. In Proceeding of Fortschritte des Akustik, DAGA, Berlin, Germany, pages 1045–1046, 2010. (Cited on page 23.)
- [Weinzierl 2015] S. Weinzierl, P. Sanvito, F. Schultz and C. Büttner. *The acoustics of renaissance theatres in Italy*. Acta Acustica united with Acustica, vol. 101, no. 3, pages 632–641, 2015. (Cited on pages 24 and 44.)
- [Weitze 2001] C.A. Weitze, C.L. Christensen and J.H. Rindel. *Computer Simulation of the Acoustics of Mosques and Byzantine Churches*. In Proceedings of the ICA, 2001. (Cited on page 25.)
- [Weitze 2002] C.A. Weitze, C.L. Christensen and J.H. Rindel. *Comparison between In-situ recordings and Auralizations for Mosques and Byzantine Churches*. In proceedings of the BNAM, 2002. (Cited on page 25.)
- [Winkler 1865] E. Winkler. *Die akustik in elementarer darstellung*. Dresden: Wolde-mar Türk, 1865. (Cited on page 146.)
- [Wright 2008] C. Wright. *Music and ceremony at the notre dame of paris*. Cambridge university press, 2008. (Cited on page 70.)
- [Yang 2007] W. Yang and M. Hodgson. *Validation of the Auralization Technique: Comparative Speech-Intelligibility Tests in Real and Virtual Classrooms*. Acta Acustica united with Acustica, vol. 93, no. 6, pages 991–999, 2007. (Cited on pages 3, 79, 80 and 98.)
- [Zeng 2006] X. Zeng, C.L. Christensen and J.H. Rindel. *Practical Methods to Define Scattering Coefficients in a Room Acoustics Computer Model*. Applied Acoustics, vol. 67, pages 771–786, 2006. (Cited on pages 23 and 48.)

## Titre : Auralisations Sérieuses

**Mots clefs :** Auralisations, Réalité virtuelle, Calibration, Direction Voix, Multi-modalité

**Résumé :** L'objectif de cette thèse était d'examiner l'utilisation d'auralisations acoustiques de salles, basées sur l'acoustique géométrique (GA) comme outil scientifique et visant à aider à la création d'auralisations historiquement exactes plus écologiquement valables.

Aujourd'hui, si l'on veut créer des auralisations, l'enregistrement anéchoïque réalisé préalablement doit être convolué avec une réponse impulsionnelle ambiante, mesurée ou simulée (RIR). Les logiciels GA sont souvent utilisés pour calculer numériquement la RIR de géométries compliquées.

Cette thèse vise à améliorer la qualité des auralisations entièrement calculées. Des mesures acoustiques de la pièce ont été réalisées dans quatre salles et des modèles d'acoustique géométrique ont été créés des mêmes espaces. Une procédure méthodique de calibration du modèle a été proposée,

validée au préalable par comparaison d'estimation de paramètres et par tests d'écoute. Par la suite, un cadre permettant d'inclure la directivité vocale dynamique a été présenté. Les résultats des tests d'écoute ont montré des différences perceptuelles entre la directivité vocale dynamique et la directivité de source statique.

L'amélioration de la validité écologique des auralisations a permis d'étudier l'influence des visualisations sur l'expérience acoustique, avec un degré de confiance raisonnable que les effets perçus sont également applicables dans des situations réelles. À cet effet, un cadre a été établi qui a permis des évaluations multimodales de pièces de théâtre et de concerts. Les résultats d'un test d'écoute indiquait qu'avec une distance source-récepteur accrue, les auralisations sont perçues acoustiquement plus éloignées et plus élevée.

## Title : Serious Auralizations

**Keywords :** Auralizations, Virtual Reality, Calibration, Voice directivity, multi-modality

**Abstract :** The goal of this thesis was to examine the use of room acoustical auralizations based on geometrical acoustics (GA) as a scientific tool and aimed to aid in the creation of more ecologically valid historically accurate auralizations.

If one wishes to create auralizations, generally anechoic recording need to be convolved with either a measured or simulated room impulse response (RIR). GA software are often employed to numerically compute the RIR of complicated geometries. This thesis aims to enhance the quality of fully computed auralizations. First room acoustical measurements were carried out in four rooms and geometrical acoustics models were created of the same spaces. A methodical calibration procedure was proposed, first validated by means of parameter estimation comparison and second by listening tests.

Second, a framework was presented which enabled the inclusion of dynamic voice directivity. The results of listening tests indicated that perceptual differences between dynamic voice directivity and regular static source ordinations.

With the improved ecological validity of the auralizations it was possible to study the influence of visualizations on the acoustical experience, with a reasonable degree of confidence that perceived effects are also applicable in real-life situations. For this purpose, a framework was established which enabled multi-model assessments of theater plays and concerts. Results of an experiment indicated that with increased visual source-receiver distance auralizations are perceived acoustically more distant and louder.