The Pennsylvania State University

The Graduate School

A SPHERICAL MICROPHONE AND COMPACT LOUDSPEAKER ARRAY MEASUREMENT DATABASE FOR THE STUDY OF CONCERT HALL PREFERENCE

A Dissertation in

Acoustics

by

Matthew T. Neal

Submitted in Partial Fulfilment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2019

The dissertation of Matthew T. Neal was reviewed and approved* by the following:

Michelle C. Vigeant
Associate Professor of Acoustics and Architectural Engineering
Dissertation Adviser
Chair of Committee

Matthew Reimherr
Assistant Professor of Statistics

Daniel A. Russell
Teaching Professor of Acoustics

Victor W. Sparrow
United Technologies Corporation Professor of Acoustics
Director of the Graduate Program in Acoustics

David A. Dick
Acoustical Engineer, Applied Research
Special Member

*Signatures are on file in the Graduate School.

Abstract

Concert hall acoustics is an inherently conflicted, perceptual art form that meets the practical, engineering-based design process. What began as a trade learned on the job from experts has gradually evolved to an engineering-rooted discipline through computer simulations, impulse response measurement techniques, and auralization. Despite these advances, engineering design practices and processes must be fundamentally related to human perception that has driven the field for many centuries.

The goal of this current work was to study concert hall perception using spherical array processing techniques in a wide variety of real concert halls. For the study, the concert hall orchestral research database, or CHORDatabase, was generated, consisting of spherical microphone and compact spherical loudspeaker array (CSLA) room impulse response (RIR) measurements in 21 concert halls. The concert halls in the database include a wide variety in shape, size, reverberance, and geography, including 15 North American and 6 European halls.

RIR measurements were made using a 32-channel spherical microphone array and a three-part omnidirectional sound source, which enabled high-resolution spatial beamforming analysis and standard room acoustic metric calculations. Additionally, the microphone array enables spatially accurate auralizations using higher-order Ambisonics over a 30-loudspeaker auralization array. Finally, a 20-channel CSLA was built to flexibly reconstruct the frequency-dependent radiation patterns of different orchestral instruments. Using this source, repeatable and realistic full-orchestral auralizations were generated for each concert hall.

A subjective study investigated which factors were most important regarding concert hall perception and preference. A factor analysis revealed three to four factors of importance, relating to clarity, strength / spaciousness, and brilliance. Average preference best correlated with the perception of proximity, and for individuals, preference varied substantially, showing correlations with both the clarity and strength / spaciousness factors. Finally, spherical array beamforming analysis revealed time and spatial RIR regions that strongly correlated with envelopment, proximity, source width, and average preference. All of these perceptual attributes either lack a standardized metric or showed little correlation with the currently proposed metric to predict their perception in ISO 3382. Promising outlook is seen for developing new, perceptually informed metrics using spherical array processing techniques.
This project was supported through the National Science Foundation, Award #1302741.

Table of Contents

List of Tables

List of Figures

List of Acronyms

Acronym	Description
ACN	Ambisonic Channel Numbering
ambiX	Ambisonics exchangeable (data format)
AURAS	Auralization and Reproduction of Acoustics Sound fields (facility)
BQI	Binaural Quality Index (metric)
BR	Bass Ratio (metric)
BRIR	Binaural Room Impulse Response
C80	Clarity Index for Music (metric)
CHORD$_{atabase}$	Concert Hall Orchestral Research Database
CHRG	Concert Hall Research Group
CplxN3D	Complex Orthonormal Spherical Harmonic Normalization
CSLA	Compact Spherical Loudspeaker Array
DirAC	Directional Audio Coding
DirRIR	Directional Room Impulse Responses
EDT	Early Decay Time (metric)
FIR	Finite Impulse Response
FuMa	Furse-Malham Spherical Harmonic Normalization
G	Strength in the 125 Hz octave band (metric)
HOA	Higher-Order Ambisonics
HRTF	Head-Related Transfer Function
IACC	Interaural Cross-correlation
ILD	Interaural Level Difference
IR	Impulse Response
ITD	Interaural Time Difference
ITDG	Interaural Time Delay Gap (metric)
J_{LF}	Early Lateral Energy Fraction (metric)
LE	Lateral Efficiency (metric)
L_J	Late Lateral Energy Level (metric)

MATLAB	Matrix Laboratory
MicRIR	Microphone (array) Room Impulse Response
MLS	Maximum Length Sequence
MUSHRA	Multiple Stimulus Test with Hidden Reference and Anchor
N3D	Real-valued Orthonormal Spherical Harmonic Normalization
NFC	Near-Field Compensation (higher-order Ambisonics)
PCA	Principal Components Analysis
RIR	Room Impulse Response
RT	Reverberation Time
SDI	Surface Diffusivity Index
SDM	Spatial Decomposition Method
SH	Spherical Harmonic
ShRIR	Spherical Harmonic Room Impulse Response
SID	Single Index Designation
SN3D	Schmidt Semi-normalized Spherical Harmonic Normalization
SNR	Signal-to-noise Ratio
SPRAL	Sound Perception and Room Acoustics Laboratory (Penn State)
ST1	Support Factor (metric)
T30	Reverberation Time measured from the first 30 dB of the decay (metric)
TDOA	Time Delay of Arrival
TR	Treble Ratio (metric)
VBAP	Vector-Base Amplitude Panning
WFS	Wave Field Synthesis

Subscript Indicators for metrics (X is a placeholder):

X_{125}	Indicates a metric value in a particular octave band (here for 125 Hz)
X_{low}	Indicates a metric averaged over the 63 – 250 Hz octave bands
X_{mid}	Indicates a metric averaged over the 500 – 1000 Hz octave bands
X_{high}	Indicates a metric averaged over the 2000 – 4000 Hz octave bands
X_{early}	Indicates a metric for the early part of the room impulse response before 80 ms
X_{late}	Indicates a metric for the late part of the room impulse response after 80 ms

Acknowledgements

This project was quite a large effort, bringing together support to generate the resources required to complete this work. Funding for the initial and larger portion of the project was made possible through the National Science Foundation, Award #1302741. Support for the extension of the project was made possible through Penn State, Penn State's College of Engineering, and Penn State's Graduate Program in Acoustics. Additional thanks to the Acoustical Society of America and the Leo and Gabriella Beranek scholarship for supporting in part my educational costs of the PhD. Without any of these sources, the project and its completion would not have been possible.

The biggest effort in the project, the generation of the measurement database, requires many, many thanks. First, thank you to all of the consultants and researchers who helped fill out the concert hall survey. Additionally, thanks to the consultants and researchers who specifically worked to connect us with many of the halls we visited. Without these essential connections, these requests would likely not have been as streamlined. A very important thanks goes to the staff and teams at each of the 21 concert halls that were measured. Most all halls were provided free of charge, or at highly reduced rates, which was very generous. The significant amount of measurement time required in each hall was no small ask.

Beyond help gaining access to each hall, I owe a huge debt of gratitude to my fellow students and colleagues who assisted during these concert hall measurements. These thanks include: Martin Lawless, Fernando del Solar Dorrego, Zane Rusk, Andrew Kinzie, Andrew Doyle, Nick Ortega, Peter Moriarty, Will Doebler, Molly Smallcomb, Pranay Muchandi, Tom Blanford, Mark Langhirt, Nathan Tipton, Ingo Witew, Vahid Naderyan, and Maryam Landi. A special thanks to Ingo Witew, Marco Berzborn, and Prof. Michael Vorländer from ITA at RWTH Aachen for important equipment assistance prior to the European measurements. Finally, thanks to Katie Krainc for helping to organize and conduct the final subjective study as I focused all of my efforts on writing; the scope of this dissertation would not have been possible without her help. Overall, this project would not have been possible without any one of you! Know that I am forever grateful.

A PhD is not something you accomplish alone. It is a large, time consuming, draining, exhausting experience, and this difficulty can only be accomplished with the help of others. First and foremost, my parents have spent the last 28 years of their lives sacrificing for me, and my opportunities, education at Penn State, and whatever happens next all directly flow from their example. Dad, every day I work hard because of the work ethic you modeled for me. This work ethic is independent of circumstances, and it has always helped me push through, even when the situation looks difficult (which certainly happened in the PhD). Mom, you have given me my personality, wit, smile, and strength, and you have demonstrated a confidence mixed with selflessness that is rare. Every relationship I build, both personal and professional, is built upon watching and learning from your positive, gentle way with people. You always focus your time and effort on the people involved in the process, rather than simply the task at hand. Oh, you also threw a few dashes of stubbornness into the mix, so thanks for that too…

To my older siblings Adam and Jessica, you both have always been people I look up to, you kept life fun and interesting growing up, and somehow you both still put up with my stubbornness to this day. To my triplet brothers, Andrew and Benjamin, it is no question that I'm a much different person than I would have been without growing up next to you both. We three know that we can't really answer the question "So, what is it like to be a triplet," because we have no idea what it is like to not be a triplet! That said, I think I can venture a few guesses. I've always liked the idea of working on a team, and honestly, it is probably because I've always had a three-man team growing up. We've done much of life together, and much of who I've become has been shaped and changed by both of your personalities. Although we don't know what the alternative of not being a triplet would be like, I wouldn't have it any other way.

Apart from my family, I have a much larger extended family and extended friend family. Thank you to all of my people in life, whether from growing up, high school, Penn State undergraduate years, or grad school years. You have all made life worth living and given value far beyond anything that work alone could satisfy. Natalie, thanks for being here for me always, especially through this PhD. Whether we were 15 hours or 15 minutes apart, you were and continue to be a constant source of support, for which I am now and always grateful. Also, thanks for learning way more about acoustics than you probably bargained for…! Also a special thanks to some personal and faith mentors throughout my ten years at Penn State. To Nate and Dave, you both are wonderful examples of faithful men, and my life was strongly shaped and guided by your wisdom. Thanks for devoting your lives to serving and mentoring; I can think of few more noble ways to spend a career and devote your time.

I'd be remiss to not mention some of the key people that helped me accomplish my dissertation work at Penn State. Dave, thanks for being on my PhD committee, but mostly, thanks for leaving a great job to come to Penn State. The project would truly not have been possible without your practical signal processing expertise and unconventional teaching techniques. Martin, you put up with me throughout the PhD experience, and I wouldn't have wanted to go through this trying process with anyone else. Dave and Martin, you both are good friends, and I wouldn't want to be O.G. SPRAL with anyone else. Zane, Molly, and Evan, you three were some of the best undergrads I could have asked for, as beyond that, you are also just great people. Thanks for working hard, giving up some of your summers to do some (most likely slightly misguided) research and help me accomplish this large project.

Finally, thank you Michelle. It is hard to imagine that it was over seven years ago that we met at your interview presentation, and it is even harder to imagine all that has happened since then. After three years of searching for architectural acoustics at Penn State, your arrival was more than a coincidence, and beyond that, we have been a great team. Thank you for your openness, letting me see all aspects of the academic world, culture, the tenure process, and the process of starting a lab. Your trust in me was essential, and it allowed me to try new things, fail, try again, and ultimately, accomplish something we probably didn't fully think was possible. All too often in academia, professors are very sure of their own opinions and ideas on a problem, and these ideas are administered to the students to implement. Your trust to allow me to generate my own ideas, think independently, and your willingness to trust me in these changes has allowed me to develop into so much more than a technically competent individual. It has allowed me to develop my problem solving and research skills in ways I did not realize. That will serve me well for the rest of my life. On top of it all, you are also a good friend. It has always been good to know I can come to you with anything, and you will respect me not only as a researcher but also as a person and friend. I'm glad to have our friendship, and I am glad to know we will be friends and colleagues for many years to come. :)

This Page is Intentionally Left Blank

Chapter 1
Grand Introduction

Concert hall acoustics sits in the subtle place between art, music, science, architecture, engineering, perception, psychology, and many other disciplines. When designed with the utmost care, as was done for famous halls that have withstood the test of time, the room acts as an extension of the performer. A room can enhance the most subtle of pianissimos of a solo cellist and emphasize the most grandiose of fortissimos at the height of a symphony. On the contrary, a poor room can leave even the most skilled performer uninspired, dulling the impact and passion behind the music. The needs of a concert hall are many. Halls must support such a wide variety of performance types, and for what ultimate goal? The ever so elusive concept of preferring, liking, or feeling that the room subjectively enhanced the performance. It is this very subjective nature of the field that makes success, and the entire problem of hall design, uniquely difficult. And if success is deemed as a summation of many judgements of preference, made by many individuals, how are such a criteria defined?

Other disciplines of engineering have clear criteria for success, about which everyone agrees. Most everyone would likely agree that a structural engineer's job is successful if the building remains intact and sturdy through its lifetime, even through earthquakes and other natural disasters. Such a task is not easy by any means, but the task at hand is at least clear. Preference is a tricky topic, as it can change with many factors. In acoustics, it can change with musical genre, individual, and even day-to-day. Silly factors like a meal eaten before a concert or a troublesome day at the office can impact how the concert environment is approached. Although hall design cannot attempt to predict the quality of a meal, the importance of individual preference and preference across musical genres is essential to the goal of a project. Currently, the role of dictating the design goals of a concert hall falls upon the principal or senior acoustic consultant. From many years of practice and experience, design goals are set that are known to have previously produce successful results. These goals are informed and tailored to the needs, desires, and budget of the client, but the client must trust the acoustician. As the expert, the acoustician builds upon previously successful room designs and uses their experience to achieve success again.

With the advent of new technology in the field, auralization, or an aural rendering of what a room will or currently sounds like, is integrating its way into real-world projects and designs. Currently, its use is mainly simulation-based or recording-based from first-order Ambisonic microphones. These techniques are not new to the field of acoustics; spherical array processing and Ambisonics were proposed almost 50 years ago, but as with most scientific discoveries, time is needed for them to find practical and tangible use. The work done in this dissertation has sought to walk the line between the subjective and the objective, the measured and the perceived, the scientific and the practical, and between art and engineering. All of the extremes must exist in proper balance. Scientists can do all of the science they would like, but if the science is not done in a way that it is accessible to practitioners and consultants, it will not find any use. This fact may be less important in other disciplines, but in the built environment, research is only impactful if new concert halls can be designed differently, more efficiently, or using new techniques or methods not previously available.

This dissertation will present how existing scientific methods based upon spherical array processing techniques can impact and extend the currently knowledge of concert hall acoustics. Spherical array processing can allow for accurate measurement-based auralizations of existing rooms. Also, spherical array beamforming can be used to accurately capture the spatial sound field in a room, even with spatially accurate source radiation properties. This level of spatial accuracy allows for large increases in realism compared to what is possible using less sophisticated measurement techniques. With spatially accurate auralizations, realistic auralizations can form the basis of subjective research using measured rooms. The use of high resolution, spatial measurements in a wide variety of real rooms helps to ensure that subjective testing can be done in a way that is very close to the real experience. Finally, advanced spherical microphone array beamforming analyses techniques give access to the full time, frequency, and spatial information that impacts an individual's perception of rooms.

This dissertation will cover an extensive research project, highlighting the use of spherical array processing techniques in a comprehensive, large scale objective measurement and subjective testing effort. First, chapter 2 provides the background on subjective research focusing on overall room perception and the multi-dimensional space of concert hall perception. This chapter also includes mention of current standardized hall measurement techniques and their limitations. Chapter 3 highlights the mathematical framework that forms the basis of spherical array processing for spherical microphone array beamforming, source radiation beamforming control, and virtual acoustics using higher-order Ambisonics.

The next three chapters, 4 – 6, each represent a key component of the large body of this dissertation. Chapter 4 describes in detail the design and construction of a compact spherical loudspeaker array (CSLA) that provides accurate reconstruction of instrument radiation patterns. Chapter 5 outlines the development and processing techniques that were used to generate the concert hall orchestral research database, or CHORDatabase. This database contains extensive measurements in 21 concert halls from around North America and Europe, containing a total of 242 room impulse responses (RIRs) and forms the basis of this work. Finally, chapter 6 describes a subjective study aimed to increase the current understanding of perception in room acoustics and analyzes the potential for studying concert hall preference at the individual level. Chapter 7 provides final summary and conclusions of the document. This work has been a labor of love (most of the time) and it is the author's sincere hope that this work will not simply gather dust, as another study to place on the bookshelf. Through the extensive visualization and explanations, hopefully some new understanding, pieces of knowledge, or shifts in perspective can begin to change how the design of a concert hall is approached in new, innovative ways resulting in acoustically beautiful spaces.

This Page is Intentionally Left Blank

Chapter 2
Concert Hall Acoustics

This chapter presents the key concepts of concert hall and room acoustics, along with background information on room acoustic measurement, auralization, and objective analysis. Summary information of the relevant literature from past research in overall preference of concert hall acoustics is presented. These studies are grouped into six categories: architectural and acoustic measurement, interview and survey-based, laboratory, live listening, measurement-based, and simulation-based auralization approaches. Each of these categories utilized different techniques to study concert hall perception, and each technique contains unique pros and cons in terms of study realism, non-acoustic influences, auralization accuracy, simulation accuracy, and much more. These different study techniques will be discussed, and overviews of the methods used for taking room acoustic measurements and generating auralizations are provided.

2.1 Concert Hall Acoustics

The field of room acoustics has gradually evolved over many years. Although the idea of acoustics, sound, and the impact of space upon sound has been used for thousands of years, it was not until some tinkering physicists sparked an interest in the field that scientific studies began. The often quoted *father of architectural acoustics*, Wallace Clement Sabine, applied a scientific approach to studying a lecture hall on Harvard's campus, Fogg Auditorium.[1] By studying the impact of adding sound absorptive material into the room, he was able to derive a fundamental equation for the reverberation time (RT) of the auditorium, relating the room's volume, surface area, and absorptive properties of different materials to the time it took the sound level to decay 60 dB. This fundamental equation remains the most well-known objective metric to date, and it helped transition the study of architectural acoustics from an experience-based practice and art form to a balance between the study of art and science.

As the mid-20th century began, many in the field expressed their opinion that RT alone did not full-dictate the quality of a room's acoustics. A search began for new parameters, to help better understand the multi-dimensional nature of room acoustic perception. The current author has been told a quote from a past Penn State Professor of Acoustics, Dr. Jiří Tichý, who

commonly referenced the following when describing the role of Sabine's RT in a course on architectural acoustics:

> *Proper reverberation time is a necessary condition, but not sufficient for achieving good acoustics.*[*]

The quote above well-summarizes this sentiment, and the author (with some Penn State pride in his heart) found it quite timely due to the recent passing of Dr. Tichý (1927-2019). It should be noted that although Sabine was made famous for his derivation of RT, he was aware that other factors impacted the acoustic quality of a room. For example, in his *Collected Papers on Acoustics*, he gives mention to the perception of loudness, the shift of localization due to early reflections, room shape, focusing of sound energy, echoes, and modal room effects.[1] Kuusinen provides a well-selected quote from Sabine in his MS thesis:[2]

> *In order that hearing may be good in any auditorium, it is necessary that the sound should be sufficiently loud; that the simultaneous components of a complex sound should maintain their proper relative intensities; and that the successive sounds in rapidly moving articulation, either of speech or music, should be clear and distinct, free from each other and from extraneous noises."*
>
> *– From Reverberation (The American Architect, 1900)*[1]

2.2 Concert Hall Acoustic Quality Studies

Scientific research to identify other aspects of room acoustic quality continued with Somerville, who studied three overall parameters to predict the acoustic quality of studies: the mean RT across frequency, the divergence of the frequency-dependent RT from the mean, and the divergence of the decay of a steady tone in the room from a linear slope.[3] The latter two parameters were defined graphically, by integrating the shaded area between measured data and the mean RT and integrating between the actual room's decay and its linear best-fit decay. The study focused on recording or broadcast studios for the British Broadcasting Corporation, but measurements of various concert halls were also included. A subsequent study involved a subjective questionnaire completed by 117 respondents (which was quoted as disappointingly small).[4] A new investigation, now also using a fourth metric related to room volume, was used to predict a hall as having subjectively "good", "average", or "bad" acoustics. None of these

[*] Quote provided courtesy of Richard Talaske, May 2nd 2019

decay irregularities caught widespread use in the field, though this could be due to the fact that this study focused primarily on studio spaces and not concert halls.

Music, Acoustics, and Architecture, published by Leo Beranek in 1962, was the cornerstone publication on this topic.[5] This publication combined subjective data collection efforts made by Beranek himself. He conducted surveys of "conductors, music critics, and aficionados of concert music" regarding their experience listening to full-orchestral works of a symphony orchestra in various halls. From this experience, he developed a list of 18 subjective terms he found to represent the acoustical quality of concert halls. This was the first formalized, large-scale compilation of subjective attributes to describe room acoustics. Although the scientific methods employed by Beranek at this time can be critically examined, this work provided a springboard for publications on the topic. Table 2.1 lists the key details of the most prominent works on the subject, mainly focused on work following Beranek's publication in 1962. The present investigation is focused on the overall perception of quality or preference for room acoustics, and as such, only studies with a focus on this overall perception are included. More specifically, studies focusing on a single perception, such as clarity, apparent source width, listener envelopment, intimacy, etc., without relating back to overall subjective quality, are not included. These studies are of primary importance in describing each individual perception but do not provide contextual importance related to room acoustic quality. Further study details will be provided in sections 2.2.1 through 2.2.5.

Table 2.1: Summary table of significant studies related to overall room acoustic quality. Studies listed in this table include works that aimed to explain overall room acoustic quality by identifying key items of importance, where they be objective, subjective, architectural, etc. in nature. Along with publication details, study type is listed, corresponding to the literature summaries organized by study methods provided in section 2.2.1 through 2.2.6.

First Author	Journal	Year	Study Type
Sabine	Harvard Press	1930	Measurement-based
Somerville	Acustica	1953, 1957	Measurements & surveys
Beranek	Springer	1962	Surveys
Nickson & Muncey	JS&V	1964a, b	Literature review
Marshall	JS&V	1967	Geometric approach
Jordan	JASA	1970	Experience-based assessment
Hawkes & Douglass	Acustica	1971	Live performance survey
Yamaguchi	JASA	1972	Measurement-based auralizations
Schroeder et al.	JASA	1974	Measurement-based auralizations
Kimura & Sekiguchi	JASJ	1976	Measurement-based auralizations
Wilkens	Acustica	1977	Recorded live auralizations
Reichardt & Lehmann	Appl. Acoust.	1978	Laboratory studies
Jordan	Appl. Acoust.	1981	Experience-based assessment
Barron	Acustica	1988	Live performance survey
Lavandier	Diss., IRCAM	1989	Laboratory studies
Kahle	Diss., IRCAM	1995	Live performance survey
Sotiropoulou et al.	Acustica	1995a, b	Live performance survey
Beranek	Springer	1996	Measurements & interviews
Beranek	ACTA Acust.	2003	Measurements & interviews
Lokki et al.	JASA	2011, 2012, 2014, 2016	Measurement-based auralizations
Skålevik	ICSV	2017	Online surveys
Weinzierl et al.	JASA	2018	Simulation-based auralizations

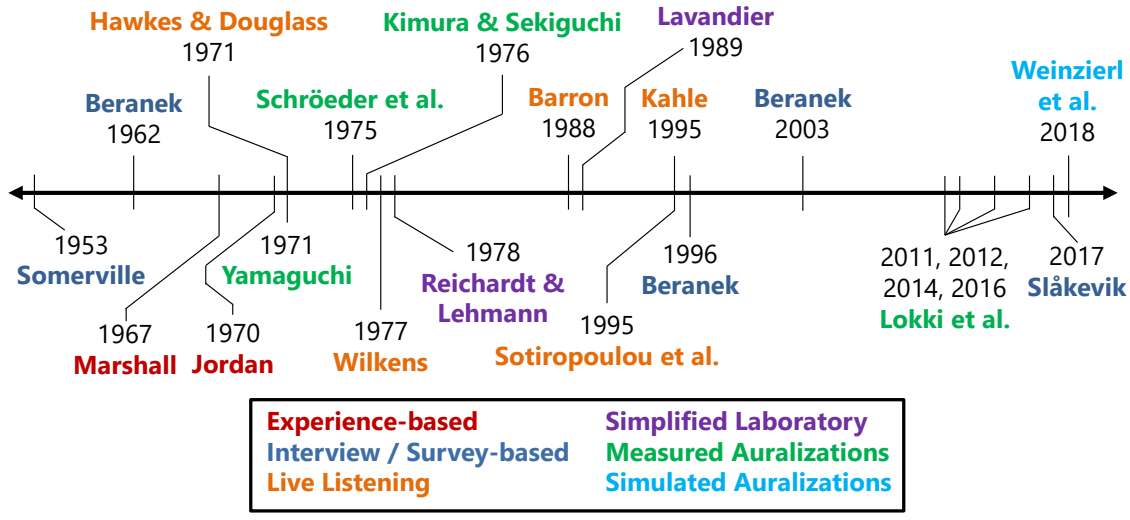


Figure 2-1: A timeline containing the most prominent studies that focused on the perception of overall acoustic quality in concert halls, many of which are also listed in Table 2.1.

2.2.1 Architectural and Acoustic Measurement Approaches

The first group of studies focused on explaining concert hall quality with measurements other than RT and did not include any specific subjective components. These studies include other RIR-derived acoustic metrics and architectural hall measurements. Nickson and Muncey co-wrote two review articles on the subject of concert halls, looking at previous studies which had investigated hall quality.[6-7] Although they believed no single condition or acoustic environment would be ideal for all performances, a small set of criteria was suggested, which if satisfied, would ensure good results. These included:

1. Proper RT be met, based upon recommended targets
2. Wall, ceiling, and overhead reflectors ensure all positions receive a substantial reflection within 35 ms of the direct sound.
3. Concave hard surfaces should be avoided
4. External noise should be minimized
5. Seating should be comfortable and *not squeak*

These simple, yet practical guidelines are clearly defined, and the importance of each seems practically valid. Some attempt at quantifying acoustic effects, such as the timing of the first reflection, is provided but without clear subjective basis.

Marshall (1967) investigated the impact of room cross-sections and the arrival timing and amplitudes of sidewall and ceiling reflections on the perception of hall quality.[8] Marshall suggested the perceptions most important to quality include loudness, spatial responsiveness, and envelopment which he defines as "direct involvement with a room". He presents all of these concepts in the context of room reverberation but places a far more focused importance on the subjective importance of early reflections and the importance of their arrival order, strength, and timing in the *spatial responsiveness* of a room. Marshall based his approach on psychoacoustic masking research performed by Seraphim based on speech perception.[9] Extending this reflection sequence masking analysis to hall geometries, Marshall suggested that low, strong ceiling reflections can create a masking of beneficial side wall reflections. Additionally, he proposed the existence of ideal room cross section ratios to ensure spatial responsiveness: when a cross-section is kept narrow, the proper arrival time of side wall reflections is prioritized.

Other than geometric recommendations, Jordan proposed various metrics that he felt outperformed Sabine's RT, and more sufficiently explained room acoustical quality. First, Jordan proposed the use of Early Decay Time (EDT) to predict the sense of reverberance in

1970.[10] Ten years later, Jordan proposed a set of objective criteria, including EDT, Clarity Index (C80), lateral efficiency (very similar in definition to lateral energy fraction – J_{LF}), and a parameter for tonal balance proposed to be the slope of frequency-dependent EDT values from 250 to 2000 Hz.[11] Much of Jordan's work is based upon his experience in the field, along with many conversations with consulting and research colleagues.

As measurement advances continued into the 1980's and 1990's, the need for more standardized measurements in a number of concert halls was highlighted in a meeting called together by Chris Jaffe, which started the Concert Hall Research Group (CHRG).[12] This group funded measurements made by three teams, led by Anders Christian Gade from the Technical University of Denmark, John Bradley from the National Research Council of Canada, and Gary Siebein from the University of Florida. These teams were already collecting data for halls in North America and Europe, mainly using a starter pistol sources and a diffuse-field omnidirectional microphone for recording RIRs. From these measurements, Chiang made comparisons between geometric hall properties and RIR-based metrics of EDT, C80, G, bass ratio (BR), treble ratio (TR), and early inter-aural cross correlation ($IACC_{early}$).[13] His study included 22 rooms, half of which were lecture halls and churches located in the Gainesville, FL area. Chiang highlights how specific architectural dimensions correlate with metric quantities, but no subjective ratings or testing is included in the work.

Both Gade and Bradley were already collecting measurements in concert halls prior to the CHRG funded effort, and that measurement effort helped increase the breadth of their work. Gade published about a collection of measurements that had been collected between himself and Bradley's team in 53 halls, including 9 from the US CHRG measurements.[14] In an effort to provide general recommendations in hall geometric properties, regression analyses were conducted. Gade focused on studying strength (G), lateral energy fraction (J_{LF}), Clarity (C80), and the change in strength per 10 m increase in source-receiver distance (ΔG). Regression analyses predicted the values for these parameters from geometric information and expected values of C80 and G were calculated from the room's RT.

Various other groups have performed measurement tours of famous halls, with similar goals of collecting data to determine what makes the most well-known halls exceptional. Despite the many efforts, measurement setups are often non-standardized, or do not collect adequate levels of spatial information required for room sound-field analysis and auralization. All these studies provide better understanding of room acoustic metrics and their link to practical architectural design considerations, but important information regarding subjective impression was not able to be studied.

2.2.2 Interview- and Survey-based Approaches

The initial, most prominent work implementing interviews to gather subjective efforts was completed by Beranek in 1962, previously mentioned in section 2.2.[5] Beranek conducted interviews with a number of musicians, conductors, and consultants to propose a list of 18 subjective attributes that better represented the multidimensional nature of concert hall perception. These 18 factors included intimacy, liveness, warmth, loudness of the direct sound, loudness of the reverberant sound, clarity (or definition), brilliance, diffusion, balance, blend, ensemble, immediacy of response, texture, freedom from echo, freedom from noise, dynamic range, tonal quality, and uniformity throughout the hall. These terms prove fairly comprehensive in nature, but a high amount of correlation exists between the categories. This list mixes descriptors of the sound field, such as uniformity and direct sound loudness, with descriptors of the subjective perception of the room, such as warmth, brilliance, intimacy, etc. Beranek provides a final summary of this work in a 100-point rating scale to judge the quality of a concert hall, with final rating scale categories shown in Table 2.2.

Table 2.2: A summary of Beranek's proposed concert hall rating scale, providing a number of points for different categories of perception on a linearly additive 100-point grading scale.

Attribute	Max. No. of Points
Intimacy	40
Liveness	15
Warmth	15
Loudness of direct sound	10
Loudness of reverberant sound	6
Balance and blend	6
Diffusion	4
Ensemble	4
Total:	**100**

Beranek continued to update his analysis, including more interviews, more concert halls, and eventually, many calculated metrics in his study. Two new publications were written in 1996 and 2003.[15-16] The largest benefit of the new analysis was the inclusion of objective parameters, which he related to his concert hall rankings. An additional publication in 2003 provided analysis of his subjective rank orderings against many metrics, including binaural quality index (BQI), reverberation time (T30), EDT, bass ratio (BR), strength at 125 Hz (G_{125}), strength at mid frequencies (G_{mid}), late strength following 80 ms from the direct sound (G_{late}), lateral fraction (J_{LF}), surface diffusivity index (SDI), initial time delay gap (ITDG), and support factor (ST1).[17] Each metric was correlated against subsets of the 58 measured halls, depending upon metric availability. Beranek concluded that the more important and significant metrics for concert hall acoustics include BQI_{mid}, EDT_{mid}, G_{125}, SDI, and ITDG, in this ordering of

importance. This effort was the most comprehensive objective metric and subjective analysis to that time, but the difference in measurement teams, equipment, and parameter calculation techniques, especially for early reflection dependent metrics, may have an impact on the results of certain metrics.

Using Beranek's rank orderings and hall listing as a starting point, Skålevik conducted an online survey of 84 respondents providing 822 votes as of March 28th, 2017. The ratings exhibited a linear relationship with Beranek's rank ordering, $r^2 = 0.58$, which increased to $r^2 = 0.79$ if only halls containing over 12 submitted ratings were included. He then used Beranek's architectural measurements and acoustic measures to predict optimal values for a highly regarded concert hall, using the metrics $T30_{mid}$, $T30_{125}$, C80, G_{mid}, G_{125}, G_{late}, H/W, and W, where H refers to a room's height and W refers to a room's width. A metric defined as a weighted summation of deviations of these parameters from optimal values is proposed as an indication of concert hall quality. The use of an online survey with Beranek's existing repository of data is of interest, but the lack of consistency in the collection of acoustic metrics for Beranek's book and the accuracy of subjective data collected from an online survey might be called into question. In addition, the existing well-known text by Beranek, thought by many to be the gold standard in the field, most likely biases the ratings provided online, ensuring correlation between results. It is likely that responders to the survey would be familiar with Beranek's rank ordering, biasing the correlation in rankings.

In general, survey-based approaches suffer from the main limitation of uncontrolled subjective factors. First, attempting to remember a subjective auditory experience is not an easy task, let alone making subtle acoustic judgments between two different halls. Moreover, differences in the types of music performed, and the attitude or biases of the listener are not removed in this technique. Moreover, each person's role in a concert hall can be quite different, depending if they are a performer on stage or a listener in the audience. Without providing a consistent, repeatable listening experience or test, these factors add subjective 'noise' to the data. This noise will cloud or blur potential findings regarding room acoustic quality, and in the worst case, they can bias results and lead to incorrect conclusions.

2.2.3 Synthetic Laboratory Auralization Approaches

Research from the group at Dresden University of Technology in the 1960s and 1970s focused initially on works studying the perception of clarity in concert halls, by which the commonly used clarity index for music (C80) was defined.[18] Reichardt referred to this concept as *clearness*, separate from the previously proposed *distinctness* for the perception of speech.[19]

The work of this group was done in a highly controlled setting, generating room-like reproductions with multiple loudspeakers surrounding a listener in an anechoic chamber. Discrete specular reflections were reproduced as delayed and attenuated copies of the direct sound from discrete loudspeakers. Then, reverberation was generated using uncorrelated artificial reverberation units played out of each loudspeaker. This technique allowed for fine control of the sound field, to finely study the impact of even adjusting single reflections in the sound field. Additionally, Reichardt and Lehmann proposed a metric coined the *Room Impression Index* (R, or *Raumeindrucksmass*) which combined the impact of direction and time of arrival of room reflections:[20]

$$R = 10\log_{10}\left[\frac{\int_{0.025}^{\infty} p^2(t)dt - \int_{0.025}^{0.08} p_r^2(t)dt}{\int_{0}^{0.025} p^2(t)dt + \int_{0.025}^{0.08} p_r^2(t)dt}\right]. \quad 2.1$$

The response of an omnidirectional microphone is given by $p(t)$, and the response from a measurement with a *frontally directed microphone* (*Richtmikrofon*) is denoted by $p_r(t)$. The *frontally directed microphone* is defined as responding equally to sound within ±40° in front of a listener and rejecting sound outside of this region. This requirement was not described in terms of practical measurement considerations, but rather, seems to be linked to the fundamental subjective perception. Although this metric considers time of arrival and direction of arrival, it has little been studied or implemented, most likely due to the difficulty of a practical measurement procedure.

A group of studies were conducted at IRCAM, contained in the theses of Catherine Lavandier and Eckhard Kahle. Due to the availability of these works only in French, these summaries are in large part based on the summaries by Kuusinen.[2] First, Lavandier attempted to validate commonly used objective acoustic measures in terms of their perceptual accuracy.[21] An array of loudspeakers, delays, filters, and a reverberation unit were used to simulate a hall-like effect for test subjects. By adjusting temporal, spatial, and frequency balances of parameters, a series of 17 listening tests were conducted. The study resulted in the suggestion of 14 different perceptual factors using individual differences scaling techniques. The final factors were reported in four different groups: those concerning temporal effects (5 factors), early reflection effects (3 factors), reverberation and strength effects (4 factors), and spatial effects (2 factors). Lavandier's thesis focused on laboratory-based manipulations of room-like soundscapes, but Kahle extended this work to live listening-based work, described further in section 2.2.4.[22] Both of these laboratory-based studies provide fine control for the experimenters, but it can be questioned whether or not such a simplified representation provides an accurate depiction of the complex acoustic environments found in a concert hall.

2.2.4 Live Listening Studies

Survey and interview approaches can be extended one step further by conducting surveys during live concerts. This technique ensures that subjects are analyzing the hall from the perspective of a listener, and not a performer. Additionally, although direct comparisons cannot be made between halls, listeners are rating a hall actively during the concert, rather than remembering back to an experience that occurred weeks, months, or even years later. This scenario overcomes the difficultly of remembering back to an auditory experience. The first study to implement live listening was performed by Hawkes and Douglas.[23] In this study, questionnaires contained 16 different bipolar ratings scales, and subjects were instructed to provide a rating for each hall condition. Three separate studies were done. First, factor analysis compared the subjective ratings during multiple concerts in Royal Festival Hall. At each concert, an electronic Assisted Resonance System adjusted the RT of the hall in the range of 58 to 660 Hz.

For the Hawkes and Douglas study, a multidimensional scaling technique identified six subjective categories to be most important, including balance and blend, resonance, intimacy, brilliance, proximity, and definition. A subsequent factor analysis identified five categories, including resonance, definition, proximity, balance and blend, and brilliance as important. The main limitation of this study was that room changes were only made by adjusting an electronic enhancement system, which may not provide a realistic difference compared to listening to separate rooms. To adjust for this effect, an additional study was performed using the same technique and repeated in four different halls. From this study, a factor analysis yielded the following 6 factors as being most important: reverberance, evenness, intimacy, definition, enjoyment, and brilliance. The factors identified in the study clearly demonstrate the multidimensional space of concert hall subjective impression.

German researchers from the Technical University of Berlin utilized a new technology of binaural recordings to allow for comparisons of the multiple halls in the close temporal proximity a laboratory reproduction provides. For their study, recordings were made in five seats during live orchestral (unoccupied) performances in six German concert halls. The researchers were able to travel with the Berlin Philharmonic Orchestra and made recordings of three different orchestral works by Mozart, Brahms, and Bartok. This procedure allowed for realistic comparisons using electroacoustic reproductions over headphones or loudspeakers. Despite this control, playing differences still existed between each performance, which could impact the repeatability of reproduction. From the study, Wilkens found that three dimensions could explain over 90% of the variance associated with 19 subjective scale ratings from 40

subjects. The three subjective factors deemed most important were identified to be strength, distance or clarity, and the timbre of the sound. Again, the furthering of the idea of a multidimensional auditory perception space is supported by the group. Additionally, two groups of subjects were identified, explaining differences in the preference ratings of perceptual data. The first group preferred a loud sound, and the second group of subjects preferred a clear sound.

Barron conducted another listening-based study, adopting the questionnaire developed by Hawkes and Douglas.[24] In his study, survey respondents included 27 expert listeners, most of whom were trained acousticians. Eleven different British concert halls were selected, a total of 40 listening positions were used, and the study resulted in 227 completed questionnaires. The questionnaire was a condensed version from Hawkes and Douglass's study and included categories found to be important in previous works: clarity, reverberance, envelopment, intimacy, loudness, balance (in terms of treble, bass, and singers / soloists relative to the orchestra), background noise, and overall impression. In analyzing the results, Barron found that across all subjects, overall impression was most highly correlated with reverberance, envelopment, and intimacy. Upon further analysis, subjects were divided into two groups: subjects that most highly correlated preference with reverberance, and subjects that most highly correlated preference with intimacy, indicated by the Venn diagram in Figure 2-2. Each group also had a high degree of correlation between envelopment, overall impression, and their preferred factor, but the correlation with the non-preferred factor largely decreased. These results have been repeated in Table 2.3 from Table III in Ref. [24]. Barron's study was performed using many halls, and the recruitment of more expert listeners is a clear advantage. Still, limitations exist associated with live performance factors, musical content, and other non-acoustic biases.

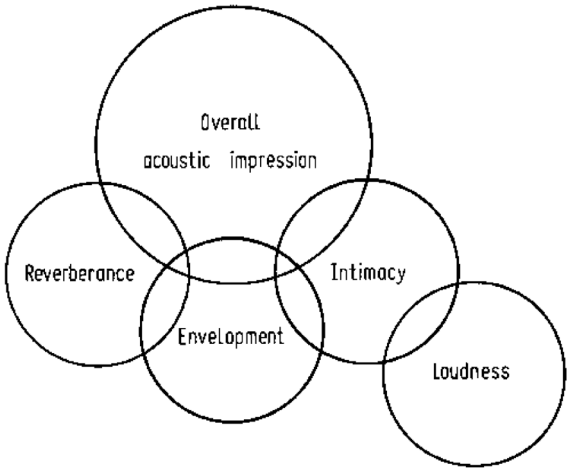


Figure 2-2: The Venn diagram illustrating the relationship between subjects' overall impression of a room, separated into two groups (reproduced from Fig. 2 in Ref. [24]).

Table 2.3: Barron's correlations between subjective factors, reproduces from Table III in Ref. [24].

	Reverberance	Envelopment	Intimacy
Subject group (a) that preferred *Reverberance*			
Envelopment	**0.74**	1	
Intimacy	0.09	0.30	1
Overall Acoustic Impression	**0.75**	**0.69**	0.34
Subject group (b) that preferred *Intimacy*			
Envelopment	0.22	1	
Intimacy	-0.02	**0.54**	1
Overall Acoustic Impression	0.18	**0.58**	**0.62**

Kahle furthered the research of Lavandier, working in conjunction with the team at IRCAM.[22] Kahle's study addressed the current research problem by using objective measurements in nine European halls and perceptual ratings judged by a group of around ten assessors during real concerts. A structured survey used by the assessors included 29 questions. Questions were grouped into one of five categories, including the perception of the hall's acoustics, the perception of the sound sources, the perception of individual sources per section, the spectral balance, preference, and personal opinions. Kahle's work resulted in the identification of eight fundamental attributes from the 29 questions to describe room perception: sound level, reverberance, general balance, contrast, level of low frequencies, level of high frequencies, muddiness, and hardness. Many more specific conclusions were also supported by Kahle's thesis, including the importance of early energy for the perception of reverberance, the separation of sound level perception into a source component and room component (source presence vs. room presence), and the need for many new metrics to describe difficult perception to assess. A summary in English of Kahle's study can be found in the thesis of Kuusinen.[2]

Sotiropoulou et al. collected subjective rating data during three live concerts in two different concert halls.[25-26] Many subjective terms in acoustics were considered for the study, and a final total of 86 terms were considered. A simple subjective experiment by listening to recorded music in a living room-type environment was used to down select these terms to a set of 27 scales. A total of 80 subjects attended the first concert, 28 of which were specifically invited and continued to attend all three concerts. A set of four descriptors was developed using a separate factor analysis for each concert, and the four overall categories were determined to be body, clarity, tonal quality, and proximity. In order of importance, they were found to be (1) clarity, (2) body, (3) tonal quality, and (4) proximity for concert A, (1) body, (2) clarity, (3) proximity, and (4) tonal quality for concert B, and (1) body, (2) tonal quality, (3) clarity, and (4) proximity for concert C. The most important factor, body, was correlated with the descriptors of full, mighty, sonorous, and voluminous. The added importance of tonal quality

in this study is important, as metrics that reliably predict these qualities are not well defined. Clarity was one of the more commonly known and reported factors of importance, and proximity was associated with distance perception and somewhat with envelopment.

Many of these studies provided a realistic presentation of the orchestra, but without measurements using a spatial microphone array, binaural mannequin, or in some case, even impulse responses, direct comparisons of halls are not possible. Some studies attempt to overcome this limitation by using trained, expert listeners, but still non-acoustic effects and performance-specific differences are not able to be held constant. The next set of studies in the following section attempted to overcome this limitation with the use of auralizations and virtual acoustic techniques, which allowed more realistic laboratory-based room comparisons.

2.2.5 Measurement-based Auralization Approaches

The first convincing large scale concert hall studies using auralization came with the invention of the binaural mannequin, or "dummy head".[27-28] By creating a model of the human head, with microphones mounted in the left and right ear canals of realistically-shaped pinnae, a binaural recording or binaural room impulse response (BRIR) can be captured. Then using either headphones or cross-talk-cancellation, a sound field can be rendered for two different halls, side-by-side in a laboratory setting. As described in the technique used by Wilkens in section 2.2.5, direct playback of recordings can be used for live performances, but this method does not ensure consistency in the musical performance in each hall. If a BRIR is measured in each hall, a virtual recreation of the acoustic listening conditions in a hall, or *auralization*, can be generated. The BRIR can be convolved with recordings made in dead or anechoic environments, free of room reflections and reverberation. The result is a virtual recreation of sitting where the dummy head was positioned, listening to a performer located where the measurement loudspeaker was placed.

Multiple studies have implemented binaural techniques, using a few different measurement methods. First, Yamaguchi conducted a series of measurements at ten seating locations in Yamaha Music Camp's auditorium.[29] Recordings were made by playing anechoic music out of an omnidirectional loudspeaker on stage and recording at various seat locations with two microphones spaced at 50 cm. No physical human head model was included in this measurement setup (no *dummy head*). Reproductions of recordings were generated using headphones directly from the recorded signals. Principal components analysis and factor rotation were used to identify the most important factors influencing the perception of both speech and light music. A total of 13 subjects participated in a comparative rating task, where

they were asked to judge the differences between different seat locations. By comparing each factor with a large set of objective metrics, three factors were found to be most important regarding perception. The first dimension was interpreted as reverberation and definition, the second dimension as sound pressure level, and the third dimension did not afford a simple interpretation. The study utilized good subjective analysis techniques, but the limitations in the binaural recording setup, without any physical head geometry, call the realism of the results into question.

Schroeder et al. also used a binaural head to record anechoic music played from two loudspeakers on the stages of 22 European concert halls, to roughly represent the spatial extent of an orchestra.[30] Recordings were reproduced using cross-talk-cancellation, and a listening test was conducted where subjects provided a binary selection of preference between two concert halls and an option to specify no preference. A total of 12 listeners participated in the study, and factor analysis revealed the importance of two dimensions, and possible marginal importance of a third and fourth dimension. In final interpretations, halls with RTs exceeding 2.2 s were excluded from the analysis, citing high unoccupied RTs. For the 11 halls meeting this criterion, the first dimension was found to relate to RT of the first 15 dB of the sound decay.

When the same analysis is performed for halls exceeding a 2.0 s RT, the first dimension highly correlated, in an opposite sense, with clarity, and RT showed high, and almost orthogonal correlation with the second dimension, also in a negative sense. This first factor was found to exhibit a consensus among all subjects' preferences. The second dimension was shown to exhibit inter-individual differences between preference ratings, and no straightforward interpretation was provided for this dimension. Individual subjective impressions, which help to inform underlying judgments of preference, were not included in this task, so interpretations were more difficult to draw. Objective parameters were correlated with the perceptual space, but this analysis relies on the accuracy of such parameters to predict an underlying acoustic impression. An important note to make is that all auralizations were level-equalized, removing the natural strength differences between halls. Although level is often a highly dominant perception in room preference, it is a naturally occurring effect. Its removal from this study separates the conclusions about preference somewhat from the realistic condition of the rooms, as effects for any subjective impression correlated with loudness, such as envelopment, are removed as well.

Kimura and Sekiguchi performed binaural recordings as well for anechoic music played out of a non-directional loudspeaker, performing recordings in 13 multipurpose auditoriums.[31]

Auralizations were presented over headphones, and pairwise comparisons were made between similar locations in all 13 halls. Also, many seat locations were measured in two of the halls. Subjects were asked to rate a wide variety of perceptions, including loudness, reverberation (for both quantity and quality), spatial impression, brilliance, definition, proximity, and preference. A resulting factor analysis revealed that the main axis related most strongly to spatial impression and quantity of reverberation when a stage enclosure was not installed. When a stage enclosure was installed, the primary dimension was most related to brilliance. Many other investigations of different objective metrics with these subjective impressions were performed as well. Some interesting results, contrary to other literature, are found in investigating their reported correlations. First, loudness showed a weak, negative correlation ($r = -0.36$) with spatial impression, quite contradictory to other literature on the topic of envelopment.[32-33] In addition, loudness was not well correlated with preference, only having a correlations coefficient of $r = 0.2$, also very much unlike many other studies on the topic.

Apart from their extensive measurement efforts with Gade, Soulodre and Bradley studied the perception of loudness, clarity, reverberance, bass, treble, envelopment, apparent source width, and preference in a paired comparison study.[34] Ten BRIRs from North American concert halls were used, selected to cover a wide range of acoustic metric values, resulting in 45 unique pairs. A total of 10 subjects completed the rating task for each of the separate parameters. The auralizations were generated by convolving anechoic music with each BRIR and rendered over a pair of near-field loudspeakers, with mechanical barriers and electronic filtering to reduce crosstalk between the left and right loudspeakers. For this study, each subjective impression was evaluated separately, focusing on developing and evaluating various room acoustic metrics, including G, C80, level adjusted C80, EDT, early bass level, and treble ratio. A final correlation analysis compared each subjective impression to preference ratings, and the highest correlation values were found with loudness ($r = 0.45$), clarity ($r = 0.73$), and treble ($r = 0.81$). Reverberance and bass both had correlations significantly different than zero, but small in magnitude ($r = -0.20$ and $r = 0.11$ respecitvely). The importance of treble is quite interesting, as no clear agreed-upon metric exists for this impression.

Lorenz-Kierakiewitz and Vercammen conducted an acoustical survey of 25 European concert halls, and recently published a short paper regarding the work.[35] Measurements were conducted over a time period from 2000 to 2007. As with many other analyses, BRIRs measured using an omnidirectional sound source on stage were convolved with anechoic music for auralization generation. Subjects were sent a CD with the selected excerpts for comparison, and parings for a total of five selected halls were included. In a printed questionnaire, subjects

were asked to compare pairs hall auralizations over headphones and select the hall they preferred. The halls included in the study were the Concertgebouw, the Musikverien, Tonhalle Zürich, Tonhalle Düsseldorf, and Royal Festival Hall. Preference orders for the five halls were generated for each subject, and this ranking information was used to cluster subjects into one of three clear preference groups. A total of four groups were found, but the fourth group simply consisted of subjects that were difficult to assign to one of the three other identifiable preference groups. This grouping typology existed for two different subjective studies conducted. Clear questions can be raised regarding the binaural reproductions, as the survey was distributed to listeners, and therefore, does not contain a consistent headphone-based auralization setup. Most notably, this would impact the perception of tonal aspects of each hall and possibly spatial perception, based upon the quality of the unequalized response of each listener's headphones.

The most prominent recent work in measurement-based auralization and perception of concert hall acoustics has come out of Tapio Lokki's research team from Aalto University. From the Ph.D. of Pätynen (2011), a *loudspeaker orchestra* was generated for use in subjective concert hall studies.[36] The orchestra consisted of 33 commercial loudspeakers (Genelec types 1029A, 1032A, and 8030) which were used to represent an orchestral layout of 24 instrument positions on stage, pictured in Figure 2-3. Directivity patterns of different orchestral instruments were made, and either a single loudspeaker or a pair of loudspeakers were arranged and oriented to best-match the original radiation pattern.[37] This loudspeaker orchestra was transported to a number of concert halls around Europe. Originally, measurements were made in eight Finnish concert halls, and various subjective studies were based upon these results.[38-40] From these studies, more well-known European halls have been measured with the same loudspeaker orchestra.[41] In each hall, a GRAS vector intensity probe (type 50 VI-1) having a 100 mm and a 25 mm capsule spacer was used to capture a spatial microphone RIR (MicRIR).

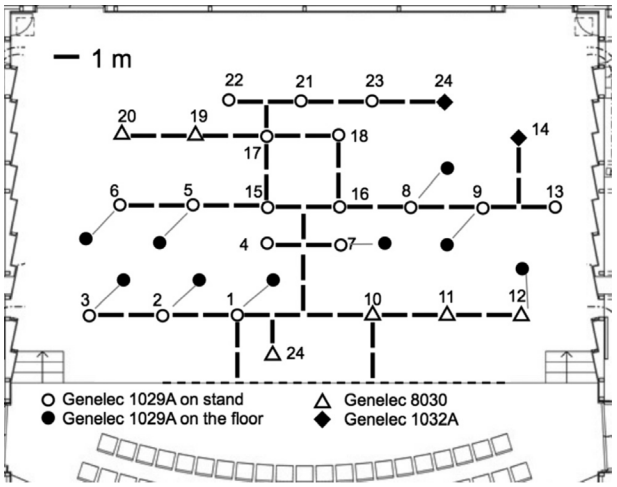


Figure 2-3: The loudspeaker orchestra used by the researchers from Aalto University, made up of commercial Genelec loudspeakers (from Fig. 1, Ref. [38]).

Once the MicRIR was obtained, it was processed for virtual acoustic reproduction. A separate MicRIR was measured for each loudspeaker in the orchestra, and each source position could be combined with anechoic music that was individually recorded for each instrument. All the measurements were then superimposed to generate a full-orchestral auralization. For auralization, the Finnish group utilized an internally generated technique called the spatial decomposition method (SDM).[42] More information regarding SDM can be found in section 2.4.2.2. An auralization array of either 14, 16, or 24 loudspeakers was used, depending upon the study. The first subjective contribution of the group was in developing a list of attributes gathered using individually-elicited descriptors.[39] Subjects performed a rating task, comparing different concert hall auralizations, but instead of using clearly defined words from previous literature, subjects provided their own vocabulary, and subsequently rated each hall on these terms for three different motifs or anechoic passages. Twenty assessors provided a total of 102 specific words, and these words were then grouped into nine categories using a statistical clustering analysis. The clustering analysis resulted in eight interpreted groups and one group lacking interpretation. The eight overall attribute group labels were reverberance in terms of size of a space, reverberance in terms of envelopment, width of sound, loudness, distance, balance, openness, and definition.

A further refinement of this study included 17 assessors with only 60 attributes, now checked in terms of the reliability of each attribute for each subject. In this study, a principal components analysis (PCA) was first run on the 60 attributes, and the first three principal components of the analysis were retained, explaining 67 percent of the variation in the attribute ratings. Then, the principal component loading values for each attribute were

calculated, and agglomerative hierarchical clustering was used to cluster each attribute based in the 3-dimensinoal principal component space. The terms developed into seven different groups from the clustering analysis, defined as definition, clarity, reverberance, envelopment and loudness, bassiness, proximity, and an undefined group. Finally, the assessor's preferences were mapped onto the perceptual factor space, and two main groups of preference were identified. The first group preferred halls with loud, enveloping, and reverberant sound fields, and the second group preferred halls with intimate, close sound fields with high definition (clarity). On average across both groups, preference best correlated with proximity, a subjective term with no current objective metric, which is shown visually in the perceptual factor space in Figure 2-4. This finding indicates the need to develop and define new metrics that match to concert hall preference, as the impression most highly correlated with preference cannot be predicted.

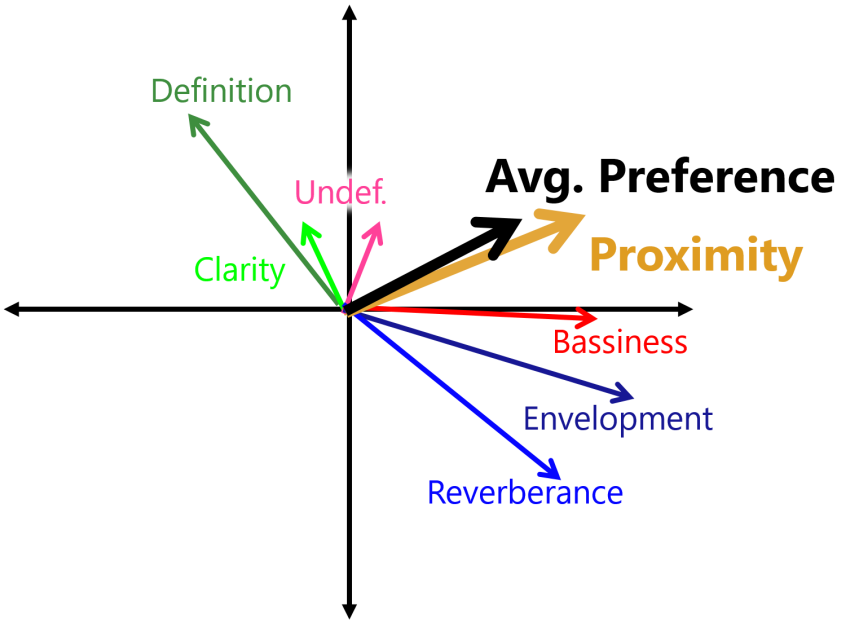


Figure 2-4: Each group of attributes along with average listener preference plotted against the first (x-axis) and second (y-axis) perceptual factors identified by Lokki et al. Based on Fig. 10a from Ref. [38].

Kuusinen et al. continued this work, attempting to generate colored maps of preference in the principal components space for each preference group and for overall preference.[40] A linear or point-based model was fit between the preference ratings for each individual and the factor scores for each hall from the descriptive sensory analysis. Then, each map overlaid the model individually fit to each subject in the group, demonstrating regions of higher and lower preference. Along with the generation of preference maps, correlations between objective room metrics and different subjective attributes were presented. A significant correlation was found between high frequency J_{LF} and preference of all subjects across all motifs. For the Beethoven

motif, correlations existed with low- and mid-frequency J_{LF} and for high-frequency L_J. When analyzed in terms of preference groups, the clarity group had a strong negative correlation between preference and mid-frequency EDT. The loud, enveloping, and reverberant preference group exhibited strong correlations with G_{mid}, J_{LF} at all frequencies, and L_J at all frequencies.

A more recent study by Lokki et al. analyzed preference using a paired-comparison technique between six more well-known European concert halls and studied the impact of seating position, musical motif, and hall on a listener's preference.[41] A total of 28 assessors completed the preference task, and again, two groups of preference were found; the first preferred a more clear condition, at the expense of loudness, and the second preferred a more reverberant condition, primarily liking louder shoebox-style halls. In this study, the choice of preference was shown to switch between halls after changing either seating position or musical style, for both preference groups. This variability indicated a difficulty in identifying one, singular hall of highest preference. The same two preference groups were identified in an earlier study, shown on a perceptual factor space with the other subjective attributes of importance for interpretation.[38]

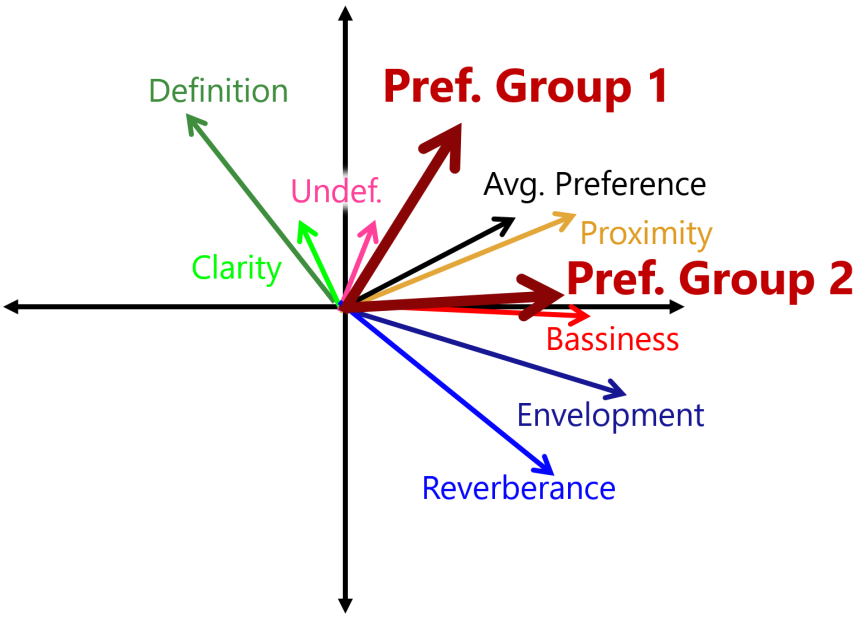


Figure 2-5: The two distinct perceptual preference groups plotted in the perceptual factor space, along with the other perceptual attributes identified by Lokki et al. Based on Fig. 10a from Ref. [38].

Despite the large amount of work done by Lokki's group, a few important limitations must be mentioned; these limitations are not adequately highlighted in the literature. Regarding the elicitation of individual attributes, self-defined by each assessor, this methodology is often employed in studying perceptions that have not been well documented

and are without a consistent vocabulary of descriptors. For the field of concert hall acoustics, a large body of research exists, and although individual musicians may use or prefer different terminology or different perceptions, a fairly comprehensive set of descriptors could be generated from the previous literature in sections 2.2.2 to 2.2.5. An inherent tradeoff exists between using consistent terminology to ensure that research comparable to one another, which may limit or bias the results towards the standard terminology. Along with the use of user-defined words to subjective ratings, the auralization methods used in this study suffer from limitations in spatial accuracy and sound field analysis, which will be further discussed in section 2.4.2.2. Finally, the source directivity treatment of the loudspeaker orchestra does not necessarily provide an accurate representation of the complex, frequency-dependent radiation of orchestral instruments. This point will be further elaborated on in chapter 4 of this dissertation.

2.2.6 Simulation-based Auralization Approaches

The most recent large-scale effort in concert hall acoustics has been the development of the room acoustical quality inventory (RAQI) by researchers at the Technical University of Berlin.[43] Weinzierl et al. conducted a subjective listening study based on room acoustic simulations of a wide variety of concert halls. A set of 35 different computer models were used to simulate BRIRs in the room acoustics software RAVEN. For both the solo trumpet and speech motifs, BRIRs in 35 rooms at 2 listening locations were generated, and for the orchestral motif, only 25 rooms were used, as stage areas were not large enough to accommodate a full orchestra. These combinations generated a total of 190 possible room acoustic conditions. A total of 190 subjects received a randomly selected sample of 14 acoustic conditions and rated each condition on 46 subjective items. Auralizations were presented over extra-aural headphones with head-tracking to allow for real-time head rotation. The 46 subject terms were selected through discussions and guided interviews with an expert focus group, consisting of room acoustical consultants and academics studying room acoustics and psychoacoustics. The focus group came up with 50 subjective terms, and four terms were removed because they were deemed as unsuitable for rating some of the stimuli.

Using exploratory factor analysis, considering the re-test reliability of different subjective terms, and considering measurement invariance, a set of either four, six, or nine descriptors were suggested as a representative, simplified subset of the 46 attributes. The smallest set, containing four factors, included quality, strength, reverberance, and brilliance. Other words were contained within these sets, including envelopment by reverberation grouped with reverberance. A slightly more specific six factor list of terms adds irregular decay

and coloration to the list. Finally, the nine-factor solution adds clarity, liveliness, and intimacy to the list. Some clear similarity with this list is consistent with previous works, such as the clear importance of strength and reverberation in the perception of room acoustical quality. It is important to note the highly placed importance of brilliance in this study, for which no widely accepted metric currently exists. Further, it should be noted that this study utilized only room acoustical simulations, allowing for a large variety in acoustical conditions. This level of variety is difficult to obtain using room acoustical measurements, but it has recently been shown that simulation-based auralizations do not fully capture the realistic experience of concert hall environments.[44] Despite this limitation in realism, the analysis of the differences between simulated rooms are still valid.

2.3 Concert Hall Impulse Response Measurements

Although room acoustical simulations can provide large variety in room environment very quickly, room acoustical simulations have been shown to lack realism compared to measurement-based auralizations. These limitations have recently been demonstrated in a round robin study on room acoustical simulation and auralization.[44] A work done by multiple research teams identified that, although plausible, simulation-based auralizations contain clear limitations from reality still exist. To overcome limitations regarding accurate material characterizations for absorption and scatting, diffusion, edge diffraction, and other software and material characterization limitations, auralizations can be made directly from room impulse response (RIR) measurements. This section is formatted to provide background on standard practices in room acoustics for RIR measurements and analysis, along with highlighting limitations of the current standard practices.

If a room is assumed to be a linear, time-invariant (LTI) system, the acoustic properties of the room can be fully classified by the RIR. The RIR consists of the room's response to an impulsive input. A graph of a standard RIR is typically divided into three regions, shown in Figure 2-6: direct sound, early reflections, and late reverberation. The direct sound (shown in red) is the energy arriving directly from the sound source, which has not interacted with any room surfaces. The early reflections (shown in green) are discrete reflections that have only reflected off of one to three room surfaces. If these surfaces are acoustically hard, the reflections will show up with a very strong, pronounced amplitude. Once the reflections have interacted with many surfaces in the room, they combine into the overall late reverberation (shown in blue). The rate of this decay is dependent upon the average absorption and scattering properties of the room surfaces and size of the room.

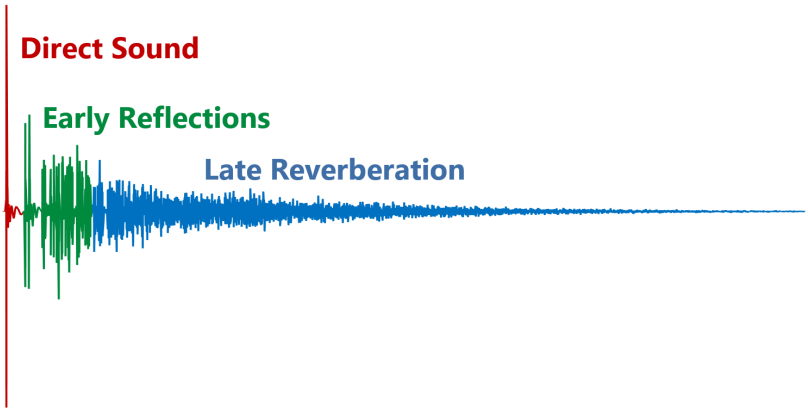


Figure 2-6: A time-domain graphic of the pressure amplitude of a RIR. Typically, the RIR is divided into the initial direct sound (red), the discrete early reflections from one to three hard surfaces (green), and the late reverberant energy (blue) which consists of many overlapping reflections.

This response is not only dependent upon the characteristics of the room, including geometry, size, and material properties, but it also depends upon the characteristics of the measurement source and microphones. Originally, these measurements were made by directly exciting a room with an impulsive source, such as the pop of a large balloon or a starter pistol firing blanks. With modern equipment, large improvements in signal-to-noise ratio can be obtained by recording measurement signals, such as swept-sines or maximum-length sequences. A RIR can be calculated using a frequency domain division, otherwise known as *deconvolution*, between the original measurement signal and the recorded measurement signal in the room.[45] This procedure results in a room's transfer function that can be related to the impulse response with an inverse Fourier transform. These techniques also allow for time-domain synchronous averaging, providing a larger SNR without requiring a louder sound source. Additionally, logarithmic (or pink-weighted) swept sine signals have a two-fold advantage of matching the common background noise spectrum in rooms and compacting any harmonic distortion effects from the loudspeaker into time-domain artifacts.[46] These artifacts can then be windowed out of the final measured RIR as they appear prior to the direct sound. The next few sections will describe measurement microphones, loudspeakers, and common calculations performed on RIRs.

2.3.1 Measurement Microphones

The simplest microphone setup for RIR measurements involves a single diffuse-field omnidirectional microphone. A single time-domain RIR is measured with this microphone, which has been calibrated to have a flat frequency response in a sound field containing an equal amount of energy arriving from all directions. The non-flat effects of the microphone's performance in a non-diffuse field can be mitigated by using a smaller diameter microphone, such as a ¼-inch vs. a ½-inch diameter microphone. The measurement standard for

performance spaces, ISO 3382, specifies that a microphone with a maximum diameter of 13 mm (1/2") be used.[47] The microphone should be equalized for random incidence, but measurements can be made with a pressure or free-field microphone if a suitable random incidence correction filter is supplied. The standard does not mention the use of microphone arrays to calculate or report any spatial information regarding the sound field of a room. Appendix A of ISO 3382 mentions the use of a figure-of-eight microphone to extract the lateral component of the RIR, but this approach provides limited spatial information.

The first improvement in capturing the spatial character of a room's sound field began with the invention of the binaural head, or "dummy head".[28] A binaural head is a physical model of the human head, ideally with an upper torso, and two ears. A microphone is located at the entrance of each ear canal, and realistic pinnae in shape and impedance are placed around the microphones. By physically building a model of the human head, binaural localization cues are naturally included in the measured BRIR.

While a binaural mannequin builds-in a specific set of spatial cues, tied to the geometry of the head's design, microphone arrays can capture spatial variation in a sound field without *decoding* the spatial information into a particular format. With the advances in computing power and capabilities, it is quite accessible to work with high-channel-count arrays, and commercially available first-order B-format microphones are common and fairly inexpensive. Beyond that, even higher-order microphone arrays are commercially available. First-order microphone arrays are capable, to a very limited extent, of reproducing a sound field and performing beamforming-style analyses. With these arrays, it is much more common to perform time delay-of-arrival (TDOA) techniques to estimate the arrival direction of specific reflections in the RIR. Analyses using TDOA techniques are nearly identical to sound intensity-based direction calculations. For early reflections that are separable in time, TDOA algorithms are accurate in identifying discrete reflections, but if reflections arrive at almost the same time in a RIR, a TDOA calculation will estimate the arrival direction as an average of the two events. This problem becomes even more significant as reflection density increases in later portions of the RIR, with many reflections arriving to a listener at the same time. The IRIS software and SDM toolbox are both algorithms that fundamentally operate off of TDOA techniques.[42,48] Higher-order microphone arrays are also available, but far less common for use in room acoustics. Although the level of complexity in data handling and processing is increased, the spatial resolution combined with the ability to spatially separate overlapping reflections in time makes these processing techniques more robust. More comments regarding SDM are provided in section 2.4.2.2.

2.3.2 Measurement Loudspeakers

For a RIR measurement, the ISO 3382 standard calls for the use of an omnidirectional sound source, but it allows for deviations from omnidirectional radiation of up to ±6 dB for the 4 kHz octave band, as illustrated in Figure 2-7. This allowance in the standard is due to practical limitations of single dodecahedron sound sources. For example, the Brüel and Kjær OmniPower sound source type 4292-L, a common industry standard for such measurements, produces directional lobes due to a spatial aliasing as low as 1000 Hz; this is shown in a plot from the company's datasheet and repeated in Figure 2-8. These directional patterns do not cause significant differences for metrics such as T30, focused on the later decay of the RIR and not significantly influenced by early reflections. For metrics that relate to the early part of the RIR, including C80, J_{LF}, and EDT, source directivity can have a large impact on metric calculations.[49]

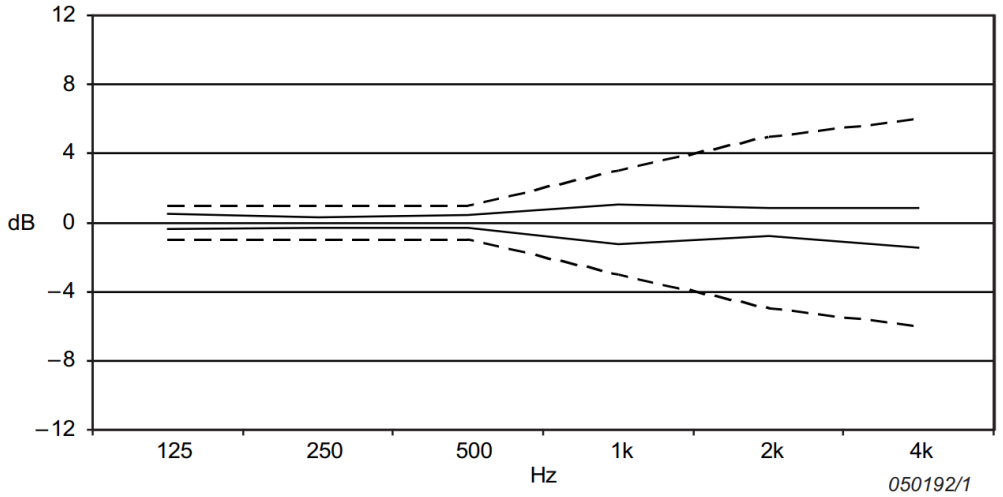


Figure 2-7: Allowed deviation for omnidirectional sound sources as suggested by ISO 3382[47], here showing the satisfaction of that criterion for the Brüel and Kjær OmniPower loudspeaker.[50]

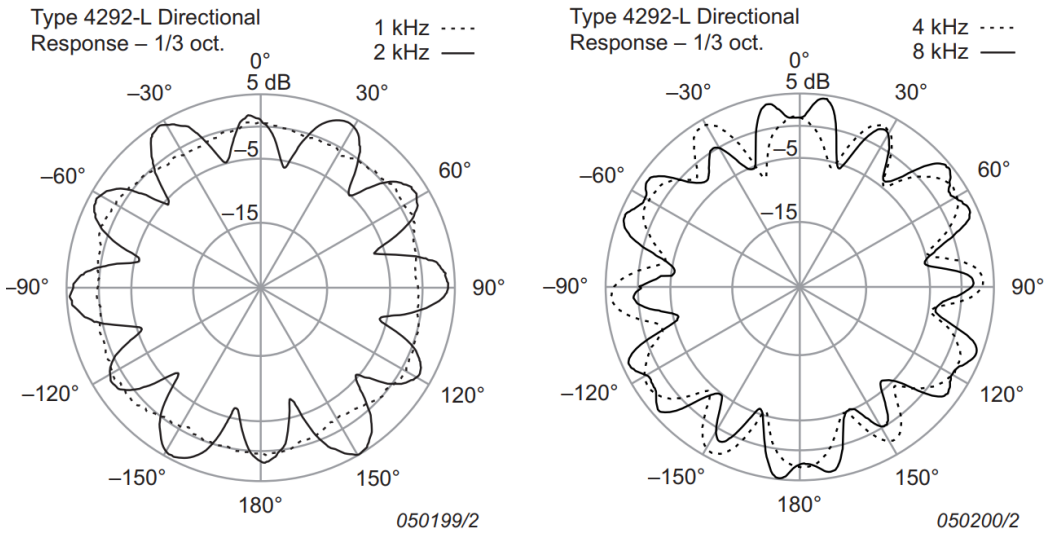


Figure 2-8: Directional radiation pattern of the Brüel and Kjær OmniPower loudspeaker, showing clear deviations from omnidirectional performance above 1 kHz. From product datasheet.[50]

Researchers at Rheinisch-Westfälishe Technische Hochschule (RWTH) Aachen University proposed a solution to this problem by designing a three-part omnidirectional sound source with each part optimized for a low-, mid-, and high-frequency range.[51] This approach was also adopted for the current setup in Penn State's Sound Perception and Room Acoustics Laboratory (SPRAL). More information on this setup can be found in section 5.9.1 or in the dissertation by Dick.[52] By combining the RIR measurements from these three sources, a wide bandwidth, truly omnidirectional RIR can be measured. Both three-part sources mentioned here exhibit a combined omnidirectional radiation pattern up to around 5 kHz. Although this provides a repeatable, simple source directivity that could be reproduced between measurement teams, this source is not realistic of any natural sound sources. This fact creates a mismatch between RIR measurements and source properties required for generating a realistic auralization of an orchestra. More information regarding orchestral source representation will be provided in section 4.6.

2.3.3 Room Acoustic Metrics

From the RIR, room acoustic metrics have been defined, some of which are properties of the decay of a room's impulse response, some energy ratios for different time and spatial ratios in the RIR, and some that include strength calibration terms. The first and most widely used metrics relate to the decay of the RIR. The metric RT is most commonly calculated using the linear fit of a backwards integrated RIR from -5 to -35 dB.[47] These levels are calculated relative to the total energy in the RIR. The estimated slope is then extrapolated to a full 60 dB of decay, and the corresponding decay time is reported as the room's T30. RT can also be calculated

using T10 and T20, calculated from -5 dB of the decay to either -15 or -25 dB, respectively. These metrics are commonly used, as they do not require the full 60 dB of decay defined by RT. The other most well-known decay-based metric is Early Decay Time (EDT), calculated from 0 to -10 dB in the backwards integrated RIR. This metric is more dependent upon the early decay of the room and has been shown to be more highly correlated to the subjective impression of reverberance as compared to T30.

The other common type of metrics are energy ratios of the RIR. The most common of these is the clarity index for music (C80) which is a decibel energy ratio of the early energy, before 80 ms from the direct sound, to the late energy after 80 ms:

$$C80 = 10\log_{10}\left[\frac{\int_0^{0.08} p^2(t)dt}{\int_{0.08}^{\infty} p^2(t)dt}\right], \tag{2.2}$$

where $p(t)$ is the omnidirectional impulse response (IR) of the room. Other integration ranges can also be used, and often the integration limit is changed to 50 ms for speech signals, called the clarity index for speech (C50).

$$C50 = 10\log_{10}\left[\frac{\int_0^{0.05} p^2(t)dt}{\int_{0.05}^{\infty} p^2(t)dt}\right], \tag{2.3}$$

Another iteration of the clarity index for speech is called definition (D50), which can be directly related to $C50$ without the use of a decibel scale:

$$D50 = \frac{\int_0^{0.05} p^2(t)dt}{\int_0^{\infty} p^2(t)dt}. \tag{2.4}$$

All of these metrics are designed to predict the subjective perception of clarity for speech or music, or both.

A final clarity metric, one which is not a ratio of energy, is called center time (Ts). This metric is a calculation of the center of gravity of the omnidirectional impulse response. The main difference with this is the lack of dividing the IR up into early and late components, as is required with the previous three metrics:

$$Ts = \frac{\int_0^{\infty} tp^2(t)dt}{\int_0^{\infty} p^2(t)dt}. \tag{2.5}$$

All of the previously mentioned room acoustic metrics do not require any level-based calibration, which allows for more straightforward calculation and comparison of results

between groups. The downside is that these metrics are insensitive to hall strength differences, which often do impact the fundamental perception of rooms. As the strength of the RIR is dependent upon the source sensitivity, amplifier gains, microphone sensitivity, microphone preamplifier gains, and digital gains, the absolute level of the measured RIR is somewhat arbitrary in units. To provide a comparable metric across all measurement teams, the metric of strength (G) was developed:

$$G = 10\log_{10}\left[\frac{\int_0^\infty p^2(t)dt}{\int_0^\infty p_{a10}^2(t)dt}\right], \tag{2.6}$$

where $p_{a10}(t)$ is the IR of the loudspeaker measured at 10 meters in a free field, or anechoic chamber. The free field measurement is to be taken using an identical setup, with identical gain settings in the entire measurement signal chain, to the measurements taken while in each concert hall.

The last group of metrics are used to predict the components of spaciousness or spatial perception in concert halls. The first metric, lateral energy fraction (J_{LF}) predicts the sense of the width of the source on stage in a concert hall. Lateral energy fraction is calculated as the ratio of early lateral energy in a RIR to the total early energy from all directions:

$$J_{LF} = \frac{\int_{0.005}^{0.08} p_L^2(t)dt}{\int_0^{0.08} p^2(t)dt}, \tag{2.7}$$

where $p_L(t)$ is the RIR as measured with a figure-of-eight microphone with the null oriented towards the measurement source. The selection of a 5 ms starting integration time for the lateral measurement is to ensure the rejection of the direct sound energy. This metric has also been proposed with a slight correction, so that the numerator has a directional variation with the cosine of the lateral angle:

$$J_{LF} = \frac{\int_{0.005}^{0.08} |p_L(t)p(t)|dt}{\int_0^{0.08} p^2(t)dt}. \tag{2.8}$$

The second component of spaciousness has been suggested to relate to listener envelopment, or the sense of being full immersed or surrounded by a sound field. The perception of envelopment has been proposed to relate to overall loudness, along with the lateral component of the late sound field. A metric for this was proposed by Bradley,[32] called the late lateral energy level:

$$L_J = 10 \log_{10} \left[\frac{\int_{0.08}^{\infty} p_L^2(t) dt}{\int_0^{\infty} p_{a10}^2(t) dt} \right]. \quad 2.9$$

This metric is the energy in the late, lateral portion of the room's response, as measured with a figure-of-eight microphone and is strength-calibrated by normalizing the response to a free-field measurement, $p_{a10}(t)$. All metrics regarding spaciousness currently use a dipole directivity pattern microphone to extract spatial information and do not consider any more complex spatial analysis. Recent work by Dick and Vigeant has used spherical array beamforming analysis to generate a new metric to predict listener envelopment.[52] This new metric integrates the energy from 60 ms to 500 ms in a RIR over the spatial range of ±20° to ±120° azimuth and 30° to 130° elevation, measured from the vertical z-axis. The energy integration is performed from the plane wave decomposition (PWD) analysis of the sound field.

Some other metrics have been proposed in the past, but most only consider the spatial properties in a limited extent. The techniques proposed by Dick and Vigeant offer a new perspective, analyzing the full three-dimensional nature of a concert hall's sound field. In addition, this analysis technique is not limited to only early parts of the RIR, when individual reflections are separable in time. Beamforming analysis can open the door to many new metric possibilities. Background on spherical array processing and beamforming analysis can be found in chapter 3. Some metrics based on the human hearing system have also been proposed, which are measured with a binaural head, like the interaural cross-correlation coefficient (IACC) or Beranek's binaural quality index (BQI).[16] Some groups utilize the metric often, but they have also found limited adoption by the wider community.

2.4 Concert Hall Auralization

Once a sound field is captured, it is often desired to reconstruct or recreate the character of the measured or simulated space. Many approaches exist for how to achieve an accurate representation of the space, and most have a complex set of benefits and shortcomings. These methods will be individually addressed in sections 2.4.1 and 2.4.2. In general, the author finds it beneficial to divide auralization techniques into one of two categories: "physically-motivated" or "perceptually-motivated". The first aims at a **physical reconstruction** of a sound field, where the second aims at generating a reconstruction that elicits the same **perceptual experience,** without perfect reconstruction of the originally measured or simulated sound field.

2.4.1 Physically-motivated Auralization Techniques

The first set of techniques is comprised of "physically-motivated" methods and aims to reconstruct the physical sound field. Often, this goal is accomplished by using a distributed array of loudspeakers that sum together to reconstruct a sound field using well-designed sampling microphone and loudspeaker arrays. Both the techniques of higher-order Ambisonics (HOA)[53] and wave-field synthesis (WFS) fall into this category.[54-55] Often, such methods are coined with the term acoustic holography.[56] These techniques can help to reconstruct a wave front, traveling in a specific direction, over an area or *sweet spot* of accuracy. Such methods can extend to highly complex sound fields, including a superposition of many plane waves and acoustic events. Binaural techniques also fall into this category, but with a different approach, attempting to physically reconstruct a sound field at the left and right ears of a listener's head.[27] Spatial qualities are assumed to be built-into the simulation or measured sound field by way of the head-related transfer function (HRTF) of the dummy head.

2.4.1.1 Binaural Techniques

The human hearing system, despite only having two sensing elements with the left and right ears, can perform amazing tasks: spatially separating listeners, focusing in on specific sound sources while tuning out others, and simply locating the position of a source in space, even when a listener closes their eyes. This processing is performed as the brain interprets interaural time differences (ITDs), interaural level differences (ILDs), and spectral notches that occur from the size and shape of the head, upper torso, and pinnae, relative to the sound source's location.[57] These measurements or recordings can be directly reproduced using headphone-based auralization or using loudspeaker-based binaural techniques, such as cross-talk cancellation.[27] If a measurement is made with a binaural head, it will naturally contain all of the built-in cues for the dummy head, and when rendered these cues will be directly input to the left and right ear, in a wide-bandwidth auralization.

In simulation, these cues must be applied by way of a head-related transfer function (HRTF). A HRTF is a measurement database of left and right ear IRs, or frequency-domain transfer functions, as a function of source position around the head.[58] These effects are properly applied in a simulation algorithm, depending upon the distance and direction-of-arrival of a particular source or reflection in a room environment. Once a simulated BRIR is generated, it can be convolved with anechoic music to generate a realistic auralization for a specific room. This process can then be repeated for different rooms, and auralizations can be used to compare changes between two rooms. This process can be used with a measured BRIR with proper diffuse-field equalization for the measurement loudspeaker and microphones. The simplest

approach to reconstruct binaurally involved headphone-based auralization. With proper headphone equalization, highly realistic auralizations can be generated with high quality pair of headphones. If it is desirable to allow for head rotation or movement in an auralization, highly important for a realistic sound field experience, a real-time synthesis algorithm or environment is needed. This is more difficult when working from measured data without taking rotated measurements of the BRIR at different head orientations.

Another binaural rendering technique is known as cross-talk cancellation (CTC).[27] CTC operates using the same principle as headphone-based binaural auralization, but this method must also account for the fact that a loudspeaker will send signals to both the left and right ears, creating *cross-talk* between channels. This effect does not occur when using headphones, as the left and right signals are inherently separated. With properly measured IRs from a left and right loudspeaker to both the left and right ears, filters can be designed so that the crosstalk from the left loudspeaker to the right ear, and vice versa, is effectively cancelled. This results in the left loudspeaker controlling only the signal at the left ear, and the right loudspeaker only controlling the signal at the right ear. Thus, an accurate pressure can be generated for both ears, and a realistic auralization is achieved. The use of loudspeakers allows for a better bandwidth of reproduction, as better performance can be found at low frequencies, and there is no physical enclosure placed over or around the ears. This approach may provide a listener with a more natural auralization presentation for even non-acoustic reasons like the perceptual impact of knowing a pair of headphones is being worn.

An important limitation of all binaural techniques lies in the assumption of a particular or average HRTF for a specific individual. Often, due to lack of a standardized individualized HRTF-matching technique, an average HRTF is used. This technique provides a highly convincing auralization, as much similarity between the geometry of a human head exists across many individuals. Despite this similarity, the human hearing system has been found to be sensitive to the inter-individual differences in ear shapes, geometries, and head sizes. Although these differences may appear small, the use of non-individualized HRTFs has been shown to cause significant errors in human localization performance, the perception of a sound image located within a listeners head (internal localization), and a lack of realism when compared to individually measured HRTFs.[58] The perceptual need for individualized HRTFs and the matching of customized HRTFs for each individual is an ongoing research topic with large push currently from the virtual and augmented reality communities.

2.4.1.2 Ambisonics / Higher-order Ambisonics

Ambisonics, a common loudspeaker-based auralization technique, was first invented in the 1970's by Gerzon.[53] Gerzon showed that directional sound could be represented in the spherical harmonic (SH) domain using the fundamentals of spherical array processing. At the time, Gerzon, only used first-order Ambisonics, containing a monopole component and three x – y – z oriented dipole components. This work led to the virtual acoustic processing and development of the SoundField microphone. As technology and time have advanced, Ambisonics was extended to the use of higher-order SH terms, and the name higher-order Ambisonics (HOA) was developed for anything using higher than first-order SH components. Open spherical arrays with cardioid microphones, along with rigid spherical microphone arrays are now commercially available, and HOA capture is more accessible than ever.

To render a HOA auralization, a distributed array of loudspeakers is required, with minimum number of loudspeakers, L, such that $L \geq (N+1)^2$. Here, N is the SH truncation order or resolution of the HOA auralization. This same minimum number of microphone criterion is applied to the capture side as well. These limits are due to the matrix inversion required to translate from a microphone capture array to the SH domain, and then back again to a loudspeaker reproduction array.[59] The process of going from a microphone array signal to a SH signal is known as *encoding*. Once in the SH domain, a *decoding* stage is required to sample the SH signals at the location of each loudspeaker in the auralization array. Both stages of processing can be implemented as matrix multiplications through the calculation of a Moore-Penrose pseudoinverse. In general, as truncation order, N, increases, so does the accuracy of the rendering. More information on the mathematical basis of HOA and spherical array processing in general can be found in chapter 3.

This field has evolved quite drastically in the past several decades, addressing issues such as the near-field compensation from loudspeakers in the auralization array, optimization for different methods of decoder matrix calculation (based upon psychoacoustic hearing differences), and the calculation of decoder matrices for non-regular loudspeaker arrays.[60-61] The benefits of HOA lie in its flexibility, but the main downside is the limited reproduction area of accuracy. HOA aims to recreate a sound field accurately in a *sweet spot*, located as the center of the array. The size of this sweet spot increases as truncation order, N, increases, but the sweet spot decreases in size as frequency increases. The processing can be quite effective over a broad frequency range, but this accuracy is only maintained at the one position where the processing is centered. Eventually, the sweet spot decreases in size until it is smaller than the head. A general criterion for the radius of this sweet spot can be expressed as $r = N/k$,

where k is the wavenumber calculated from frequency and sound speed as $k = 2\pi f/c$. For third-order Ambisonics, the sweet spot is around the size of the human head (assuming an 8 cm radius) at around 2000 Hz and is degraded at 4000 Hz. This degradation is perceptually noticeable as a change in timbre.

Finally, it should be noted that once a sound field is represented in the SH domain, it can also be directly rendered to other formats of auralization, such as binaural auralization. Just as a surrounding loudspeaker array can represent a SH domain RIR or recording, a HRTF can also be represented in the SH domain.[62] Essentially, a decoder matrix would be developed for the left and right ears, building in the SH domain representation of the spatial cues that before were naturally applied with a dummy head. The benefit of this representation is that the sound field can be measured, independent of any specific individual head geometry, and flexibly rendered to any HRTF. As long as a measured HRTF for an individual is available, a personally tailored binaural rendering can be made. Since a distributed loudspeaker is not needed, the problems associated with a shrinking sweet spot do not apply in this rendering technique. This concept borders between HOA and binaural rendering and shows highly promising future implementation once concerns of HRTF individualization are addressed.

2.4.1.3 Wave-field Synthesis

The other primary loudspeaker-based auralization technique, based upon a fundamentally different principle compared to HOA, is known as wave-field synthesis (WFS). Based on the concept of Huygens principle, WFS aims to reconstruct a sound field located within an arbitrarily defined surface or boundary.[54-55] At a high level, if pressure measurements are made at a fairly even sampling distribution over the surface, a WFS loudspeaker array can be placed at the same sampling locations as the original measurement microphones. If these loudspeakers, acting as point sources, generate the measured acoustic pressure field, the entire field will be reconstructed within the entire area of the surface. This current discussion is regarding an internal reconstruction of a field containing sources external to the measurement boundary. A more rigorous description of the field, requiring the pressure and pressure gradient measurement (related to acoustic particle velocity), can be used to not only reconstruct a field within such a boundary but also external to the boundary.

The main limitation of this approach is in the phenomena of spatial aliasing. As an analogy, when sampling a pressure field for a single location in time, temporal frequencies can only be resolved if the frequency is below the Nyquist frequency, defined by half of the sampling frequency. In the same way, a spatial Nyquist wavelength applies to WFS arrays,

where the accuracy of the sound field reconstruction is dependent upon the spatial loudspeaker sampling rate, or the spacing of the nearest loudspeakers to one another. Because loudspeakers take up physical space, WFS systems require extremely high counts of loudspeakers to produce accurate reproductions over a reasonable bandwidth. For the number of channels on multichannel audio interfaces, typically around 24 – 32, HOA would create a fairly wide-bandwidth reproduction at a single sweet up to around 3 – 4 kHz using third-order SH processing. With WFS, a distribution of 24 sources would either only surround a very small boundary or volume (let alone the need to accommodate the size of the loudspeaker elements), or it would cover a reasonable area and exhibit spatial aliasing in the low- to mid-frequency range. Due this combined channel count and loudspeaker spacing limitation, WFS systems are typically implemented only in two-dimensional setups. Practical three-dimensional setups have very low-frequency cutoffs for accurate reproduction.

2.4.2 Perceptually-motivated Auralization Techniques

The other category of auralization, coined by the current author, is "perceptually-motivated" auralizations. These methods are not directly aimed towards physical reconstruction of a wave front, as was done for the methods described in the previous section. Rather, they are targeted at creating a perceptually identical experience, even if the physical measurement of the sound field is different. These methods do not have the same high frequency limitations. If the assumptions that simplify the problem for perceptual accuracy are not violated, the accuracy might occur over the entire audible range. In general, these techniques are event-based rather than spatial sampling-based. A measurement, like a RIR, is often separated into a series of acoustic events arriving from particular directions, such as room reflections. Although many different techniques exist, the two most common in the concert hall acoustics community include vector-base amplitude panning (VBAP) and the spatial decomposition method (SDM).

2.4.2.1 Vector-base Amplitude Panning (VBAP)

One of the first convincing, yet simple, virtual acoustic techniques came from the idea of amplitude panning. Amplitude panning is a technique where correlated signals are played from multiple loudspeakers, and based upon the relative gain difference between channels, a virtual image will appear to play from a location in between the loudspeakers. The summation of the loudspeaker signals mimics the interaural cues for sound localization that a listener would hear as if a source were placed at a location between the loudspeakers. If a physical measurement would be made, the sound field would consist of two correlated loudspeakers at different locations, not a reconstructed wave front originating from the desired location. By

understanding the primary localization cues for binaural hearing, this technique provides a simple way to mimic these cues with only a gain manipulation in a loudspeaker array.

Vector-base amplitude panning (VBAP) was developed by Ville Pulkki, and this technique identifies a triplet of loudspeakers that are closest to the virtual source location for a given loudspeaker rig.[63] The gains of the loudspeaker are determined using a projection of the vector pointing toward the location of the virtual source onto unit vectors pointing in the direction of the closest loudspeakers. These projections are used to calculate appropriate gains between loudspeakers to create the virtual image between the sources. Pulkki has also suggested a directional audio coding technique (DirAC) that codes a measured or simulated sound field into directional and diffuse components.[64-65] Directional components of a sound field are coded with metadata dictating the azimuth and elevation of a particular source. Additionally, a diffuseness estimate of a sound field is determined. VBAP can be used to render the directional aspects of a sound field for each individual source. The diffuse component of a sound field can be reproduced, in a lower quality fashion, by rendering the diffuse component as a de-correlated version of the input signal to all array channels. A higher quality version is said to be obtained when the original de-correlated measured microphone signals are rendered over the auralization array, requiring the retention of higher rates of data streaming for the larger amount of information.

2.4.2.2 Spatial Impulse Response Rendering and Spatial Decomposition Method (SDM)

For auralization, a Finnish group of researchers implemented a variation of VBAP and DirAC for using in the auralization of RIRs, called the spatial decomposition method (SDM).[42] SDM is a virtual acoustic rendering technique that divides the RIR into short time segments, and for each time segment a cross-correlation analysis is performed to determine the time delay associated with a reflection arriving at different microphone array capsules. Using these time delays, the slowness vector for sound energy can be computed in a least-squares sense, using matrix inversion techniques, and an estimate for a direction of arrival is obtained. This direction of arrival is obtained for each time segment in a RIR, and the energy is placed in the loudspeaker of the auralization array closest to the arrival direction. The method is an adaptation of a method, also developed by Pulkki, termed the spatial impulse response rendering method.[66-67] The main difference lies in the use of the closest-loudspeaker technique for early reflection placement, rather than amplitude panning. This was done to avoid potential spectral coloration that was claimed to impact the timbre of the auralization when VBAP was used.[68]

The main limitation with SDM is in the assumption that all acoustic events are separable in time. In other words, no two reflections can occur in the RIR at the same time. The creators of SDM claim that early reflections are perceptually the most prominent, and since reflection density is low in the early part of the RIR, this assumption is claimed to be valid. In any room, eventually, reflection density increases such that multiple reflections occur in the same time window, and this is not uncommon. The debate is whether the spatial information in the RIR at that time is perceptually important to reproduce. Early energy often dominates perception, but later-arriving energy can be important for certain perceptions, such as listener envelopment.[32] Once a particular time in the RIR is reached, when energy is no longer separated in time, SDM randomly assigns the energy in each time bin to a loudspeaker in the auralization array, creating highly diffuse reverberation. If spatial character in the mid- to late reverberation is perceptually important, SDM will not faithfully recreate this character. More specifically, if two early reflections arrive within the same short time window, SDM will not accurately portray the spatial placement of these reflections. This situation is also quite common when performing beamforming analysis of RIRs, even before 100 ms.[69] This question has not been fully addressed in the literature and is left open to further research.

2.4.3 Source Directivity Representation in Auralization

A final note should be made regarding the directivity of sound sources used in measurement-based auralizations. Common omnidirectional sound sources used for RIR measurements do not behave as realistic sources. An auralization made from a RIR measured with a dodecahedron loudspeaker, with proper equalization, would sound as if the orchestra was emitting from a single point on stage. Although this provides a repeatable source for RIR measurement and metric calculations, it is not realistic to any performance condition. Multiple studies have demonstrated the perceptual importance of including multiple sources in auralization and including instrument directivity.[70] This representation can be implemented in simulation software, but the inclusion of source directivity in measurements is a more complex question.

As previously mentioned, the Finnish research group led by Lokki developed an orchestral source representation of commercially available loudspeakers described in section 2.2.5. This representation accomplished two goals. First, the loudspeaker orchestra provides a distributed array of sources, rather than a single point source in the center of the stage. Secondly, the loudspeakers were selected to have a directional radiation pattern somewhat similar to that of orchestral instruments.[37] Many groups have measured the directional radiation pattern of different instruments, and it is known that instrument directivity varies

quite substantially for different orchestral instruments and across frequencies. Some of these measurements include works by Meyer,[71] TU Berlin and RWTH Aachen University,[72] Aalto University,[37] and Brigham Young University.[73]

Although attempts were made to match these commercial loudspeakers with the directivity of each instrument, it is clear upon inspection that the highly frequency-dependent nature of instrument directivities cannot be easily represented. The commercial loudspeaker orchestra consisted of only three types of loudspeakers, so only three radiation patterns were available for representing the frequency-dependent nature of each instrument. Additionally, no adjustment of these radiation patterns can be made. Only the natural directivity of subwoofer or two-way loudspeakers provide the basis of directional radiation accuracy. At the highest level, it can also be understood that the mechanism which gives a boxed loudspeaker a directional response is quite different from any woodwind, brass, or any string instrument. The use of multiple sources is a vast improvement on auralization realism, but this method highly simplified the frequency-dependent radiation patterns of orchestral instruments. On a more practical note, it is also highly difficult to pack, transport, and assemble a 34-loudspeaker measurement setup, requiring a large amount of supporting equipment, and it is costly and cumbersome to transport for using a consistent measurement setup in halls located in different geographic regions.

Another approach to accurately including instrument directivity in RIR measurements is implemented through spherical array processing techniques. Just as a spherical microphone array can capture the spatial components of a measured sound field in the capture side of the RIR, the reciprocal condition of a CSLA is capable of recreating an arbitrary, frequency-dependent directional radiation pattern. In spherical microphone array processing, a separate RIR is captured by each microphone capsule, and the multi-channel spatial RIR can be processed to spatially analyze a sound field. Spherical array beamforming techniques allow for arbitrary control of the directional pick-up pattern of the microphone array for measurement analysis and auralization.

On the source end, a compact spherical array of loudspeaker drivers can be generated such that each driver can be controlled individually. If a separate RIR is measured for each source driver, these separate measurements can be post-processed and combined to generate a single measurement with an arbitrary source directivity. With separate measurement for each driver, this directional processing is fully flexible, and can even be manipulated as a function of frequency. Multiple versions of such CSLAs have been developed in the past decade, attempting to reconstruct the directional radiation patterns of musical instruments.[74-78] More

details regarding spherical array processing techniques can be found in chapter 3, and specific examples of these types of arrays are provided in section 4.2.4.

This Page is Intentionally Left Blank

Chapter 3
Spherical Array Processing

This chapter provides the fundamental mathematics and techniques behind spherical array processing. The goal of this chapter is to provide a comprehensive presentation of the mathematical framework of spherical array processing in the context of room acoustics. In the current work, this background used for spherical microphone array beamforming techniques, CSLA instrument directivity reconstruction, and higher-order Ambisonic (HOA) auralization using a surrounding, distributed loudspeaker array. These techniques are all rooted in the literature of spherical array processing, and both have evolved separately in the fields of spherical microphone array beamforming and virtual acoustics. Through this chapter, the fundamental concepts behind the representation of sound fields in terms of spherical harmonics is provided, and then extended to the application of a spherical microphone array beamforming and virtual acoustic sound field reconstruction. Specific details of the implementation of these techniques for this work will be provided in chapters 4 and 5.

3.1 The Wave Equation in Spherical Coordinates

The most common form of the wave equation, not expressed in a specific coordinate system, can be expressed in terms of the second time derivative of pressure, the sound speed, c, and the Laplacian of pressure:[79]

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0\,. \tag{3.1}$$

From this general form, the Laplacian operator can be specified in a given spatial coordinate system. The most mathematically straightforward three-dimensional space for expressing this term is in Cartesian coordinates as:

$$\nabla^2 p = \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2}\,. \tag{3.2}$$

Taking Eqns. 3.1 and 3.2, the three-dimensional wave equation becomes,

$$\left(\frac{\partial^2 p}{\partial x^2}+\frac{\partial^2 p}{\partial y^2}+\frac{\partial^2 p}{\partial z^2}\right)-\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}=0\,. \tag{3.3}$$

In this expression, the spatial dependence of the sound field is expressed as a partial derivative in the x, y, and z dimensions. If it is assumed that no sound field variation occurs in the y and z dimensions, the equations simplifies to the one-dimensional wave equation:

$$\frac{\partial^2 p}{\partial x^2}-\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}=0\,. \tag{3.4}$$

The solution to this differential equation is well known, taking the form of complex exponential functions. This approach can be applied in different coordinate systems, and different basis functions are determined. For many physically-based problems, it is convenient to consider a complex sound field centered around a specific point, dependent upon the angle or orientation about that point. It can be especially useful when considering spherical boundary conditions or distributions of microphones or loudspeakers. In spherical coordinates, the Laplacian from Eqn. 3.1 is represented as:

$$\nabla p^2=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial p}{\partial r}\right)+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial p}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2 p}{\partial\phi^2}. \tag{3.5}$$

The spherical coordination system here is shown in Figure 3-1. Elevation, θ, is the vertical angle from the upwards z axis and the azimuthal angle, ϕ, is the horizontal angle from frontal x axis, with counterclockwise defined as positive.

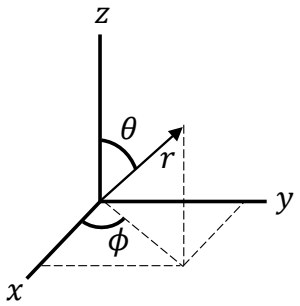


Figure 3-1: Spherical coordinate system, common in spherical array beamforming literature.

Now, using Eqns. 3.1 and 3.5, the wave equation in spherical coordinates can be expressed in full as:

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial p}{\partial r}\right)+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial p}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2 p}{\partial\phi^2}-\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}=0\,. \tag{3.6}$$

In order to find solutions to the differential equation given in Eqn. 3.6, the separation of variables technique is used. To start, it is assumed the solutions to the differential equation can be expressed separately as a product of four solutions, each separating out the dependence of each four dimensions of the problem: θ, ϕ, r, and t:

$$p(r,\phi,\theta,t) = R(r)\Psi(\phi)\Theta(\theta)T(t)\,, \tag{3.7}$$

where r is the radius from the center of the coordinate system. In this technique, each variable from Eqn. 3.6 will be independently isolated on one side the equation, separated from the other variables. Then, both sides will be set equal to a separation constant, and this constant will be used to help solve for the specific solutions for each variable. To start, the assumed form of the solution from Eqn. 3.7 can be solved by substituting it into Eqn. 3.6, and simplifying by dividing by the right side of Eqn. 3.7:

$$\frac{1}{r^2R}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right)+\frac{1}{r^2\Theta\sin\theta}\frac{\partial}{\partial\theta}\left(\cos\theta\frac{\partial\Theta}{\partial\theta}\right)+\frac{1}{r^2\Psi\sin^2\theta}\frac{\partial^2\Psi}{\partial\phi^2}-\frac{1}{c^2T}\frac{\partial^2T}{\partial t^2}=0\,. \tag{3.8}$$

3.1.1 Time and Azimuthal Solutions: Complex Exponential Functions

The time dependence can be isolated on the right side of the equation, and both sides of the equation can be set equal to a separation constant chosen to be equal to $-k^2$:

$$\frac{1}{r^2R}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right)+\frac{1}{r^2\Theta\sin\theta}\frac{\partial}{\partial\theta}\left(\cos\theta\frac{\partial\Theta}{\partial\theta}\right)+\frac{1}{r^2\Psi\sin^2\theta}\frac{\partial^2\Psi}{\partial\phi^2}=\frac{1}{c^2T}\frac{\partial^2T}{\partial t^2}=-k^2\,. \tag{3.9}$$

This separation constant was chosen in terms of the wavenumber, k, the ratio between angular frequency, ω, and the speed of sound, c. Looking at the now isolated time dependence,

$$\frac{1}{c^2T}\frac{\partial^2T}{\partial t^2}=-k^2\,. \tag{3.10}$$

and by converting from wavenumber to angular frequency, and with some algebraic manipulation, Eqn. 3.10 becomes:

$$\frac{\partial^2T}{\partial t^2}+\omega^2T=0\,. \tag{3.11}$$

This differential equation has the same form as the one-dimensional wave equation form Eqn. 3.4, whose solutions are complex exponentials in the form of:

$$T(t)=A_1e^{i\omega t}+A_2e^{-i\omega t}\,, \tag{3.12}$$

where the A_1 and A_2 are constant weights for the complex exponential basis functions. The choice of using either the $e^{i\omega t}$ or the $e^{-i\omega t}$ is a matter of time convention. The convenience of selecting the separation constant as $-k^2$ can now be seen in the final form of our time solutions. This solution can also be expressed as a sum of sine and cosine functions, but typically the sine and cosine version is used for applications involving standing waves, and the complex exponential form is used for applications involving traveling, or propagating waves.

Next, taking the left side of Eqn. 3.9 and the first separation constant, $-k^2$, the azimuthal dependence (ϕ) can be separated on a single side of the equation from both the elevation (θ) and radius (r) and set equal to a separation constant of m^2:

$$\sin^2\theta\left[\frac{1}{R}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right)+\frac{1}{\Theta\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Theta}{\partial\theta}\right)+k^2r^2\right]=-\frac{1}{\Psi}\frac{\partial^2\Psi}{\partial\phi^2}=m^2\,. \tag{3.13}$$

Now with the azimuthal dependence isolated, the right side of Eqn. 3.13 can be rearranged as:

$$\frac{\partial^2\Psi}{\partial\phi^2}+m^2\Psi=0\,. \tag{3.14}$$

This differential equation is of an identical form to the time dependence from Eqn. 3.11, meaning that the azimuthal solutions also have the form of:

$$\Psi(\phi)=B_1e^{im\phi}+B_2e^{-im\phi}\,. \tag{3.15}$$

In spherical coordinates, the azimuthal angle must maintain rotational symmetry and continuity, such that the values for $\phi=0,2\pi,4\pi,6\pi$ ... are all equal. To ensure this, the value of m must take on integer values. Using Euler's formula, Eqn. 3.15 can be rewritten as:

$$B_1e^{im\phi}+B_2e^{-im\phi}=B_1[\cos(m\phi)+i\sin(m\phi)]+B_2[\cos(m\phi)-i\sin(m\phi)]\,. \tag{3.16}$$

Looking at the sine and cosine representation of these solutions, the values of azimuth will exhibit continuity if m takes on integer values. The variable m will be denoted as the degree of this solution to the wave equation, which will become more apparent once the elevation solution is determined.

3.1.2 Elevation Solution: Associated Legendre Polynomials

Finally, the radial and elevation dependence on the left side of Eqn. 3.13 can be separated from one another. After algebraic manipulation, and setting the two remaining separated variables equal to a separation constant, C, the differential equation becomes:

$$\frac{1}{R}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right)+k^2r^2=\frac{m^2}{\sin^2\theta}-\frac{1}{\Theta\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Theta}{\partial\theta}\right)=C\,. \tag{3.17}$$

Taking the right side of Eqn. 3.17, mathematical manipulation can yield the following form of the differential equation:

$$\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Theta}{\partial\theta}\right)+\left(C-\frac{m^2}{\sin^2\theta}\right)\Theta=0\,. \tag{3.18}$$

Performing the derivative in terms of elevation,

$$\frac{\cos\theta}{\sin\theta}\frac{\partial\Theta}{\partial\theta}+\frac{\partial^2\Theta}{\partial\theta^2}+\left(C-\frac{m^2}{\sin^2\theta}\right)\Theta=0\,, \tag{3.19}$$

and using the following substitution,

$$x=\cos\theta\,, \tag{3.20}$$

$$d\theta=\frac{dx}{-\sin\theta}, \tag{3.21}$$

and using both of the following relationships,

$$1-x^2=1-\cos^2\theta=\sin^2\theta\,, \tag{3.22}$$

$$d\theta^2=\frac{dx^2}{\sin^2\theta}, \tag{3.23}$$

the differential equation from Eqns. 3.19 – 3.23 becomes:

$$-2x\frac{\partial\Theta(\mathrm{x})}{\partial x}+(1-x^2)\frac{\partial^2\Theta(\mathrm{x})}{\partial x^2}+\left(C-\frac{m^2}{1-x^2}\right)\Theta(\mathrm{x})=0\,, \tag{3.24}$$

And if we set $C=n(n-1)$, the differential equation becomes the form of the associated Legendre differential equation:

$$(1-x^2)\frac{d^2\Theta(\mathrm{x})}{dx^2}-2x\frac{d\Theta(\mathrm{x})}{dx}+\left(n(n-1)-\frac{m^2}{1-x^2}\right)\Theta(x)=0\,, \tag{3.25}$$

This equation has known solutions called associated Legendre polynomials. The reason for assigning $C = n(n-1)$ was to ensure that the solutions of the differential equation would be bounded at $x = \pm 1$. These associated Legendre polynomials are written as:

$$\Theta(x) = P_n^m(x) \quad \text{or} \quad \Theta(\theta) = P_n^m(\cos\theta)\,, \tag{3.26}$$

and defined as:

$$P_n^m(x) = (-1)^m(1-x^2)^{\frac{m}{2}}\frac{d^m}{dx^m}P_n(x) \quad \text{for } m \geq 0\,, \tag{3.27}$$

$$P_n^{-m}(x) = (-1)^m\frac{(n-m)!}{(n+m)!}P_n^m(x) \quad \text{for } m < 0\,, \tag{3.28}$$

and $P_n(x)$ are called Legendre polynomials, defined in Rodrigues' Formula by:

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}[(x^2-1)^n]\,, \tag{3.29}$$

where n is a non-negative integer denoting the order of the polynomial and m denotes the degree of the polynomial. For each order n, the values of m are integers such that $-n \leq m \leq n$. The $(-1)^m$ from Eqns. 3.28 and 3.29 is referred to as the Condon-Shortly phase term. This set of functions forms the basis of the azimuthal solutions of the wave equation. Figure 3-2 shows plots of the normalized, positive-degree associated Legendre polynomials from Eqn. 3.27.

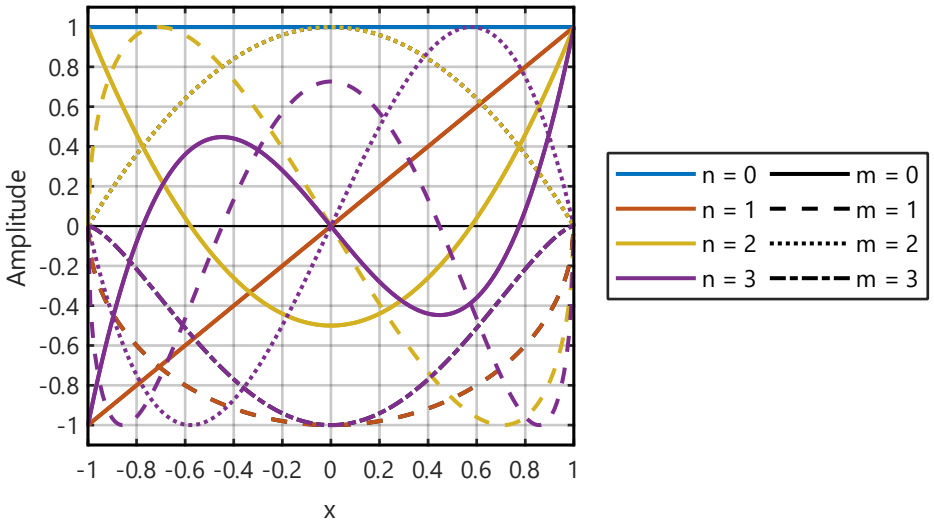


Figure 3-2: Plots of the associated Legendre polynomials (normalized) for orders n = 0 to n = 3. Here, the amplitude has been normalized to range from -1 to 1 for visual representation.

3.1.3 Radial Solution: Spherical Bessel & Hankel Functions

Looking at the left side of Eqn. 3.17, and performing some algebraic manipulation yields:

$$\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right)+(k^2r^2-C)R=0\,. \tag{3.30}$$

By taking the derivative with respect to radius and applying the chain rule, along with a substitution for the definition of C from section 3.1.1.2, the following form of the differential equation is found:

$$\frac{\partial^2 R}{\partial r^2}+\frac{2}{r}\frac{\partial R}{\partial r}+\left(k^2-\frac{n(n+1)}{r^2}\right)R=0\,. \tag{3.31}$$

This differential equation closely resembles the Bessel equation. Using the substitution,

$$R(r)=\frac{1}{r^2}u(r)\,, \tag{3.32}$$

the differential equation then takes the form of Bessel's equation:

$$\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}+\left(k^2-\frac{(n+1/2)^2}{r^2}\right)u=0\,. \tag{3.33}$$

The solutions to this differential equation are given as Bessel, $J_n(kr)$, and Neumann functions, $N_n(kr)$ of order $n+1/2$:

$$R(kr)=\frac{D_1}{(kr)^{1/2}}J_{n+1/2}(kr)+\frac{D_2}{(kr)^{1/2}}N_{n+1/2}(kr) \tag{3.34}$$

Note that this expression now depends upon the product kr, including the wavenumber in its argument. The half-integer order of the Bessel and Neumann functions have a close relationship with the spherical Bessel and Neumann functions as:

$$x^{-1/2}J_{n+1/2}(x)=\left(\frac{2}{\pi}\right)^2 j_n(x)\,, \tag{3.35}$$

$$x^{-1/2}N_{n+1/2}(x)=\left(\frac{2}{\pi}\right)^2 n_n(x)\ , \tag{3.36}$$

By grouping some of the newly introduced constants into a redefinition of our coefficients, A_5 and A_6, our solution can be more simply stated as:

$$R(kr)=D_1 j_n(kr)+D_2 n_n(kr) \tag{3.37}$$

The spherical Bessel functions of the first kind, $j_n(x)$, and of the second kind, $n_n(x)$ (also referred to as spherical Neumann functions), can be written from the Rayleigh formulas as:

$$j_n(x) = (-1)^n x^n \left(\frac{1}{x}\frac{d}{dx}\right)^n \frac{\sin(x)}{x}, \tag{3.38}$$

and

$$n_n(x) = -(-1)^n x^n \left(\frac{1}{x}\frac{d}{dx}\right)^n \frac{\cos(x)}{x}. \tag{3.39}$$

The operator n in the exponent of Eqns. 3.38 and 3.39 implies applying a derivative and multiplication by $1/x$ a total of n times. Another set of solutions to the differential equation given by Eqn. 3.31 is the set of spherical Hankel functions. These functions can be directly defined in terms of the spherical Bessel functions, as:

$$h_n^{(1)}(x) = j_n(x) + i n_n(x)\,, \tag{3.40}$$

and

$$h_n^{(2)}(x) = j_n(x) - i n_n(x). \tag{3.41}$$

The spherical Hankel functions of the first kind, $h_n^{(1)}(x)$, and of the second kind, $h_n^{(2)}(x)$, can be also be represented from Eqns. 3.38 – 3.41:

$$h_n^{(1)}(x) = -i(-1)^n x^n \left(\frac{1}{x}\frac{d}{dx}\right)^n \frac{e^{ix}}{x}, \tag{3.42}$$

and

$$h_n^{(2)}(x) = i(-1)^n x^n \left(\frac{1}{x}\frac{d}{dx}\right)^n \frac{e^{-ix}}{x}. \tag{3.43}$$

For the radial part of the wave equation, solutions can be written as a linear combination of these sets of functions. Often, they are expressed as either a linear combination of spherical Bessel functions, or as a linear combination of spherical Bessel and Hankel functions. In spherical acoustics literature, a linear combination of the spherical Bessel and Hankel functions of the first kind are often selected.

3.1.4 Full Solution to the Spherical Wave Equation

Combining the solutions from Eqns. 3.12, 3.15, 3.26, and 3.37 into 3.7, the general solution to the wave equation in spherical coordinates can be represented as:

$$p(k,r,\theta,\phi,t) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} a_{nm}(k)\begin{Bmatrix} e^{i\omega t} \\ e^{-i\omega t} \end{Bmatrix}\begin{Bmatrix} j_n(kr) \\ n_n(kr) \end{Bmatrix}\begin{Bmatrix} e^{im\phi} \\ e^{-im\phi} \end{Bmatrix} P_n^m(\cos\theta)\,. \tag{3.44}$$

or

$$p(k,r,\theta,\phi,t) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} a_{nm}(k)\begin{Bmatrix} e^{i\omega t} \\ e^{-i\omega t} \end{Bmatrix}\begin{Bmatrix} h_n^{(1)}(kr) \\ h_n^{(2)}(kr) \end{Bmatrix}\begin{Bmatrix} e^{im\phi} \\ e^{-im\phi} \end{Bmatrix} P_n^m(\cos\theta)\,. \tag{3.45}$$

Here, the previous constants have been grouped together into once combined frequency dependent weighting, A_{nm}. The bracket notation implies that the solution can take the form of any linear combination of the either function inside each bracket. Note that although angular frequency is not specified in the arguments of the pressure expression, only radius, it is defined by the choice of wavenumber and the speed of sound. The choice of $e^{i\omega t}$ or $e^{-i\omega t}$ is a selection of time convention, and for this work, $e^{i\omega t}$ will be used. It should be noted that this choice is not arbitrary, as its selection will dictate which solutions represent ingoing or outgoing traveling waves. Similarity in the azimuthal solution, this dissertation will use the convention of $e^{im\phi}$. Finally, from the conclusion of last section, the most common set of basis functions selected for the radial solution is the combinations of the spherical Bessel and Hankel functions of the first kind. This makes the complete solution the combination of,

$$p(k,r,\theta,\phi,t) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} a_{nm}(k) j_n(kr) P_n^m(\cos\theta) e^{im\phi} e^{i\omega t}\,, \tag{3.46}$$

and

$$p(k,r,\theta,\phi,t) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} a_{nm}(k) h_n^{(1)}(kr) P_n^m(\cos\theta) e^{im\phi} e^{i\omega t}\,, \tag{3.47}$$

Plots of both the spherical Bessel and Hankel functions are provided in Figures 3-3 and 3-4. The Bessel functions have dips or nulls in their response, and the Hankel functions diverge towards the origin. Due to these differences, different problems retain different combinations of spherical Bessel and Hankel functions due to physical boundary conditions or constraints. Plane waves in space cannot attain infinite amplitude, so often plane wave-related solutions involve the spherical Bessel functions, $j_n(kr)$. On the other hand, point sources do exhibit high

amplitudes at the location of the source, so the spherical Hankel function, $h_n^{(1)}(kr)$ can be useful in these formulations.

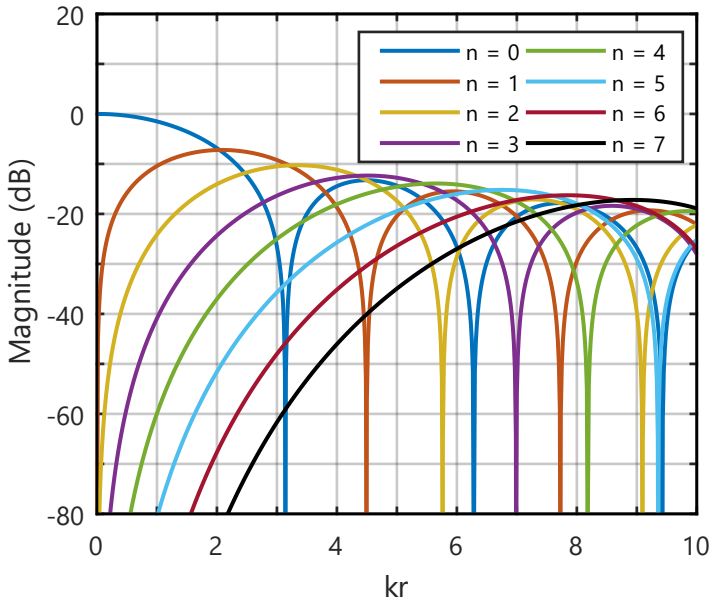


Figure 3-3: Magnitude of the spherical Bessel functions, $j_n(kr)$, for orders $n = 0$ to $n = 7$.

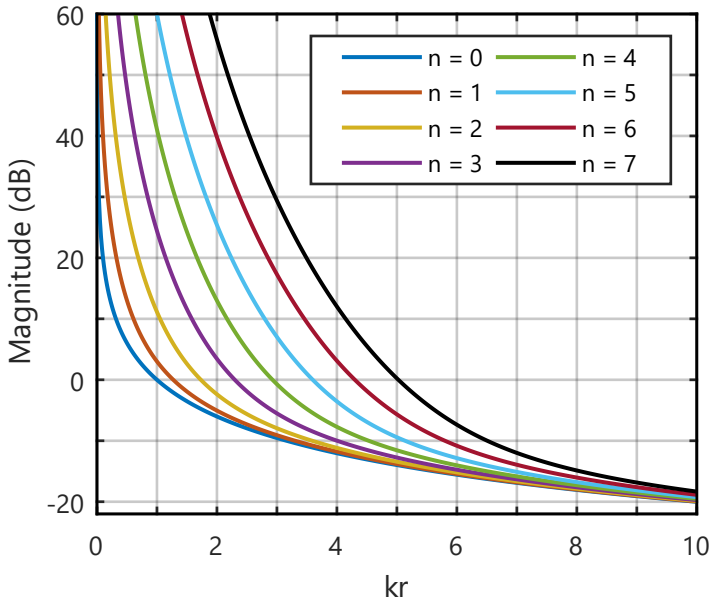


Figure 3-4: Magnitude of the spherical Hankel functions, $h_n^{(1)}(kr)$, for orders $n = 0$ to $n = 7$.

3.2 Spherical Harmonics

From the solutions to the wave equation in spherical coordinates found in Eqns. 3.46 and 3.47, the parts that depend upon azimuth and elevation are often grouped together to define a set of functions referred to as spherical harmonics (SH):

$$Y_n^m(\theta,\phi) = N_n^m P_n^m(\cos\theta)e^{im\phi}\,, \tag{3.48}$$

The SHs functions are composed of associated Legendre polynomials, complex exponentials, and a normalization term, N_n^m. Balloon-style plots of these SHs are shown in Figures 3-5 through 3-7. This set of functions forms the eigenfunctions of the angular portion of the three-dimensional wave equation using the Laplacian in spherical coordinates. The application of these functions is not only found in acoustics, as the Laplacian is common in fundamental differential equations for electromagnetism, gravity, quantum mechanics, etc. When visually inspecting these functions, a few things become noticeable. First, as you increase in SH order, the spatial complexity of the functions increases. Just as increasing the frequency of a sine or cosine wave will cause 'quicker' oscillations in the time-domain, increased SH order causes 'quicker' oscillations in the spatial domain.

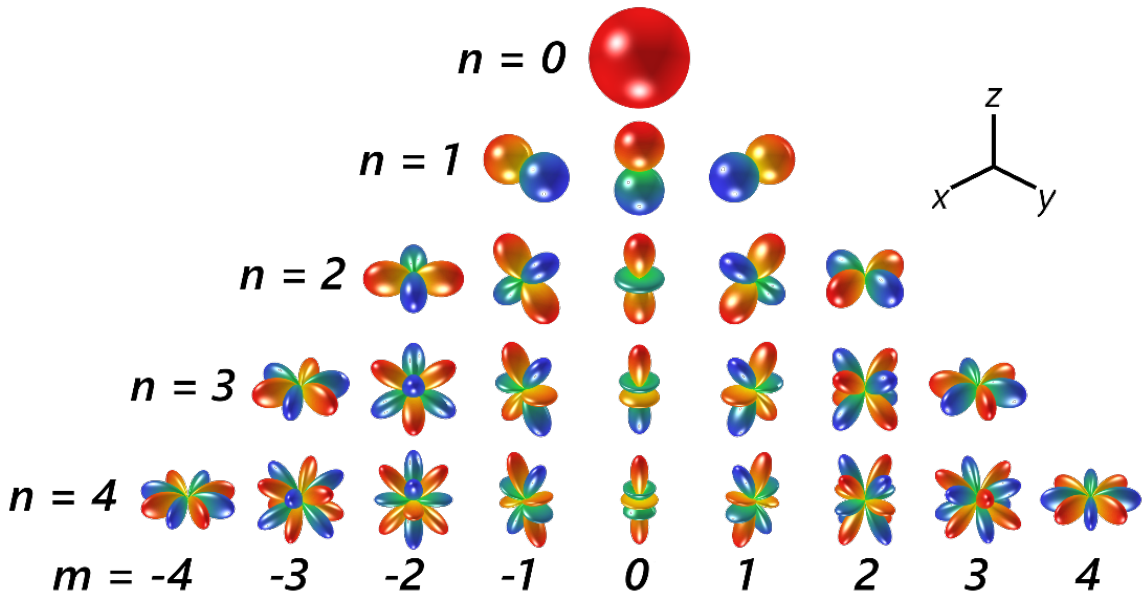


Figure 3-5: A tree-style diagram of the complex SH functions. For visual representation, the real part of the complex SHs are plotted for $n \geq 0$ and the imaginary part is plotted for $n < 0$.

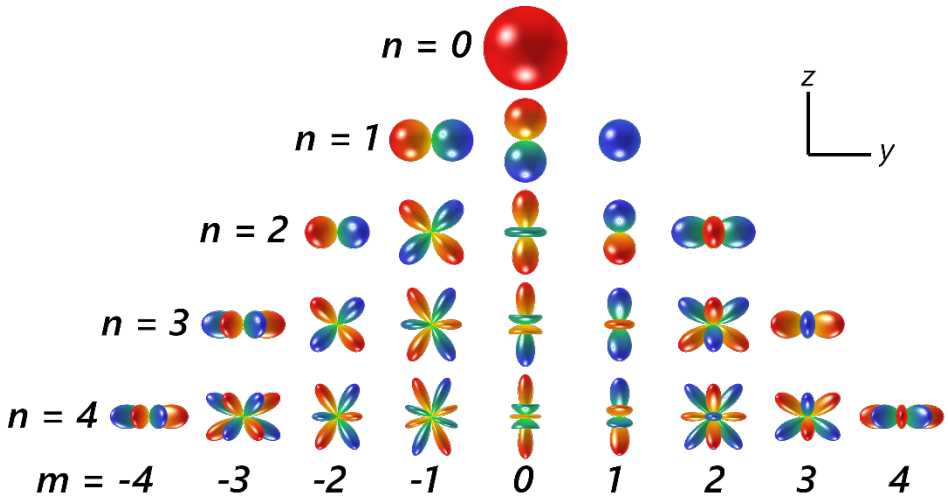


Figure 3-6: The same as in Figure 3-5, now showing a frontal view, oriented with the *x* axis.

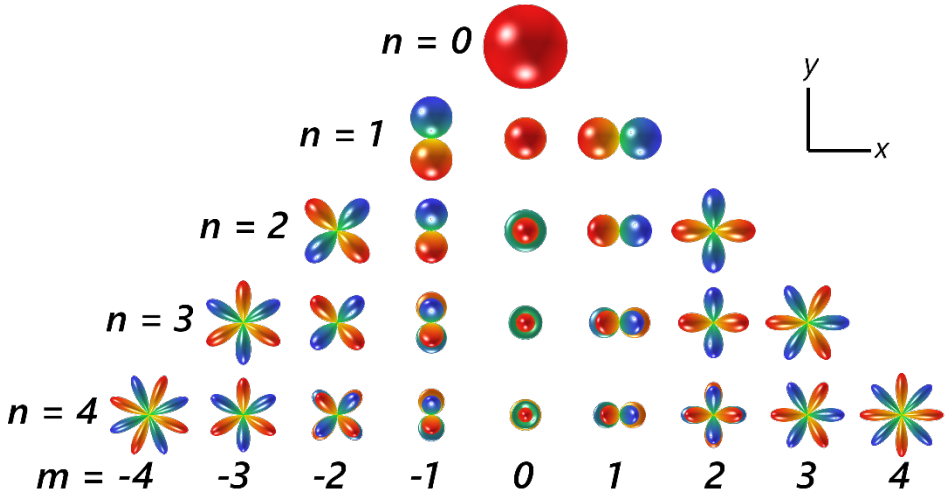


Figure 3-7: The same as in Figure 3-5, now showing a top-down view, oriented with the *z* axis.

Next, the harmonics of degree $m = 0$ are seen to be rotationally symmetric about the z axis. Azimuthal variations are found as you move away from the center of the function 'tree'. This property is most easily seen in the top-down view from Figure 3-7. It is also seen that the higher the absolute value of degree, the higher the number of azimuthal nulls in the specific harmonic. In this way, the functions can be thought of as directly analogous to vibrational modes of rectangular or circular plates, membranes, or rooms, which can be classified with indices corresponding to modal peaks and nodes. Elevational nodes appear in a similar way, shown in Figure 3-6. As the functions move from the outside of the tree to the inside of the tree, a higher degree of spatial complexity and high numbers of nodal cones of constant elevation are found.

3.2.1 Properties of Spherical Harmonics

On further inspections, some clear mathematical, integration-based properties of these functions emerge. For this section, it will be assumed that the normalization term from Eqn. 3.48 will be set to:

$$N_n^m = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}}. \tag{3.49}$$

First, except for the zeroth-order harmonic, all SH functions have an integral of zero over the unit sphere:

$$\int_0^{2\pi}\int_0^{\pi} Y_n^m(\theta,\phi)\sin\theta\, d\theta d\phi = \sqrt{4\pi}\delta_{n0}\delta_{m0}\,. \tag{3.50}$$

Here, δ_{n0} is the Kronecker delta function, having a value of 0 except when the two values in its subscript are equal, when it takes on a value of 1. In this case, it is only equal to 1 when $n = 0$ (implying as well that $m = 0$). This integral property can be more broadly extended to the orthogonality property of SHs:

$$\int_0^{2\pi}\int_0^{\pi} [Y_n^m(\theta,\phi)]^* Y_{n'}^{m'}(\theta,\phi)\sin\theta\, d\theta d\phi = \delta_{nn'}\delta_{mm'}\,. \tag{3.51}$$

Orthogonality is inherently a mathematical property, but its physical importance is fundamental to the usefulness of representing sound fields in the SH domain. Equation. 3.51 compares any two of the SH functions over the sphere by integrating one harmonic multiplied by the complex conjugate of another harmonic. In the cartesian coordinate system, two vectors are said to be orthogonal if their dot product is equal to zero. Equation. 3.51 can be thought of as a dot product of two spatial functions in spherical coordinates. This property states that all harmonics are orthogonal, and therefore bring unique spatial information not contained in any other harmonic up to an infinite order, $n = \infty$. This property's usefulness will become clear in section 3.3. Using the current normalization scheme, the SH functions also demonstrate the property of orthonormality, as the right side of Eqn. 3.51 is equal to 1 when $n = n'$ and $m = m'$.

Another property of the set of SH functions is the completeness property:

$$\sum_{n=0}^{\infty}\sum_{m=-n}^{n} [Y_n^m(\theta,\phi)]^* Y_n^m(\theta',\phi') = \delta(\cos\theta - \cos\theta')\delta(\phi - \phi')\,. \tag{3.52}$$

In Eqn. 3.52, it is much easier to approach this analogy by considering the time-domain basis functions of sine and cosine functions. This time-domain analogy will be used again in section 3.3 to discuss the spherical Fourier series. In the time domain, an impulse can be constructed by summing together an infinite number of in-phase cosine functions, with an infinitely small frequency resolution. Although the individual cosine functions have energy at times other than $t = 0$, across the infinite sum, the positive and negative values cancel and sum to zero at any time where $t \neq 0$. This superposition of sinusoids, summing to create an impulse in time, is shown in Figure 3-8. From here, any time series of pressure can be thought of as a series of samples, or impulses, with different weights and time delays (phase). The *completeness* of sine and cosine basis functions is expressed in their ability to recreate any function in the time domain.

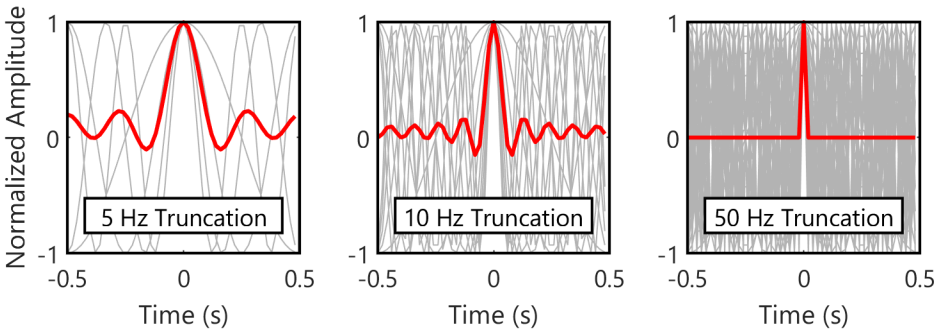


Figure 3-8: A plot showing the summation of cosine functions at a 50 Hz sampling rate with 1 Hz frequency resolution, truncated at 3 different frequencies. As truncation frequency increases to the sampling rate, the summation of in-phase cosines creates a time-domain impulse.

This directly translates to the spatial domain. Equation. 3.52 shows that when summing up SHs, up to an infinite order and across all degrees, this superposition will result in a spatial impulse, or Dirac delta with an infinite amplitude. When you shift, or rotate slightly in space, making $\theta \neq \theta'$ or $\phi \neq \phi'$, then the superposition and phase of the SH functions will cancel to a value of 0. Similar to the time domain, if a spatial Dirac delta can be constructed, any bounded spatial function could also be considered a superposition of series of these spatial impulses, with different weights and individually rotated in space (spatial phase). Thus, any bounded spatial function can be represented using the set of SH functions as a basis. As an interpretation, this expresses the completeness property in the spatial domain. A more complete representation of a spatial direct delta function, otherwise called a plane wave, is presented in section 3.3.

3.2.2 Real-valued vs. Complex-valued Spherical Harmonics

Up to this point, this dissertation has presented the SH functions in their complex-valued form, but these functions can also be expressed in a real-valued form. This representation can be more convenient when working with time-domain data, as the evaluation of the function results in a real-valued result, as opposed to a complex-valued result. The real-valued SHs can be derived from the complex valued harmonics as:

$$Y_{nm}(\theta,\phi) = \begin{cases} \frac{(-1)^m}{\sqrt{2}}\left(Y_n^{|m|}(\theta,\phi) + iY_n^{-|m|}(\theta,\phi)\right) & \text{if } m > 0 \\ Y_n^0(\theta,\phi) & \text{if } m = 0 \\ \frac{(-1)^m}{i\sqrt{2}}\left(Y_n^{|m|}(\theta,\phi) - iY_n^{-|m|}(\theta,\phi)\right) & \text{if } m < 0\,. \end{cases} \quad 3.53$$

And with evaluation of the full expression:

$$Y_{nm}(\theta,\phi) = \begin{cases} (-1)^m\sqrt{2}\sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}}P_n^m(\cos\theta)\cos m\phi & \text{if } m > 0 \\ \sqrt{\frac{2n+1}{4\pi}}P_n^m(\cos\theta) & \text{if } m = 0 \\ (-1)^m\sqrt{2}\sqrt{\frac{2n+1}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}}P_n^{|m|}(\cos\theta)\sin|m|\phi & \text{if } m < 0\,. \end{cases} \quad 3.54$$

Note that both the order and index have been moved to the subscript of the SH function shorthand, $Y_{nm}(\theta,\phi)$, to indicate the difference between real-valued and complex-valued SHs. By expressing the complex exponential function that represented the azimuthal variation now in sine and cosine functions, the positive degree functions become cosine-type SHs, while the negative-degree functions become sine-type SHs. Figures 3-9 through 3-11 show a comparison of the real and complex parts of the complex-valued SHs and the real-valued SHs from Eqns. 3.53 and 3.54.

Comparing the two, the non-negative degree real-valued SHs (right side) are taken from the real component of the non-negative complex-valued SH functions. The negative degree real-valued SHs are taken from the imaginary component of the positive-degree complex-valued harmonics. The phase differences that result from comparing Figures 3-9 and 3-10 to 3-11 are due to the $(-1)^m$ factor. The relationship between the real and complex harmonics, more intuitive to the visual comparison, can also be expressed as:

$$Y_{nm}(\theta,\phi) = \begin{cases} \sqrt{2}(-1)^m \mathrm{Im}\left[Y_n^{|m|}(\theta,\phi)\right] & \text{if } m > 0 \\ Y_n^0(\theta,\phi) & \text{if } m = 0 \\ \sqrt{2}(-1)^m \mathrm{Re}[Y_n^m(\theta,\phi)] & \text{if } m < 0\,. \end{cases} \tag{3.55}$$

The real-valued SHs in this section contain a factor known as the Condon-Shortly phase term, expressed as a multiplication by $(-1)^m$. In acoustics, real-valued SHs are most common in the field of higher-order Ambisonics (HOA). Conventions may or may not include this factor, but the most commonly agreed upon convention suggests that this factor not be included in the definition of SH. This factor can easily be removed by again multiplying each SH function by the $(-1)^m$ factor. For more complete information on the channel ordering and normalization schemes used in acoustics for SHs, see section 3.2.3

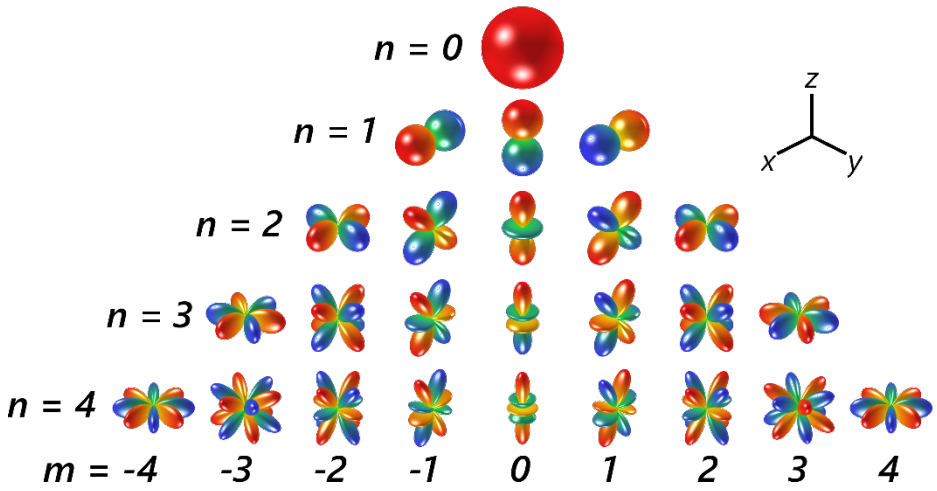


Figure 3-9: The real part of the complex-valued SH functions from Eqn. 3.48.

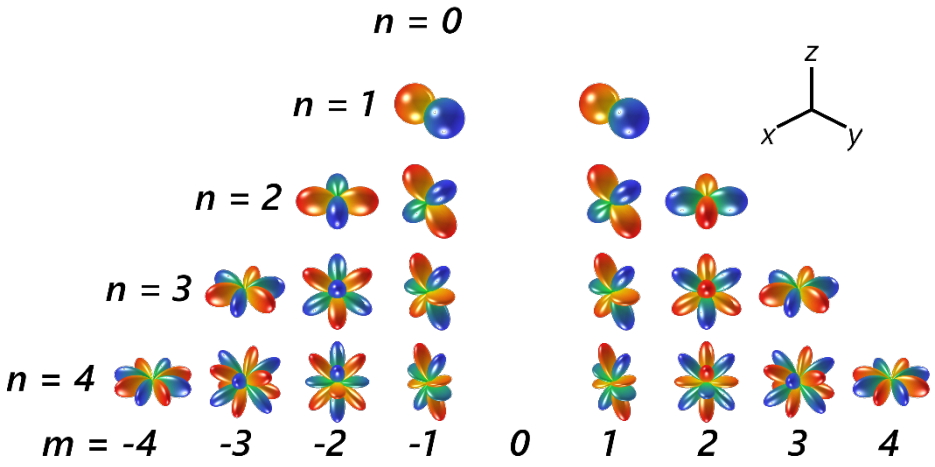


Figure 3-10: The imaginary part of the complex-valued SH functions from Eqn. 3.48. The 0th degree SHs have no imaginary component.

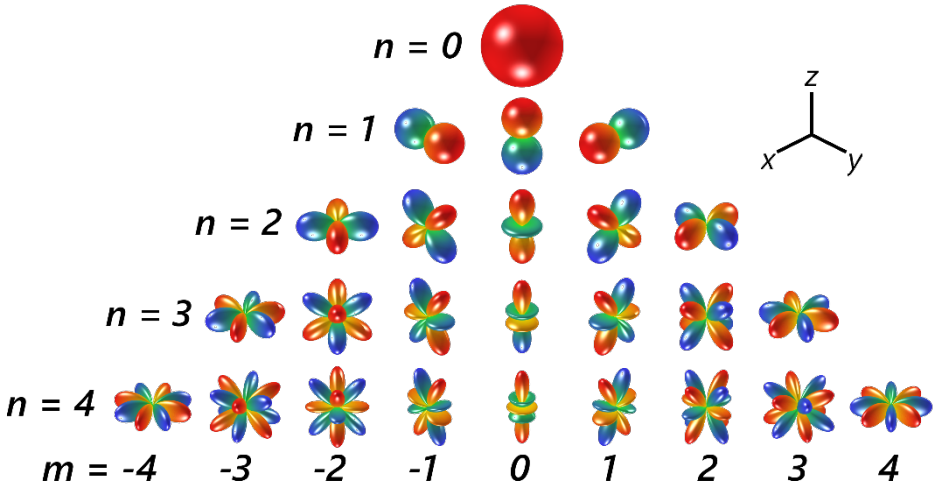


Figure 3-11: Real-valued SHs from Eqn. 3.54. The non-negative degree SHs (right & center, above) come from the real part of the complex SHs (Figure 3-9, right & center), and the negative degree SHs (left, above) come from the imaginary part of the positive degree complex SHs (Figure 3-10, right).

3.2.3 Normalization, Channel Ordering, and Condon-Shortly Phase Term

Up to this point, the SH functions have been defined using a specific normalization term, N_n^m, and the issue of channel ordering has not been addressed. When defining the set of SH functions, it is possible to choose many different normalization schemes, all with different benefits and shortcomings. The most common normalization term across different fields is the orthonormal scheme, expressed in Eqn. 3.49. This normalization scheme will ensure that the orthonormality criterion for the complex-valued SHs, Eqn. 3.51, is satisfied. If another normalization scheme is used, the right side of Eqn. 3.51 will be equal to the multiplication of the two Kronecker delta functions along with a constant, not equal to unity. The complex SH functions defined using other normalization schemes are orthogonal and complete, but they are not orthonormal. In the spherical acoustics world, this property is not always essential, so it is the choice of the artist, researcher, or engineer to be consistent and careful in selection.

The SH functions can be defined using different normalization schemes, but the most common forms are complex orthonormal normalization (CplxN3D), real orthonormal normalization (N3D, sometimes referred to as full 3D normalization), Schmidt semi-normalization (SN3D), and MaxN normalization.[80] These different normalization schemes are given in Eqns. 3.56 to 3.58:

$$N_n^m{}_{CplxN3D} = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}}. \tag{3.56}$$

$$N_n^m{}_{SN3D} = \sqrt{(2-\delta_m)\frac{(n-|m|)!}{(n+|m|)!}}, \quad for\ \delta_m = \begin{cases} 1 & \text{if } m = 0 \\ 0 & \text{if } m \neq 0 \end{cases}. \tag{3.57}$$

$$N_n^m{}_{N3D} = \sqrt{2n+1}N_n^m{}_{SN3D}\,. \tag{3.58}$$

The MaxN normalization scheme is performed by separately normalizing the maximum value of each SH function to a value of 1. Although this scheme does prevent the occurrence of higher values at higher-order SH components, an explicit mathematical definition, determined for any given value of n and m, is not known. CplxN3D and N3D both provide normalization that satisfy the orthonormality criterion given in Eqn. 3.51 for complex-valued SHs and real-valued SHs, respectively.

The second selection is in the channel ordering of the SH functions. This set of functions is defined by index values for order and degree, n and m respectively, and they are often visually organized in tree-like diagrams. When working with practical audio signals, and sharing audio data, each SH function must be assigned to a specific channel or column in a matrix, forcing the tree-like diagram into a one-dimensional ordering. The initial ordering scheme was developed by the inventor of first-order Ambisonics, Michael Gerzon, and consisted only of the first four functions ordered as W – X – Y – Z. The indication of W is for the zeroth-order monopole component, and the remaining letters correspond to the dipole components oriented along each axis, shown in Figure 3-11. This ordering scheme is the B-format or the Furse-Malham (FuMa) channel ordering. The letter designations are provided in Figure 3-12, along with the corresponding channel numbers (indexed from zero). To continue the letter-based designation, second-order channels uses the letters R – S – T – U – V, third-order channels uses the letters K – L – M – N – O – P – Q, and although not often seen, fourth-order channels could have the letters B – C – D – E – F – G – H – I – J.

For the FuMa (B-format) ordering, it can be seen in Figure 3-12 that the channel index, while traversing from first to second-order, switches patterns. To keep a consistent pattern, an outside-to-inside traverse of the SH tree, Daniel implemented a new ordering scheme coined as the Single Index Designation (SID). Shown in Figure 3-13, this scheme only differs in from FuMa in correcting this traverse to a consistent pattern. Using the previous letter labels, the SH functions are now ordered as W – X – Y – Z for first-order, U – V – S – T – R for second-order, P – Q – N – O – L – M – K for third-order, and I – J – G – H – E – F – C – D – B for forth-

order. One very simple problem with both FuMa and SID is that after fourth-order, the alphabet runs out of letters to assign to every channel. This method of traversing through the SH tree does not have a clear explicit relationship between channel number, SH order, and degree. This also makes the ordering method difficult to extend to higher orders.

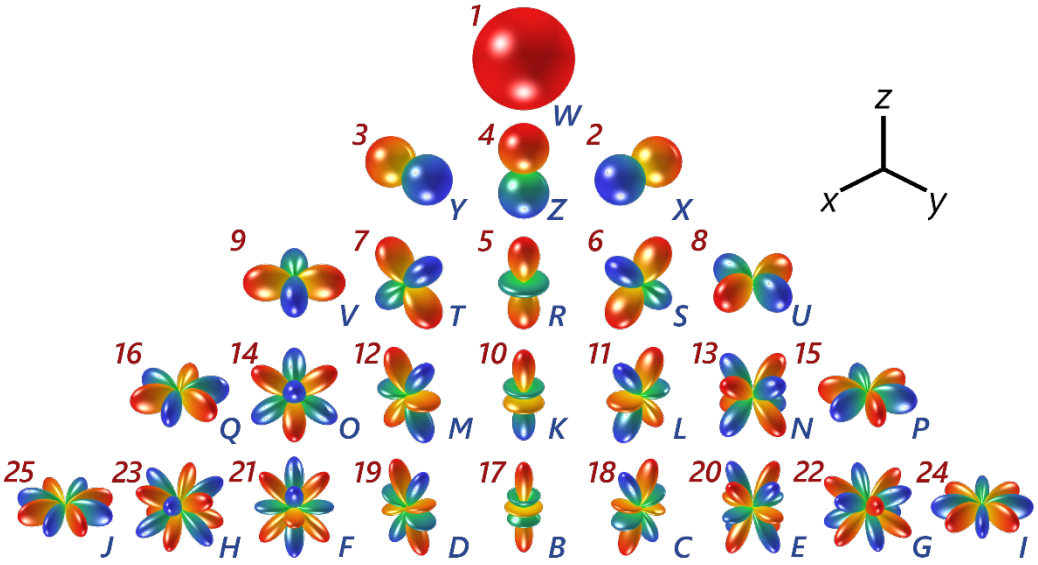


Figure 3-12: Complex SH functions shown with their common letter designations and channel index values for the Furse-Malham channel ordering convention.

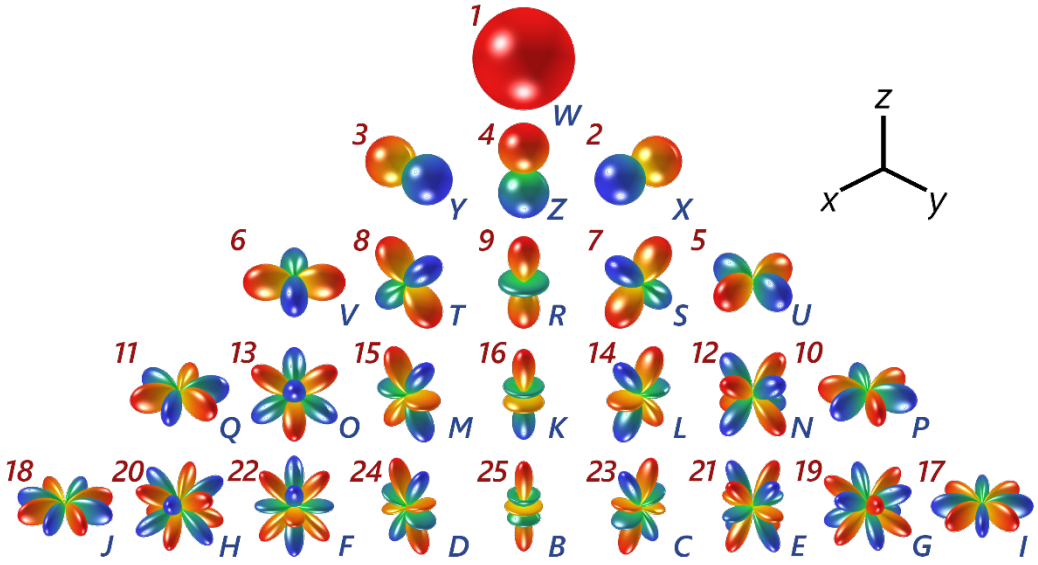


Figure 3-13: Complex SH functions shown with their common letter designations and channel index values for the Single Index Designation (SID, from Daniel) channel ordering convention.

To overcome this limitation, the Ambisonic Channel Number (ACN) ordering scheme was proposed.[80] This scheme orders the SH functions reading left to right across the SH tree,

and progressing to a new order (row) once the previous order is complete. Just as one would read a book, this method is much more visually intuitive than previous schemes. More importantly, ACN also allows for the channel number to be explicitly calculated in terms of both the SH order, n, and degree, m, such that:

$$ACN = n^2 + n + m\,. \tag{3.59}$$

Note that ACN is indexed starting from 0 instead of 1 in terms of channel number. This explicit definition of the channel number allows for easier implementation into code-based algorithms and processing. The ACN ordering scheme is pictured below in Figure 3-14, showing the channel index for each SH function.

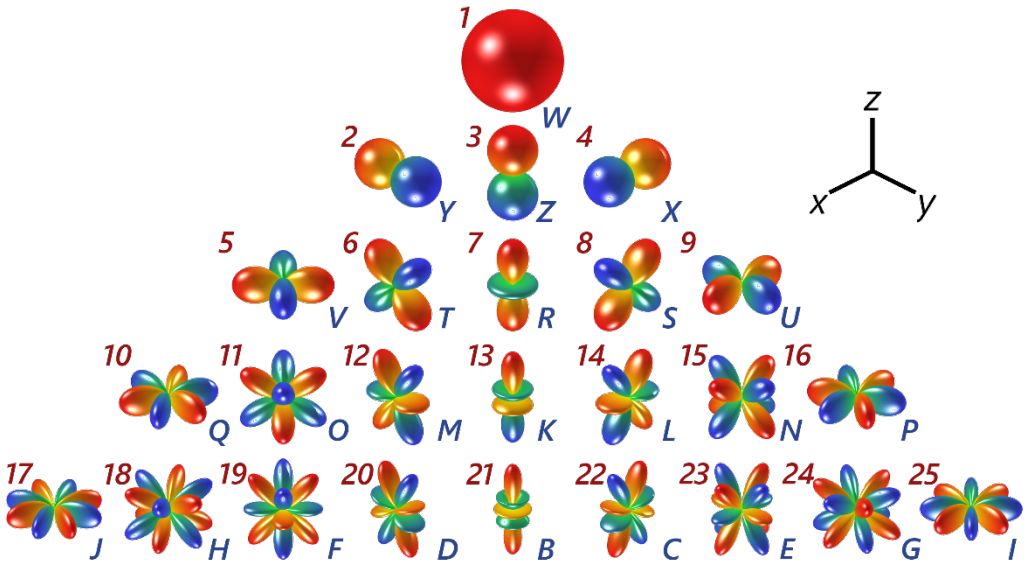


Figure 3-14: Complex SH functions shown with their common letter designations and channel index values for the Ambisonic Channel Number (ACN) channel ordering convention. Compared to the visual, the ACN index values intuitively progress from left to right, down each row. Note that the channel indices are given as the $ACN + 1$, compared to the common zero indexing in Eqn. 3.59.

The final element that can adjust the definition of the SH functions is the use of the Condon-Shortly phase term. This term appears as a factor of $(-1)^m$, and it reverses the sign of every other SH function when using the ACN convention. This factor is commonly included in the definition of the associated Legendre polynomials, found in Eqn. 3.27, so it is not seen in the definition of the SH functions in Eqn. 3.48 or in the normalization factors found in Eqns. 3.56 to 3.58. It is important to be conscious of the software packages being used to calculate values for either the SH functions or the associated Legendre Polynomials and how internal or built-in functions define these polynomials. It can be excluded by either removing this factor from the definition of the associated Legendre polynomials or multiplying the final SH definition again by a factor of $(-1)^m$.

Overall, different areas and fields use different combinations of these schemes. Table 3.1 presents a summary of the SH type, normalization scheme, channel ordering, and Condon-Shortly phase term inclusion for different formats. In general, the orthonormal complex format is the most common format in the spherical microphone array and beamforming literature. When looking at literature based around HOA, there is no clear reference textbook, and as such, multiple formats have found use. Historically, the FuMa convention was the most widely used, but recently, to standardize a convention, the ambiX convention seems to be the most agreeable.[80] It is also important to be considerate of the coordinate system, especially in terms of the elevation angle's definition. Ambisonics literature will commonly define the elevation angle as the angle from any point to the horizontal, $x-y$ plane. Before, Figure 3-11 showed the real-valued SHs using the orthonormal real format. As this is not the most common format for HOA, Figure 3-15 shows real-valued SH function using the ambiX format (now excluding the Condon-Shortly phase term). This exclusion of the $(-1)^m$ factor causes a polarity reversal in some of the SH functions.

Table 3.1: A summary table of the different SH or Ambisonic conventions in terms of SH type, normalization scheme, channel ordering convention, Condon-Shortly phase term inclusion, and some additional notes.

Ambisonic Format	SH Type	Norm.	Chan. Order	CS?	Notes:
Orthonormal	Complex	CplxN3D	ACN	Yes	Common in beamforming
Orthonormal	Real	N3D	ACN	No	Less common for HOA
ambiX	Real	SN3D	ACN	No	Recommended for HOA
Furse-Malham	Real	maxN	FuMa	No	-3dB on W, hist. common
Daniel	Real	maxN	SID	No	Not in widespread use

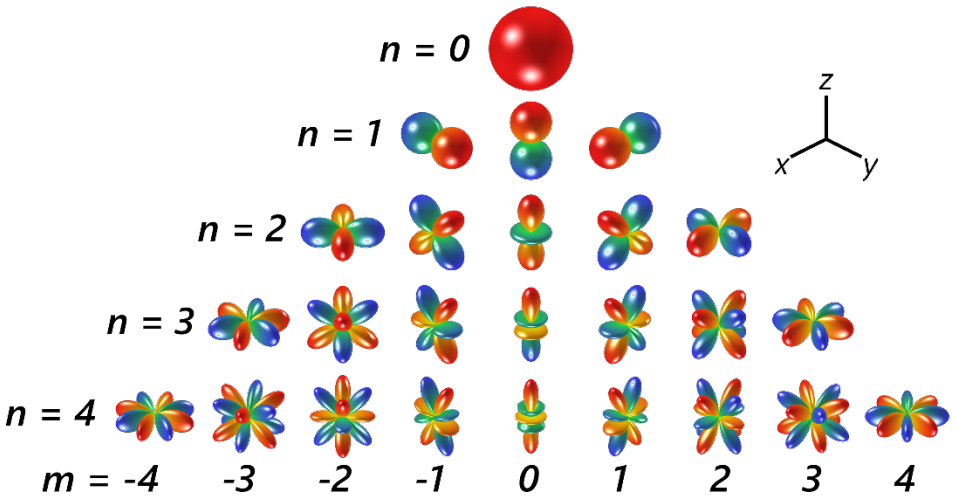


Figure 3-15: Real-valued SHs up to order $n = 4$.

Most importantly, all processing into the SH domain (referred to as encoding) and processing out of the SH domain (referred to as decoding) should be done using a consistent scheme. Just as speaking a foreign language to a speaker of another language can cause confusion and misinterpretation, processing a SH signal with algorithms designed for another normalization scheme will create a confusing and chaotic result. It is essential that we understand what language our SHs functions are defined in and what language our SH processing algorithms expect to ensure proper results. If there are mismatches between conventions, this can be corrected by the appropriate translation between different conventions.

3.3 Spherical Fourier Analysis

The key application of SH functions is found in spherical Fourier analysis. A direct analogy can be made between SHs and sine and cosine functions for the spherical and time-frequency Fourier series, respectively. When measuring a pressure time series, time-frequency analysis relies upon the fundamental assumption that any time series can be represented as a weighted sum of sine and cosine functions, or complex exponentials:

$$p(t) = \sum_{n'=-\infty}^{\infty} X[n'] e^{in'\omega_o t} . \tag{3.60}$$

From this formulation, the coefficients, X, can be defined as:

$$X(n') = \frac{1}{2\pi} \int_{-\pi}^{\pi} p(t) \left[e^{in'\omega_o t} \right]^* dt , \tag{3.61}$$

where n' is an integer representing the frequency index of the coefficient $X(n')$, an " * " denotes a complex conjugation, and ω_o is the fundamental angular frequency, making $n'\omega_o$ the n'^{th} harmonic of ω_o. This summation shows how sine and cosine functions become the building block of a time series, $p(t)$, which can be reconstructed by determining the appropriate weighting for each individual sine and cosine function, $X(n')$, found from Eqn. 3.61. To visualize this time-domain reconstruction, Figure 3-16 shows how a target sawtooth wave can be reconstructed with the proper weighting factors, $X(n')$, determined using a time-frequency Fourier transform, shown in Figure 3-17. As can be seen, the more harmonics that are used in the reconstruction, the closer the reconstructed function approximates the original function. Due to the selection of a sampling rate, a spatial Nyquist frequency limits the maximum frequency that can be used in the reconstruction, and the signal length determines the frequency resolution.

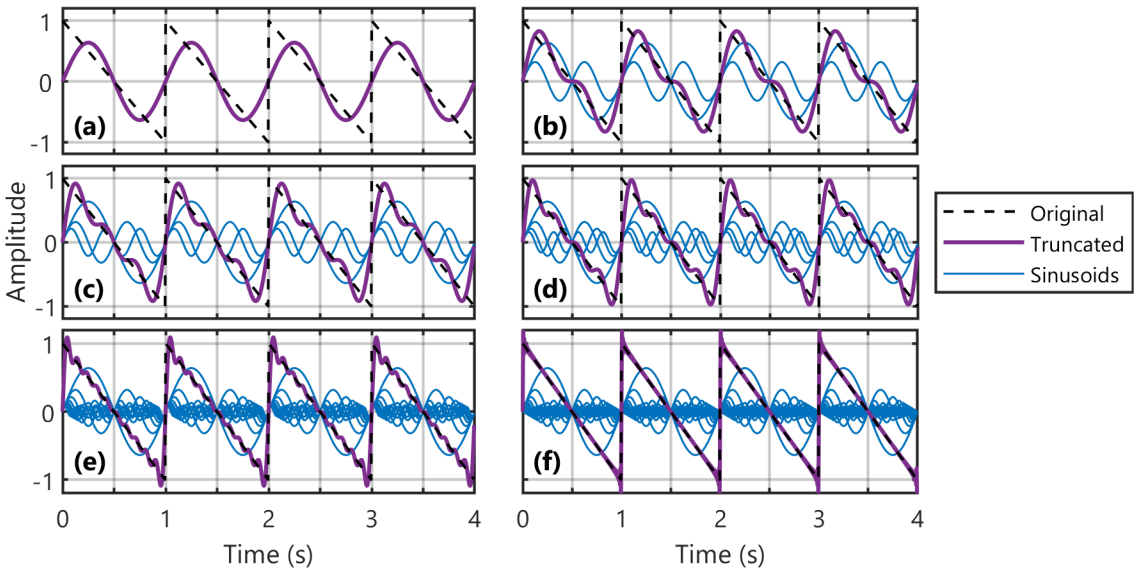


Figure 3-16: A frequency-truncated reconstruction of a 1 second period sawtooth wave, for truncation frequencies of (a) 1, (b) 2 , (c) 3, (d) 4, (e) 10, and (f) 50 Hz.

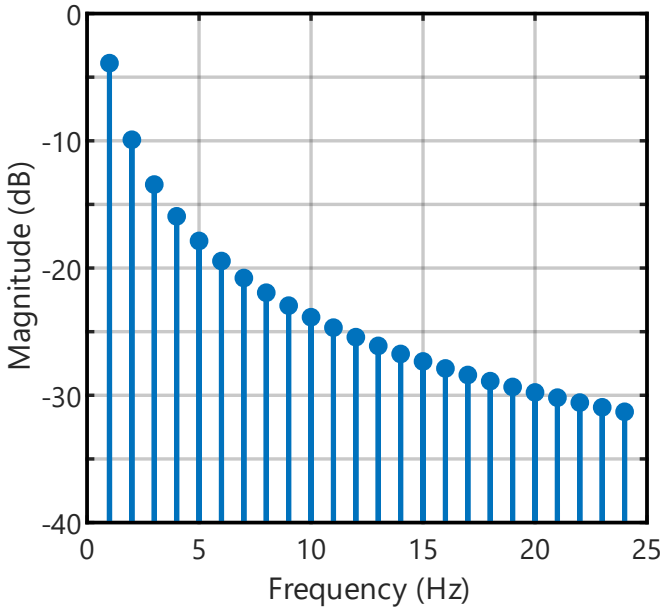


Figure 3-17: The time-frequency Fourier transform of the sawtooth wave, providing frequency weights for the reconstructions performed in Figure 3-16.

This same analogy holds for the spherical Fourier series such that any bounded spatial function, $p(\theta, \phi)$, can be represented as a weighted sum of the SH functions:

$$p(k, \theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} a_{nm}(k)\, Y_n^m(\theta, \phi)\,, \qquad 3.62$$

where the SH weights are determined from:

$$a_{nm}(k) = \int_0^{2\pi} \int_0^{\pi} p(k,\theta,\phi)[Y_n^m(\theta,\phi)]^* \sin\theta \, d\theta d\phi \, . \tag{3.63}$$

Just as summing up the appropriately weighted frequency harmonics reconstructed the original sawtooth wave, summing up appropriately weighted SH functions will result in a reconstruction of a spatial function. Just as a truncated frequency causes reconstruction errors in the time-domain, a truncated-order representation in the SH domain will result in errors in a spatial-domain reconstruction. The basis of spherical array processing, which follows from the spherical Fourier series is the spherical Fourier transform. The spherical Fourier transform takes a discretely sampled function in space and transforms that data from a spatial spherical coordinate system to the SH domain. Just like a time-frequency Fourier transform results in frequency domain weights, the spherical Fourier transform results in SH domain weights. With this transform, SH domain weights to reconstruct any bounded function can be determined.

To continue with a common and useful example, a plane wave can be represented in the SHs domain. A plane wave is defined to be a traveling wave, with only one direction of travel towards the center of a spherical coordinate system. Since the wave only travels in one direction, the ideal spatial representation of the wave over the unit sphere would be a spatial impulse, having a finite value at one point and having a zero value elsewhere. A balloon-style representation of an ideal plane wave is provided in Figure 3-18a. Just as an impulse in the time domain can be reconstructed using an infinite sum of in-phase cosine functions, a plane wave can be reconstructed using a summation of SHs up to an infinite order, n. For complex orthonormal SHs, a plane wave originating from the (θ_l, ϕ_l) direction can be represented as (Eqn. 2.37 in Ref. [81]):

$$p(k,r,\theta,\phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} 4\pi i^n j_n(kr)[Y_n^m(\theta_l,\phi_l)]^* Y_n^m(\theta,\phi) \, , \tag{3.64}$$

and the plane wave coefficients, $p_{nm}(k,r)$, can be defined as

$$p(k,r,\theta,\phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} p_{nm}(k,r) \, Y_n^m(\theta,\phi) \, , \tag{3.65}$$

making (from Eqn. 2.41 in Ref. [81]),

$$p_{nm}(k,r) = 4\pi i^n j_n(kr)[Y_n^m(\theta_l,\phi_l)]^* . \qquad 3.66$$

In practice, it is impossible to sample with infinite SH order, so truncation errors occur. If sufficient SH order resolution is used, these truncation errors can be assumed to be negligible. An order truncated plane wave can be calculated at $r = 0$ by setting the ∞ to a truncation order of N. Figure 3-18 shows a plane wave for truncations orders of $N = 1,\ 3,\ 5,$ and 7.†

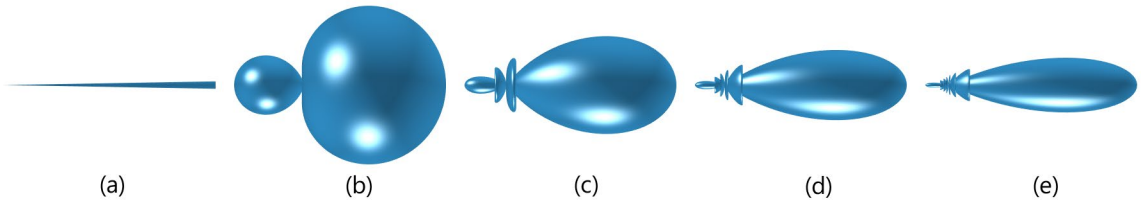


Figure 3-18: An ideal representation of a plane wave is shown in (a), along with the order truncated representations of a plane wave for orders 1, 3, 5, and 7 in (b) – (e), respectively.

An order truncated plane wave is a useful function in both spherical array beamforming and HOA. In beamforming, a plane wave represents the most directional beam pattern that can be achieved by a microphone array. In HOA, if a monaural signal is copied into multiple channels of data, one for each SH component for a given SH truncation order, it can be weighted with coefficients defined by the spherical Fourier transform of a plane wave. Then, if rendered over a HOA auralization system, it creates a plane wave sound source playing the original signal from the desired direction. This source can then be steered or panned flexibly in full 3D, as long as the auralization system supports full 3D coverage.

3.4 Spherical Array Processing

In order to practically implement the SH domain processing, both microphone and compact loudspeaker arrays are required to sample or reconstruct these SH components field or radiation pattern. Just as a microphone discretely samples a time waveform, spherical arrays sample the spatial components of a sound field or sound radiation pattern. This sound field or radiation pattern is a complex superposition of individual SH components. This section will describe the sampling strategies and processing techniques for representing measured

† An animation of a truncated order plane wave generated for presentations by the author can be found online at: https://sites.psu.edu/spral/files/2019/06/Microphone-Plane-Wave.gif.

sound fields using SH components, and reconstructing these sound fields from their SH components.

To capture a three-dimensional sound field, a spherical microphone array samples the spatial variations of the acoustic pressure. Each individual microphone records a discrete time sampling of a pressure, and the overall array of multiple time-synchronized microphones provides a discrete sampling in space. To provide an unbiased estimate, it is important to design a microphone array that evenly samples the spatial field. This criterion is ensured by using geometries from regular polyhedral, such as a tetrahedron, octahedron, dodecahedron, icosahedron, etc. These known platonic solids provide an even coverage around the sphere. Other sampling schemes can be determined using optimization techniques, often involving distributing electrical point source charges over a sphere and minimizing the total potential energy. Quadrature weighting factors can be calculated that correct for an uneven sampling scheme, such as equiangular sampling. This process of sampling is commonly known as encoding and will be further discussed in section 3.4.1.

In the same way a spherical microphone arrays can capture a spatial sampling of a sound field incident upon an array, spherical loudspeaker arrays can reconstruct a measured or simulated sound field that has been represented in the SH domain. The most common example of this is with distributed loudspeakers in a surrounding spherical array around a listener. If the SH components of a sound field are known and represented as SH signals, the loudspeakers can be used to individually sample each SH component at their loudspeaker locations. This process is known as *decoding* in the HOA community and will be discussed more in section 3.4.3. This technique recreates a sound field radiating into the center of an array, but the same techniques hold in the reverse direction.

Sound sources exhibit a radiation pattern or directivity that can be approximated in the SH domain as well. Radiation patterns can be defined as frequency-dependent SH weights. Then, a compact loudspeaker array, made up of individual transducers mounted in a rigid housing, can sample each of the SH components of the radiation patterns at the individual driver locations. The superposition of each driver signal sampling each SH component will reconstruct the radiation pattern of the desired source. In the next few sections, the encoding, equalization, and decoding steps will be described, but note that the same techniques hold for the reciprocal cases of microphone and loudspeaker arrays. First, a compact microphone array can measure an incident sound field, and a distributed surrounding loudspeaker array can reconstruct that sound field. The reciprocal case that holds is a surrounding, distributed

microphone array can sample the radiation pattern of a source, and a compact spherical loudspeaker array (CSLA) can reconstruct a source with the same radiation pattern.

3.4.1 Encoding into Spherical Harmonic Functions

The first step involves the capture of a sound field. This step can either be done for an incident sound field using a spherical microphone array or for a sound field radiation from a source using a surrounding microphone array. In either case, each microphone provides an individual directional sample at each microphone location, (ϕ_q, θ_q). A time-frequency Fourier transform is used to generate complex pressures in individual frequency (or wavenumber, k) bins for each microphone. This problem discretizes the spherical Fourier series from Eqn. 3.62 as:

$$p(k, r, \theta, \phi) \approx \sum_{n=0}^{N} \sum_{m=-n}^{n} \tilde{a}_{nm}(k)\, Y_n^m(\phi_q, \theta_q), \tag{3.67}$$

where the q^{th} microphone of Q total microphones samples each SH component of the acoustic field. It is noticed that now, a truncation order N is introduced into the equation, as it is practically impossible to reconstruct a sound field with an infinite order of SH components. This equation can be extended to represent all of the Q total microphones as:

$$\boldsymbol{P} = \boldsymbol{Y_Q}\widetilde{\boldsymbol{A}}\,, \tag{3.68}$$

or written more specifically as:

$$\begin{bmatrix} p(\phi_1, \theta_1, k) \\ p(\phi_1, \theta_1, k) \\ \vdots \\ p(\phi_Q, \theta_Q, k) \end{bmatrix} = \begin{bmatrix} Y_0^0(\phi_1, \theta_1) & Y_1^{-1}(\phi_1, \theta_1) & Y_1^0(\phi_1, \theta_1) & \dots & Y_N^N(\phi_1, \theta_1) \\ Y_0^0(\phi_2, \theta_2) & Y_1^{-1}(\phi_2, \theta_2) & Y_1^0(\phi_2, \theta_2) & \dots & Y_N^N(\phi_2, \theta_2) \\ & \vdots & & \ddots & \vdots \\ Y_0^0(\phi_Q, \theta_Q) & Y_1^{-1}(\phi_Q, \theta_Q) & Y_1^0(\phi_Q, \theta_Q) & \dots & Y_N^N(\phi_Q, \theta_Q) \end{bmatrix} \begin{bmatrix} \tilde{a}_{0,0}(k) \\ \tilde{a}_{1,-1}(k) \\ \tilde{a}_{1,0}(k) \\ \vdots \\ \tilde{a}_{N,N}(k) \end{bmatrix}. \tag{3.69}$$

$\boldsymbol{P}$ is a $Q\ x\ N_{samps}$ matrix, $\boldsymbol{Y_Q}$ is a $Q\ x\ (N+1)^2$ matrix, and $\widetilde{\boldsymbol{A}}$ is a $(N+1)^2\ x\ N_{samps}$ matrix. The matrix $\boldsymbol{P}$ represents the pressure field measured by each microphone, and the term N_{samps} is the number of frequency (wavenumber) bins resulting from the time-frequency Fourier transform. The matrix $\boldsymbol{Y_Q}$ represents each SH function evaluated at the direction of each microphone, and the matrix $\widetilde{\boldsymbol{A}}$ represents the spherical Fourier weights of the measured sound field. Currently, a tilde is introduced as notation to indicate that the weights are unequalized for the design and geometry of the microphone array. More regarding this equalization will be discussed in section 3.4.2. The goal of *encoding* a sound field is to determine the SH weights from the microphone signals. To do this operation, Eqn. 3.68 is solved for the SH weights matrix in terms of the measured pressure signals using a pseudoinverse:

$$\widetilde{\boldsymbol{A}} = \boldsymbol{Y}_{\boldsymbol{Q}}^{\dagger}\boldsymbol{P} = \left(\boldsymbol{Y}_{\boldsymbol{Q}}^{\boldsymbol{T}}\boldsymbol{Y}_{\boldsymbol{Q}}\right)^{-1}\boldsymbol{Y}_{\boldsymbol{Q}}^{\boldsymbol{T}}\boldsymbol{P}\,. \tag{3.70}$$

This step performs a least-square fit on the measured pressure field to calculate the SH weights. A matrix pseudoinverse is used because the matrix $\boldsymbol{Y}_{\boldsymbol{Q}}$ is not a square matrix, and therefore, not directly invertible. Practically, this step is done using either a Moor-Penrose pseudoinverse or singular value decomposition (SVD). Looking back at this operation, the solutions can only be determined if the number of microphones is greater than or equal to the number of SH components for a given truncation order, N:

$$Q \geq (N+1)^2\,. \tag{3.71}$$

For example, a microphone array of 20 elements can only reproduce up to third-order SH functions. Otherwise, the problem represents an underdetermined system of equations, and it is therefore not possible to solve this system of matrix equations. If this criterion is met, it is possible to estimate up to N^{th} order SH components of a sound field. The matrix $\boldsymbol{Y}_{\boldsymbol{Q}}^{\dagger}$ from Eqn. 3.70 is known as the *encoder matrix*, and this matrix performs the discrete spherical Fourier transform, to represent a measured pressure field as SH weights from the individual microphone signal samples, all performed in the frequency or wavenumber domain. This same transformation is possible for a distributed microphone array along with a compact microphone array.

3.4.2 Array Design Equalization

When using a spherical microphone array to measure the sound field incident upon a point in space, it is also important to correct for effects due to the design of the microphone array. The goal of a spherical microphone array is to measure the spatial sound field at the point located at the center of the array, as if the microphone was not present in the field. Two common design typologies exist. The first typology is an open microphone array. First, consider a plane wave represented in spherical coordinates. Section 3.3 shows the SH representation of a plane wave, but often, sound fields are more complex than containing a single plane wave. This expression can be extended to any sound field, represented as a superposition of many plane waves, such that:

$$p(k,r,\theta,\phi) \approx \sum_{n=0}^{N}\sum_{m=-n}^{n} 4\pi i^n a_{nm}(k) j_n(kr) Y_n^m(\theta,\phi), \tag{3.72}$$

Where $a_{nm}(k)$ is the spherical Fourier transform of a function that defines the directional amplitude density of the continuum of plane waves, $a(k,\theta_a,\phi_a)$. The benefit of this technique

is that the sound field's spherical Fourier coefficients, $a_{nm}(k)$, can be estimated by a measurement at a single radial location, $r = r_a$. Once determined at this location, assuming plane wave behavior, the sound field can be extrapolated to any point in space. This general solution can also be reduced back to the case of a plane wave by setting the amplitude density function to be equal so a spatial Kronecker delta,

$$a(k, \theta, \phi) = \delta(\cos\theta - \cos\theta_a)\delta(\phi - \phi_a), \tag{3.73}$$

so that,

$$a_{nm}(k) = [Y_n^m(\theta, \phi)]^*, \tag{3.74}$$

If a case where an open microphone array is used to sample the sound field, it is assumed that microphone array elements are sufficiently sub-wavelength and do not disturb the sound field due to scattering. It is desired to estimate the spherical Fourier coefficients, $a_{nm}(k)$, but what is measured are these coefficients, plus an added term due to the size of the array, $b_n(kr)$:

$$\tilde{a}_{nm}(k) = a_{nm}(k)b_n(kr)\,. \tag{3.75}$$

The term $\tilde{a}_{nm}(k)$ is taken as the unequalized Fourier coefficients, determined from a direct spherical Fourier transform (encoding) of the measured microphone signals. Comparing Eqn. 3.75 to Eqn. 3.72, and evaluating at the microphone array's radius, $r = r_a$, the added effect of the sampling with an open spherical microphone array is characterized by the factor:

$$b_n(kr_a) = 4\pi i^n j_n(kr_a). \tag{3.76}$$

In order to calculate the desired spherical Fourier transform coefficients, the measured coefficients must be *equalized* using a multiplication by $1/b_n(kr_a)$. To correct for this, a filter can be designed that has a magnitude and phase response matching this factor. For the open sphere, some problems emerge when inspecting the magnitude response of this factor, shown in Figure 3-19. Large dips and nulls occur in the array's response to a plane wave. When using measured signals, large boosting requirements also cause signal-to-noise ratio (SNR) problems in boosting noise contained in a measurement. In addition, it is difficult to design stable, efficient filters that could provide the necessary boosting required by this factor.

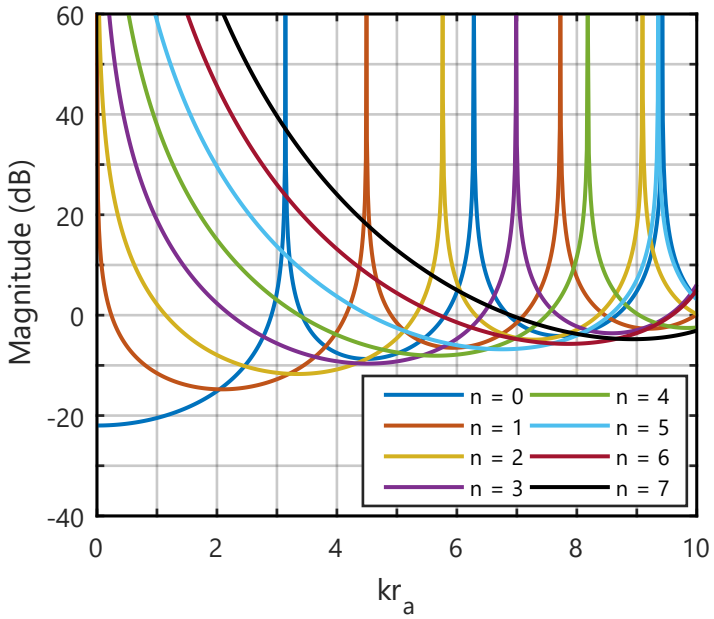


Figure 3-19: Magnitude of the equalization factor, $1/b_n(kr_a)$, for a spherical microphone array designed using a open configuration, shown in terms of the product of array radius and wavenumber.

To avoid this issue, a combination of two open arrays can be used in a dual-concentric array design, with two different radii, so that the summation of both signals, with different r_a values, will remove the nulls from these responses. Another trick comes from using cardioid directivity pattern transducers, which can be related to the SH coefficients as well but remove the nulls in an arrays response. This technique has even been shown to work at higher orders of SH processing, but is most commonly used in first-order B-format microphones. The most common solution is to mount the microphones within a rigid sphere. This enclosure allows for hardware to be concealed within the array but maintains a consistent geometry with a known equalization term. For this array, the incident sound field composed of plane waves, p_i, is still given in a similar fashion to Eqn. 3.72 as:

$$p_i(k,r,\theta,\phi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} 4\pi i^n a_{nm}(k) j_n(kr) Y_n^m(\theta,\phi). \qquad 3.77$$

This is the component of the sound field that is desired to be measured, containing the spherical Fourier weights $a_{nm}(k)$. This pressure field is not the measured sound field, as the presence of the rigid sphere creates a boundary condition at r_a and a scattered pressure field. The scattered pressure field can be written as a summation of SHs such that:

$$p_s(k,r,\theta,\phi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} c_{nm}(k) h_n^{(2)}(kr) Y_n^m(\theta,\phi)\,. \qquad 3.78$$

Here, spherical Hankel functions of the second kind are used, as the scattered pressure term is composed of energy radiating outwardly from the sphere. Due to the presence of the rigid sphere, the radial component of the total particle velocity in the sound field must equal zero at the surface of the array, $r = r_a$.

$$u_r(k, r_a, \theta, \phi) = 0\,. \tag{3.79}$$

The particle velocity can be equated to the pressure using Euler's equation (equation of conservation of momentum) as:

$$i\rho_o ck u(k, r, \theta, \phi) = \nabla p(k, r, \theta, \phi)\,. \tag{3.80}$$

The radial solution of particle velocity from the Laplacian in spherical coordinates from Eqn. 3.5 and the radial component of the particle velocity can be isolated, making:

$$i\rho_o ck u_r(k, r_a, \theta, \phi) = \frac{\partial}{\partial r}[p(k, r_a, \theta, \phi)\,]. \tag{3.81}$$

also computing the total pressure as the sum of the incident pressure and scattered pressures from Eqns. 3.77 and 3.78, and applying the boundary condition from Eqn. 3.79 at $r = r_a$, the expression becomes:

$$0 = \frac{\partial}{\partial r}[p_i(k, r, \theta, \phi) + p_s(k, r, \theta, \phi)]\Big|_{r=r_a}\,. \tag{3.82}$$

After carrying out the derivative in terms of r, the SH coefficients, $c_{nm}(k)$, can be solved for in terms of the coefficients of the incident sound field, $a_{nm}(k)$, such that:

$$c_{nm}(k) = -a_{nm}(k) 4\pi i^n \frac{j_n'(kr_a)}{h_n^{(2)'}(kr_a)}\,. \tag{3.83}$$

Now, the total measured pressure, p_t, at any point outside of the rigid sphere can be represented by summing the incident and scattered pressures from Eqns. 3.77 and 3.78 and substituting in for the SH coefficients of the scattered pressure field from Eqn. 3.83:

$$p_t(k, r, \theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} a_{nm}(k) 4\pi i_n \left[j_n(kr) - \frac{j_n'(kr_a)}{h_n^{(2)'}(kr_a)} h_n^{(2)}(kr) \right] Y_n^m(\theta, \phi)\,. \tag{3.84}$$

Taking the pressure now evaluated at the surface of the microphone array, $r = r_a$, a similar term representing the impact of the array design on the measured sound field can be defined from Eqn. 3.75 as:

$$b_n(kr_a) = 4\pi i^n \left[j_n(kr_a) - \frac{j_n'(kr_a)}{h_n^{(2)'}(kr_a)} h_n^{(2)}(kr_a) \right]. \qquad 3.85$$

Just as in the open microphone array case, in order to compensate for these effects of the microphone array, a multiplication by the factor $1/b_n(kr_a)$ is required. This equalization factor is shown in Figure 3-20. Comparing this response to the response for the open array design from Figure 3-19, a few key differences are noticed. First, the rigid boundary condition does not contain any nulls, which were found in the response of the open array design. Without these dips, the equalization for the rigid sphere is much more robust to measurement noise, and much more practical in terms of designing stable equalization filters. Additionally, there is a benefit seen at the boosting required for low frequencies. The case of the rigid sphere required less boosting for the low frequency roll-off, meaning the array will not reach the noise floor of a measurement until a lower frequency. Both plots have been overlaid in Figure 3-21 for close comparison.

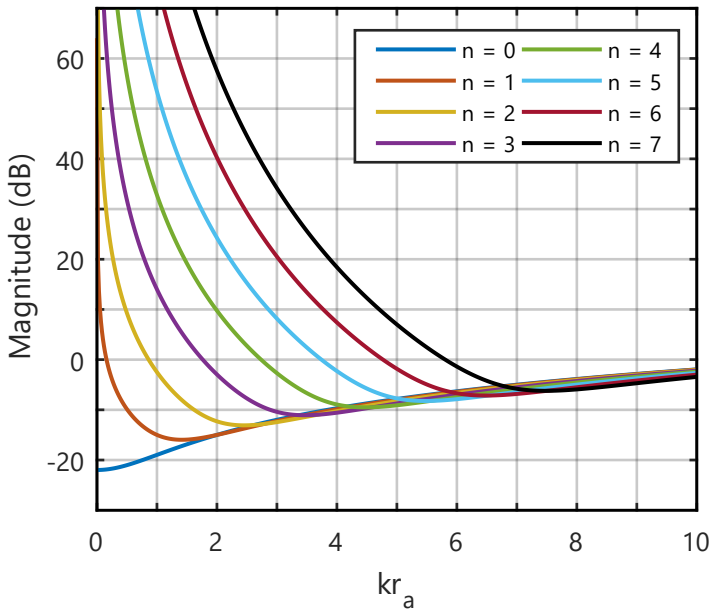


Figure 3-20: The equalization factor, $1/b_n(kr_a)$, for a spherical microphone array designed using a rigid configuration. Shown in terms of the product of array radius and wavenumber.

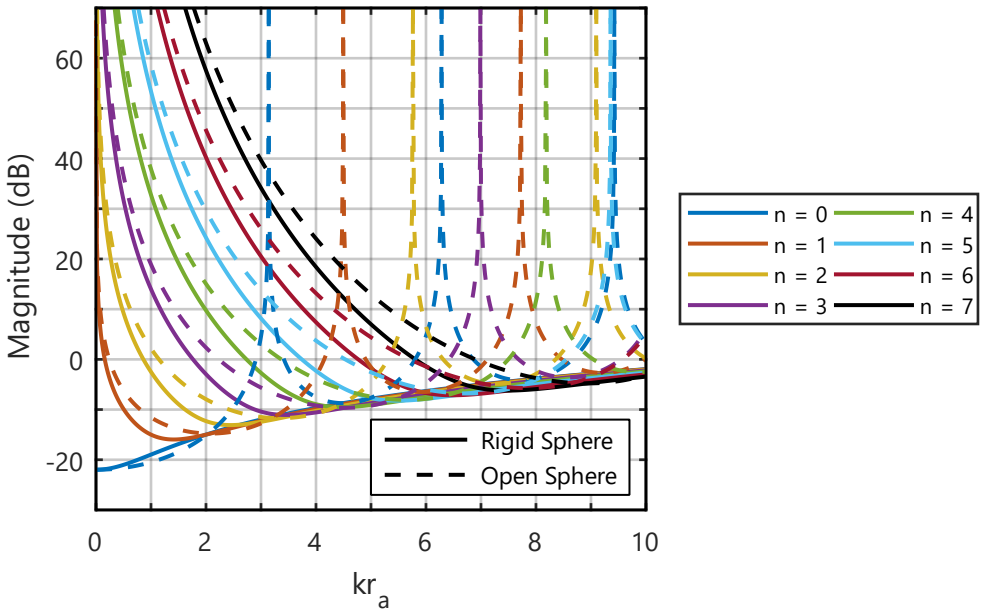


Figure 3-21: The equalization factor, $1/b_n(kr_a)$ for both rigid and open microphone arrays, overlaid.

A last, and very important point addresses the mathematical simplicity behind this array equalization technique. For open and rigid spheres, the equalization factor, $1/b_n(kr_a)$, depends only upon the array size, frequency (given as wavenumber), and SH order. Due to the problem's symmetry and the ability to represent a sound field as a summation of plane waves, the equalization is completely independent of direction arrival of individual plane waves. Once array parameters are known, the processing can be seamlessly integrated into the signal chain, applied a single signal for each SH order. After applying this correction, the proper spherical Fourier weights are known for a point at the center of the microphone array, as if the array was not present in the field. In the literature, this step is coined with the term *radial filtering*, correcting for the impact of a microphone array's boundary conditions and size on a measurement.

3.4.3 Decoding into Loudspeaker Signals

Once a sound field is represented in the SH domain, it is often desirable to reconstruct this measured sound field. This technique is most commonly used in the field of HOA, where the goal is to use a distributed array of loudspeakers to sample the SH components of a sound field to provide a reconstruction of the original acoustic field. This summation can be represented mathematically as:

$$a_{nm}(k) \approx \sum_{l=1}^{L} g_l(k) Y_n^m(\phi_l, \theta_l)\,, \qquad 3.86$$

where g_l is the signal of the l^{th} loudspeaker, as a function of frequency (wavenumber), the quantity (ϕ_l, θ_l) is the direction of the l^{th} loudspeaker in the distributed auralization loudspeaker array, and $a_{nm}(k)$ are the SH weights that are being reconstructed. By substituting this reconstructed pressure solution into Eqn. 3.62, it is shown that:

$$p(k,\theta,\phi) \approx \sum_{n=0}^{N} \sum_{m=-n}^{n} \left[\sum_{l=1}^{L} g_l(k) Y_n^m(\phi_l, \theta_l) \right] Y_n^m(\theta, \phi)\,. \tag{3.87}$$

Here, the loudspeaker array produces an order-truncated representation of the measured sound field. The use of the approximate equality is due to both errors from the order truncation and errors due to the design of the loudspeaker array. In effect, the loudspeaker array samples each SH component independently, and the superimposed rendering of each component, across all frequencies, creates a full reconstruction of the wave field. Equation. 3.86 can also be represented in matrix notation as:

$$\boldsymbol{A} = \boldsymbol{Y_L G}\,, \tag{3.88}$$

or more completely as,

$$\begin{bmatrix} a_{0,0}(k) \\ a_{1,-1}(k) \\ a_{1,0}(k) \\ \vdots \\ a_{N,N}(k) \end{bmatrix} = \begin{bmatrix} Y_0^0(\phi_1,\theta_1) & Y_0^0(\phi_2,\theta_2) & \cdots & Y_0^0(\phi_L,\theta_L) \\ Y_{-1}^1(\phi_1,\theta_1) & Y_{-1}^1(\phi_2,\theta_2) & \cdots & Y_{-1}^1(\phi_L,\theta_L) \\ Y_0^1(\phi_1,\theta_1) & Y_0^1(\phi_2,\theta_2) & \cdots & Y_0^1(\phi_L,\theta_L) \\ \vdots & \vdots & \ddots & \vdots \\ Y_N^N(\phi_1,\theta_1) & Y_N^N(\phi_2,\theta_2) & \cdots & Y_N^N(\phi_L,\theta_L) \end{bmatrix} \begin{bmatrix} g_1(k) \\ g_2(k) \\ g_3(k) \\ \vdots \\ g_L(k) \end{bmatrix}. \tag{3.89}$$

$\boldsymbol{A}$ is a $(N+1)^2 \times N_{samps}$ matrix, $\boldsymbol{Y_L}$ is a $(N+1)^2 \times L$ matrix, and $\boldsymbol{G}$ is a $L \times N_{samps}$ matrix. The matrix $\boldsymbol{A}$ represents the spherical Fourier transform coefficients, and the term N_{samps} is the number of frequency or wavenumber bins from the time-frequency FFT. The matrix $\boldsymbol{Y_L}$ represents each SH function up to a given truncation order, N, evaluated at the angular location of each loudspeaker in the auralization array. Finally, $\boldsymbol{G}$ is a matrix of the loudspeaker signals for reconstruction of the sound field. In order to solve for the loudspeaker signals, $\boldsymbol{G}$, in terms of both $\boldsymbol{A}$ and $\boldsymbol{Y_L}$, another matrix pseudoinverse, $\boldsymbol{Y_L^+}$, is required:

$$\boldsymbol{G} = \boldsymbol{Y_L^+ A} = \left(\boldsymbol{Y_L^T Y_L}\right)^{-1} \boldsymbol{Y_L^T A}\,. \tag{3.90}$$

This process is known in the HOA community as *decoding*, going from the SH domain into loudspeaker driving signals. The matrix $\boldsymbol{Y_L^+}$ is commonly referred to as the *basic decoder matrix* or mode-matching decoder matrix, performing a least-squares fit between the

loudspeaker signals and the spherical Fourier transform weights. A direct comparison can be made between the encoding process from Eqns. 3.68 to 3.70 and the decoding process from Eqns. 3.88 to 3.90. Both processes are identical in nature, with only subtle differences in application. Finally, just as was found with encoding from microphone signals, in order to solve this matrix equation and prevent it from being underdetermined, an auralization array must have at least the same number of loudspeakers as SH components, such that:

$$L \geq (N+1)^2 \,. \tag{3.91}$$

In designing a HOA auralization array, it is important to sample the sound field around a listener as evenly as possible. As an extreme example, if an array met the criterion from Eqn. 3.91, but only placed loudspeakers to the left side of a listener, this array would not provide good reproduction of the acoustic field. The matrix inversion for the decoder matrix calculation would be mathematically possible, but the results would be poor. Not only is a sufficient number of loudspeakers needed, but a sufficiently even distribution of the loudspeakers is needed. Ambisonics does allow for flexibility, not requiring a perfectly uniform spacing of loudspeakers, but any arrangement should attempt to sample all parts of the sphere around a listener's head as evenly as possible. It is a common misconception in Ambisonics that loudspeaker placement is entirely flexible, with no impact upon accuracy. Finally, sections 3.4.3.1 through 3.4.3.3 will discuss additional considerations when designing a HOA decoder matrix.

Just as this technique works for a surrounding auralization array, it also applies to the reverse case of a CSLA, reproducing a radiation pattern composed of the spherical Fourier transform weights, $a_{nm}(k)$. The basic format of decoding equations and recommendations all apply just the same as they do for a surrounding array recreating a measured sound field at a sweet spot. It is important to note that since this is effectively a reciprocal case, the same radial filtering before the decoding stage in the rigid spherical microphone array processing must also be applied before the decoding stage of the CSLA processing. This step can be thought of as a pre-filtering, to account for the effects of the rigid boundary condition and size of the CSLA which occurs during playback. This processing will be further discussed in Chapter 4.

3.4.3.1 Dual-band decoding for improved localization

In the human auditory system, binaural hearing cues help a listener determine the direction from which a sound originates. At lower frequencies, the hearing system's primary cue is the time difference between a signal arriving at the left and right ears, termed the interaural time difference (ITD). As frequency increases, and wavelength decreases, the

human head becomes around the same size, or much larger than the wavelength of sound. Because of this change, high amounts of masking from wave fronts travelling across the head creates large amplitude differences that were not present at lower frequencies. The hearing system relies on this interaural level difference (ILD) cue for high frequency localization.

When designing a HOA loudspeaker array, the decoder attempts to reconstruct a plane wave at the center of the array. Ideally, this plane wave arrives from one direction, but as can be seen in (a) – (e) in Figure 3-22, the artifacts from calculating an order-truncated plane wave create side lobes. When a basic decoder matrix is used, the loudspeakers sample this plane wave shape at their locations, and if positioned in a side lobe location, a loudspeaker will sample this artifact. This artifact is not related to the ideal, infinite order plane wave and is merely a result of the order truncation and subsequent sampling. This artifact has been shown to not impact perception at low frequencies, where the ITD cue dictates localization, but it can degrade high frequency localization. A loudspeaker sampling a high-amplitude side lobe, angularly far from the main plane wave radiation direction, can alter the desired ILD cue.

Gerzon defined two quantities, r_V and r_E, that represent the quality of localization at low and high frequencies.[82] The r_V vector corresponds to the exact reconstruction of the acoustic pressure and particle velocity at the center of the array and corresponds with the quality of low frequency localization. This criterion of reconstructing the measured pressure field at the center of the array is satisfied typically at lower frequencies, up to around 350 or 400 Hz, where the auditory system relies on the low frequency localization cues. As frequency increases, the *sweet spot* of accurate reproduction also decreases in size. Typically, the sweet spot is still much larger than a human head at these frequencies, so a psychoacoustically plausible environment is still ensured. Above this range, Gerzon developed a series of weights that can be applied to provide a better high frequency localization, which were mathematically derived by Daniel,[83] and have been computed and provided by Heller et al. and are provided in Table 3.2.[84]

Table 3.2: Max-Re weighting factors as per-order gains for periphonic regular polyhedral arrays (recreated from Ref: [84]).

SH Order	Max r_E Gain	Order gains per truncation order
1	0.57735	1, 0.57735
2	0.774597	1, 0.774597, 0.4
3	0.861136	1, 0.861136, 0.612334, 0.304747
4	0.90618	1, 0.90618, 0.731743, 0.501031, 0.245735
5	0.93247	1, 0.93247, 0.804249, 0.62825, 0.422005, 0.205712

To demonstrate the effect of these weights, an ideal order truncated plane wave for each order is presented in Figure 3-22. Comparing the ideal plane waves to the max-r_E plane waves, the new order-dependent weightings help to minimize the side-lobe amplitudes relative to the main lobe amplitude. Thus, the high frequency ILD locations cue is better preserved. As a tradeoff, it is seen that the main beam does become wider, to accommodate the reduction in side lobe amplitude. A similar phenomena in spherical microphone array beamforming regarding Dolph-Chebyshev beam shapes will be presented in section 5.9.4.2.

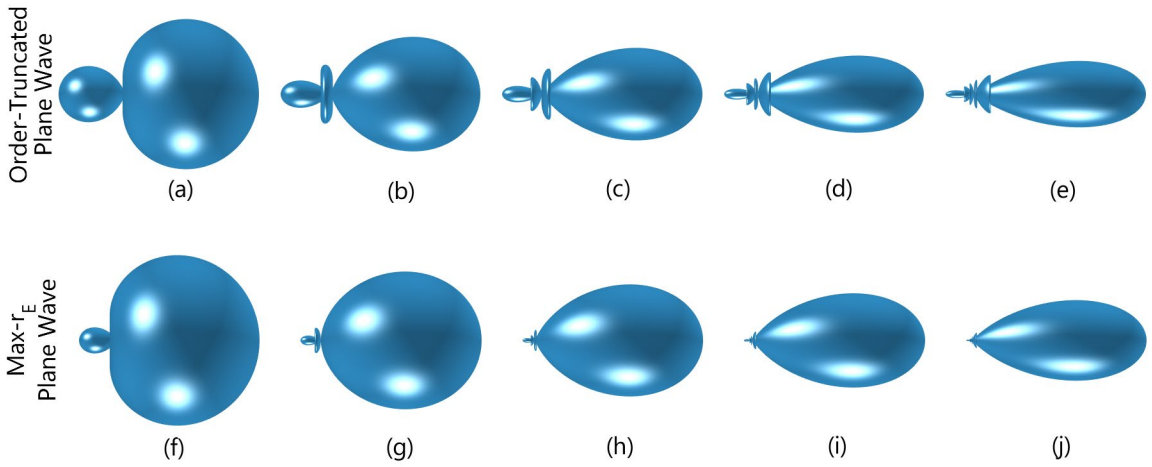


Figure 3-22: Order-truncated plane waves for orders 1 – 5 are shown in (a) – (e), respectively. The corresponding max-r_E plane waves are given for orders 1 – 5 as well in (f) – (j), respectively.

3.4.3.2 Near-field compensation

Another assumption in the HOA processing mathematics is that loudspeakers used in a HOA auralization array are in the far field, acting as plane waves. In acoustics, plane wave-like behavior can be assumed from a simple spherical source when the product of wavenumber and distance from a source is much greater than one, $kr \gg 1$. These locations are in the far field, and the energy received from the source is entirely resistive, only having energy associated with outward propagation from the source. As you move into the near field of the source, when $kr \leq 1$, the near field of a source also contains a reactive component, associated with energy that is being stored locally around a source. This energy is sometimes colloquially associated with energy *sloshing* back and forth around a source, thus being locally stored in the acoustic medium. Since loudspeakers are at a fixed distance from the center of the array, the near field becomes prominent as frequency decreases. The reactive component of the sound field creates a boosting phenomenon, over-emphasizing low frequency energy.

To correct for this near-field effect, first, an ideal plane wave (far-field) source is represented in the SH domain, p_{PW}:

$$p_{PW}(k,r,\theta,\phi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} 4\pi i^n j_n(kr)[Y_n^m(\theta_l,\phi_l)]^* Y_n^m(\theta,\phi)\,. \tag{3.92}$$

If a source is in the near field, it cannot be assumed to be plane wave-like, and it must be represented as a point source. Just as plane waves can be represented in the SH domain, point sources, p_{PS}, at a specific radius from the origin, r_s, can be represented in the SH domain as:

$$p_{PS}(k,r,\theta,\phi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} 4\pi(-i)kh_n^{(2)}(kr_s)j_n(kr)[Y_n^m(\theta_l,\phi_l)]^* Y_n^m(\theta,\phi)\,. \tag{3.93}$$

The representation in Eqn. 3.93 is for a point source at a radius of r_s measured at a location r such that $r < r_s$ (See, for example, Eqn. 2.49 in Ref. [81]). In other words, the point source must be external to the location of the measurement sphere at r. To correct for this effect, a near-field correction factor can be defined by dividing the desired response of a plane wave, p_{PW}, by the actual response exhibited of a point source, p_{PS}. This correction factor is what is applied in a process proposed by Daniel commonly referred to as *near-field compensation* (NFC).[85] After some algebraic reduction, the near-field correction factor, $NFC(k,r_s,n)$ can be defined for each SH order as:

$$NFC(k,r_s,n) = \frac{-i^{(n-1)}}{kh_n^{(2)}(kr_s)}. \tag{3.94}$$

To demonstrate this effect, the magnitude of both the near-field strength of a point source and far-field strength of a point source (assumed to act like a plane wave) are shown in Figure 3-23 for a source located at $r_s = 1$ m. The factor derived to correct for the non-far field behavior of a near field source from Eqn. 3.94 is shown in Figure 3-24. As can be seen, this correction factor is essential to produce the correct low frequency level balance in reproducing a sound field accurately when $kr_s < 10$. To provide a more physically useful description, clear deviations between far field and near field behavior start around $kr_s = 10$, and greatly increase for $kr_s = 1$. For a HOA auralization rig with radial loudspeaker spacing of $r_s = 1$ m in air ($c = 343$ m/s), $kr_s = 10$ occurs at 545 Hz, and $kr_s = 1$ occurs at 55 Hz.

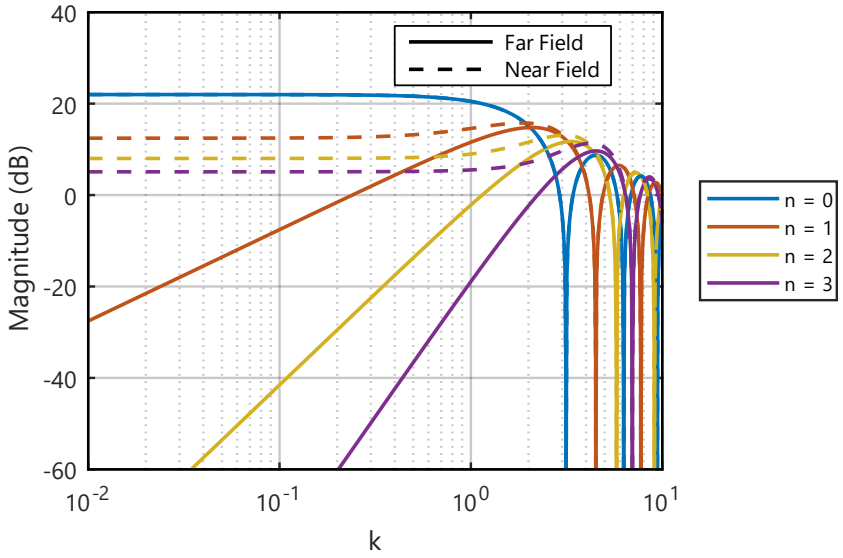


Figure 3-23: A comparison of a near field source at $r_s = 1$ m and a far field source (plane wave).

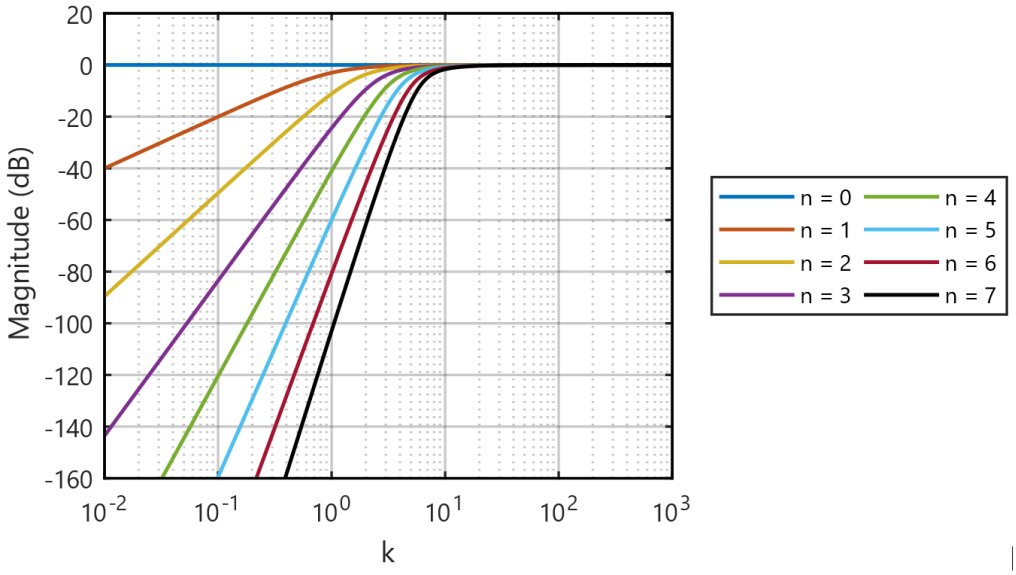


Figure 3-24: The near-field corrections factor derived from the division of a far field source (plane wave) over a near-field source at $r_s = 1$ m. Clear effects begin when $kr_s \leq 10$.

3.4.3.3 Sweet-spot size in terms of spherical harmonic order and frequency

As the SH truncation order increases, not only does the accuracy of the plane wave increase in directional response, but the extent of the spatial region of accurate reproduction also increases. This phenomenon can be visualized by looking at order-truncated plane waves at different frequencies over a Cartesian grid since the representation of an incident order-truncated plane wave also includes a radial term dependent upon kr. The sound field for an order-truncated plane wave is given as a finite summations of SH functions:

$$p(k,r,\theta,\phi) \approx \sum_{n=0}^{N} \sum_{m=-n}^{n} p_{nm}(k,r) Y_n^m(\theta,\phi)\,, \tag{3.95}$$

where,

$$p_{nm}(k,r) = 4\pi i^n j_n(kr)[Y_n^m(\theta_l,\phi_l)]^*\,. \tag{3.96}$$

To show the practical limitations of SH representation in the context of HOA using the speed of sound in air, $c = 343$ m/s, and assuming a plane wave traveling from 90 degrees elevation and 20 degrees azimuth, the real part of the pressure field for a plane wave at 500, 1000, 2000, and 4000 Hz has been generated for four different truncation orders of $N = 1,\ 3,\ 5$, and 7 in Figure 3-25. In general, where $kr < N$, accurate reproduction of the plane wave exists, and when $kr > N$, accuracy of the plane wave is degraded. In Figure 3-25, a small black ellipse is plotted, corresponding to an average head width of 14.8 cm and breadth of 17.7 cm. Also shown in Figure 3-25 is a dashed white circle defining the boundary of accurate reproduction where $kr = N$ or $r = N/k$, calculated from the frequency and truncation order.

For first-order Ambisonics, no true sweet spot exists, and the extent of a plane wave traveling through space is not reproduced. The jump to third-order representation creates considerable advantages, having a sweet spot much larger than the head at 500 and 1000 Hz, somewhat on the same order as the head around 2000 Hz, and smaller than the head after 4000 Hz. Fifth-order Ambisonics has a sweet spot larger than the head, which shrinks to roughly the size of the head around 4000 Hz, while seventh order appears to maintain accurate reproduction in a sweep spot that could fit an average human head, even up to 4000 Hz. Effectively, as frequency is doubled, the SH truncation order must also be double to maintain the same size of sweet spot. If a minimum sweet spot size of $r = 8$ cm is defined to maintain accurate reproduction larger than the human head, a cutoff frequency can be calculated. This frequency has been calculated for various orders and is shown in Table 3.3.

Table 3.3: Calculated cutoff frequencies to ensure that the sweep spot has a minimum radius of 8 cm, as defined by the criterion $N = kr$ for truncation orders one through nine.

Cutoff	Spherical Harmonic Truncation Order (N)								
	1	2	3	4	5	6	7	8	9
Frequency (Hz):	682	1365	2047	2730	3412	4094	4777	5459	6141

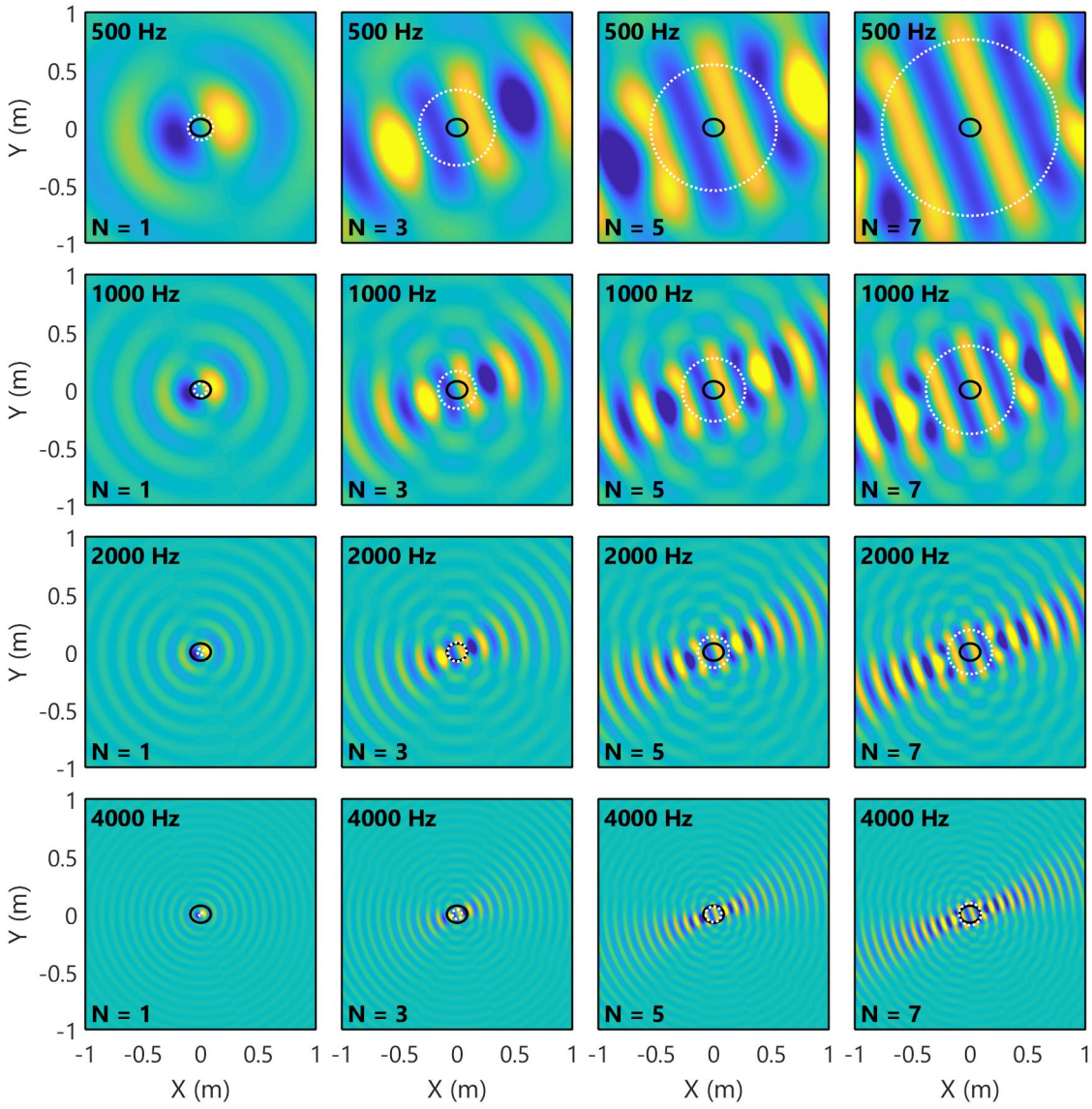


Figure 3-25: Spatial calculations of order truncated plane waves for SH truncation orders of 1, 3, 5, and 7 (left to right) at 500, 1000, 2000, and 4000 Hz (top to bottom) in air. The black ellipse is displayed as an average human head, and the white dashed circle represents the *sweet spot*, where $r = N/k$.

3.5 Application to the Present Work

In the current dissertation, these topics of spherical microphone array processing, spherical array beamforming, and virtual acoustic reconstructions using HOA are implemented into a measurement database of concert hall acoustics measurements. Chapter 4 presents the spherical array processing techniques from the source side using a CSLA This array allows for arbitrary control of the directional radiation pattern of the measurement source, matching it to realistic instrumental sources found in an orchestra. This application of spherical array techniques provides realistic measurements as a basis for auralization and subjective testing. Measurements were captured using a 32-channel spherical microphone array, so that measurements could be auralized over a surrounding loudspeaker array using HOA. The spherical microphone array processing techniques used to generate a measurement database of 21 US and European concert halls are presented in Chapter 5. Along with HOA techniques, the use of a spherical microphone array allows for spatial analysis of the sound field using plane wave decomposition (PWD). Finally, the combination of objective spherical array beamforming analysis and subjectively realistic auralizations using spherical array processing techniques are integrated into a subjective study regarding individual preference in Chapter 6.

Chapter 4
A Compact Spherical Loudspeaker Array

This chapter contains information written for external publication, which is intended for submission to a major peer-reviewed acoustics or audio engineering-related journal. As such, this chapter also contains its own introduction, background, and results section. The overall introduction (Ch. 1), background (Ch. 2 and 3), and results (Ch. 7) chapters of the dissertation fill in the greater picture of the entire dissertation work.

This paper provides the details on the design and construction of a compact spherical loudspeaker array used to accurately represent the directivity of different musical instruments. As well, the design of filters that control the radiation pattern produced by the sound source is demonstrated. This loudspeaker array and processing techniques are used in the concert hall measurement database that is further described in chapter 5. The directional accuracy of this array, along with its radiation pattern flexibility, allow for measurements than can realistically represent an orchestra performing in a particular concert hall.

A Compact Spherical Loudspeaker Array for Generating Full-orchestral Concert Hall Auralizations

Matthew Neal and Michelle Vigeant
Graduate Program in Acoustics, The Pennsylvania State University, University Park, PA 16802

Abstract:

High levels of realism can be achieved when using measurement-based auralizations of concert halls, but typical measurement loudspeakers limit source auralization realism. A compact spherical loudspeaker array (CSLA) was designed to reconstruct the frequency-dependent radiation patterns of different orchestral instruments. A set of filters for each instrument were designed to pre-process measurement signals with built-in source radiation during RIR measurement

4.1 Introduction

When taking RIR measurements in a concert hall, dodecahedral loudspeakers are typically used as standardized measurement sources to provide consistency and repeatability across group. The resulting RIR can be used to calculate common room acoustic metrics, and if desired, to generate auralizations from anechoic music. Although standard, this type of measurement source is far removed from a realistic performance condition. For example, an orchestra consists of a distributed array of sources with unique, frequency-dependent radiation patterns. These radiation patterns have been found to be perceptually important concert hall auralizations.[70] Although this directivity information can be modeled using commercially available computer modeling techniques, software-based auralizations can have significant differences from auralizations based upon measurements in an existing room.[44]

The goal of this study was to generate realistic, repeatable measurement-based orchestral auralizations using spherical array processing techniques. Previous studies have represented instrument radiation pattern weights for the set of SH basis functions.[72] This set of functions provides a mathematical framework for defining the frequency-dependent radiation patterns of orchestral instruments. Once the weights are known, spherical array processing techniques can be used to reconstruct this radiation pattern using a compact spherical loudspeaker array (CSLA). The design details of such an array and radiation control filters demonstrate how time-efficient source directional RIR measurements can be made in a concert hall to generate full orchestral auralizations.

4.2 Background

The unique radiation patterns of different orchestral instruments has been studied in depth, most notably by Meyer,[71] but also by Lokki,[37] TU Berlin & RWTH Aachen,[72] BYU,[73] and others. Furthermore, investigations have shown that auralizations accounting for both source distribution and directivity have clear subjective differences from the reference case of omnidirectional radiation and a single source.[70] Pätynen and Lokki developed a loudspeaker orchestra, consisting of a 24-channel system comprised of 31 commercially available loudspeakers, using three different models. Efforts were made to select and orient the loudspeakers to match the radiation of each instrument, and in some cases, two loudspeakers were paired for separate low- and high-frequency radiation representation. Auralizations generated from the loudspeaker orchestra's RIRs are reproducible between halls but are limited in terms of how realistically commercial loudspeakers match the frequency-dependent radiation patterns of each unique instrument. Additionally, such a large system is resource-intensive in terms of cost, transportation, setup time, and complexity.

4.2.1 Spherical Harmonic Representation of Instrument Directivity

After deriving the wave equation in spherical coordinates using the separation of variables technique,[56,81,86] the directional dependence of the solution is represented by the set of SH functions, pictured in Figure 4.1:

$$Y_n^m(\theta,\phi) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}}P_n^m(\cos\theta)e^{im\phi}, \tag{4.1}$$

where θ is the elevation angle measured from the vertical axis, ϕ is the azimuthal angle, measured in the counterclockwise direction from the forward axis, and $P_n^m(\cos\theta)$ are the associated Legendre polynomials of order n and degree m.

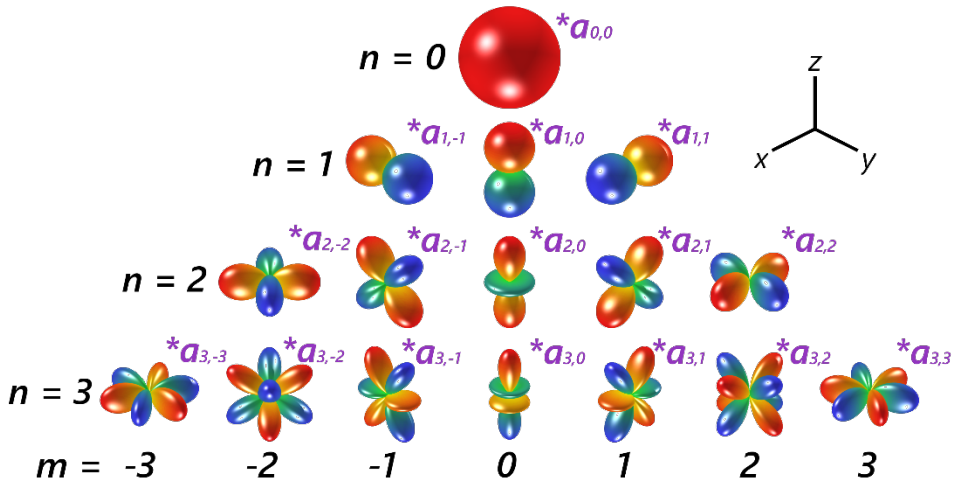


Figure 4.1: The set of SH functions up to order $n = 3$. The plot shows the real part of the non-negative degree functions ($m \geq 0$) and the imaginary part of the negative degree functions ($m < 0$). Yellow and blue indicate positive and negative values respectively.

This set of functions forms a spatial basis for the spherical Fourier series:[81]

$$p(\theta,\phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{m=n} a_{nm} Y_n^m(\theta,\phi), \tag{4.2}$$

where $p(\theta,\phi)$ is a bounded spatial function that is square-integrable on the unit sphere. In other words, a spatial function, such as an instrument's radiation pattern at a specific frequency, can be represented as a weighted sum of SHs. These weights, a_{nm}, are determined from the spherical Fourier transform of $p(\theta,\phi)$,

$$a_{nm} = \int_0^{2\pi} \int_0^{\pi} p(\theta,\phi)[Y_n^m(\theta,\phi)]^* \sin\theta \, d\theta d\phi. \tag{4.3}$$

Just as a time-frequency Fourier transform can result in a complex frequency bin weight, weights for these SH functions can also be complex.‡

4.2.2 Database of Instrument Radiation Patterns

These complex SH weights were calculated for 41 orchestral instruments in a comprehensive measurement database generated from RWTH Aachen and TU Berlin.[72] For each instrument, a surrounding array of 32 calibrated electret condenser microphones (5 mm diameter) 2.1 m from the player was used to measure the frequency-dependent radiation

‡ An animation of an order truncated representation of an arbitrary directivity-like pattern generated by the author for presentations can be found online at: https://sites.psu.edu/spral/files/2019/06/CSLA-Truncation.gif.

pattern of each instrument for chromatic tones played at different dynamic levels. From the measurements, phase related source-centering algorithms were used to center the radiation pattern for each one-third octave band. The centered measurements were then averaged to calculate a set of 25 complex fourth-order SH weights in each one-third octave band for every instrument. From the radiation database, the surrounding microphone array captures the full complexity of an instrument's radiation pattern in the SH domain. This measurement includes the sound generation mechanisms and scattered pressure terms due to the shape and design of each instrument. This set of weights forms the basis of the directional reproduction patterns used in the present study.

4.2.3 Spherical Array Processing Techniques

Spherical array processing techniques can reconstruct these radiation patterns, but the physical presence and size of the CSLA in the sound field will impact the accuracy of this reconstruction. This equalization is commonly referred to as *radial filtering* in the spherical microphone array literature. Rafaely[81] demonstrates that a measured sound field spherical Fourier weights, incident upon a rigid sphere, $p_{nm}(k,r)$, can be related to the sound field's spherical Fourier weights, $a_{nm}(k)$, as:

$$p_{nm}(k,r) = a_{nm}(k)b_n(kr), \tag{4.4}$$

where for a rigid enclosure design for the transducer array,

$$b_n(kr) = 4\pi i^n \left[j_n(kr) - \frac{j_n'(kr_a)}{h_n^{(2)'}(kr_a)} h_n^{(2)}(kr) \right]. \tag{4.5}$$

The term $b_n(kr)$ is found by representing the total pressure of a measured sound field as a summation of the incident sound field, composed of plane waves, and the scattered pressure field from a rigid sphere. Applying the zero radial particle velocity (rigid) boundary condition at the surface of the sphere, $r = r_a$, yields the result for $b_n(kr)$ obtained in Eqn. 4.5. In simple terms, this factor represents the relationship between the measured sound field at the center of the array, as if the rigid sphere were not present, $a_{nm}(k)$, to the measured pressure field on the surface of the rigid sphere, $p_{nm}(k,r_a)$. To correct for these differences, the measured pressure field at the surface is multiplied by $1/b_n(kr_a)$. Figure 4.2 illustrates these correction factors across frequency for a rigid sphere of radius r_a = 7.6 cm, the radius of the CSLA designed in this study, assuming a sound speed in air of 343 m/s.

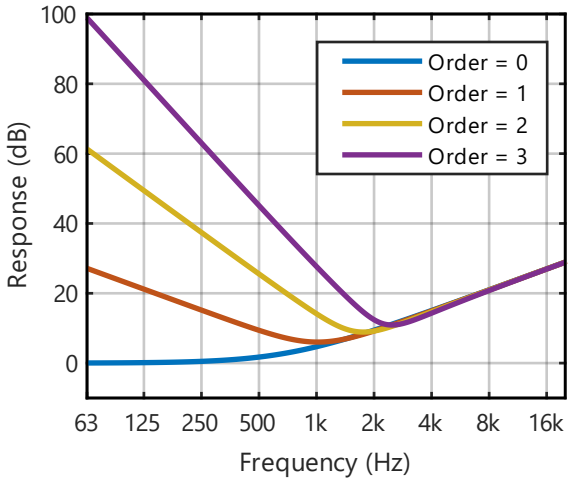


Figure 4.2: The radial filter correction factor, $1/b_n(kr_a)$, for a rigid sphere equal to the size of the compact array ($r_a = 7.6$ cm).

An important mathematical result of this correction is the dependence only upon frequency (or wavenumber), the radius of the rigid sphere, and SH order. Since the dependence can be represented independent of direction-of-arrival, it can be applied in the SH domain in this simple formulation for any spatial sound field, $a_{nm}(k)$, composed of plane waves from any direction. Although this correction has been documented in terms of spherical microphone arrays, by the principle of reciprocity, the identical formulation holds for CSLA.

4.2.4 Previous Compact Spherical Loudspeaker Arrays

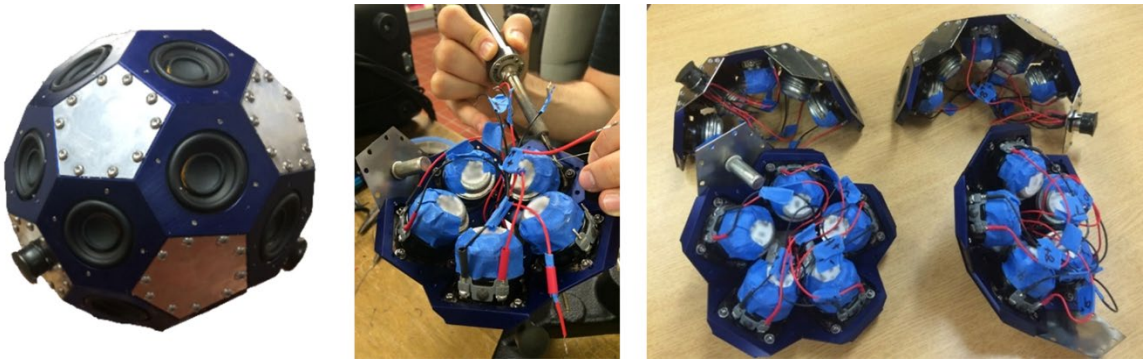
CSLAs have been developed by multiple groups in the past including IRCAM,[74] the Center for New Music and Audio Technologies (CNMAT) at UC Berkeley,[75] the Institute of Technical Acoustics at RWTH Aachen,[78,87] and the Institute of Electronic Music and Acoustics (IEM) at the University of Music and Performing Arts Graz.[76-77] At IRCAM, theoretical and practical studies regarding directivity control were carried out and studied, using a dodecahedron geometry.[74] The control of these arrays were more practical in nature, controlling patterns with a gain adjustment between channels based upon first-order SH patterns. The researchers from CNMAT generated multiple arrays, with the largest consisting of a 120-channel array using 32 mm drivers.[75] The work at RTWH Aachen first centered around converting dodecahedral omnidirectional sources into 12-channel directional arrays.[87] Further work has evolved into an array with different sized drivers that can be mechanically rotated by a turntable for arbitrarily higher-order SH resolution specification.[78] Zotter created a 20-channel icosahedron sound source, having a 28.5 cm radius with 16.5 mm diameter drivers[76] and a 16-channel source with 50.8 mm drivers.[77]

From all of these works, proof of concept of such arrays has been validated. The focus of each previous work can be divided into two categories: accuracy-focused and practicality-focused. Some of the original sources were often limited to only first-order reproduction, and instrument directivities have been found to exhibit radiation patterns with significant higher-order SH representation.[72] Secondly, some of the larger sources, such as the icosahedron made at IEM with a 28.5 cm diameter, exhibited spatial aliasing and directional radiation breakdown at relatively low-frequencies, around 700 Hz. Some more CSLAs have been designed, such as the 120-element array from CNMAT, but the drivers used were silk-dome tweeters, which are not designed to operate well in low- to mid-frequency regimes. The balance between selecting drivers that exhibit good low-frequency SNR and maintaining close spacing for high-frequency spatial aliasing limits is difficult to attain.

RWTH Aachen first considered using multiple dodecahedron sources, with optimized design and driver selection for different frequency ranges.[87] More recently, Aachen's turntable-controlled source allows for arbitrary sampling resolution using multiple sizes of drivers, providing high spatial sampling to avoid spatial aliasing at high frequencies, while allowing for larger drivers to provide good low frequency SNR.[78] The downside to this thorough technique is practical in nature, taking multiple hours to sample a high resolution RIR for a single source position. For the present study, the aim was to develop a single spherical loudspeaker array that could produce sufficient low-frequency SNR with small, high performance 40-mm drivers and a compact 15.2 cm overall diameter to limit effects of spatial aliasing to high frequencies. The decision to use a single source was made due to the practical needs of having only limited measurement time in each concert hall.

4.3 Compact Spherical Loudspeaker Array (CSLA) Design

A 20-channel CSLA was designed and developed to accurately reconstruct instrument radiation patterns, shown in Figure 4.3. The driver arrangement was taken from the face center angles of an icosahedron, providing a uniform sampling scheme. The icosahedron was truncated to allow for rear-mounting of the drivers and the integration of connector plates. This number of drivers enables up to third-order SH representation of radiation patterns. Also, cost-effective digital-to-analog converters often have up to 24 channels, providing a reasonable channel limit for practical considerations.



(a) (b) (c)

Figure 4.3: In-process finished enclosure for the CSLA (a), along with assembly photographs (b – c).

4.3.1 Physical Construction Details

The array was designed to be as compact as possible with an overall diameter of 15.2 cm (6 in.) to increase the frequency at which spatial aliasing occurs. The individual drivers are 40 mm in diameter with a resonance of 200 Hz and a high linear excursion. An inherent competition exists between constructing an array with larger drivers, to ensure a good low frequency signal-to-noise ratio (SNR) and minimizing the spacing between drivers to maintain directional accurate at high frequencies. This driver's compact size and low frequency performance provided a proper balance for the design of a broadband array. The enclosure was built out of machined aluminum panels for the driver mounting panels and waterjet-cut steel connection plates. Four custom connectors and a custom 4 m cable was created to connect the CSLA to a mobile custom hardware rack.

4.3.2 Loudspeaker Driver Performance and Equalization

Once the array's construction was completed, on-axis measurements were made of each driver to characterize their individual response and perform individual driver equalization. As seen in Figure 4.4, the drivers have significant response down to around 200 Hz. To improve the array's performance, individual equalization filters were designed for each driver. These filters were designed by inverting the time-aligned complex transfer functions for each driver's on-axis response. A linear-phase band-pass filter was also applied to limit excessive amplification below the 160 Hz one-third octave band and above the 16 kHz band. The same band-pass filter was used for each driver, to ensure that each driver had the same phase response.

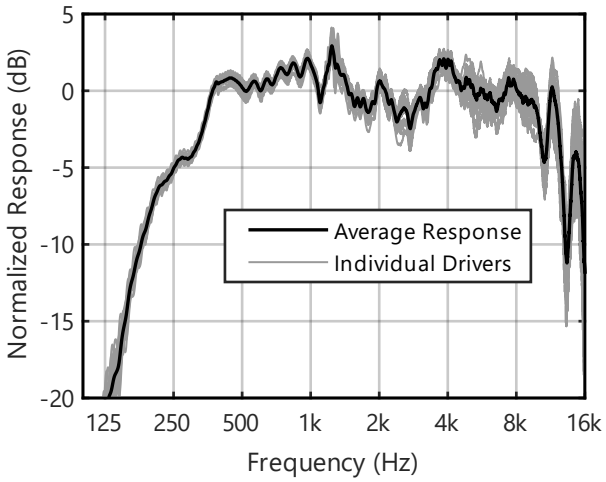


Figure 4.4: The on-axis response of each driver in the CSLA. The response extends low in frequency for a driver of this size, down to the 200 Hz one-third octave band.

4.3.3 Hardware and Software Control

To control the array, a 24-channel MOTU 24Ao audio interface was used to provide the necessary 20 channels of audio. A diagram of the complete hardware setup for the orchestral RIR measurements is pictured in Figure 4.5. Each channel was connected to four 6-channel Sure Electronics Class-D TDA7498 amplifier boards with 100 W per channel of power output. A custom hardware box was also designed to make the system portable and provide direct connections for USB, word clock, and power to the hardware interfaces shown in Figure 4.6. Casters were integrated into the hardware box so that it could be easily moved around the stage with the sound source. Finally, a second MOTU Ultralite AVB interface was purchased to allow connection to the MOTU 24Ao via AVB Ethernet connection. A 15 m cable providing Ethernet, BNC word clock syncing, and power was used to connect the computer system at the front, center of the stage to movable source hardware box.

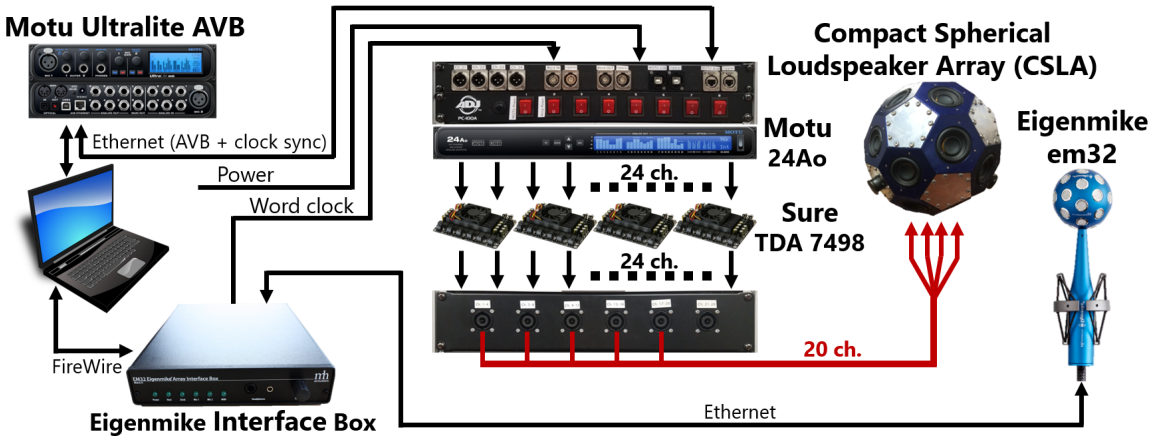


Figure 4.5: A schematic layout of the hardware setup used for the CSLA orchestral RIR measurements.

The source was controlled by sending 20 channels of audio from a custom-designed user interface in Max7.[88] The 20 channels signals were logarithmic swept-sine signals, replicated over all 20 channels. Each channel of the measurement sweep was then filtered with a set of 20 individual driver filters. Each filter channel was designed for a specific driver for each orchestral instrument. For example, after any measurement signal was filtered with the 20-channel trumpet filter bank, the measurement signal radiated from the array with the frequency-dependent radiation pattern matching that of a trumpet. The design of these filters will be further discussed in section 4.4.

(a) (b)

Figure 4.6: Photographs of the custom hardware box to control the CSLA. The USB MOTU audio interface and powered CLSA outputs are shown in (a) and the power switches, ethernet, USB, word clock, and four low level (unamplified) XLR outputs are shown in (b).

4.4 Instrument Radiation Filter Design Methodology

To provide flexible control of the directional array, a set of 20 filters, corresponding to each driver, was designed for each instrument. These filters contain the complete processing required to take any single-channel signal, such as a logarithmic swept-sine signal, and play that signal from the array with the frequency-dependent radiation pattern of a specific instrument. The following sections describe encoding the instrument's directivity into one-third octave bands, equalization for the CSLA's design, decoding into individual driver signals, and normalization to limit driver distortion.

4.4.1 Encoding Directivity into One-third Octave Bands

To start, linear-phase one-third octave band FIR filters were designed to have steep transitions between frequency bands, but also summed to have a flat frequency response and

a purely impulsive time response. The frequency domain behavior of these filters and their summed response is shown in Figure 4.7. To encode the instrument directivities, a time-frequency Fourier transform of each filter IR resulted in a complex-valued frequency spectrum for each filter. Then, 16 copies of each one-third octave band filter's spectrum were made, one for each SH component of the radiation pattern.

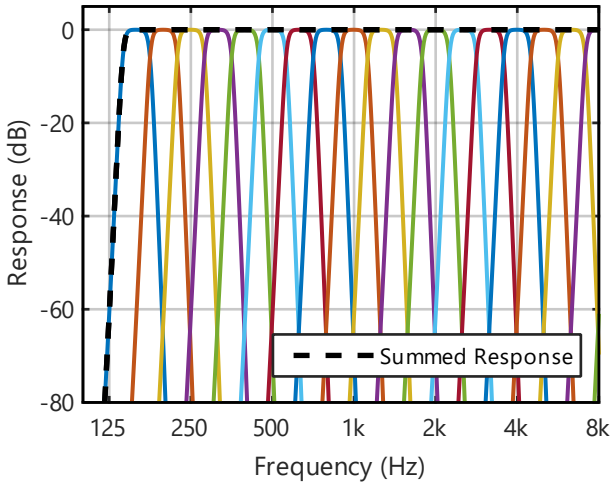


Figure 4.7: Linear-phase one-third octave band filters designed to the encoding the radiation patterns. The filters are designed to sum flat, having the same roll-off as a 20th order Butterworth filter.

Finally, for each frequency band, the complex SH weights for a specific instrument were taken from the instrument radiation pattern database, described in section 4.2.2.[72] Step-by-step, the 16 complex-valued SH weights for a single octave band were multiplied with the 16 copies of the complex frequency spectrum of the current one-third octave band filter. This step was repeated for each frequency band, "encoding" the radiation information into the filters. For the current study, this process was repeated for the database's radiation patterns at a *forte* dynamic using modern-style instruments for the violin, viola, cello, double bass, transverse flute, clarinet, oboe, bassoon, trumpet, French horn, tenor trombone, and tuba.

4.4.2 Radial / Modal Array Equalization Filters

The next filter design step accounted for equalization due to the physical presence of the CSLA in the sound field, which has an impact upon the reproduced field. The previously encoded frequency domain one-third octave band signals represent $a_{nm}(k)$, and without radial filtering, $p_{nm}(k, r_a)$ would be the reconstructed spherical Fourier weights from the loudspeaker array from Eqn. 4.4, including the scattered pressure term at the surface of the array, $b_n(kr_a)$. To implement a correction for the scattered pressure at the sphere's surface, $b_n(kr_a)$, linear phase FIR filters were designed based upon the filter targets from Figure 4.2, up to order $n = 3$. As excessive amplification is needed for higher orders at low frequencies, band

limitations were applied to each order. The different SH order processing was truncated below the 315, 397, and 630 Hz one-third octave bands for orders 1, 2, and 3, respectively. These cutoffs can be seen in the lower limits of the designed filters for each SH function in terms of order in Figure 4.8. These frequencies were chosen by inspecting the IR of the filters for all instruments and selecting a cutoff frequency that kept the filter ring-down time within 1 ms. Excessive boosting of low-frequency energy can result in a very wide, or 'ringy' response of the filter in the time domain. This frequency-dependent order truncation allowed for control of filter ring, which could generate unwanted and unnatural time-domain artifacts. Informal subjective listening was performed, playing directionally filtered anechoic music from the CSLA, and no temporal artifacts or degradation was heard. The radial filters were designed to have the same number of samples as the one-third octave band FIR filters, and they were multiplied in the frequency domain, generating encoded, array equalized SH directivity filters.

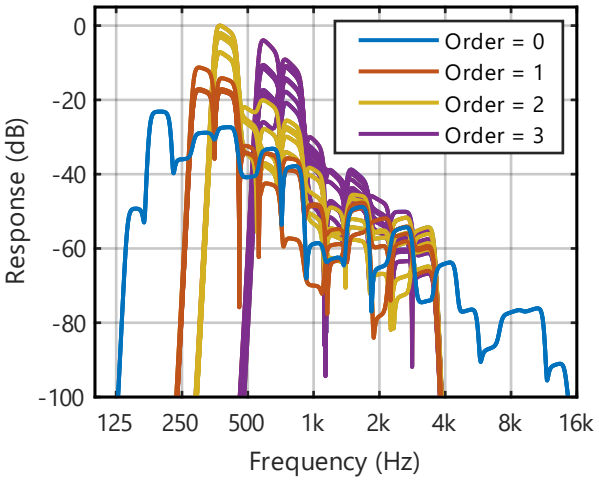


Figure 4.8: The encoded instrument directivity for each SH function where different colors were used to represent the SH order of each function. These functions include one-third octave band SH directivity weights, radial equalization filters, and crossover to lower SH order truncation at low frequencies.

4.4.3 Ambisonic Decoding to Array Driver Signals

Once the SH signal is encoded for a given radiation pattern, each of the SH functions can be sampled at driver locations to determine the loudspeaker signals required to reconstruct the original spatial function. This practical step discretizes the problem, as the array has a finite number of drivers. In the SH domain, this limitation results in an order-truncated representation of the original spatial function, $f(\theta, \phi)$, because the matrix inversion to sample the SH functions requires that the number of drivers, L, be greater than the number of SH components for a given truncation order (N), $L \geq (N+1)^2$. With 20 driver elements, the current array has a maximum truncation order of $N = 3$.

This step is commonly referred to in the literature as *decoding* for higher-order Ambisonics (HOA), a virtual acoustic reproduction technique. Using a CSLA, each driver samples each SH function, and summed together, provides a reconstruction of the spherical Fourier weights:

$$a_{nm}(k) = \sum_{l=1}^{L} g_l(k) Y_n^m(\phi_l, \theta_l), \tag{4.6}$$

where k is wavenumber, related to frequency. This step illustrates the reconstruction of the weights, $a_{nm}(k)$, found in the discrete spherical Fourier series. Each driver signal can be weighted with each SH function evaluated in the look-direction of the driver, (ϕ_l, θ_l), and the resulting summation reconstructs the desired radiation pattern weights. This summation can also be expressed in matrix notation for spherical Fourier weights up to a truncation order of $n = N$ as:

$$\boldsymbol{A} = \boldsymbol{Y_L} \boldsymbol{G_L}, \tag{4.7}$$

$$\begin{bmatrix} a_{0,0} \\ a_{1,-1} \\ a_{1,0} \\ \vdots \\ a_{N,N} \end{bmatrix} = \begin{bmatrix} Y_0^0(\phi_1, \theta_1) & Y_0^0(\phi_2, \theta_2) & \cdots & Y_0^0(\phi_L, \theta_L) \\ Y_{-1}^1(\phi_1, \theta_1) & Y_{-1}^1(\phi_2, \theta_2) & \cdots & Y_{-1}^1(\phi_L, \theta_L) \\ Y_0^1(\phi_1, \theta_1) & Y_0^1(\phi_2, \theta_2) & \cdots & Y_0^1(\phi_L, \theta_L) \\ \vdots & \vdots & \ddots & \vdots \\ Y_N^N(\phi_1, \theta_1) & Y_N^N(\phi_2, \theta_2) & \cdots & Y_N^N(\phi_L, \theta_L) \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ \vdots \\ g_L \end{bmatrix}. \tag{4.8}$$

If N is the number of frequency bins for k, then $\boldsymbol{G_L}$ is a 20 x N matrix, $\boldsymbol{Y_L}$ is a 20 x 16 matrix, and $\boldsymbol{A}$ is a 16 x N matrix. Using a matrix pseudoinverse, the loudspeaker signals can be calculated from the spherical Fourier weights as,

$$\boldsymbol{G_L} = \boldsymbol{Y_L^\dagger A} = \left(\boldsymbol{Y_L^T Y_L}\right)^{-1} \boldsymbol{Y_L^T A}, \tag{4.9}$$

where $\boldsymbol{Y_L^\dagger}$ is known as the decoder matrix, a 20 x 16 matrix calculated from the pseudoinverse of $\boldsymbol{Y_L}$. In practice, this pseudoinverse is typically performed using singular-value deposition (SVD) or a Moore-Penrose pseudoinverse. This type of decoder is referred to as a basic or mode-matching decoder matrix in HOA literature.[84] This process allows the final filters that correspond to each driver, $g_L(k)$, to be calculated in terms of the SH signals, $a_{nm}(k)$. This calculation was repeated for each discrete frequency, mapping each of the 16 channels of the SH domain filter to a 20-channel filter bank for each instrument. Finally, each of these 20 channels was filtered with an individually designed equalization filter for each driver and was discussed previously in section 4.3.2.

4.4.4 Filter Normalization for Distortion Prevention

Using all of the techniques listed in sections 4.4.1 through 4.4.3, the designed filters containing one-third octave band instrument radiation information, array equalization, Ambisonic decoding, and individual driver equalization were generated. Initially, a few observations can be made. First, large gain differences are observed between the different one-third octave bands from the encoded signals in Figure 4.8. These differences are due to the combination of frequency-dependent strength of each instrument and the radial filtering. If this filter were used for a measurement signal, it would produce high signal in some frequency bands, and little signal in other bands.

During measurements, overall gains are typically set as high as possible to ensure good SNR, while still preventing any noticeable driver distortion. To provide good broadband SNR and minimize distortion, all driver signals were normalized to the driver with the maximum driving amplitude within each one-third octave band. Due to the decoding process and frequency dependence of the radiation patterns, this driver often changed between each frequency band. Figure 4.9 shows the resulting driver filters from this normalization designed for the oboe, now with similar peak signal amplitudes broadband. Another important note is that all driver signals within the same one-third octave band were scaled with the same factor, maintaining the same relative differences. If the relative differences between drivers within each frequency band is constant, radiation accuracy is still preserved.

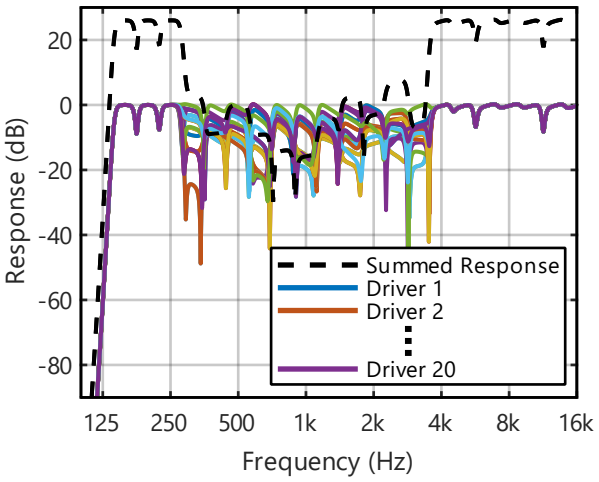


Figure 4.9: Individual array driver filters for an oboe instrument directivity. The filters have been normalized for the peak amplitude across all 20 drivers within each one-third octave band. The dashed black line shows the summed response for all 20 drivers.

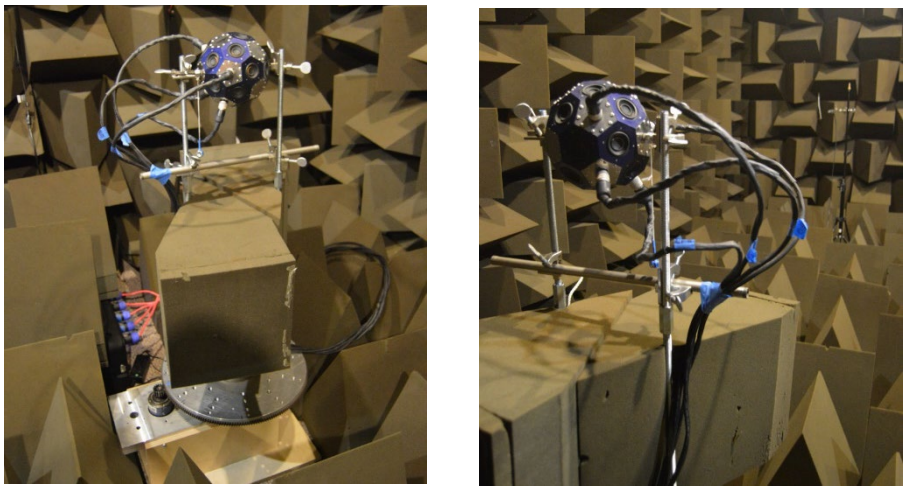
Secondly, large dips in the frequency response are observed in between some of the bands. These dips occur when the radiation pattern in adjacent one-third octave band signals

from the oboe are not in-phase with one another. This limitation occurs when a full frequency phenomenon, such as an instrument directivity, is represented in a highly discretized resolution. Observing the total summed response across all 20 driver signals for the instrument, the severity of these dips is heavily reduced. This demonstrates that despite these dips, the final reproduced measurement signal still exhibits a good broadband SNR for the total measurement.

Finally, as will be shown in section 4.5, above 4000 Hz the array was found to exhibit significant spatial aliasing, which degraded the directional reproduction accuracy. Since the array was no longer directionally accurate, the filters were transitioned back to zeroth-order reproduction at and above 4000 Hz. This provided a higher signal energy from the array for increased SNR. This same phenomenon of higher amplitude signal is seen at low frequencies, where zeroth-order reproduction is used to prevent excessive low-frequency radial filter boosting. For zeroth-order reproduction at low- and high-frequencies, all drivers sum coherently, thus, creating higher levels. For higher order reproduction, drivers contain amplitude and phase differences relative to one-another, which can cause incoherent summation for accurate radiation pattern reconstruction. From these differences, only the maximum amplitude driver in each band will be driving at the same level as the zeroth-order reproduction case. Section 4.6 will discuss the diffuse-field equalization to remove strength differences for realistic auralization generation.

4.5 Radiation Reconstruction Results and Discussion

To assess the directional accuracy of the array, IR measurements were made in an anechoic chamber using a full, three-dimensional equiangular sampling scheme. The loudspeaker was mounted on its side, so that a horizontally spinning turntable rotated in the loudspeaker's elevation direction. Measurements with the turntable provided 5-degree elevation (θ) resolution, and manual rotation of the array provided 10-degree azimuthal (ϕ) resolution. The IRs were measured using a logarithmic swept-sine and time-domain averaging, with sweep lengths and frequency ranges similar to those used in-situ during concert hall measurements. At each measurement direction, the array was sent a 20-channel swept-sine signal that was separately filtered for each of the thirteen instruments. The array cycled through an IR measurement for each instrument used in the orchestral setup, generating 1,332 total IRs per instrument.



(a) (b)

Figure 4.10: Photographs of the measurement turntable used to directionally sample IR measurements from the CSLA. The speaker was sampled by the turntable at a 5-degree elevation resolution (mounted on its side) and manually rotated in 10-degree steps in azimuth.

To generate radiation plots for the array's instrument reconstruction, a time-frequency Fourier transform converted each IR into a frequency-domain transfer function (TF). The energy in the discrete frequency bin closest to the center frequency of the one-third octave band was isolated, and that energy was converted from a pressure quantity into a sound pressure level. Balloon plots were generated showing the directional radiation pattern at each frequency, with a lower radius cutoff of 20 dB (re: max). The radiation performance for the array reconstruction an oboe's directivity is shown in Figure 4.11. Each letter represents a different one-third octave band center frequency, from 315 – 4000 Hz. The upper plots for (a) – (l) are the mathematically calculated directivities from the radiation database using the order-truncated version of Eqn. 4.2, and the lower plots are the reconstructed radiation patterns by the loudspeaker array. In general, the patterns demonstrate high levels of accuracy at lower frequencies. Subtle deviations occur in the range from 630 – 2000 Hz, but a high visual accuracy is still achieved. Once the 3175 and 4000 Hz frequency regions are reached, spatial aliasing-based degradations are first observed in the shapes of the smaller amplitude lobes of the directivity patterns, and eventually, full degradation results in a typical flower petal-like radiation plot, common to commercial dodecahedron loudspeakers at mid- to high-frequencies. Additionally, the accuracy of the directional radiation reconstruction for the viola is shown in Figure 4.12. Comparing results for both instruments, a similar behavior and spatial aliasing frequency can be seen.[§]

[§] A complete set of the radiation patterns for all frequency bands can be found in Appendix A at the end of this dissertation.

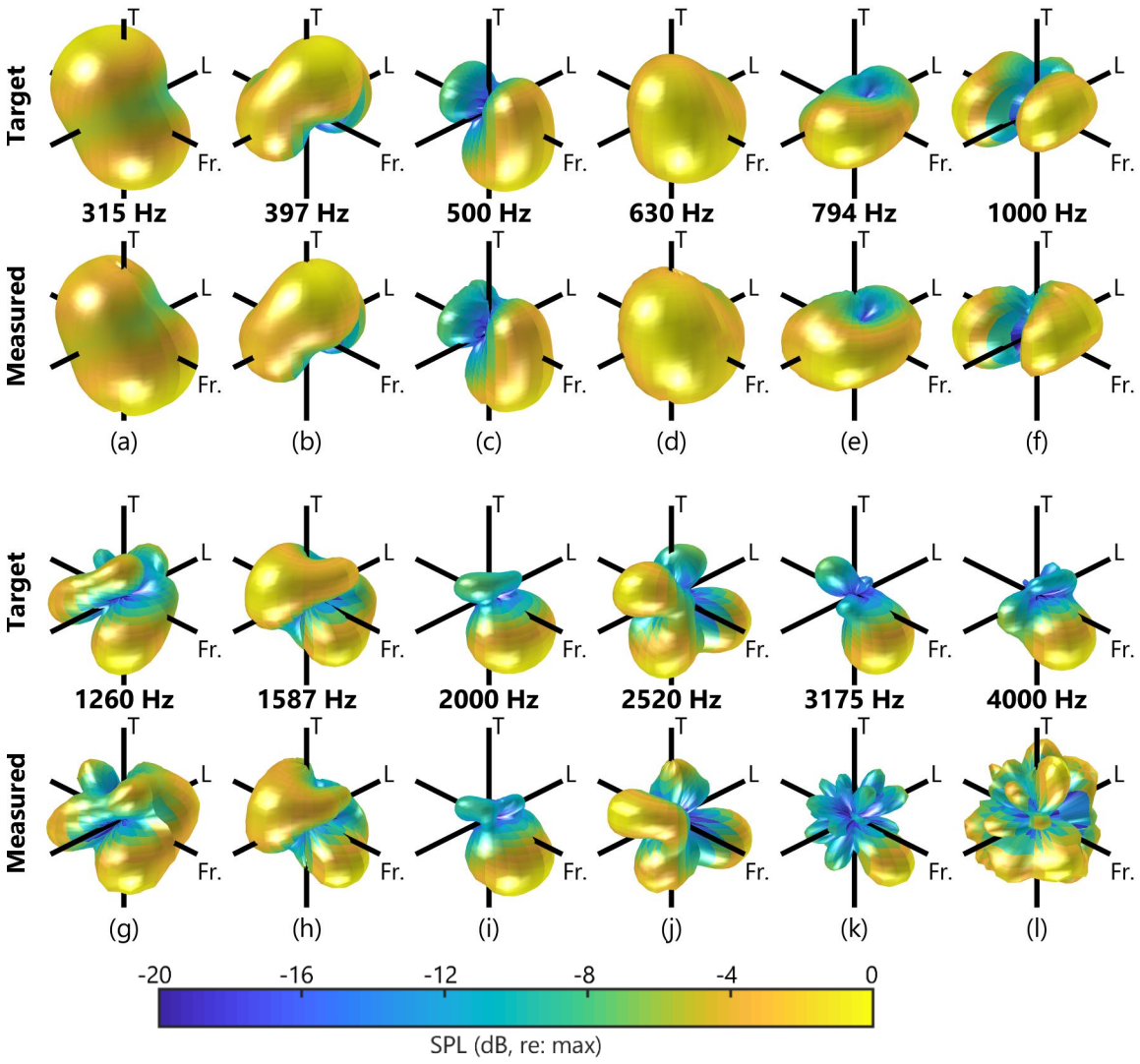


Figure 4.11: Balloon style directivity plots for the source operating with the oboe source filters. For each plot, (a) – (l), the upper plot represented the directivity calculated from the turntable measurements and the lower plot represents the target pattern, calculated from the order-truncated summation of each SH component with the proper weights from the radiation database in each one-third octave band.** The labels Fr., L, and T denote the front, left, and top directions from the instrument.

To demonstrate the performance of the array across all instruments, the spatial coherence was calculated between the target pressure function and the reconstructed pressure function. The spatial coherence function can be calculated for each direction at each center band frequency as,

$$C(f,\theta,\phi) = \frac{|S_{tr}(f,\theta,\phi)|^2}{S_{tt}(f,\theta,\phi)S_{rr}(f,\theta,\phi)}, \qquad 4.10$$

** Note: due to the order dependent crossovers described in section 3.2, (a) applies 1st order truncation, (b) and (c) apply 2nd order truncation, and (d) – (l) apply 3rd order truncation, the maximum SH order of which the current array is capable.

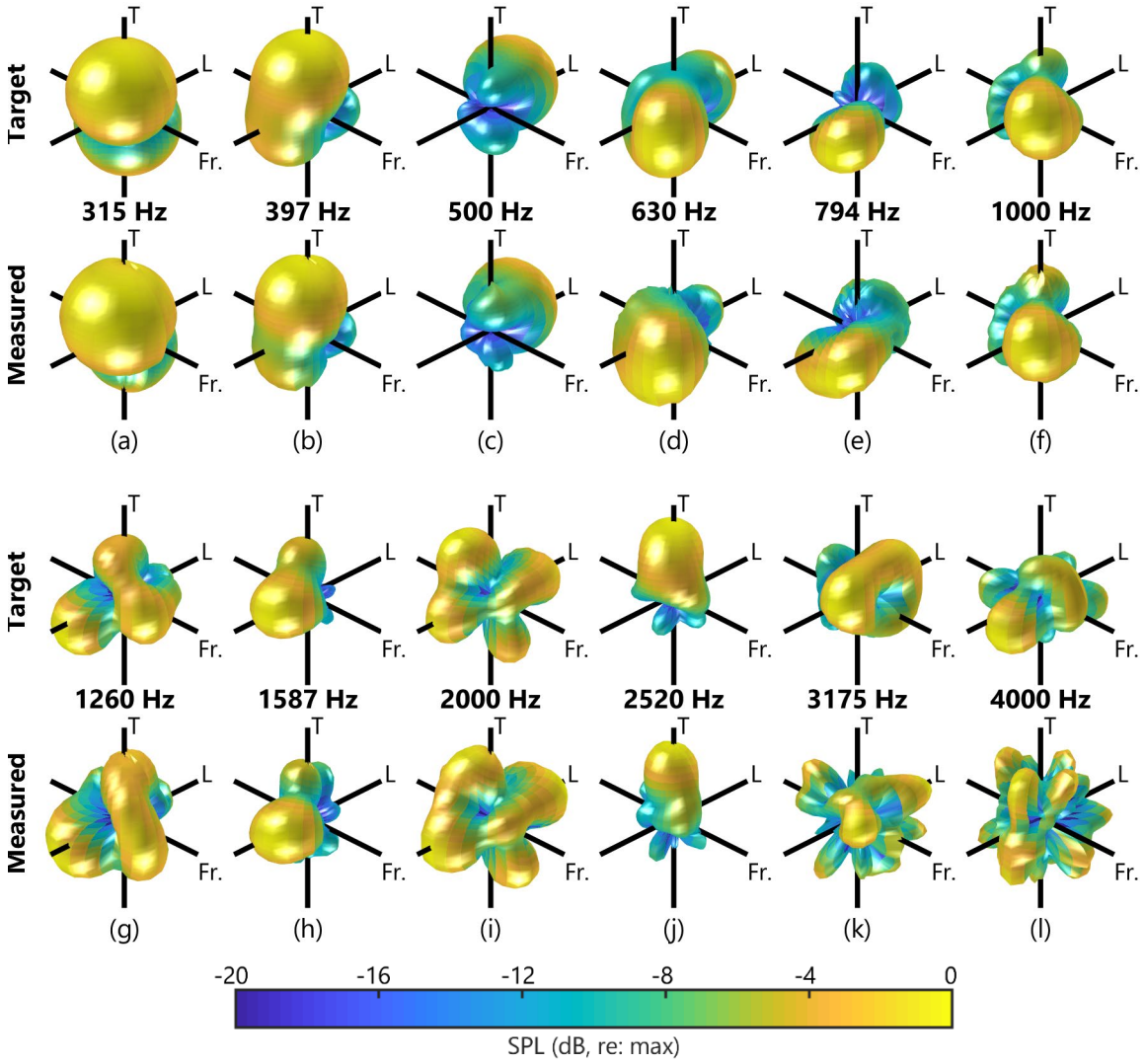


Figure 4.12: Balloon style directivity plots for the source operating with the viola source filters. Layout is identical to Figure 4.11.

where S_{py} is the cross spectral density between the target pressure field, $p_t(\theta, \phi)$, and the reconstructed pressure field, $p_r(\theta, \phi)$. S_{tt} and S_{rr} are the power spectral density functions. Inspecting the coherence function, if $S_{tt} = S_{rr}$, then $C = 1$, but if S_{tt} and S_{rr} are uncorrelated functions, C will approach 0. The spatial coherence function was calculated at each one-third octave band center frequency across all directions. Since the coherence has a unique value in each direction, to reduce this performance to a single value at each frequency band, for each instrument, the 1st-percentile of the coherence across space was calculated. Effectively, it is a measure of the maximum coherence across space that is robust to single directional sample peaks, which sometimes occur in frequency bands with overall low performance. Figure 4.13

shows the 1st percentile coherence, or the coherence that was exceeded for only one percent of all directional samples.

Looking at the coherence, in general, good coherence is observed across all instruments at lower frequencies, up to about 2.5 – 3 kHz. The gradual degradation of coherence between 250 and 630 Hz corresponds directly to the changes in SH truncation order at the 315, 397, and 630 Hz one-third octave bands. These decreases are not indicative of degraded performance, but rather, are indicative of the increasing spatial complexity of the target patterns. The coherence is stable after 630 Hz, averaging a value of 0.85 up to 2000 Hz. Then, between 2520 and 4000 Hz the coherence degrades quite drastically, indicating the point of spatial aliasing. Above this frequency, the coherence of individual instruments is generally lower, below 0.8, with some random nature due to the aliased character of the signal. When averaged across all instruments, the degradation is more consistent and apparent.

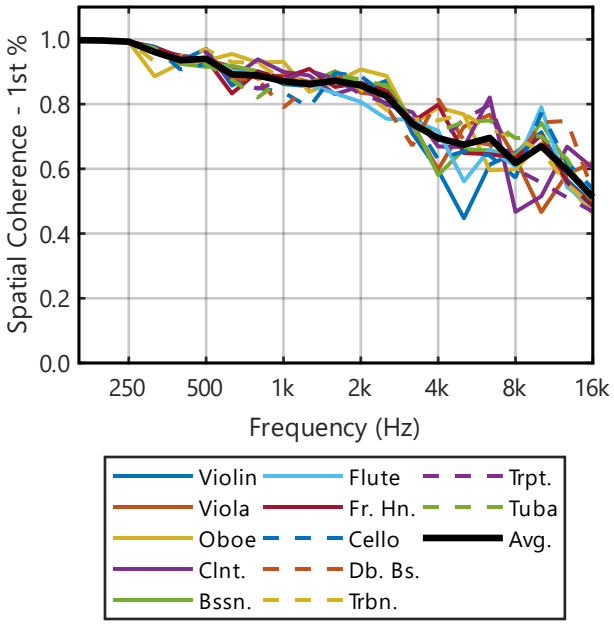


Figure 4.13: The 1st percentile maxima of the spatial coherence between the target pressure field and the measured pressure field reproduction for each instrument. The coherence across all instruments severely degrades for all instruments between the 2520 and 4000 Hz one-third octave bands.

4.6 Processing for Full-orchestral Auralizations

Using this custom setup, measurements have been made in nineteen concert halls around North America and Europe. In each of the halls, a pre-defined measurement grid was used to ensure a consistent orchestral arrangement between measurements. Figure 4.14 shows the layout of these measurement locations, and Table 4.1: lists the locations of these

measurement positions and their corresponding instrument. The central grid 'conductor' position was typically selected to be 1 m from the edge of the stage, but this position was adjusted up- or down-stage to accommodate arrangement of risers in certain halls.

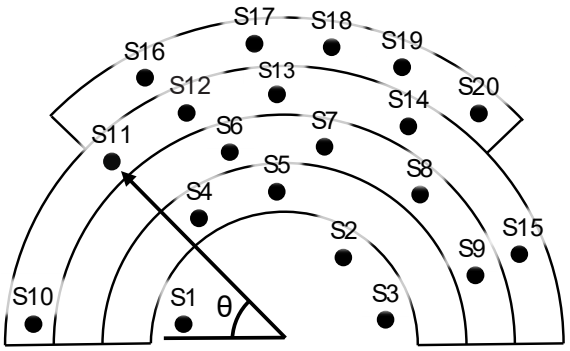


Figure 4.14: Layout of the source measurement grid used to take consistent orchestral measurements in each concert hall. Table 4.1: provides more detailed source location information.

Table 4.1:: Locations of the 20 source measurement positions, along with the built-in radiation patterns in each position.

Source	Row	r (m)	θ (degrees)	Instrument
S1	1	3.06	10.6	Violin
S2	1	3.06	124.4	Viola
S3	1	3.06	169.4	Cello
S4	2	4.45	54.9	Violin
S5	2	4.45	86.7	Viola
S6	3	5.84	73.8	Clarinet
S7	3	5.84	101.2	Oboe
S8	3	5.84	132.2	Viola
S9	3	5.84	160.8	Cello
S10	4	7.34	4.4	Violin
S11	4	7.34	46.1	Violin
S12	4	7.34	67.0	Flute
S13	4	7.34	87.9	Bassoon
S14	4	7.34	119.2	Trumpet
S15	4	7.34	159.7	Double Bass
S16	5	8.83	61.9	Percussion
S17	5	8.83	83.9	French Horn
S18	5	8.83	98.5	French Horn
S19	5	8.83	113.2	Trombone
S20	5	8.83	130.3	Tuba

In each hall, the array was moved to each position, a measurement signal was pre-processed with the filter bank for the appropriate instrument, and a measurement was taken with built-in source directivity. Additionally, a separate measurement was taken with the subwoofer component of a three-part omnidirectional sound source to obtain good low frequency SNR for auralization. This resulted in 40 unique measurements, which took roughly

1.5 – 2 hours to complete. It is possible to take a separate measurement with each driver of the array individually, which would allow for flexible post-processing of the directional source radiation properties. The downside to this measurement is that it takes 20 times as long to complete, which quickly proves to be impractical when multiple source locations are desired. Even with the efficiently of built-in radiation patterns, this measurement was still time-intensive, so only one receiver position was measured in each hall. The selected seat was in the center of the hall, 15 meters back from the conductor position in the source grid.

Finally, coloration due to the non-flat nature of the measurement signal and the loudspeaker's response must be accounted for to produce a natural auralization. For a typical measurement source, this can be done by measuring spatial samples of the frequency response of the loudspeaker in an anechoic environment and creating a filter with a spectrum matching the inverse of the diffuse-field averaged spectrum. For the CLSA, this correction is also dependent upon instrument, as the diffuse-field response of the array changes with each instrument's pre-processing filter bank. Figure 4.9 shows a representation of the total summed response for an oboe, in the dashed black line. Due to the directivity of the loudspeaker and the normalization techniques, a non-flat response exists.

To correct for this effect, a diffuse field-averaged response was taken for each instrument from the turntable measurements described in section 4.5. These measurements were made at the same driving levels as the measurements were made in each concert hall. Figure 4.15 shows the diffuse-field averaged response the CLSA operating with the oboe directivity filters applied. This response was inverted, and a minimum phase FIR filter was designed to remove the frequency coloration directional array for each instrument. The target inverted response (smoothed) and the designed FIR equalization filter are shown in Figure 4.15. A similar filter was designed for all 13 instruments in the orchestral setup.††

†† A complete set of these diffuse-field equalization filters can be found in Appendix B.

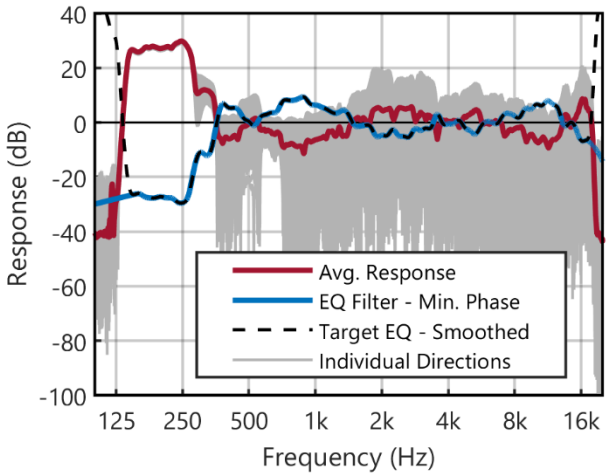


Figure 4.15: Designed minimum-phase FIR filter to compensate for the non-flat response of the array, operating in each instrument condition. Shown above in blue is the design filter for an oboe.

4.7 Conclusions and Future Work

A CSLA was designed using compact drivers with good low frequency response, balancing low-frequency power with the upper spatial aliasing limit. To allow efficient full orchestral measurements in concert halls, a set of 20 driver filters was designed for each orchestral instrument. These filters enabled the source to be flexibly operated for any instrument's full-frequency radiation pattern. The filters included the encoding of instrument radiation patterns,[72] array design equalization (radial filtering), decoding to individual driver signals, and individual driver equalization. Additionally, the filters were designed to control for individual driver distortion through normalization to each maximum amplitude driver within each one-third octave band.

Analyzing the directivity results, the array demonstrates high directional accuracy until the spatial aliasing limit of the array, which significantly breaks down around 3175 or 4000 Hz. By looking at the 1st-percetile of the spatial coherence function at each one-third octave band, this trend also continues across all instruments. Using these filters, a measurement grid of 20 consistent source locations enabled consistent measurements, which were subsequently used to generate a full-orchestral auralization in each hall. Finally, diffuse-field equalization filters have been designed to correct for the non-flat response of the array for each individual instrument filter configuration. This allows for a natural, repeatable method for measurement-based orchestral auralization. Nineteen halls were measured with this setup described in section 4.6. This setup takes an average of 1.5 to 2 hours to measure these 20 source locations for a single listener location. This setup took an average of 1.5 to 2 hours to measure in each hall for the single listener location.

Future work should first investigate further optimization of the designed filters for this array. Currently, the discrete one-third octave band representation of the array allows for phase-discontinuities in the frequency domain, which can result in individual peaks or notches between bands. This representation of the data could be improved by either processing the instrument directivity measurements with finer frequency resolution or attempting a frequency domain interpolation during filter design. This technique was attempted in the current study but proved to be difficult to implement while maintaining filter stability. The current filters could be improved with consideration of the mutual impedance between each loudspeaker driver.[87] Forces due to the pressure exerted from one driver on adjusted drivers can impact the measured radiation pattern. Additional compensation for these effects might reduce deviations from the target pattern at lower frequencies, below the spatial aliasing limit.

4.8 Acknowledgements

The authors would like to acknowledge Dave Dick for his advice regarding the design of the CSLA and hardware setup for the array's control, Molly Smallcomb for her assistance in the physical construction of the array as a summer undergraduate research assistant, Andrew Coward for his assistance in the physical design and fabrication of the enclosure of the compact array, and Ben Steers for his assistance with the orchestra layout. This work was supported by the National Science Foundation award #1302741.

This Page is Intentionally Left Blank

Chapter 5
The Concert Hall Measurement Database

This chapter contains information written for external publication, which is intended for submission to a major peer-reviewed acoustics or audio engineering-related journal. As such, this chapter also contains its own introduction, background, and results section. The overall introduction (Ch. 1), background (Ch. 2 and 3), and results (Ch. 7) chapters of the dissertation fill in the greater picture of the entire dissertation work.

Chapter 5 provides details regarding the measurement setup and generation of the concert hall acoustics orchestral research database. This database contains spherical microphone array measurements in twenty-one concert halls in North America and Europe. This database also implements the CSLA described in chapter 4 to produce realistic full-orchestral auralizations. This database was obtained in part using the basis for the psychoacoustic experiment described in chapter 6.

The CHORDatabase: a spherical microphone and compact loudspeaker array impulse response measurement database

Matthew Neal and Michelle Vigeant
Graduate Program in Acoustics, The Pennsylvania State University, University Park, PA 16802

Abstract:

A tension inherently exists between using standardized protocols for assessing concert halls and representing a hall in a realistic and ecological way. Standard room acoustic measurements provide somewhat repeatable measurement conditions across groups, but these methods do not dictate how to provide a realistic auralization with high spatial accuracy. To accomplish both, a database has been generated using measurements in 21 concert halls. In each hall, measurements were made at multiple seats using a three-part omnidirectional loudspeaker and a 32-element spherical microphone array. Additionally, for a single receiver, 20 orchestral source position measurements were made using a compact spherical loudspeaker array (CSLA) for accurate instrument directivity representation. Spherical array processing techniques provided accurate frequency-dependent source radiation and high spatial resolution for omnidirectional room impulse response (RIR) analysis. A total of 242 individual RIRs were measured, and a correlation analysis of standard metrics in ISO 3382 identified the relationship between each parameter. Furthermore, spherical beamforming analyses demonstrated the differences in early and late energy between halls of different shapes. Finally, full video-based time animation of the RIR provides a highly intuitive representation of this new type of spatial RIR analysis.

5.1 Introduction

In the study of concert hall acoustics, a struggle exists between balancing two goals: objective accuracy and subjective realism. As a design discipline that is increasingly studied from a physics, math, or engineering perspective, the need for objective truth or facts is known. Engineers desire well-documented, standardized design practices to help ensure a high-quality acoustic product on all projects. Well-defined room acoustic metrics, if linked to the fundamental psychoacoustic perception of rooms, can also provide opportunity to ensure a quality acoustic environment in a wide variety of new, innovative architectural forms and ideas. The problem comes when the metrics the field relies upon does not tell the entire story.

Objective metrics are always an indirect measure of the simply stated subjective question that dictates a project's success "How much do I like this hall?" The question,

although simple, is packed with a highly complex auditory and neurological system, with countless highly correlated subjective acoustic factors at play. To further complicate the question, the concept of 'liking' is not inherently consistent or well-defined and may vary between individuals. In defining objective metrics, it must be ensured that these metrics work well, and are tied back to the inherent subjective, psychoacoustic attributes they intend to measure. It must be ensured that these metrics work for all types of halls, whether a common hall geometry or new and innovative form, to ensure a successful design.

To perform this subjective validation, auralization provides a listener with the ability to virtually transport between rooms. The direct comparison of spaces is a powerful tool in performing such studies, but it is essential that the auralizations be realistic and repeatable. Currently, the standards set forth for RIR measurements do not ensure repeatable results between different measurement teams for established metrics, let alone auralization. Differences in source characteristics, microphone characteristics, processing techniques, equalization, and calibration make results between different measurement teams difficult to compare. Beyond reduced comparability between teams, directional limitations of omnidirectional loudspeakers and low spatial microphone resolution can call into question the subjective realism of current practices in concert hall acoustics.

The goal of this study was to generate a RIR measurement database using a spherical microphone array, three-part omnidirectional sound source, and a compact spherical loudspeaker array (CSLA) to satisfy the need for subjective realism and high-resolution objective sound field analysis. A total of 242 RIRs were measured using an omnidirectional sound source and a spherical microphone array in 15 North American and 6 European concert halls. Using state-of-the-art spherical array beamforming analysis, higher resolution spatial analysis of a sound field can enable the calculation of standard metrics and the proposal of new metrics with advanced spatial sensitivity. Next, a 20-channel directionally controllable CSLA was used to reproduce 13 unique musical instrument radiation patterns at 20 source locations on stage, building-in proper instrument directivity into RIR measurement. These measurements enabled the generation of realistic, full-orchestral auralizations. This paper will discuss previous measurement practices, the new Concert Hall Orchestral Research Database (CHORDatabase), and the application of spherical array processing techniques.

5.2 Previous Concert Hall Measurements

The history of room measurements of concert hall acoustics has gradually evolved over the past century. Sabine first brought the idea of an objective measurement of a room's acoustic

properties through his development of reverberation time (RT).[1] Eventually, acousticians felt RT alone did not explain the entire story of subjective perception in rooms. Room impulse response (RIR) measurement techniques were further developed in the 1960s and 70s, and with more than a decay curve alone, many new possibilities for metrics were proposed in the following decades (and even to this day). These included energy ratios between early and late reflections, sound strength in concert halls, and even the isolation of lateral sound with figure-of-eight microphones.[47]

Measurements of room metrics were then paired with subjective data, using interviews as done by Beranek,[17] surveys during live concerts as done by Hawkes and Douglass,[23] Barron,[24] and Kahle,[22] or using reproduced recordings of a live orchestra as used by the group at the Technical University of Berlin (TU Berlin).[89] These conditions ensure realism, but it can be difficult to control non-acoustic factors during live performances, and direct, seamless comparisons of two rooms is not possible. The invention of the binaural mannequin and the virtual acoustic technique of crosstalk cancellation enabled a new way of directly comparing two halls, using either the convolution of anechoic orchestra music with measured binaural room impulse responses (BRIRs) or reproduced binaural recordings. Thus, direct comparisons in a laboratory setup allowed controlled and reliable subjective ratings. Measurements and subjective ratings using binaural techniques had been done by the group from Berlin,[89] the group from Göttingen,[30] and Soulodre & Bradley.[34] The Göttingen group was able to measure 22 halls around Europe using two loudspeakers placed on stage and binaural mannequins as the receivers. A few other groups did more fundamental, controlled laboratory studies using arrays of loudspeakers that were fed with artificial reverberation and discrete delayed early reflections to simulate a controlled room-like effect.[20-21]

After Beranek invited leading consultants to pool-together resources to fund concert hall measurements from the Concert Hall Research Group (CHRG), multiple researchers, including Bradley, Gade, and Chiang, were funded to measure a number of halls in North America.[13] Along with these measurements, Gade and Bradley have pooled-together measurements of 53 different halls, and subsequent comparisons were made between room acoustic parameters and overall geometric hall measurements.[14] Much of this data has also been compiled in Beranek's book, including measurements from many other research and consulting teams.[17] Despite the extensive measurements collected, only a few metric-based comparisons can be made between halls due to setup differences. Also, the measurements do not allow for consistent auralizations across groups.

Even though many advances in the field have been made, these prior studies contain a few consistent limitations. For objective metrics, many room acoustic metrics are sensitive to changes between measurement setups and source properties. Limitations from the non-omnidirectional radiation of standard dodecahedron loudspeakers or starter pistols can cause large changes, preventing compatibility of results.[49] For auralizations, source reconstruction in each study was either uncontrolled, coming from a live performance of an orchestra, or was limited to a small number of loudspeakers (often a single, omnidirectional loudspeaker) on stage. Although binaural mannequins do provide plausible auralizations, they do not allow for flexible post-processing of the spatial sound field.

Much work by the Finnish group led by Tapio Lokki from Aalto University has been done to improve source realism in auralization. The team developed a loudspeaker orchestra, consisting of 33 commercial loudspeakers distributed across the stage to represent 24 orchestral instruments. An effort was made to position and pair commercial loudspeakers to match the directivity of each orchestral instrument. This work has provided significant advances for source realism, representing an orchestra with highly increased accuracy from previous attempts with one or two loudspeakers on stage. It should be noted that musical instruments produce a highly complex, frequency-dependent radiation pattern, which can be quite subjectively different from a commercial loudspeaker.

For spatial sound field capture and subsequent auralization, the Finnish group used a 6-element GRAS intensity probe with 10-cm or a 2.5 cm cross-array microphone spacing. Spatial auralization of the sound field is based upon time-delay of arrival (TDOA) techniques, identifying the arrival of sound energy in discrete time segments of the RIR. This method can identify the directions of arrival of discrete early reflections, but such algorithms are unable to separate two reflections arriving to a listener at the same time. This limitation may prevent accurate auralization of the spatial properties of the mid-to-late arriving reverberant energy. Additionally, the array is only capable of measuring the first-order Ambisonic components of the sound field. First-order representation is a very coarse resolution, which has shown increased errors in sound localization-focused studies.[62] Similarly, beamforming analysis with only first-order components provides very ambiguous spatial energy maps of the sound field, due to the wide main lobe width of a first-order plane wave.

5.3 Concert Hall Selection Process

An online survey was generated to collect suggested candidates for hall measurements. The survey was distributed to many researchers and consultants in room acoustics. Survey

participants were asked to provide recommendations of a wide variety of rooms, both of high and low quality, large and small, reverberant and dry, etc. If available, responders could provide metric results, subjective impressions, and contact information for arranging measurement trips. The survey collected over 70 responses, and in all, there was surprisingly little overlap in suggested halls, resulting in 130 unique hall suggestions.

After removing halls that could not accommodate a full-orchestral stage setup and limiting to cost-effective locations in North America and Europe, 35 candidates were identified. For final selection, values for the seat count, volume, average unoccupied T30, and room shape designation were recorded from existing literature.[16] Halls were assigned one of the following five shape categories pictured in Figure 5-1: historic shoebox (H. Sh.), modern shoebox (M. Sh.), fan, vineyard (Vnyd.), and other. Considering geographic location, acoustic quality, T30, and hall size, a final subset of 30 halls was identified, representative of the overall list of hall suggestions and located within geographic proximity of one another to provide for cost-effective travel arrangements. Each of these halls were contacted regarding measurements, and in the end, measurements were able to be made in 21 halls.

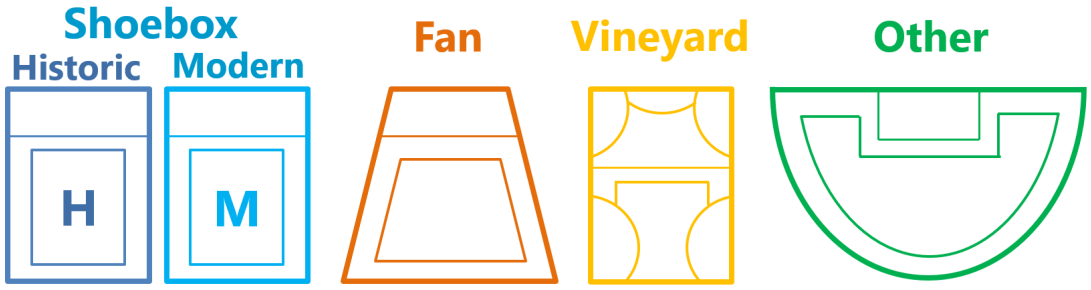


Figure 5-1: Diagrams representing the different categories for assigning hall shapes.

Figure 5-2: The overall shape distribution for the 21 concert halls included in the CHORDatabase. The database includes 15 North American and 6 European halls.

Out of the 21 measured halls, 15 halls were located in North America and 6 were located in Europe. The distribution of the shape categories of the 21 halls is pictured in the color ring in Figure 5-2. A complete listing of the hall information for all 21 halls is provided in Table 5.1. For each hall, a shape designation and its relative size in terms of volume is provided. To maintain anonymity of each hall, only a designation of extra small (XS: 15,000 – 17,500 m^3), small (S: 17,500 – 20,000 m^3), medium (M: 20,000 – 22,500 m^3), large (L: 22,500 – 25,000 m^3), or extra-large (XL: 25,000 – 27,500 m^3) is given. Halls falling outside of this range are given an XXS or an XXL indicator. If variable acoustic settings (VAS) were available (indicated by the VAS column), other settings of the hall were measured. For all halls, the VAS labeled A is the main orchestral setting used by the hall during unamplified classical concerts. Other settings are given a letter B, C, etc. If all acoustic conditions of each hall are considered unique hall environments, a total of 33 hall environments were measured. The mid-frequency (500 and 1000 Hz) unoccupied T30, EDT, C80, and G averaged across all measured receivers is also provided.

To demonstrate the database's variety, Figure 5-3 shows a scatter plot of the hall-averaged unoccupied mid-frequency reverberation time (T30) vs. mid-frequency strength (G) for the 23 hall environments with at least 5 measurement positions. Variety across hall shape is indicated by separate colors and markers on the scatter plot. All of the 242 separate RIR measurements are also indicated by small grey dots. The halls represent a wide range of data, covering a large range of the typical metric coverages suggested in Annex A of ISO 3382, which are represented by the black dashed line. The lack of lower T30 data and higher G values is due to the focus on halls used to study orchestral music, which tend to be medium to large rooms.

Table 5.1: Summary data for each hall included in the CHORDatabase. Each hall is assigned a shape and relative size indication. If multiple hall settings were measured using variable acoustic elements, the variable acoustic setting (VAS) is indicated with a unique letter. The letter A always represents the configuration used for unamplified orchestral performance. Additionally, the number of receivers and the mid-frequency hall average parameters are listed for T30, EDT, C80 and G. The Orch. column indicates if the full orchestra was measured at the R2 location for a given hall environment.

Hall	Shape	Size	VAS	T30 (s)	EDT (s)	C80 (s)	G (dB)	Rec.	Orch.?
1	H. Sh.	S	–	2.50	2.55	-2.76	3.01	11	Yes
2	H. Sh.	XS	–	2.17	2.13	-2.51	2.78	9	Yes
3	H. Sh.	S	–	2.51	2.43	-3.25	2.84	12	Yes
4	H. Sh.	M	A	2.56	2.92	-2.53	3.11	11	Yes
			B	1.97	2.38	–	–	3	No
5	H. Sh.	XXS	A	2.27	2.27	-3.21	5.43	5	No
			B	2.37	2.43	–	–	3	No
6	M. Sh.	S	A	1.61	1.62	0.41	2.97	4	Yes
			B	1.71	1.64	–	–	1	No
7	M. Sh.	L	A	2.73	2.54	-2.79	0.52	12	Yes
			B	1.54	1.74	–	–	1	No
8	M. Sh.	L	A	2.86	2.09	-1.33	2.78	7	Yes
			B	2.38	1.93	-0.76	2.12	6	Yes
			C	2.80	2.11	–	–	1	No
			D	2.74	2.12	–	–	1	No
			E	2.62	2.07	–	–	1	No
			F	2.65	2.05	–	–	1	No
9	M. Sh.	S	–	2.12	2.07	-3.15	1.70	14	Yes
10	M. Sh.	L	A	2.81	2.35	-1.80	3.78	7	Yes
			B	1.78	1.57	-0.28	2.78	6	Yes
11	Fan	XL	–	1.58	1.71	-0.35	-0.06	7	No
12	Fan	S	–	1.79	1.75	0.81	1.60	14	Yes
13	Fan	XXL	–	2.29	2.21	-2.76	-0.16	10	Yes
14	Fan	S	–	1.78	1.76	-0.53	2.43	10	Yes
15	Vnyd.	M	–	2.09	2.06	-0.59	0.88	13	Yes
16	Vnyd.	XS	–	2.09	1.85	-2.34	2.49	5	Yes
17	Vnyd.	S	–	2.21	2.09	-0.40	3.56	15	Yes
18	Vnyd.	XXL	–	3.28	2.80	-1.32	1.01	12	Yes
19	Other	XL	–	1.73	1.44	0.71	0.75	12	Yes
20	Other	M	–	2.41	2.40	-3.84	2.99	11	Yes
21	Other	L	A	2.09	2.00	-2.22	1.44	11	Yes
			B	1.75	1.76	–	–	3	No
			C	1.58	1.68	–	–	3	No
		Totals:	**33**	**–**	**–**	**–**	**–**	**242**	**21**

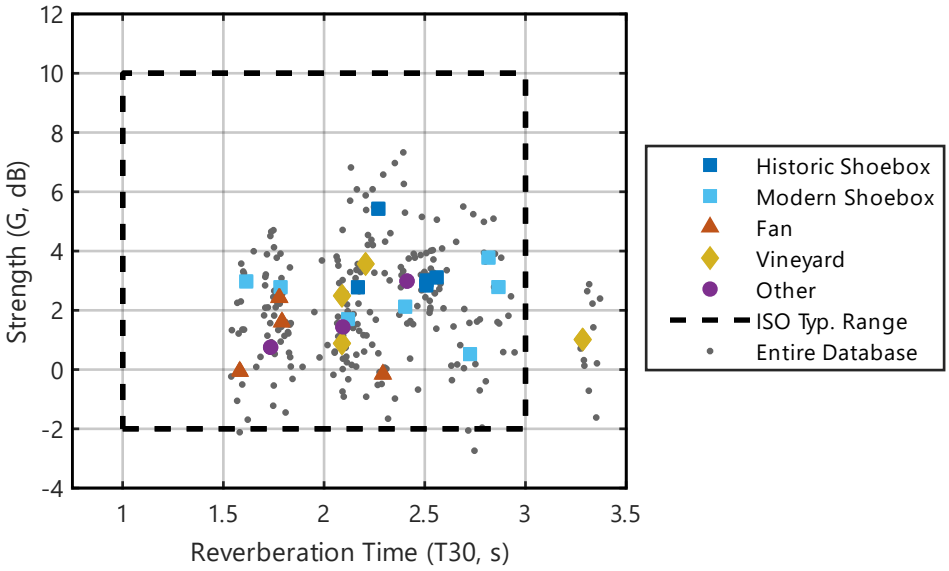


Figure 5-3: A scatter plot showing the mid-frequency hall average values for T30 and G across the entire database, where the dashed line is the typical range of these metrics according to ISO 3382. Hall averages are shown per hall shape as colored indicators, and every individual RIR measurement is indicated by a small grey dot, showing the coverage of the entire database.

It can be noted that there is a slightly higher sampling of the modern shoebox halls, and this is largely due to the more common availability of these halls located in the US. Also, shoebox-style halls are much more common, for they are known to perform well acoustically. Overall, quite a wide range is seen in halls included in the database, spanning from 1.6 to 3.2 mid-frequency unoccupied T30. This range is representative of the natural acoustics of most halls used for orchestral performance. It is also clear that more halls tend to be clustered around the central range of the distribution, from 2.1 to 2.5 seconds. Depending upon a hall volume, this range is quite common for target T30 set points, and it is of no surprise that more halls cluster in this range.

5.4 Measurement Setup

In order to meet both the objective and subjective goals of this database, two different measurement protocols were implemented in each hall. The purpose of the first measurement protocol was to acquire objective measurements, which could be used to calculated measured properties of the sound field. This included standardized room acoustic metric calculations from ISO 3382 and third-order spherical array beamforming analysis of the RIR.[47] This dataset can be used to propose new metrics that could be measured by other research teams or consultants with a comparable setup. The purpose of the second measurement protocol was to acquire subjectively-motivated measurements used to generate realistic, full-orchestral

auralizations. Although these measurements are not the primary focus of this paper's results, they are included here to show the comprehensive coverage of the database.⌗

5.4.1 Measurements for Objective Sound Field Analysis

In each hall, objective measurements were made using a 32-channel spherical microphone array and a three-part omnidirectional sound source. The sound source was comprised of a subwoofer, mid-frequency dodecahedron (similar to commercially available sources), and a high frequency dodecahedron. Each source was designed to provide high signal for measurements over the audible range of human hearing, and the high frequency dodecahedron maintained omnidirectionality up to the 5000 Hz one-third octave band.[52] Full details on these sources is available in section 5.9.1. At each of 242 seat locations, a separate measurement was made with each source, to ensure that the sources were co-located, and to prevent any mutual scattering of the sources found when using a stacked arrangement. After applying appropriate crossover filters, a wide-band omnidirectional RIR was obtained for standard objective analysis of the RIR. This measurement was done using 5-s logarithmic sine sweeps with eight time-domain averages, taking a total of five minutes per receiver.

To provide consistency between each hall, a standard receiver grid was established as shown in Figure 5-4. The source was placed close to the front of the stage, typically 1 m from the stage front, in a "conductor" location, at a height of 1.5 m. This "conductor" position was the relative center of orchestral arrangement described in section 4.6, so if necessary, it was moved to ensure the orchestral layout fit on each hall's orchestra risers. Once the conductor position was set, shown as a red dot in Figure 5-4, a grid of seven seats was determined relative to this point. Receivers R1 – R4 were placed at seats in the center of the hall at distances of 10, 15, 20, and 25 m. These seats provided comparable receiver locations across all halls, despite the geometric variety between halls.

Receivers R5 – R7 were set 5 m from the central seat in the same rows as R2 – R4. Due to differences in seating curvature, these seat locations exhibit differences in SR distances between halls. Apart from this grid, additional receiver seats were set throughout the hall in an effort to obtain a complete sample of all unique seating areas, as time permitted. In each hall, a minimum of four (R1-R4) and a maximum of 15 receiver locations were measured, as shown in Table 5.1. Across all hall environments, the total number of unique individual seats resulted in 242 omnidirectional source RIRs. This database is the largest formal RIR database

⌗ More information on the processing techniques behind these measurements can be found in chapter 4, sections 4.2 and 4.4.

that uses consistent measurement equipment across such a variety of halls. The coverage of the database across both the U.S. and Europe is also not provided in any other database of this extent. Finally, the database contains a level of spatial accuracy that is not common in room acoustic measurements by use of a higher-order spherical microphone array.

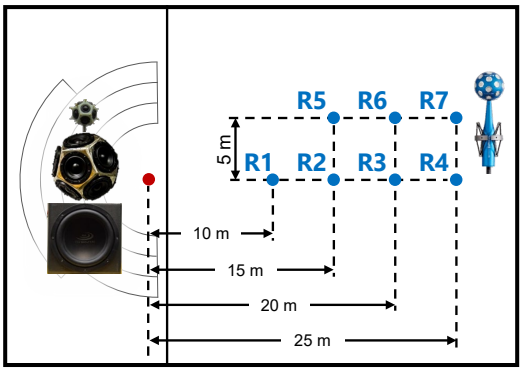


Figure 5-4: Standardized receiver layout for each hall. Receivers R1 – R4 were measured in all halls, and time permitting, receivers R5 – R7 and other unique locations were selected to well-sample seating areas in each hall.

5.4.2 Measurements for Subjective Auralizations

In standard RIR measurements, an omnidirectional loudspeaker is placed at the center of the stage and one single measurement is taken for each source, as described in section 5.9.1. Although repeatable, this method is not a realistic representation of a full orchestra consisting of a distributed array of sources with unique, frequency-dependent radiation patterns. To generate a more realistic auralization, an orchestral source measurement grid was generated for capturing a set of full-orchestral RIRs. The grid consisted of 20 source positions, placed in a consistent location between all measured halls, shown as red circles in Figure 5-5. For each measurement, a CSLA was used to build-in the radiation patterns of specific instruments, assigned to each measurement location.

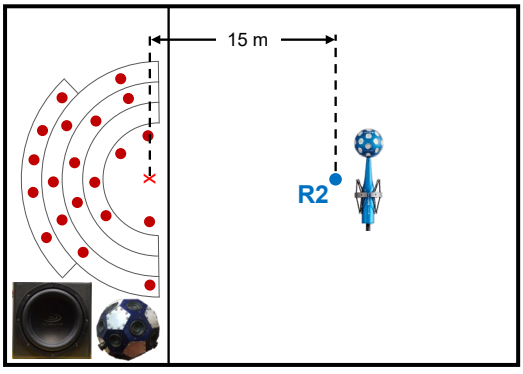


Figure 5-5: Orchestra source distribution, along with the single receiver at which the orchestral measurements were made. This measurement took 1.5 – 2 hours for the single receiver.

The CSLA is an array of 20 individually controllable drivers mounted in a rigid enclosure as described in Chapter 4. For each instrument, a filter bank was designed to process a measurement signal into a 20-channel measurement signal, resulting in a radiation of the measurement signal with the desired directivity characteristics. It is possible to take separate RIR measurements with each individual driver in the CSLA, allowing for flexible post-processing of source directivity, but this technique would increase the measurement time by a factor of 20. In each hall, the microphone array was placed in the 15 m source-receiver distance position, and a separate RIR measurement was taken using the CSLA at each orchestral position. Due to the lower power output of the CSLA, 16 averages were used for a five second sweep signal.

This array can achieve adequate SNR down to 200 Hz, so an additional measurement was taken with the subwoofer component of the omnidirectional source at each instrument location. This extra measurement provided adequate SNR at low frequencies for auralization purposes. This measurement routine took a total of 1.5 to 2 hours for all source positions for both the directional array and the subwoofer. Even with increased efficiency from built-in radiation patterns, the longer time required for this measurement only allowed for one receiver location. After diffuse-field equalization and overall sensitivity difference equalization across all instruments, a crossover filter combined the subwoofer measurement with each individual instrument RIR. Then, each RIR in the orchestral setup was individually convolved with single instrument orchestral anechoic recordings, and after superimposing all auralizations on top of one another with balance adjustments, a realistic, full-orchestral auralization was generated in each hall.

5.4.3 Measurement Software and Hardware Setup

To accommodate RIR measurements with 32 channels of input and 20 channels of output, all acquired simultaneously, a custom measurement hardware and software setup was developed. MATLAB was used to generate logarithmic swept-sine signals with tapered ends for either a single channel or twenty channels, depending upon the sound source being used. If the CSLA was in use, the appropriate instrument-specific filters were used to provide the proper built-in source radiation pattern. To accommodate this multiple input multiple output (MIMO) system with high channel counts, Max7 was used to simultaneously play and record 20 and 32 channels of audio respectively.[88] Measurements were made using MacOS, as core audio drivers allow for the separate specification of input and output devices directly in the Max7 environment. Once the measured sweep signals from the microphone array are recorded, they are saved as 32-channel WAVE audio files and reloaded in the MATALB environment for

subsequent time-averaging, frequency domain transfer function calculation, and inverse transformation back into a 32-channel microphone RIR (MicRIR).

For the equipment setup, a 32-channel spherical microphone array was used along with a custom-built 3-part omnidirectional sound source and a custom-built CSLA. The microphone array has ½" microphones distributed over a rigid sphere ($r = 4.2$ cm). The array has a proprietary hardware interface box, allowing for 32-channel of audio to be sent over a FireWire audio stream. To control each of the different sound sources, a Mark of the Unicorn (MOTU) 24Ao was used to provide 24 channels of digital-to-analog conversion. For each component of the omnidirectional source, a single channel from the MOTU 24Ao was sent to a high-power audio amplifier (Crown XLS 2500). For the CSLA, 20 channels from the MOTU 24Ao were sent to four 6-channel Sure Electronics TDA7498 class-D amplifier boards. This setup provided simultaneous amplification of 20 separate channels of audio. These boards and the MOTU 24Ao were all mounted in a custom hardware box with custom built cabling for connection to the 20-channel source, shown in Figure 5-6. For the European measurements, a power amplifier from ITA at RWTH Aachen was used to power the omnidirectional sound source, and electrical full-frequency transfer function was measured to correct for the amplification differences of the European measurements. The power amplifier had a virtually flat response over the audio bandwidth, but small deviations from a flat response were also corrected using a linear-phase FIR filter.

Due to the high number of powered audio channels running from the hardware box to the CSLA, the cable was limited to 4 meters in length, minimizing potential signal loss, while still providing separation from the measurement array and the hardware box on stage. Absorptive blankets were draped over the hardware box on stage to minimize unwanted scattered energy and provide some attenuation of the amplifier cooling fan noise. Casters allowed for easy movement of the hardware box around the stage with the CSLA. Another MOTU Ultralite AVB interface was used to connect to the 24Ao device over Ethernet, instead of USB, to allow for longer cable runs. These two devices were clock synced over Ethernet, and the Eigenmike interface box was clock synced with the 24Ao over a word clock BNC connection.

(a) (b)

Figure 5-6: A picture of the CSLA hardware box (a) and the CSLA's mobile setup, for easy movement between orchestral source positions (b).

5.5 Spherical Beamforming Analysis of Impulse Responses

Conventional RIR measurements use a single diffuse-field omnidirectional microphone and potentially a figure-of-eight microphone to measure lateral sound energy. Such measurements can provide information of the temporal and spectral properties of the room, along with some information regarding the spatial properties of the RIR. However, with only the lateral dipole component of the RIR, much information about the spatial character of a sound field is lost. It has become quite common to use B-format, or first-order Ambisonic microphones for RIR measurements. These measurements can provide spatial information using TDOA techniques, but for RIR analysis, these techniques are only effective when reflections are separated in time. This assumption may hold true for early reflections, but the increase in reflection density with time causes multiple reflections to arrive simultaneously at mid to later time segments in the RIR.[90] If multiple reflections arrive simultaneously, a TDOA technique will estimate a direction, roughly, as a weighted average of the two reflections. The failure to satisfy the time-separable assumption will create a spatially averaged, potentially unnatural result.

Now, commercially available higher-order microphone arrays are capable of spatially segregating room reflections, even if occurring at the same time, using plane wave decomposition (PWD).[91] This technique allows for the directional response of the microphone array to take on a beam-like directionality that can be flexibly steered in full 3D space. Sections 5.5.1 and 5.5.2 will describe the techniques required to calculate and analyze the spatial components of a RIR using spherical microphone arrays.

5.5.1 Encoding into the Spherical Harmonic Domain

The basis of spherical microphone array processing lies in representing a RIR in the SHs domain. When the three-dimensional wave equation is solved in spherical coordinates, using the separation of variables technique, the directional solution in terms of azimuth, ϕ, and elevation, θ, are known as the set of SH functions:[56]

$$Y_n^m(\theta,\phi) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}}P_n^m(\cos\theta)e^{im\phi}, \tag{5.1}$$

where n and m are denoted as the order and degree of the SH functions and associated Legendre polynomials, $P_n^m(\cos\theta)$. In this definition, azimuth is measured from the frontal x-axis with counterclockwise as positive, and elevation is measured from the vertical z-axis.

The key concept behind this representation of spatial functions lies in the properties of the SH functions. The set of SH functions forms an orthogonal basis for any bounded function on the surface of the unit sphere, such as a sampled pressure field in a room. In other words, we can represent any bounded spatial function (such as a spatially sampled pressure field) as a weighted summation of SH functions. This fact is directly expressed by the spherical Fourier series as:

$$p(k,\theta,\phi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} a_{nm}(k)Y_n^m(\theta,\phi)\,, \tag{5.2}$$

where $a_{nm}(k)$ are the frequency-dependent weighting factors. After multiplication with each SH function and summation across order and degree, these factors reconstruct the frequency-dependent pressure field, $p(k,\theta,\phi)$.

Using a spherical microphone array, the pressure field is spatially sampled in different directions around the sphere, and these samples provide a time-domain RIR at each microphone location. First, a traditional time-frequency Fourier transform of each RIR provides a spatial sample of the room transfer function, now a function of frequency or wavenumber, k. A spherical Fourier transform is taken from the frequency domain transfer function sampled at each spatial microphone location. This step is done discretely for each frequency or wavenumber bin, generating a set of frequency-dependent spatial Fourier weights for each SH function. Sections 5.5.1.1 – 5.5.1.6 will describe the multiple-step processing for calculating a spherical harmonic room impulse response (ShRIR) from a MicRIR.

5.5.1.1 Loudspeaker Diffuse-field Equalization and Three-way Source Crossover Filters

Since separate measurements were made using the three-part omnidirectional sound source, crossover filters were required to combine the three separate MicRIRs into a single broadband MicRiR. Diffuse-field equalization filters were developed for each omnidirectional source to correct for the non-flat response of the drivers, including compensation for the different sensitivities of each driver. Then, corrections for the differences in gain settings of the power amplifier matched the relative sensitivities between all three measured MicRIRs from each part of the omnidirectional source. After the levels had been matched, a linear-phase FIR crossover filter was designed with a response matching that of a 7th order squared Butterworth roll-off, smoothly combining the three MicRIRs into a single, broadband MicRIR. More details on diffuse field equalization and crossover filters can be found in section 5.9.1.

5.5.1.2 Microphone Capsule Gain Adjustments

Next, individual corrections were applied to each microphone channel of the Eigenmike, em32.[49] This correction adjusted for individual sensitivity differences between microphones. Although high quality capsules that exhibit in-phase, flat frequency responses were used, they still contain broadband gain differences between capsules. To calibrate the capsules, multiple low frequency tones were recorded with the microphone array in an anechoic chamber and outdoors for low-frequency, free-field behavior. Frequencies for tones were selected where the scattering and spacing of the microphone capsules produced negligible differences. Capsule sensitivities were calculated from the average between-capsule differences for each tone.

5.5.1.3 Spherical Harmonic Encoding

Next, the broadband MicRIR was *encoded* into a ShRIR using a spherical Fourier transform. Looking at Eqn. 5.2, the measured pressure field, $p(k,\theta,\phi)$ can be represented as an weighted summation of SHs, $Y_n^m(\theta,\phi)$ with weights $a_{nm}(k)$. This summation can also be represented in matrix notation as:

$$\boldsymbol{P} = \boldsymbol{Y_Q}\widetilde{\boldsymbol{A}}\,, \tag{5.3}$$

where $\boldsymbol{P}$ is a $Q \times N_{pts}$ matrix, $\boldsymbol{Y_Q}$ is a $Q \times (N+1)^2$ matrix, and $\widetilde{\boldsymbol{A}}$ is a $(N+1)^2 \times N_{pts}$ matrix. Q represents the number of microphones in the array, N represents the SH truncation order, and N_{pts} is the number of frequency bins from the time-frequency Fourier transform. The tilde symbol is introduced to represent the estimated SH weights, before correction for the array design, discussed in section 5.5.1.4.

In order to solve for the weights, $\widetilde{\boldsymbol{A}}$, a matrix inversion is required. Since often $Q \neq (N+1)^2$, the inversion requires a Moore-Penrose pseudoinverse or singular value decomposition (SVD) to solve for these weights,

$$\widetilde{\boldsymbol{A}} = \boldsymbol{Y}_{\boldsymbol{Q}}^{\dagger}\boldsymbol{P} = \left(\boldsymbol{Y}_{\boldsymbol{Q}}^{\boldsymbol{T}}\boldsymbol{Y}_{\boldsymbol{Q}}\right)^{-\mathbf{1}}\boldsymbol{Y}_{\boldsymbol{Q}}^{\boldsymbol{T}}\boldsymbol{P}\,. \tag{5.4}$$

The matrix $\boldsymbol{Y}_{\boldsymbol{Q}}^{\dagger}$ is referred to as the *encoder matrix* by the Ambisonics community.[84] As long as the microphone array has a nearly uniform sampling scheme, and does not contain any large holes or gaps in the array, the matrix should be well-conditioned. If non-uniform but complete coverage sampling schemes are used, such as equiangular sampling, this process can still be used, but proper adjustment weights for each microphone must be calculated to ensure quadrature. For the current study, the MATALB-based Sound Field Analysis (SOFiA) Toolbox was used to perform the spherical Fourier transform.[91]

5.5.1.4 Radial Filtering

Once the MicRIR has been encoded into a ShRIR, proper equalization must be made to represent this ShRIR as if the measurement was made without the microphone array present. Effectively, corrections are made for the scattering due to the rigid sphere and for the size of the array, sampling at a given radial location, $r = r_a$ (the radius of the array), instead of sampling at the center of the array, $r = 0$. It can be shown that after representing both the incident plane wave and the scattered pressure from the rigid sphere as a summation of SHs, the relationship between the estimated coefficients, $\tilde{a}_{nm}(k)$, and the desired coefficients, $a_{nm}(k)$, is:

$$\tilde{a}_{nm}(k) = a_{nm}(k)b_n(kr)\,, \tag{5.5}$$

where, for a rigid spherical microphone array,

$$b_n(kr_a) = 4\pi i^n\left[j_n(kr_a) - \frac{j_n'(kr_a)}{h_n^{(2)'}(kr_a)}h_n^{(2)}(kr_a)\right]. \tag{5.6}$$

To correct for this effect, radial filters can be designed which match the response of the inversion of this factor, $1/b_n(kr_a)$. As this correction factor can create large required boosting at lower frequencies for higher orders, boosting limits must be placed on the response to ensure filter stability. Again, the SOFiA Toolbox was used to design the radial filters, and built-in soft limiting can be used to control for excessive boosting at low frequencies. Figure 5-7 shows the filters designed for the Eigenmike array, having a radius of $r_a = 4.2$ cm. This boosting sets

practical low frequency limit on the usable range of each order of SH processing, based upon SNR limitations for each measurement.

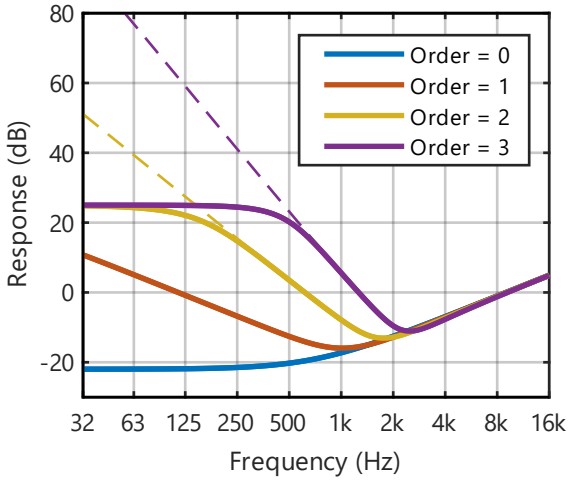


Figure 5-7: The target equalization filters for radial filtering (dashed) and the designed, soft-limited radial filters for a rigid array, $r_a = 4.2$ cm (solid) up to order $N = 3$.

5.5.1.5 Spherical Harmonic Rotation of Look Direction

Once the equalized (radial filtered) SH signals are known, a correction can be made for the orientation of the microphone. The benefit of representing the RIR using SH functions is that the directional response pattern of the microphone can be adjusted to any order-truncated pattern. Along with this flexibility, these patterns can be rotated in full three-dimensional space. For each ShRIR, the direct sound in the ShRIR was time-windowed, to contain only the first 2 ms of response. This window length was found to include all relevant energy in the direct sound without introducing any other first-order reflections, often found to be off of the stage floor. After the orientation of the direct sound was determined, the ShRIR was rotated so that the 'look' direction of each receiver was oriented towards the source. This rotation was performed using a MATLAB-based SH toolbox created by Archontis Politis.[92]

5.5.1.6 Diffuse-field Equalization of Measurement Microphones

Finally, the first channel of the ShRIR can be extracted, and this channel is equivalent to the omnidirectional RIR, as measured with a diffuse-field omnidirectional microphone. Although corrections have been made for the scattered pressure around the array from the radial filtering step, non-flat responses can still exist, due to non-ideal effects of the array design and directional response effects of the microphone array above its spatial aliasing limit of 8 kHz. The spherical microphone array was placed in an anechoic chamber, and IR measurements were made at all directions around the microphone, covering the full sphere. A reference measurement was also made with a ½" free field microphone, and it was used as the

denominator in the transfer function calculation to remove the on-axis response of the loudspeaker from the measurement. Each measurement was processed using the steps in sections 5.5.1.2 – 5.5.1.5, and a spatial average of the microphone's TF was determined using a surface integral of the omnidirectional response. This response was inverted, and a minimum-phase FIR filter was designed to provide a final correction of the non-flat response of the omnidirectional, zeroth-order SH component of the array. More information about this filter design can be found in section 5.9.2.2.

5.5.2 Beamforming Analysis using Plane-wave Decomposition

Once a ShRIR is calculated, the different SH components of this spatial representation can be weighted and summed to flexibly control the directional response of the array. One very common technique for analyzing the spatial properties of a room is known as plane-wave decomposition (PWD).[91] This technique breaks down a spatial function into its plane wave components for directions of arrival all around a sphere. A schematic diagram of the present technique for applying PWD to a MicRIR is shown in Figure 5-8. This process is used to generate a spatial energy map over any given time range or frequency range of a RIR. The five steps highlighted in this map are each explained in sections 5.5.2.1 through 5.5.2.5, respectively. These include (1) a spherical Fourier transform and radial filtering, (2) application of plane wave coefficients and SH rotation, (3) time windowing, (4) bandpass filtering, and (5) energy integration.[§§]

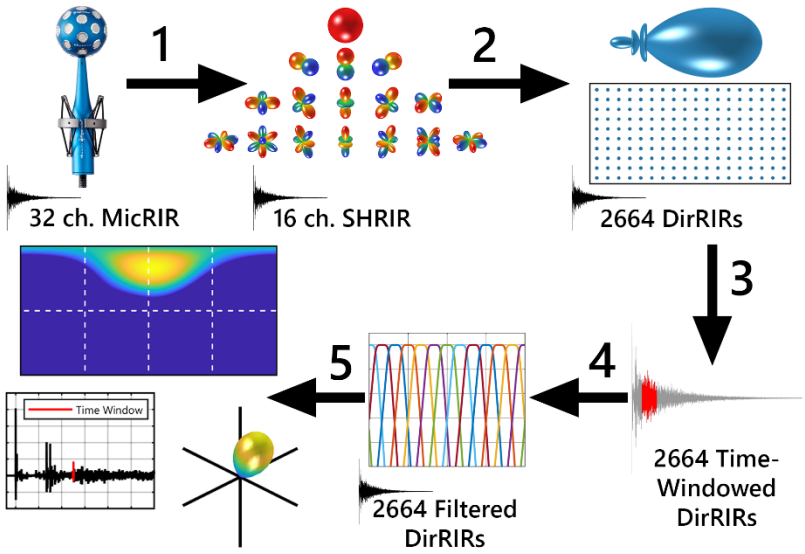


Figure 5-8: A schematic process diagram for the beamforming analysis explained step-by-step throughout section 5.5.2, where MicRIR is the microphone room impulse response, ShRIR is the spherical harmonic RIR, and DirRIRs is the directional RIRs.

[§§] An animation of planewave decomposition created for presentations by the authors can be found online at: https://sites.psu.edu/spral/files/2019/06/Beamforming-Map.gif.

5.5.2.1 Spherical Fourier Transform (encoding)

This step is the transformation from the MicRIR to the ShRIR. This step includes all of steps 1 – 6 described in section 5.5.1.

5.5.2.2 Generating Directional Room Impulse Responses (DirRIRs)

Once a ShRIR is calculated, the individual SH components of the sound field can be weighted and summed to produce a beam-shaped directivity pattern. The most common beam shape is identical to an order-truncated plane wave.[81] This beam shape maximizes the directivity index (DI) of the possible patterns for a given SH truncation order. Effectively it makes the most spatial impulse-like beam possible for a given order. For an ideal, infinite-order plane wave, the beam becomes a spatial Kronecker delta function, responding to energy only arriving from a single direction. In practice, an order-truncated plane wave has a main beam width and side lobes, artifacts of the SH order truncation, but as truncation order, N, increases, the beam width becomes smaller, and the side lobe amplitudes decrease.

Once a beam pattern has been generated in one direction, the response of the beam can be rotated in the SH domain, to generate a directional RIR in all directions around the sphere. Conceptually, it can be thought of as generating a RIR in each direction, with the microphone having a beam-shaped directivity oriented in each particular direction. This processing results in a two-dimensional grid of DirRIRs. Both of these steps for beam generation and rotation are performed by applying SH weights, which can be formulated into weights for axis-symmetric beamforming, w_{nm}, such that:

$$w_{nm} = d_n Y_n^m(\theta_l, \phi_l)\,, \tag{5.7}$$

where to generate an order-truncated plane wave, oriented at the look direction $(\theta_l, \phi_l) = (0,0)$, with an on-axis amplitude of 1,[81]

$$d_n = \frac{4\pi}{(N+1)^2}\,, \tag{5.8}$$

and $Y_n^m(\theta_l, \phi_l)$ is a term that steers the axis-symmetric beam pattern in other look directions.

Along with the generation of plane-wave shaped beam patterns, other optimizations of beam shapes can be used. For the current study, Dolph-Chebyshev beam patterns are used, to minimize the interference of side lobes on the final spatial energy maps.[93] When a single reflection is present, the effect of side lobes are less pronounced, but when multiple reflections are present in a sound field or RIR, side lobe interference and interactions can cause errors in

the relative amplitudes of individual reflections or generate artifacts when multiple side lobes sum together, appearing as nonexistent reflections. Using Dolph-Chebyshev beam patterns, side lobe amplitude, L_{SL}, is a parameter that can be used to tune the beam shape for any given truncation order, N. Practically, these coefficients can be simplified into order-dependent weights, $d_n(L_{SL}, N)$, with a mathematically intricate formulation, determined by equating the plane wave beam shape to Chebyshev polynomials with additional order-dependent factors. The spherical array beamforming was completed using the SOFiA Toolbox in MATLAB with computationally efficient externals written in C and C++.[91]

5.5.2.3 Time-windowing of DirRIRs

Once a grid of the DirRIRs was generated, specific time regions of the IR were isolated using a rectangular time window. Just as common room acoustic metrics are used to analyze the early or late components of a RIR's energy, time windowing can isolate the early component of each DirRIR in the beamforming analysis. Not only can large averages of just early (typically 0 to 80 ms) or late (80 ms to ∞) energy be analyzed, but even small, 1 ms width time windows can isolate the spatial arrival direction of individual early reflections. For the representation of time energy in small windows, it is important to first time-window the DirRIRs, before bandpass filtering, as the bandpass filters spread energy in the time domain, especially at low frequencies. This spreading of energy will distribute energy from a single reflection over wider time range, based upon the IR of each octave band filters.

5.5.2.4 Bandpass Filtering of Time-windowed DirRIRs

Once the full set of DirRIRs had been consistently time-windowed, a bandpass filter was applied to each RIR to isolate a particular frequency range of the RIR. This analysis can be done on the broadband DirRIR, but often, large boosting at low frequencies and high SH orders limits the lowest frequency of accurate directional representation. Additionally, the spatial aliasing limit of an array will set a high frequency limit on directional performance. For the current study, third-order DirRIRs were used with band limits set to include the 1 – 4 kHz octave bands.

5.5.2.5 Energy Integration and Normalization of DirRIRs

Finally, each time-windowed, bandpass filtered DirRIR can be calculated by integrating the squared pressure of each DirRIR, summed to a single energy value for each direction. These energy values can then then be converted to decibel quantities, and typically, they are normalized to the maximum energy in the spatial map. To visually demonstrate these styles of plots, a 2D **unwrapped** spatial energy map of a reflection arriving from the ceiling is shown

in Figure 5-9. Effectively, this image is a heat map of the spatial energy in a room over a given time and frequency range. The horizontal directions are labeled for visual interpretation as front (F), left (L), right (R), and back (B). The top (Up) and bottom (Down) are spread over the entire top and bottom of the plot, due to the visual artifact of this representation of the data.

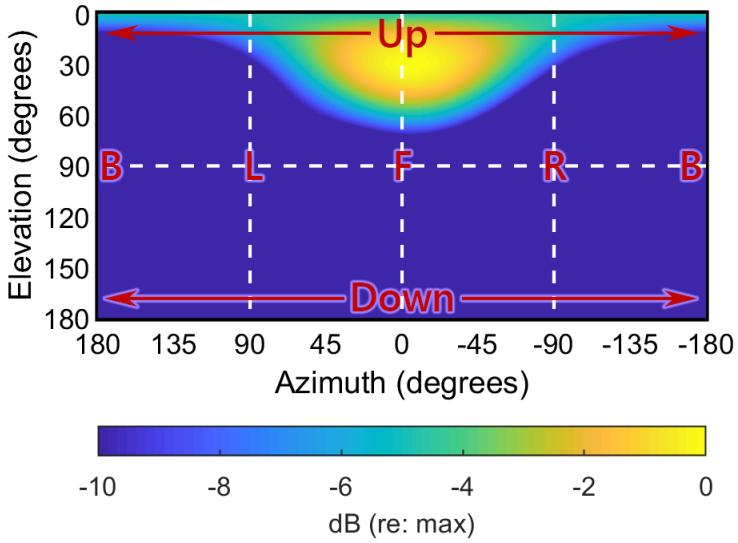


Figure 5-9: Example of a two-dimensional (2D) beamformed spatial energy grid for a RIR. This grid was created using a 1 ms window for a single reflection. The front (F), left (L), right (R), and back (B) directions have been labeled. Due to the unwrapping of a spherical function onto a grid, the single up and down points are stretched across the top and bottom of the plot, a visual artifact of this style of representation. This artifact is not seen in the balloon-style representation in Figure 5-10.

Since this map is made by unwrapping a spherical function onto a rectangular plot, it suffers similar issues found when generating 2D plots of a globe. This data can also be presented on the surface of a sphere, or in balloon-style, 3D polar plots as well. Figure 5-10 shows the same data from Figure 5-9, now represented in a balloon style plot.

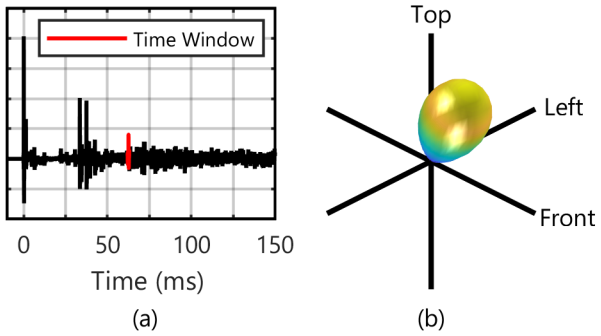


Figure 5-10: The same as in Figure 5-9, now showing the time-domain windowed RIR (a) in red and a balloon-style beamforming plot (b). This 3D representation is more intuitive than the 2D plot, and although direction of arrival is more difficult to precisely identify, no unwrapping artifacts occur.

5.5.3 Calculation of Standard Room Acoustic Metrics

Along with more advanced spatial analysis, spherical microphone arrays can be used to estimate both the omnidirectional and figure-of-eight RIR to calculate standard room acoustic metrics. Spherical microphone arrays have even been found to be more accurate at measuring lateral sound metrics, such as J_{LF} and L_J, since both responses are simultaneously measured around the same point, and the null of the figure-of-eight response can be rotated in post-processing to directly face the omnidirectional loudspeaker.[94] The omnidirectional microphone response can be extracted from the first (zeroth-order) channel of the ShRIR, and the figure-of-eight response can be extracted using the proper beamforming weights. If represented as real-valued SH functions using the Ambisonic channel number ordering convention, the figure-of-eight response corresponds to the second channel of the ShRIR. It is important to remove the SH normalization convention used in defining the ShRIR, so that the relative amplitudes between the omnidirectional and figure-of-eight channel is matched.

5.6 Results

From the database, MicRIRs were processed to calculate both standard room acoustic metrics and spatial energy maps of the early and late sound fields. This section will first provide distributions of standard metric calculations across all 242 RIR measurements across all hall environments and seat locations within the database. Distributions of each parameter at low-, mid-, and high-frequencies will be shown for all parameters from ISO 3382 and a few more not in the standard.[47] Then, beamforming analysis of discrete, early reflections are provided for one measurement as an example. Finally, beamforming analysis of average early energy and average late energy for different hall geometries will be provided to show the physical intuition of this method of RIR analysis for different hall geometries.

5.6.1 ISO 3382 Metric Distribution Across the Database

First, overall descriptive statistics were calculated for each of the metrics provided in ISO 3382. All RIRs were calculated according to the methods provided in the standard. To ensure that noise did not impact the calculation of room acoustic metrics, the noise floor of the octave-band filtered RIRs was replaced with octave-band filtered noise that was matched in slope to the RIR. More details on the specifics of this RIR cleaning procedure and metric calculations can be found in section 5.9.

Table 5.2 shows the mean, standard deviation, maximum, and minimum for eight metrics: reverberation time (T30 and T20), early decay time (EDT), clarity index for music

(C80) and speech (C50), center time (Ts), strength (G), lateral energy fraction (J_{LF}), and the late lateral energy level (L_J). All metrics are presented using the broadband averaging listed in ISO 3382 (most often the arithmetic mean of the 500 and 1000 Hz octave bands). As measured RIRs were used, some metrics occasionally resulted in extreme values, so outliers at a distance greater than 2.5 times the interquartile range were removed from the descriptive statistic calculations. These extreme values could be due instabilities in the metric's calculation procedure or non-ideal results in the cleaning procedure of the RIR. The removal of the outliers prevented bias in the estimates for means and standard deviations for each metric. Table 5.2 presents the number of points deemed outliers for each metric, which may indicate how robust certain metrics are to producing extreme or impractical results.

Table 5.2: Descriptive statistics for the metric calculations across all 242 source-omnidirectional RIRs. Outliers were defined as having a distance from the mean exceeding 2.5 times the interquartile range.

Metric	Mean (ISO)	Std. Dev. (ISO)	Min. (ISO)	Max. (ISO)	# Outliers			
					Low	Mid	High	ISO
T30 (s)	2.25	0.42	1.54	3.37	0	0	0	see Mid
T20 (s)	2.23	0.42	1.53	3.46	0	0	1	see Mid
EDT (s)	2.13	0.41	1.20	3.33	0	0	3	see Mid
C80 (dB)	-1.99	1.82	-6.29	3.48	0	0	0	see Mid
C50 (dB)	-5.21	2.19	-10.05	0.81	0	0	0	see Mid
Ts (ms)	158	32	92	257	0	0	1	see Mid
G (dB)	2.10	1.88	-2.74	7.33	1	0	0	see Mid
J_{LF}	0.19	0.10	0.01	0.55	11	5	6	5
L_J (dB)	-4.75	2.46	-11.97	1.17	0	9	1	0

Figure 5-11 shows the statistical distribution of each metric using violin-style plots. To enable visual representation, all metrics have been mean-centered and normalized to the standard deviation values from Table 2. Additionally for each metric, the calculations are now provided for low-, mid-, and high-frequency ranges, corresponding to 63-250 Hz, 500-1000 Hz, and 2000-4000 Hz respectively. Each metric was normalized to the mid-frequency values, except for J_{LF} and L_J, which are defined to be broadband averaged over the range from 125-1000 Hz.

By visual inspection of Figure 5-11, all metrics have a large range of coverage across the database, also seen in the minima and maxima values from Table 5.2. All metrics mostly cover the typical ranges provided in Table A.1 of Annex A in ISO 3382.[47] The ranges of EDT, C80, D50, and Ts associated with smaller, more dry halls is not quite covered, as smaller rooms did not accommodate the full orchestra stage setup required for this study. Along with coverage, clear expected relationships between each frequency range emerges. In general, T30 and EDT are always lower at high frequencies, and tend to be slightly larger at low frequencies. The low frequency emphasis is much less pronounced for EDT than for T30. C80 and D50 follow a

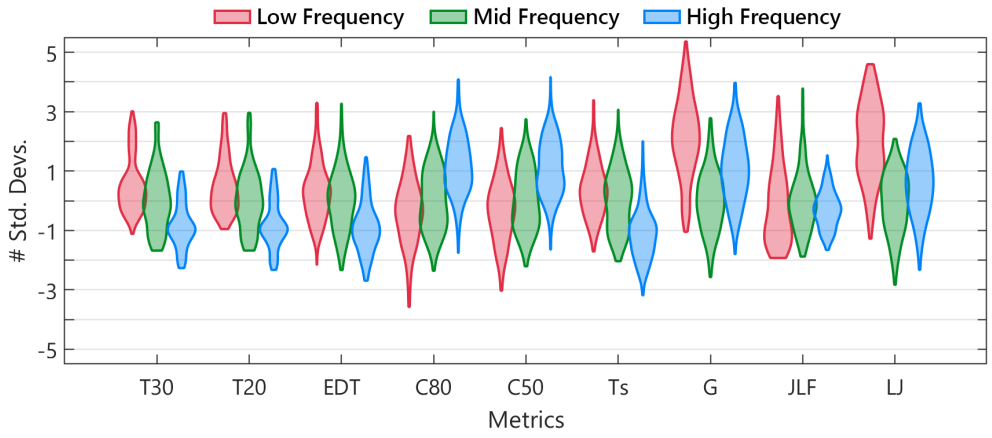


Figure 5-11: Violin plots showing the low- (63-250 Hz, red), mid- (500-1000 Hz, green), and high-frequency (2000-4000 Hz, blue) distributions for each metric. The width of each plot is normalized to the maximum, and the shapes show the relative distributions. The metrics were mean-centered and normalized to the standard deviations of the mid-frequency measurements calculated in Table 5.2 for visual representation. The distributions against their original *y*-axes are shown as histograms in Figure 5-12, providing a clear indication of their ranges.

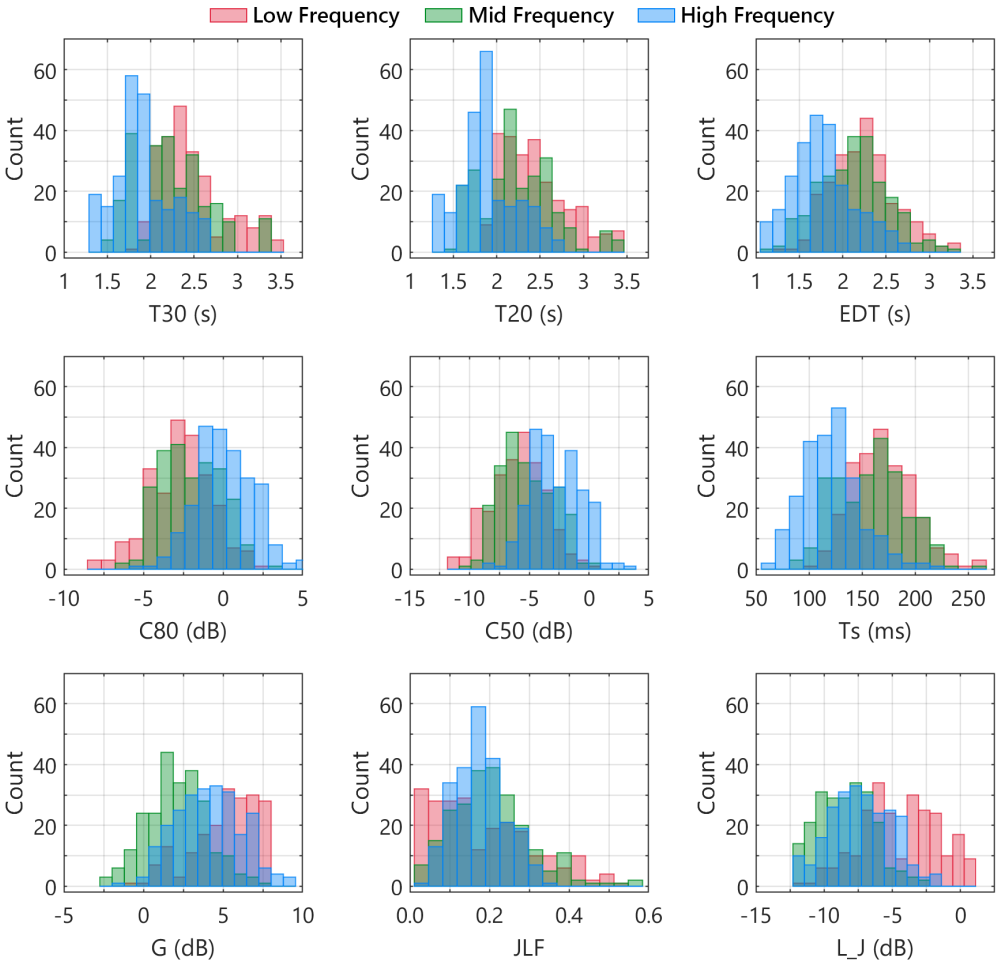


Figure 5-12: Histograms for the low- (red), mid- (green), and high-frequency (blue) distributions of the 242 RIRs in the CHORDatabase. These represent the same distributions shown in Figure 5-11, shown as histograms instead of violin plots.

reverse trend. Some of the metrics exhibit distributions that exhibit skew, including J_{LF}, D50, and Ts. A potential bimodal distribution seems to appear for T30 and T20. In the end, very little variation occurs within a hall for T30 and T20, so the effective sample size is only as large as the number of independent room settings, $n = 33$. Thus, this distribution would likely approach a normal distribution with the inclusion of more halls. Distributions for all of the other metrics vary considerably within the same hall, and appear to be well sampled, having a true effective sample size of $n = 242$. The frequency-dependent distributions are also shown as traditional histograms in Figure 5-12. The same colors from Figure 5-11 represent the low-, mid-, and high-frequencies.

5.6.1.1 Correlation Analysis between Metrics

Along with descriptive statistics, it can be helpful to analyze the correlation between different room acoustic metrics. Correlation can help determine, in general, which metrics are related, and which metrics provide new or independent information. This database, with a high sample size, enables much larger statistical power that has been possible in previous databases for correlation between standardized metrics. This consistency is especially important for metrics that are highly sensitive to source directivity, such as C80, D50, EDT, and J_{LF}.[49] Table 5.3 presents the Pearson correlation coefficients between metrics that were significantly different from zero.

Table 5.3: Pearson correlation coefficients for metrics calculated using the 242 measured RIRs. Coefficients found to be not significant ($p \geq 0.05$) are indicated with an *n.s.* Higher values, falling above a Person correlation coefficient of 0.5 have been bolded for visual emphasis.

	T30	T20	EDT	C80	C50	Ts	G	J_{LF}
T20	**0.99**	–	–	–	–	–	–	–
EDT	**0.79**	**0.81**	–	–	–	–	–	–
C80	-0.39	-0.40	**-0.59**	–	–	–	–	–
C50	-0.28	-0.30	-0.39	**0.84**	–	–	–	–
Ts	**0.67**	**0.70**	**0.85**	**-0.88**	**-0.76**	–	–	–
G	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	–	–
J_{LF}	0.35	0.34	0.27	-0.30	-0.22^	0.32	*n.s.*	–
L_J	0.32	0.33	0.36	-0.30	-0.16*	0.32	**0.66**	0.47

If not indicated, $p < 0.001$. Otherwise, *^indicates $p < 0.01$, *indicates $p < 0.05$, & n.s. indicates $p \geq 0.05$.*

As might be expected, high degrees of correlation exist between the standard room acoustic metrics. First, many of the early-energy based metrics, including EDT, C80, C50, and Ts, demonstrate very high degrees of correlation. Beranek reports correlations from measurements in 42 concert halls, with in general, much higher correlations between RT, EDT, and C80.[17] The current study found a correlation between T30 and EDT of 0.79, compared to Beranek's finding of 0.99. For T30 and C80, the current correlation of -0.39 is far reduced from

Beranek's finding of -0.84. EDT and C80 have a correlation of -0.59 compared to Beranek's -0.88. It is not reported if Beranek's analysis was performed on hall average parameters or individual seats measurements, which may explain some differences. In general, the current dataset suggests that C80 provides new information not explained by T30 and EDT, contrary to Beranek's suggestion.[17] Most notably, Ts is highly correlated with all parameters, except G and the lateral parameters, J_{LF} and L_J. Both lateral parameters show relatively little correlation with other parameters, except the expected correlation between G and L_J.[32,95] Although not as widely measured, due to the need of a separate figure-of-eight RIR measurement, lateral energy metrics appear to contain new information not explained by the other parameters.

5.6.2 Spherical Array Beamforming Analysis Results

5.6.2.1 Identification of Individual Reflections

First, the beamforming analysis techniques described in section 5.5.2 can be used to identify discrete early reflections in small time windows. Using the SOFiA Toolbox, third-order spatial energy maps from the 1 – 4 kHz octave bands have been generated for 1 – 2 ms time windows for a single ShRIR, to demonstrate such results. For example, reflections measured at a seat in the first balcony of hall 3 have been visualized in Figure 5-13.[91] These short time windows highlight early reflections using Dolph-Chebyshev beam shapes with 15 dB side lobe rejection. Such a detailed analysis allows for the identification of discrete early reflections, including both highly pronounced reflections, like the rear-ceiling reflection at 35 ms in Figure 5-13 (d), and even less pronounced reflections, like the front-ceiling reflection at 16 ms in Figure 5-13 (c). Additionally, even when multiple reflections occur at the same time, shown in Figure 5-13 (f) and (g), this type of analysis is capable of identifying both. Most all previous spatial RIR analysis techniques rely upon TDOA calculations, which cannot separate two reflections that overlap in time due to spatial processing limitations.

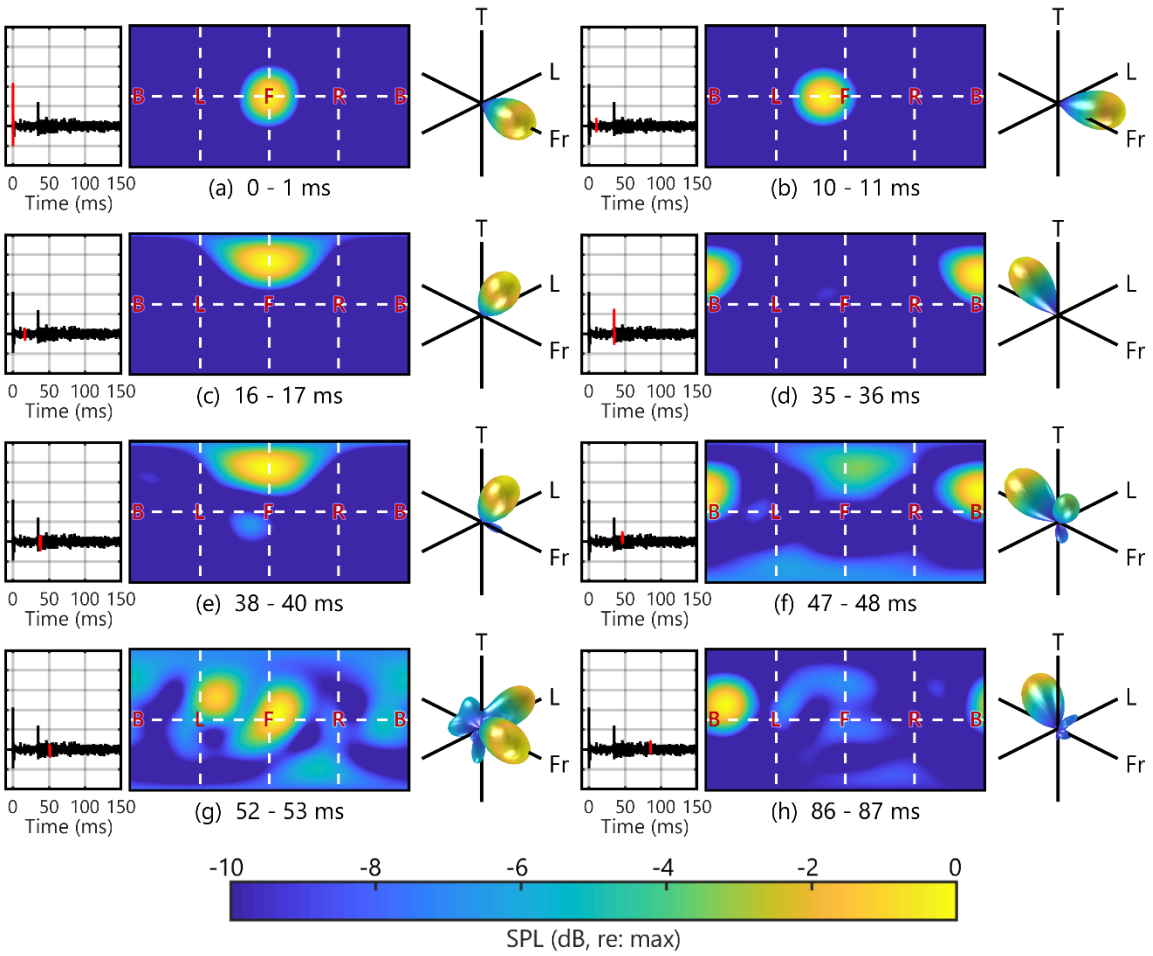


Figure 5-13: Spatial energy maps of individual early reflections for a receiver in the first balcony of hall 3. Individual 1 ms time windows allow for clear identification of individual reflections, even when they overlap in time.

5.6.2.2 Hall Comparisons using Energy Averages for Early and Late Time Windows

Spatial energy maps can also be used to compare the average energy across larger time ranges in the ShRIR. Spatial energy maps generated for two fan, two vineyard, two historic shoebox, one modern shoebox, and one non-standard hall shape, shown in Figure 5-14. The maps show the average early spatial energy from 5 ms to 100 ms using Dolph-Chebyshev beams with 25 dB side-lobe rejection. Plots for the later reverberant energy from 100 ms to 1000 ms in the same halls is also provided in Figure 5-15. All analyses are for the R2 seat position (15 m).

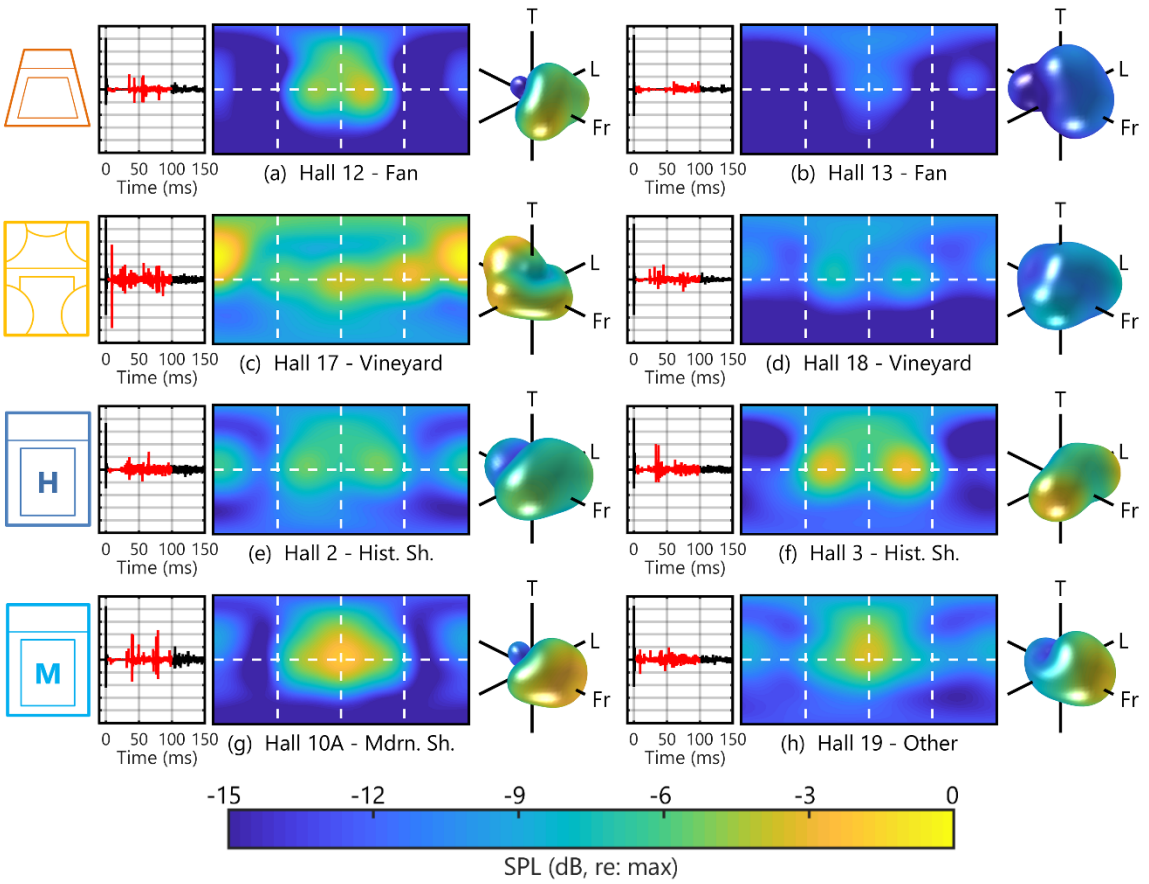


Figure 5-14: Spatial energy maps of the average early reflections for 8 different halls from Table 5.1.

From these maps, some intuitive results appear that match classical wisdom. For both early and late energy, a lack of lateral energy is found for fan-shaped halls (a and b), which is especially apparent in the later reverberant energy, shown in Figure 5-15. Although the early energy for fan-shaped hall number 13 (Figure 5-14, b) appears to have some spatial character, the overall early energy in this hall is much less than any of the other eight halls. This effect can be seen in the overall magnitude of the time-domain RIR, but as each spatial energy map is normalized to its maximum direction, the spatial maps do not visually indicate overall level differences. Conversely, the historic shoebox halls have pronounced early energy (Figure 5-14, e and f), along with containing very strong lateral reflections in both early and late energy. The modern shoebox hall also contains lateral energy it its early and late reflections, but this energy is not quite as lateral in the early energy and not quite as diffuse in the later reverberation.

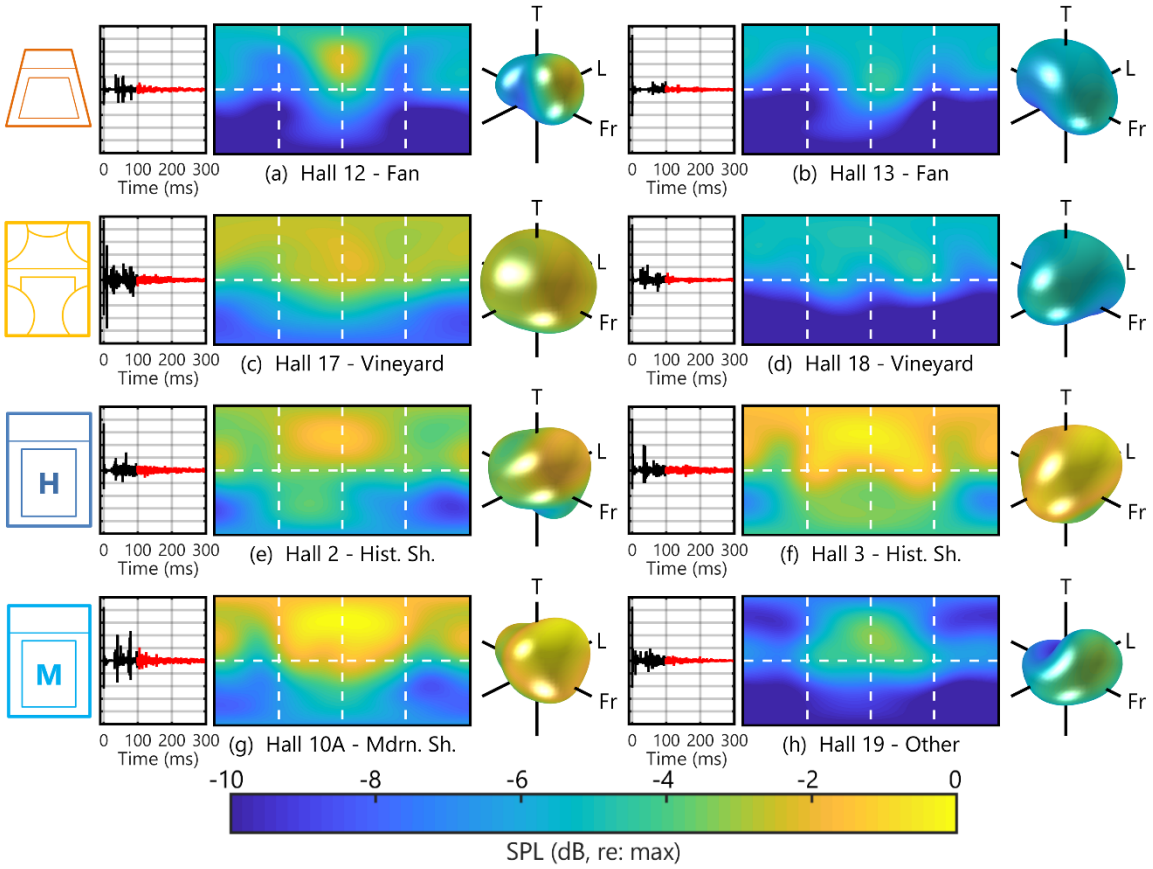


Figure 5-15: Spatial energy maps of the average late reverberation for 8 different halls from Table 5.1.

The vineyard halls exhibit high uniformity in later spatial energy (Figure 5-15, c and d), similar to the historic shoebox halls, but they also exhibit the most uniformity in their early energy as well (Figure 5-14, c and d). Although the early spatial energy map for hall 17 has a strong upwards, rear component, this is due to the strong reflection occurring only 10 ms after the direct sound (Figure 5-14, c). In the design of vineyard halls, walls of adjacent seating pods are often shaped to optimize reflection coverage to each seating location. This optimization appears to be evident in the nice spatial distribution of early reflections in both halls 17 and 18. Such a distribution is not present in either fan shaped hall (halls 12 and 13, a & b) and the non-standard hall (hall 19, h). The intuition that results in these comparisons helps to demonstrate the utility of this analysis. Such techniques could be used to derive new spatially informed metrics, which connect psychoacoustic judgements with time, frequency, and spatial regions that best correlate with perception.

5.6.2.3 Time Animation of the Spatial Room Impulse Response

Finally, just as discrete plots can be made using these techniques, small time windows, as were made in Figure 5-13, can be stitched together into a video animation of the RIR. An example of such a video has been generated and can be found in Appendix C of this dissertation

or at https://sites.psu.edu/spral/animations-and-data/beamforming_video/. In this video, a small 1-ms wide time window (red) slides through the ShRIR in 0.25 ms time steps. As it moves, balloon-style plots show the spatial analysis of the sound field. For each balloon, the radius of the balloon shows the amplitude in the given time window, normalized to the peak spatial level in the current time window. As the energy in each time window can contain quite drastic changes, especially early in the RIR, the color of balloon plot corresponds to the amplitude relative to the direct sound in the RIR. This representation provides a method to visualize a large amount of temporal and spatial information in a single ShRIR. Such a tool could be highly useful in understanding a room's sound field in the time and spatial domains, simultaneously.

5.7 Conclusions and Future Work

A comprehensive measurement database for a wide variety of concert hall shapes has been established using measurements in 21 concert halls. With variable acoustic elements, 33 unique hall environments were measured. Using spherical array processing techniques, spatially accurate processing was applied to both the source and receiver sides of the measurements. Using a 20-channel CSLA, radiation directivities were built into a standard measurement grid of 20 orchestral source locations, measured at a single seat in each hall. This measurement provides a basis for a highly realistic and repeatable full-orchestral auralization across 19 different halls and two additional variable acoustics settings. A spherical microphone array allows for spatially accurate auralizations of these measurements.

A three-part omnidirectional loudspeaker allowed for repeatable, standardized measurements from the 63 – 4000 Hz octave bands. Measurements in a number of seat locations were made with the omnidirectional source and spherical microphone array, totaling 242 unique RIRs. Using these measurements, standard room acoustic metrics have been calculated from ISO 3382.[47] The wide distribution in calculated room acoustic metrics demonstrates the wide variety included in the database. With high numbers of consistently measured RIRs, correlation analyses were performed between room acoustic metrics. High degrees of correlation were found between metrics, indicating which metric provide unique, new information and which metrics appear to contain information mostly explained by the other parameters. The highest correlations of metrics were found between T20 and T30 and between all of the clarity-based measures. The lateral energy-based metrics showed the most independence from other metrics.

Additionally, the usefulness of plane wave decomposition techniques was demonstrated in the generation of spatial energy maps of RIR time regions. Larger time windows of RIRs can provide details for the average spatial character of the early energy in a specific hall, or the average character of the later reverberant energy. Comparisons were made between eight different halls for both regions in the RIR, and results matched with classical understanding regarding hall shapes. Fan-shaped halls demonstrated a lack of early and late energy from the lateral directions, while historic, well-liked shoebox halls exhibited strong early and late lateral energy. Vineyard-style halls exhibited a good distribution of late energy, and the spatial character of the early reflections was unique to the specific design of each hall.

Finally, individual early reflections can be visualized using these techniques over small 1 – 2 ms time windows, even capable of identifying multiple reflections in the same time window. Using these small windows, visual animations were generated by stitching together plots into a time-based video. This visualization presented the complete spatial analysis of the RIR in a highly intuitive and useful manner. Clear potential can be seen in these new spatial analysis techniques for understanding the spatial character of a sound field, defining new metrics using spherical array beamforming techniques, and generating high spatial resolution auralizations for subjective assessment.

5.8 Acknowledgments

The authors would like to acknowledge Martin Lawless, Fernando del Solar Dorrego, Zane Rusk, Andrew Kinzie, Andrew Doyle, Nick Ortega, Peter Moriarty, Will Doebler, Molly Smallcomb, Pranay Muchandi, Tom Blanford, Mark Langhirt, Nathan Tipton, Vahid Naderyan, and Maryam Landi for assistance during the concert hall measurements. Special thanks to the many researchers and consultants who assisted in contacting many of the halls and all of the hall staff and employees who helped arrange and facilitate the measurements. Special thanks to Ingo Witew, Marco Berzborn, and Prof. Michael Vorländer from ITA at RWTH Aachen for equipment assistance prior to the European measurements. This work was supported by the National Science Foundation (NSF) Award #1302741.

5.9 Additional RIR Processing and Auralization Details

***Note**: The purpose of this section is to provide additional details and specifics for the processing and equalization of the measured RIRs included in the CHORDatabase. These details are not intended for inclusion in the manuscript to be submitted to a peer-reviewed journal but are presented here for comprehensive documentation. These descriptions will provide further details on diffuse-field equalization of both loudspeakers and microphones, the RIR cleaning techniques, the HOA auralization methods, and some additional details on the beamforming analysis.*

5.9.1 Omnidirectional Sound Source

In room acoustic measurements, it is common to use an omnidirectional dodecahedron loudspeaker with twelve 4-inch drivers, installed in a rigid enclosure. This design ensures good signal-to-noise in measurements down to the 125 Hz octave band. Due to the spacing of the drivers, the source begins to exhibit spatial aliasing around 1 – 2 kHz, resulting in a flower pedal-shaped directivity pattern. The source still produces energy up to higher frequencies, through the entire audible range, but the non-omnidirectional source directivity will impact the measurement, especially the early part of the RIR.[49] These effects are most noticeable for parameters that depend upon the early part of the IR, including C80, C50, EDT, and J_{LF}. Depending on the frequency-specific placement of the lobes and nulls in the loudspeakers directivity, early reflection amplitudes will shift as a result of these reflections. As a more diffuse field is reached later in the RIR, this impact is less noticeable.

(a) (b) (c)

Figure 5-16: The three-part omnidirectional loudspeaker used in this measurement database, consisting of a low- (a, 40 – 120 Hz), mid- (b, 120 – 1300 Hz), and high-frequency (c, 1300 Hz – 20 kHz) component.

To provide as broadband a measurement as possible, a three-part omnidirectional loudspeaker was constructed. The three components consisted of two 12-sided dodecahedron loudspeakers and one 2-driver element subwoofer, all shown in Figure 5-16. The mid-frequency dodecahedron contains 4-inch full range drivers, the high-frequency dodecahedron contains 12 ¾-inch drivers, and the subwoofer is a one-foot cube with two 10-inch subwoofer drivers

mounted on opposite sides of the enclosure. Dick provides a thorough description of this sound source in Appendix B of his dissertation.[52] He demonstrated the impact of source rotation on each of the three-part omnidirectional source components, showing changes in early and late energy in the measured RIR. It was found that the differences in late energy were within ±1 dB across frequency, while differences in early energy were found to be as high as ±10 dB as frequency increased past the source's omnidirectional limits. Differences in the late and early energy are shown in Figures 5-17 and 5-18, repeated from Figures B-11 and B-12 from Ref. [52].

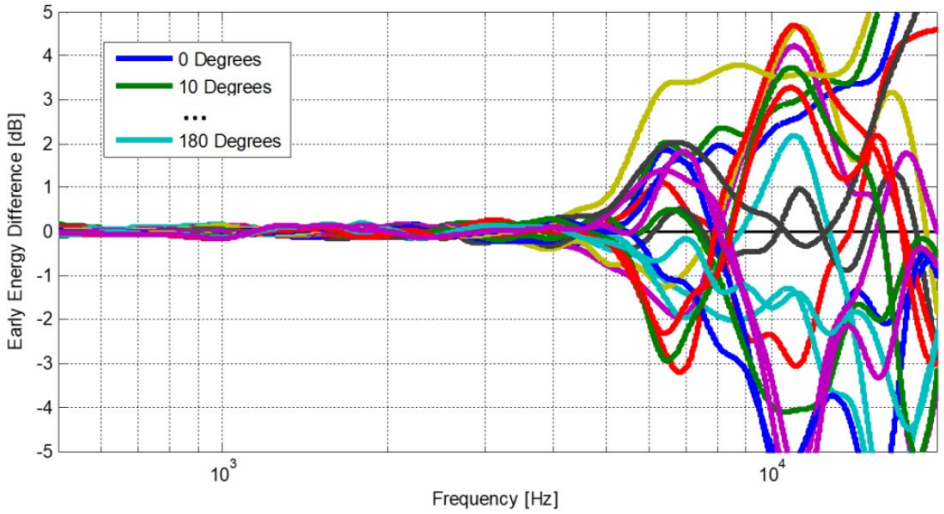


Figure 5-17: (from Ref. [52], Fig. B-11) The frequency response as a functions of source rotation angle for the early part of the RIR measured in a 2500-seat performance hall.

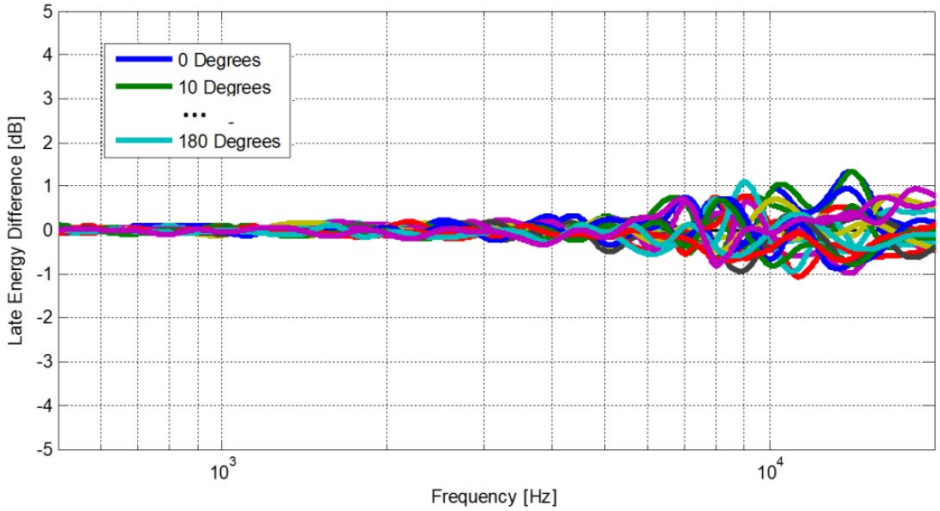


Figure 5-18: (from Ref. [52], Fig. B-12) The frequency response as a functions of source rotation angle for the late part of the RIR measured in a 2500-seat performance hall.

Dick also investigated the impact of taking a stacked-configuration measurement vs. three separate measurements, with each source located in the same position. In this measurement, differences in both the early and late energy. Differences in the early energy were again found to be as high as −10 dB to +6 dB, but differences were found across all frequencies, even when the source exhibits omnidirectional radiation, shown in Figure 5-19. These effects are due to both the non-central location of the sound source, which can create phase mismatches between the measurements, and the radiation and scattering effects from the presence of the other sources in the sound field. This physical presence and the scattering from the other loudspeaker enclosures also inhibit the omnidirectionality of the loudspeakers, especially at higher frequencies. Although this effect might not produce a subjective difference in auralizations based upon RIR measurements, noticeable differences are found in objectively calculated sound field parameters. The differences between the stacked and the concentric configurations of the sound sources are shown for the worst case (closest) receiver location for both early and late sound, Figures 5-19 and 5-20 respectively. Worst case differences were found for the closest measured receiver during test measurements in a 2500-seat auditorium on Penn State's campus.

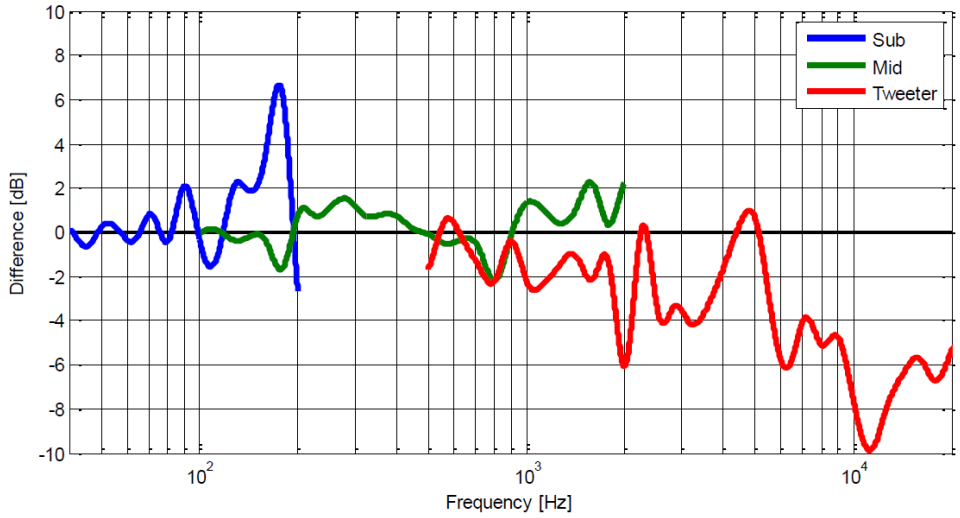


Figure 5-19: (from Ref. [52], Fig. B-20) Difference in early energy between stacked and coincident configuration for the omnidirectional loudspeaker.

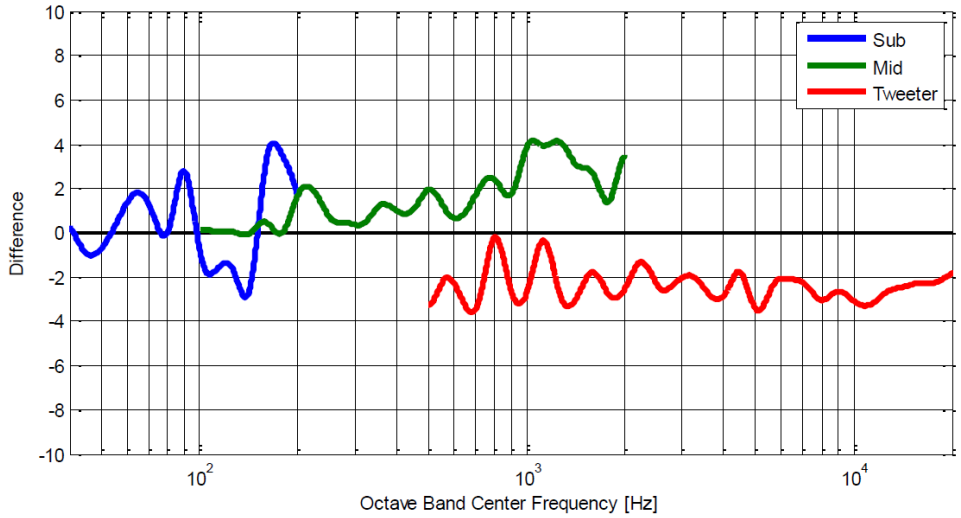


Figure 5-20: (from Ref. [52], Fig. B-21) Difference in late energy between stacked and coincident configuration for the omnidirectional loudspeaker.

5.9.1.1 Omnidirectional Loudspeaker Equalization Filters

Additionally, if results are to be compared between different measurement teams, diffuse-field equalization must correct for the non-flat response of the loudspeaker. For the current study, diffuse-field equalization filters were designed for each of the three-part omnidirectional sound sources. Each filter was generated by inverting the complex pressure response of the transfer function of the loudspeaker. This transfer function was calculated using the Fourier transform of the loudspeaker's spatially averaged time-domain IR. High- and low-frequency amplification limits on this inversion were set, to ensure stable filter responses when the loudspeaker has poor power output.

The subwoofer equalization filter was based upon outdoor measurements made in an empty parking lot, to ensure anechoic measurement conditions at low frequencies. The diffuse field filter corrected for this non-flat response, and as the subwoofer was only used at low frequencies, the outdoor baffled measurement was valid over this frequency range. Multiple directional measurements sampled in a 90-degree measurement arc were made, and the average of these measurement was the basis of the filter's design. Finally, a halving of the pressure was applied, to correct for the pressure doubling due to the baffling effect of the parking lot. The equalization filter was designed as a minimum-phase FIR filter, with a spectrum matching the inverted average response of the loudspeaker. Normalization was performed to account for the sensitivity difference between the low-frequency subwoofer and the mid-frequency dodecahedron.

The equalization filter for the mid-frequency dodecahedron was based upon a combination of measurements made outdoors and in an anechoic chamber. The outdoor measurements were adjusted to match anechoic conditions by halving the pressure, accounting for the baffling effect of the parking lot. These measurements provided a clean response up to 1 kHz, where the ground reflection from the parking lot began to generate incoherent comb-filter effects. The indoor anechoic measurements were made using a single two-dimensional sampling plane of the source. As this loudspeaker is only used in a range where it is omnidirectional, its frequency response is not a function of direction. Thus, a complete three-dimensional spatial sampling was not required, and the average of the two-dimensional measurements was used for source equalization. At 600 Hz, a crossover filter was applied between the outdoor and indoor anechoic measurements using linear-phase crossover filters, matching the magnitude spectrum of squared Butterworth filters. This combined response was inverted, and an equalization filter was also generated for this source, using the same process from the subwoofer. The source filter was normalized to the loudspeaker's response at 1000 Hz. This normalization was applied to the other source components as well, to remove sensitivity differences between sources.

Finally, a diffuse-field equalization filter was designed for the high-frequency dodecahedron. Three-dimensional turntable measurements of IRs from the loudspeaker were made using resolutions of five-degrees in elevation and ten-degrees in azimuth. The diffuse-field response of this loudspeaker was generated using a weighted average of the IRs, weighted with the area integration factor from the spherical equiangular sampling scheme. Again, the smoothed, average spectral response of the loudspeaker was used as a target for a linear-phase FIR filter using MATLAB's fir2 function in the signal processing toolbox. This filter was then converted to a minimum phase filter and normalized to the response of the mid-frequency dodecahedron at 1000 Hz.

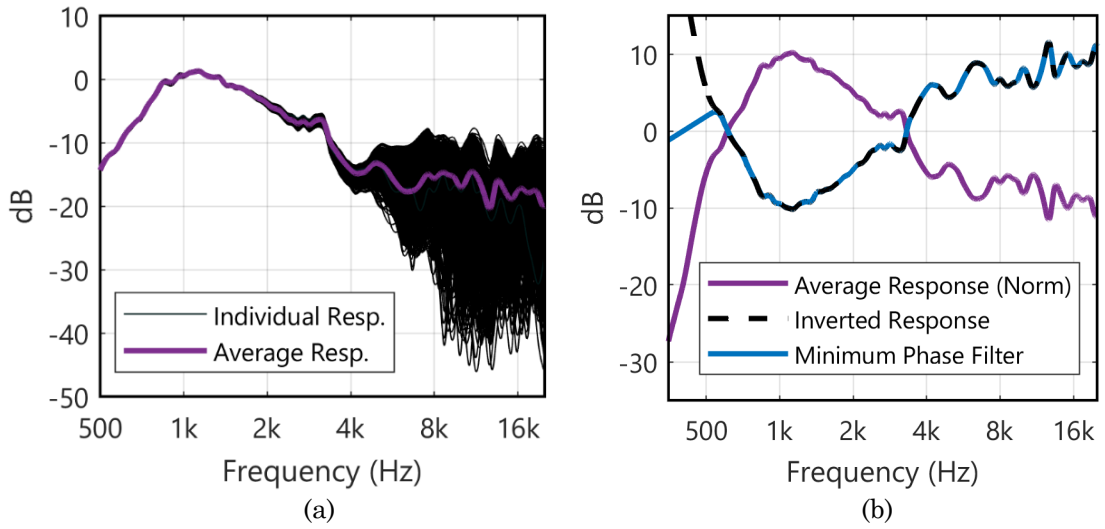


Figure 5-21: The spatially-averaged diffuse-field response of the high-frequency dodecahedron (a), and the design result of a minimum phase FIR filter, inverting the magnitude of the diffuse-field average response from the sound source (b).

5.9.1.2 Three-way Crossover Filters for the Omnidirectional Loudspeaker

To combine the three RIR measurements into a single, broadband RIR, a set of three linear-phase crossover filters were designed. These filters were designed to have a response magnitude matching that of the squared spectrum of a 7th order Butterworth filter, providing a smooth transition that also has the desirable property of a flat summed spectrum. The diffuse-field equalization filters described in section 5.9.1.1 are applied first to each individual measured RIR. Next, correction factors for the knob setting of the power amplifier's 'clicks' are applied, to remove any loudness differences due to the amplifier gain. This provides three equalized RIRs that can be combined into a broadband RIR. One additional check is made to ensure that each of the measured RIRs is time-aligned and will in-phase with one another. The direct sound for the omnidirectional RIR is isolated from each measurement, and the pure delay of the direct sound is calculated using a linear fit of the direct sound's phase. Due to potential latency differences in the software between measurements, the time delay was adjusted to a consistent value for all sources, and the crossover filters were used to combine the measurement into a single RIR. A plot of the crossover filters is provided in Figure 5-22.

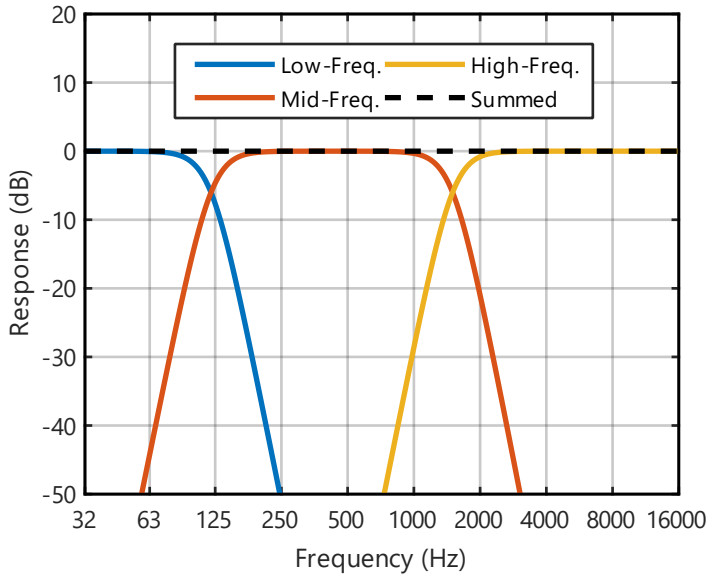


Figure 5-22: The crossover filter designed to combine the three separate omnidirectional source measurement into a single, broadband RIR.

5.9.2 Spherical Microphone Array Processing

The main details regarding the spherical array processing implementation in this study are presented in section 5.5.1. This section presents the process from going from a MicRIR and transforming to a complex-valued orthogonal representation of the ShRIR. Once the RIR is represented in the SH domain, the measurement is first processed to remove the noise floor from the RIR for more accurate RIR metric calculations and clean auralization. Additional details are then given on converting to an ambiX formatted ShRIR, diffuse-field equalization for the microphone array, and estimation of the omnidirectional and figure-of-eight RIR from the ambiX ShRIR.

5.9.2.1 Cleaning the Room Impulse Response

After a MicRIR is encoded into a ShRIR, it still contains the frequency-dependent noise floor from the signal-to-noise ratio (SNR) in the measurement. Measurements in the database were found to have good SNRs, but this noise will still impact the calculation of room acoustic parameters and can cause undesirable artifacts in the RIR. For a single-channel omnidirectional microphone, techniques have been developed to truncate the noise floor and provide correction terms for the proper calculation of metrics. Another technique is to replace the noise floor of the IR with a decaying noise signal, with a slope matching that of the linear fit of the room's decay. This process is most accurate when the noise decay is matched separately in individual frequency bands, rather than a broadband sense. This problem has not been formally approached for microphone-array related processing. The problem becomes

somewhat involved, as individual microphone channels from microphone array outputs are highly correlated at low frequencies, due to small capsule spacings and long wavelengths in the measured field. As frequency increases, these correlations diminish. Additionally, the large number of channels make the problem very time-intensive to manually calculate and clean each channel of the MicRIR separately.

To overcome these challenges, an automated cleaning procedure was developed to clean RIRs measured with the spherical microphone array. First, to overcome the problem associated with the frequency-dependent correlation of the later reverberant energy, the cleaning procedure was done on the ShRIR. Since each of the SH functions are orthogonal, this inherently means that the signals should be uncorrelated once encoded into the SH domain. As such, the later decay of each SH function was replaced with a separately generated uncorrelated noise signal in each of the 16 third-order SH channels. The next issue regarded the robustness of the calculation algorithm. Since each RIR measurement contains 16 SH channels, measured for 242 RIRs in the CHORDatabase (not to mention the orchestral CSLA RIRs), an automated, robust procedure was needed.

First, for each SH channel of the RIR, the backwards integration of the RIR was calculated. Plotted on a decibel scale, most time-domain raw RIR signals resemble the plot shown in Figure 5-23 (a), comprising of an initial delay, direct sound, linear decay, and ultimate noise floor. This noise floor also occurs before the direct sound, and the level of the SNR is a function of frequency and a function of SH channel. When this shape of signal is backwards integrated, the decay becomes smooth in character, and exhibits three main regions, shown in Figure 5-23 (b). First, before the direct sound, all of the energy has already been integrated in the RIR, so the backwards integral is merely a flat line at 0 dB when normalized to the maximum energy in the RIR. Then, once the direct sound occurs, the initial decay of the RIR is somewhat non-linear, and depends largely upon the relative strength and arrival time of the direct sound and early reflections, all relative to one another. After 5 – 10 dB of decay, depending largely upon SR distance, the decay smooths out to a more linear regime, and eventually, the noise floor of the measurement is reached. Since the noise floor's amplitude is a constant with time, it has a backwards integral on a logarithmic decibel scale that resembles a flat line, but eventually, decays exponentially with time. This shape is very consistent with most all measured RIRs, except for transient noises that sometimes occur in the calculation of the RIR.

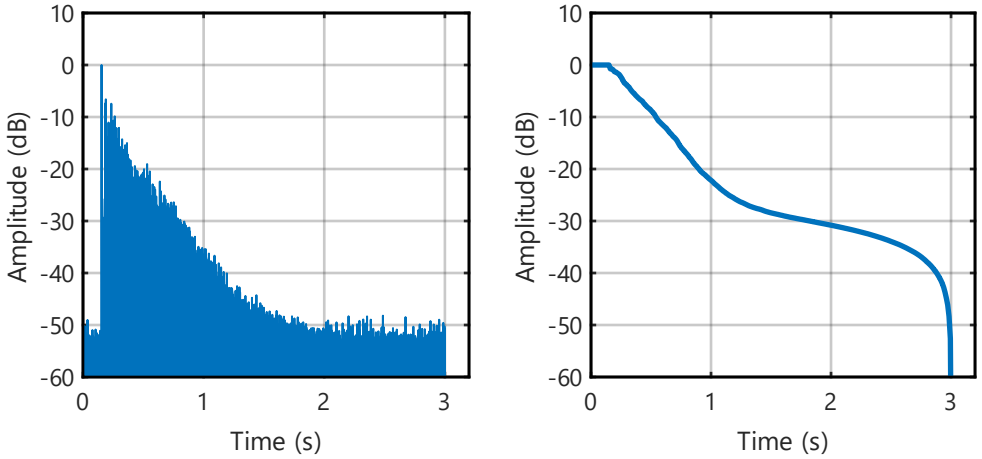


Figure 5-23: A measured RIR (a) and its corresponding backwards integration (b).

A mathematical model of this common shape was developed by assuming the RIR could be represented as a superposition of two time-domain functions: a decaying exponential function and a time-invariant constant noise floor. These functions can be fully defined by three parameters: the time of the direct sound (t_d), the decay amplitude at $t = t_d$ (A_o), the decay rate of the complex exponential (β), and the noise floor amplitude (N_o). The superposition of these two functions is shown mathematically in Eqn. 5.9 and visually in Figure 5-24.

$$IR(t) = \begin{cases} A_o e^{-\beta(t-t_d)} + N_o & \text{for } t \geq t_d \\ N_o & \text{for } t < t_d \end{cases}. \qquad 5.9$$

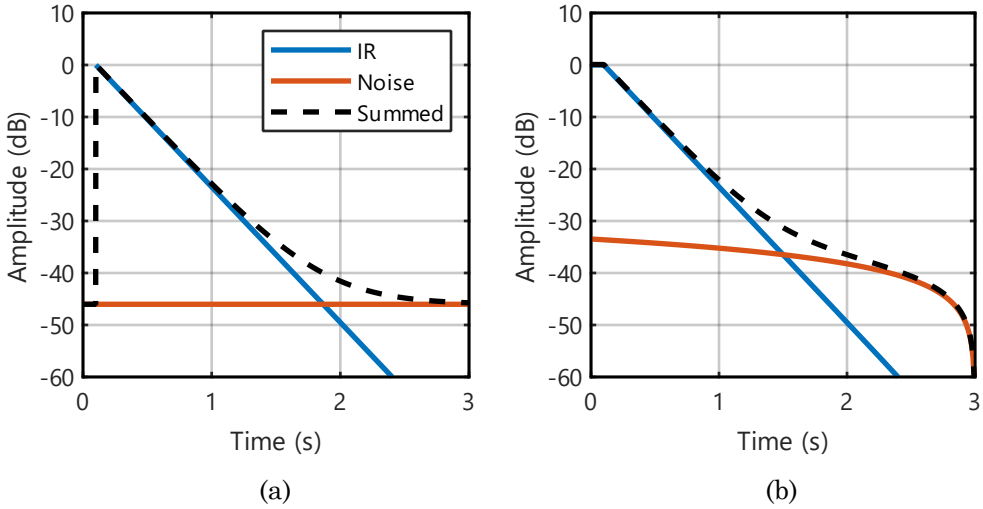


Figure 5-24: An image of the estimated decaying exponential function with a $t_d = 0.1$, $A_o = 1$, $\beta = 3$, and $N_o = 0.005$. The summed response resembles the shape of a typical measured RIR. The result is shown as the time-domain RIR in (a) and its corresponding backwards integration in (b).

To simplify this model, the backwards integral of each SH channel in each ShRIR was calculated. The noise floor before the direct sound was first zeroed out, to remove unwanted effects. Next, the point of the backwards integral that was 5 dB below the total integrated energy in the RIR was identified. The backwards integral of the measured RIR was then fit to the backwards integral of the model provided in Eqn. 5.9 separately for each octave band. This fitting was done using non-linear least squares optimization over the parameters A_o, β, and N_o using MATALB's lsqnonlin() function. This analysis resulted in a decay rate and noise floor for each SH channel and each octave band for a single RIR. As RIRs can be quite variable in the first initial part of their decay, depending largely upon source-receiver distance, the region before the -5 dB down point was not used in the non-linear least squares fitting, as this region produced less robust and reliable results. With the decay rate, direct amplitude, and noise floor amplitude estimates, a transition time or elbow point between the noise floor and the room's decay could also be directly calculated.[96]

For the slope estimate, this value should not be a function of receiver directivity and should be consistent between each SH channel. To assist with robustness, the median value for the decay rate across all SH channels was taken as the estimated decay rate. Since decay rates do vary with frequency, this was also done separately in each frequency bin. Finally, a decaying octave-band filtered noise signal was stitched onto the original measured RIR at a point that was 10 dB above the amplitude of the backwards integral at the calculated decay-noise transition time or elbow point. This ensured that effects of the noise floor were adequately removed. The elbow points used the median of the decay noise floor times across each SH order and octave band, as this time was found to be consistent within the same SH order. This result is mainly due to the noise floor's dependence upon the radial filtering step, an order-dependent operation. Many other robustness checks and tolerances were built into a complex RIR cleaning algorithm, also including many warning flags to allow for manual checking of RIRs that might have produced results outside of the typical range of possibilities. These manual flags helped to ensure that RIR cleaning across the entire database was done in a consistent yet robust manner.

5.9.2.2 Diffuse-field Microphone Array Equalization

Even after radial filtering and compensation for the boundary conditions associated with the rigid microphone array's housing and scattered pressure field, the microphone still exhibited a non-flat response, mainly at higher frequencies and especially above the spatial aliasing limit at 8 kHz. To compensate for these effects in the estimate of the omnidirectional response of the array, a diffuse-field equalization filter was designed. Microphone IR

measurements were sampled from around the array using a consistent sound source in an anechoic chamber and a turntable, taking measurements with 5-degree elevation resolution and 10-degree azimuth resolution. The spatial average of the omnidirectional output of the microphone array was calculated, including individual capsule equalization, SH encoding, radial filtering, and the proper integration factors based upon the equal-angle sampling scheme. In the same location, in the same room, with the same loudspeaker, a ½" free-field microphone was oriented towards the loudspeaker to take a frequency response of the loudspeaker. The difference between the free-field microphone array and the spatial average of the microphone array's response was determined to remove the non-flat response of the loudspeaker from this measurement.

Finally, the spatial average was taken, it was centered around its mean value to a reference of 0 dB, and the response was inverted to generate a target for a diffuse-field equalization filter. A minimum-phase FIR filter was designed to match this target response, correcting for the non-flat diffuse field response of the array, and ensuring that results would not be colored or impacted by the array itself. This step of correction is already built into a measurement setup when a diffuse-field omnidirectional microphone is used. The response of a diffuse-field microphone is already corrected for the effects of the scattered pressure off of the microphone's housing. As the array is inherently neither diffuse- nor free-field equalized, this step is important to consider for proper measuring RIRs with an accurate frequency spectrum corresponding to that of the room, and not the measurement setup.

5.9.2.3 Conversion to Real-valued ambiX Ambisonic Format

All of the higher-order Ambisonics processing is typically done in the ambiX format, consisting of real-valued SHs, instead of the complex-valued harmonics typical to spherical array beamforming literature. To provide as consistent a processing framework in the study, the encoding of the RIR, radial filtering, ShRIR direct sound rotation, and RIR cleaning was all performed on the complex-valued ShRIR using the SOFiA Toolbox in MATLAB, a spherical array beamforming toolbox.[91] Once the ShRIR was properly represented in it's complex-valued format and cleaned, it was directly compatible with built-in PWD functions in the SOFiA toolbox. To then generate an ambiX real-valued ShRIR that was compatible with the Ambisonic decoder toolbox for auralization, the conversion function between complex and real-valued SHs created by Archontis Politis was implemented.[92] This function was designed to take an orthonormal complex-valued SH function and convert it to a real-valued orthonormal SH function. The relationship between real and complex-valued SH functions is somewhat complex and was described in chapter 3. It is important to note that Politis's function

automatically built-in the removal of the Condon-Shortly phase term factor, $(-1)^m$, and adjusted to the N3D normalization scheme, which is different than the complex-valued orthonormal normalization scheme. Finally, manual conversion from the N3D to the SN3D normalization was performed to produce an ambiX format ShRIR, compatible with auralizations generated using the Ambisonic decoder toolbox for MATLAB created by Aaron Heller et al.[61,84]

5.9.2.4 Omnidirectional and Figure-of-eight Room Impulse Responses

Once the ambiX ShRIR was determined, the extraction of the omnidirectional and laterally oriented figure-of-eight microphone response is more straightforward. The omnidirectional response can be simply extracted as the first channel of either the complex-valued or the real-valued ShRIR. It is important to note that this measurement was diffuse-field equalized so that this estimate is as accurate as possible. In the ambiX format, it is convenient also that second channel of the ShRIR directly corresponds to the laterally oriented dipole response of the microphone array. In this case, this is only true due to the fact that before conversion to the real-valued ambiX ShRIR, the ShRIR was rotated so that the direct sound was oriented directly in the frontal, x-axis direction of the RIR. This step ensured that the y-axis-oriented dipole (contained in the second channel) is pointed in the proper direction. A final note, it is important to remove any normalization effects from the ShRIR so that the on-axis left-right amplitude response of the omnidirectional and dipole RIRs are compatible in terms of amplitude. Not all SH formats have the same normalization factor between these two components. It is also possible to extract the dipole response from the complex-valued ShRIR, but it would be calculated as a combination of various first-order harmonics to generate the desired beam-formed array output.

5.9.3 Higher-order Ambisonics Auralization

The higher-order Ambisonic auralization techniques were created using the Ambisonic decoder toolbox in MATLAB, created by Aaron Heller.[61,84] The facility at Penn State used for higher-order Ambisonic-based auralization is known as the Auralization and Reproduction of Acoustic Sound fields (AURAS) facility. The AURAS facility is a 30-loudspeaker auralization array located in an anechoic chamber with 18" foam wedges. This provides a controlled mostly free-field environment, so that the array can flexibly and accurately simulate a wide variety of sound fields. An image of the AURAS facility is provided in Figure 5-25.

Figure 5-25: The AURAS Facility, a 30-loudspeaker and 2-subwoofer higher-order Ambisonics auralization array located on Penn State's campus.

The facility contains 30 custom-built two-way sealed box loudspeakers with a crossover frequency around 1.8 kHz. The speakers have a low frequency resonance at 60 Hz, so two subwoofers located in the corners of the facility are used to reproduce the low-frequency range of auralizations. The subwoofers were also custom-built using 18" drivers with high linear excursion and a low-frequency resonance of 20 Hz. The facility allows for direct reproduction of multiple concert halls, side-by-side, and has been implemented in a number of subjective acoustic tests relating to concert hall, office noise, supersonic flight, and traffic noise perception. Full details on the design and construction of the facility, loudspeakers, hardware, and software-based control can be found in the thesis by Neal.[33] The complete details regarding the higher-order Ambisonics auralization can be found in the dissertation by Dick.[52] A few key details regarding the higher-order Ambisonic processing are presented in the next sections, regarding order-dependent SH crossover filters and decoding techniques.

5.9.3.1 Order-dependent Crossover Filters

When performing the radial filtering equalization step on a measured ShRIR, dependent on the size of the microphone array, large boosting is required as frequency decreases and order increases. Eventually, the boosting required approaches the SNR of the measurement, so accuracy is lost. To adjust for this issue, order-dependent crossover filters were generated. For these filters, once the boosting for a given SH order was too great, a high-pass filter was generated to zero out the higher-order terms below a certain cutoff frequency. Essentially, if the third-order components are high-pass filtered above 1 kHz, below that frequency, the auralization would be truncated to only second-order reproduction. Since the SH functions

satisfy orthogonality, this truncation can be done by simply zeroing out these signals. By inspecting the radial filters, a reasonable maximum level of boosting for the measurement setup in this study was found to be around 25 dB, without producing noticeable noise artifacts in the auralization. Crossover frequencies to second-, first-, and zeroth-order SH truncation were set based upon this criterion at 1300 Hz, 500 Hz, and 40 Hz, respectively.

A four-band linear-phase crossover filter was designed with these three transition frequencies. For the third-order SH filter, only the high pass filter above 1300 Hz was used, with a pass-band amplitude of 0 dB. When truncating the order of the auralization, this will also cause a reduction in energy in the auralization below that frequency. Common practice to account for this effect is to consider an order-truncated plane wave at each SH order, and normalize a plane wave at each order to have equal on-axis amplitude, or to have equal energy integrated around the sphere. In the current application, to prevent artificially high energy differences in between plane-wave truncation orders, the lower-order plane waves were normalized to have the same spatially integrated energy as the third-order plane wave. Thus, the low-pass filter, below 90 Hz, was only used for the zeroth-order SH component, and its pass band gain was set to normalize the integral of the monopole component to match that of the third-order plane wave. The second band-pass filter from 40 to 500 Hz was used for the first- and zeroth-order SH components, set with a pass band to normalize the first-order truncated plane wave to have the same response as the third-order. This pattern was continued for the second-order plane wave from 500 to 1300 Hz, and the four filters were summed together

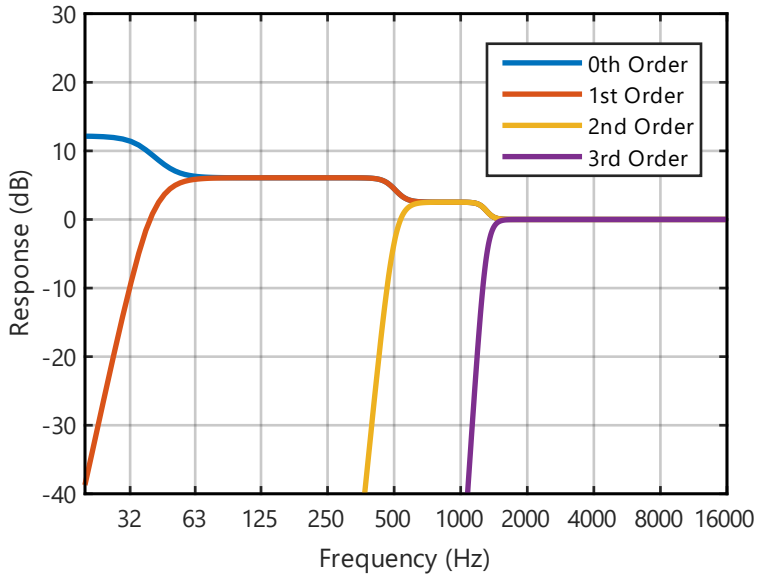


Figure 5-26: The order-dependent crossover filters for truncation orders of 0 through 3, to preserve the spatially integrated energy of a plane wave as SH order was truncated to prevent excessive boosting of low frequency noise in the RIR measurement.

separately for each order, zeroing out the filters below the cutoff frequency for their respective order. A plot of the filters to cross between SH order, while preserving the total integrated energy for a plane wave, is provided in Figure 5-26.

5.9.3.2 Decoding to Loudspeaker Signals

For auralizations, a dual band decoding scheme was used, that was explained in section 3.4.3.1 from the background on spherical array processing in chapter 3. Basic or mode-matching decoding was done for low frequencies, below 400 Hz, which has been shown to work well for the low-frequency localization of the human hearing system. At higher frequencies, side-lobes associated with the order-truncation of a plane wave in the SHs domain can cause unwanted artifacts that cause errors in the interaural level difference (ILD) cue that dominates high-frequency localization. To minimize these artifacts, max-Re weights that depend upon order are applied to the decoder matrix, to minimize the amplitude of the side lobes in the representation of the plane wave. This minimization does result in a wider main-lobe of the order-truncated plane wave, but this tradeoff helps to prevent the more noticeable psychoacoustic concern of a mismatch between high- and low-frequency localization. This dual-band decoder was designed for the AURAS facility and has been implemented to generate auralizations for the current full-orchestral measurements. Finally, the AURAS facility's array is not entirely spherical in nature, so individual time delays and level corrections for the distance differences between loudspeakers are included in the processing chain. Further, near-field compensation is built-into the Ambisonic decoder design, correcting for the near-field effect of the loudspeakers in the array at low frequencies. Full details on the design of the Ambisonic decoder using the MATLAB-based Ambisonic decoder toolbox can be found in the dissertation by Dick.[52]

5.9.4 Further Details on Beamforming of Room Impulse Responses

5.9.4.1 Influence of Truncation Order on Beamforming Resolution

It is also important to note that the order of spherical array beamforming can have a clear impact upon the results, resolution, and usefulness of the analysis. To demonstrate this effect, beamforming analysis from the same measured RIR, for both the early and late parts of the RIR, have been performed by truncating the resolution of analysis in post processing. The same results for third-, second-, and first-order processing are shown in Figure 5-27. As can be seen in both the early and late sound field, extremely poor spatial resolution is found for first-order beamforming analysis, and results are almost incompatible with the higher-order analyses. For the late part of the sound field, the analysis is being performed over many

different arriving reflections, and it is inherently a large-scale average. Because of this, less noticeable changes are observed when increasing from second- to third-order.

On the other hand, when an analysis is performed on a set of discrete early reflections, the third-order beamforming analysis helps to more precisely locate individual reflections, compared to the second-order analysis. Additionally, some of the reflections clearly shift in the direction of arrival or in relative amplitude compared to one another. This effect is mainly due to the minimization in side lobe amplitude relative to the main lobe. Due to sampling limitations, rear and side lobes always exist in spherical array beam patterns. As order increases, the amplitudes of these lobes, relative to the main lobe, decrease, and these effects are minimized. Although the amplitudes might seem small in comparison, less than –12 dB at third-order reproduction, when multiple reflections are being analyzed in the same time range, arriving from different directions, coherent and incoherent summation and interactions of competing main and side lobe energy occurs. This can result in improper direction of arrival estimates and even improper amplitude estimates of specific reflections. This effect of side lobes is less pronounced, but still occurs in the later energy beamforming. Due to these effects, a large benefit in analysis accuracy occurs with an increase from second- to third-order beamforming analysis resolution.

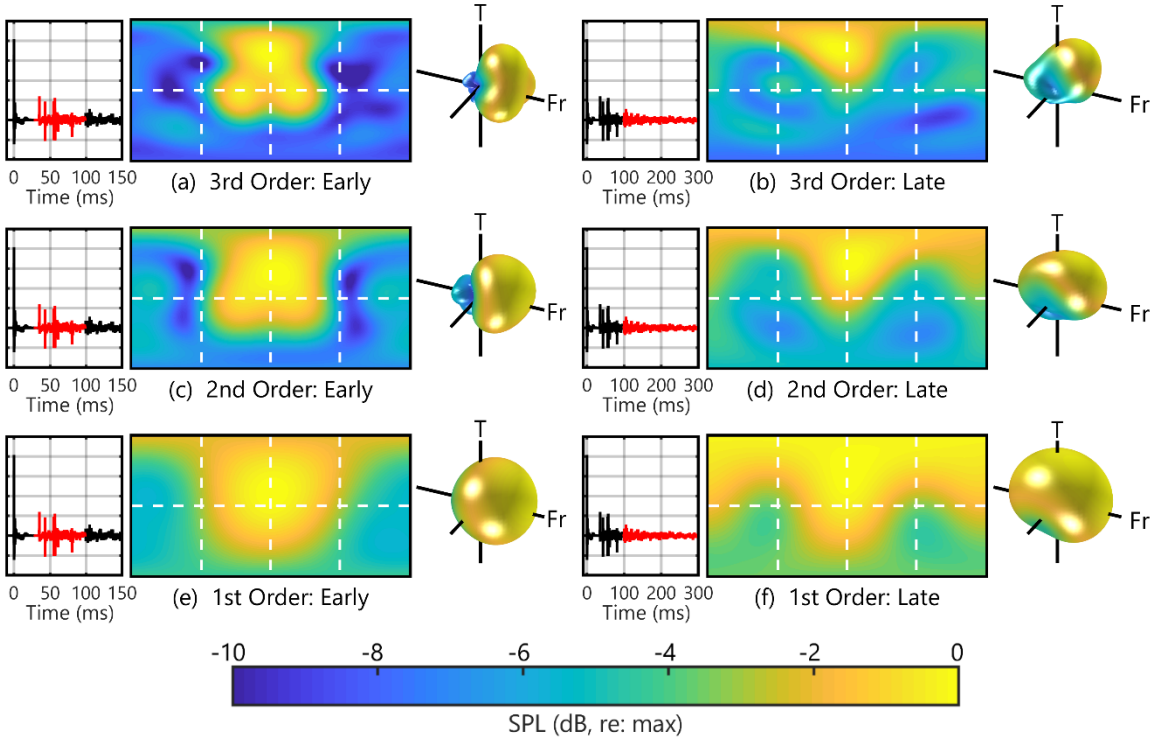


Figure 5-27: Beamforming analysis comparing third-order (a) & (b), second-order (c) & (d), and first-order (e) & (f) beamforming analyses for early and late energy, respectively, in the RIR.

5.9.4.2 Dolph-Chebyshev Beam-shaped Decomposition

Another way to mitigate the effects of side- and rear-lobes upon the beamforming analysis is to use another beam shape or topology. The most common beam pattern, which is identical to an order truncated plane wave, is designed to maximize the directivity index of the resulting beam pattern. Essentially, it maximizes the beam response on-axis while minimizing the beam patterns response in all other directions, providing the best approximation of a spatial Kronecker delta function (or a plane wave). Mathematically, it contains unequal rear and side lobe amplitudes that result from the SH order truncation. A beam shape adjustment can be made by using a Dolph-Chebyshev beam pattern instead of the standard plane wave beam pattern.[93] Chebyshev polynomials have a desirable mathematical property that either the side-lobe rejection level or the main-lobe beam width can be specifically set or adjusted to produce a flexible beam pattern. Essentially, the plane wave beam pattern can be slightly modified to create equal-amplitude side- and rear-lobes, matching the shapes of Chebyshev polynomials that can then be adjusted to any arbitrary side lobe amplitude. This factor can then be adjusted and selected for the particular needs of the specific beamforming scenario.

A plot of a third-order plane wave beam shape as a function of angle is provided in comparison to a Dolph-Chebyshev beam pattern set to have different side- and rear-lobe rejection values of –15 and –25 dB in Figure 5-28. As can be seen, all of the Dolph-Chebyshev beam patterns have a larger minimum rejection compared to the –12 dB rear-lobe of the third-order plane wave. A clear trade off exists though; as the side-lobe rejection amplitude is increased, the main beam width becomes wider and wider. This inherent tradeoff between side lobe rejection amplitude and main beam width is a tunable design choice.

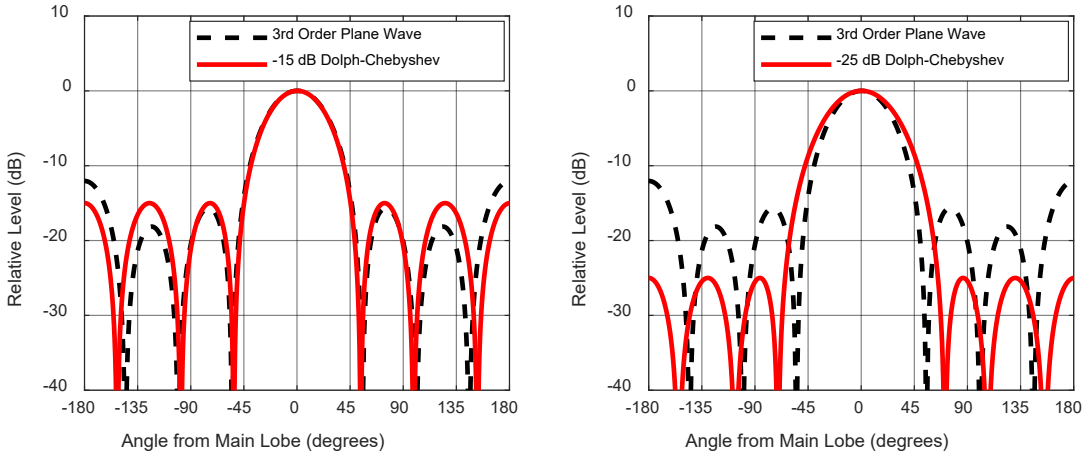


Figure 5-28: Comparison of a third-order plane wave beam patterns with Dolph-Chebyshev beam patterns for a -15 dB side-lobe level in (a) and a -25 dB side-lobe level in (b).

To further demonstrate this effect, as was done in section 5.9.4.1 for the effect of beamforming order, the early and late parts of the same RIR have been generated using third-order plane wave decomposition in Figure 5-29. Then, the same regions of the same RIR are analyzed with –15 dB, –20 dB, and –25 dB Dolph-Chebyshev beam patterns in Figure 5-29. Comparing these results, it is quite astonishing how much the patterns change and the effects of side- and rear-lobe amplitudes on the beamforming map are reduced with increasing rejection amplitude. This effect is quite noticeable in the early energy, where the amplitudes and arrival directions of multiple early reflections shift quite dramatically. Also, energy peaks in the standard plane wave decomposition map disappear as rejection amplitude is increased, indicating these are most likely not actual reflections. Rather, they are simply locations where multiple side lobes from different reflections have summed coherently to create a reflection-like artifact in the map, not present in the actual sound field. Even the large-scale average of the later reverberant energy is highly impacted by these side lobe effects, and the importance of using beam shapes other than a plane wave in RIR beamforming is clearly necessary.

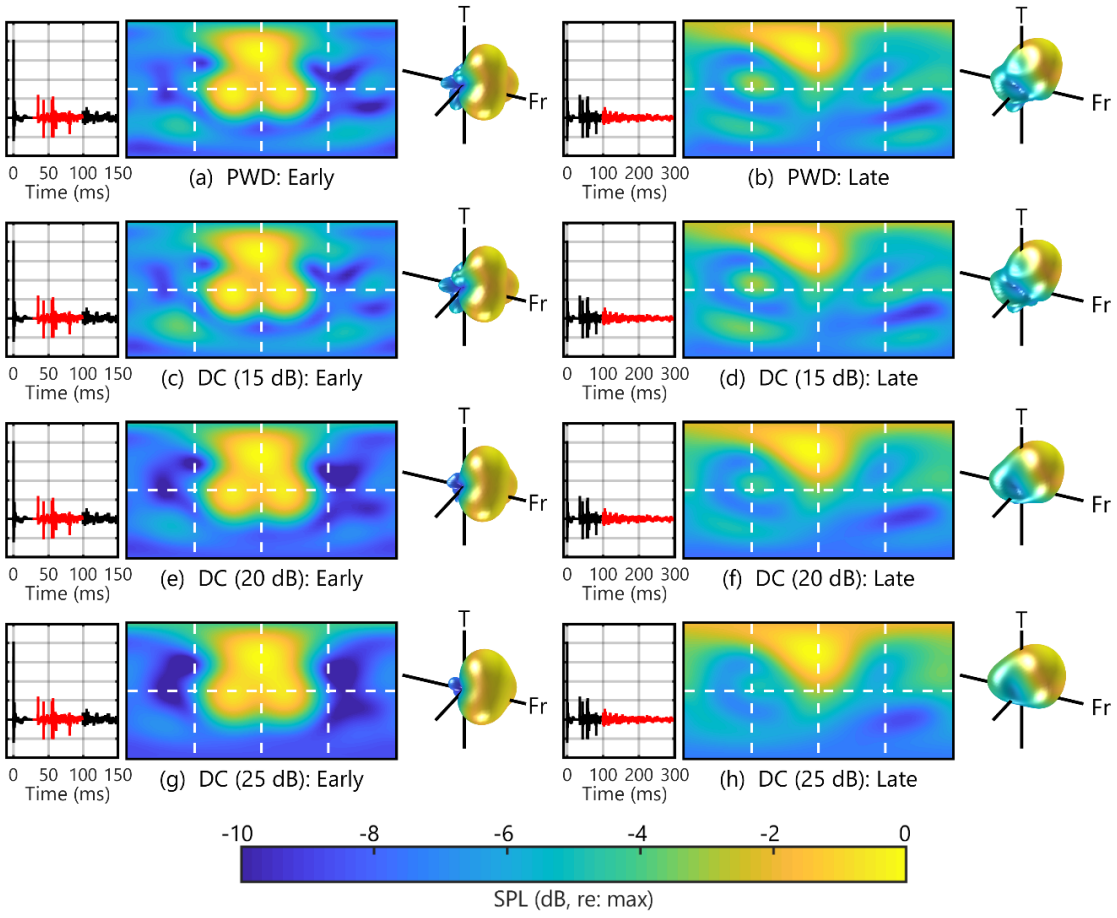


Figure 5-29: Beamforming analysis comparing beamforming results for ideal plane waves (a) & (b) to 15 dB rejection Dolph-Chebyshev beam patterns (c) & (d), 20 dB rejection Dolph-Chebyshev beam patterns (e) & (f), and 25 dB rejection Dolph-Chebyshev beam patterns (g) & (h) for early and late energy, respectively, in the RIR.

Chapter 6
Individual Preference in Concert Halls

This chapter contains information written for external publication, which is intended for submission to a major peer-reviewed acoustics or audio engineering-related journal. As such, this chapter also contains its own introduction, background, and results section. The overall introduction (Ch. 1), background (Ch. 2 and 3), and results (Ch. 7) chapters of the dissertation fill in the greater picture of the entire dissertation work.

This chapter describes the psychoacoustic experiment designed using the realistic auralizations generated using the CSLA and CHORDatabase measurements from chapters 4 and 5, respectively. An analysis of the subjective responses from 16 individuals is provided, which includes connecting these subjective responses with the spherical beamforming techniques described in chapter 5. Correlation analyses in this chapter demonstrate the temporal and spatial regions of the room impulse response that best correlate with specific subjective perceptions. Factor analyses also suggest three to four independent factors that explain a majority of the variation in subjective perception of concert halls.

Subjective Attributes in Concert Hall Acoustics and their Impact on Individual Preference

Matthew Neal and Michelle Vigeant
Graduate Program in Acoustics, The Pennsylvania State University, University Park, PA 16802

Abstract:

The subjective impression of concert halls is multi-dimensional in nature, any many previous efforts have approached the problem using different techniques. The goal of the current study was to use a spherical microphone and compact loudspeaker array (CSLA) measurement database to conduct a subjective experiment covering subjective acoustic attributes and preference. Repeatable and realistic full-orchestral auralizations were generated using CSLA measurement at 20 source positions. Measurements in 21 hall environments were narrowed to a representative set of 14 using statistical clustering techniques. Using a comparative task, subjective ratings were collected for the ten attributes found important in previous literature, along with preference. Multiple factor analysis found that perceptual attributes could be mostly explained by three to four factors interpreted as clarity, strength and envelopment, strength and source width, and brilliance. Proximity and clarity best correlated with overall average preference, but specific individuals showed strongest correlation between preference and either the clarity factor or the spaciousness factor. Proximity was not highly related to one factor, but rather, showed somewhat strong correlations with most all factors. Clear importance for considering preference at the individual level was observed, and the standard two preference groups in concert hall acoustics may not accurately represent individual taste.

6.1 Introduction

The central goal in the design of a concert hall is to satisfy the end users of the space: the musicians, the conductor, the audience members, the board members, the donors, and many others. The challenge with satisfying preference is its inherent undefined nature. Preference can be defined by a panel of experts. It is a choice, made by each individual, based upon whatever they deem to be desirable for a specific context. It can be impacted by the performers, the room, the music, the meal eaten before the concert, the fight before the concert between spouses… (and the list could continue with increasingly obscure examples). Preference will undoubtedly vary between individuals and musical piece, but this variation does not inherently indicate that trends, patterns, or threads cannot be identified. This field is not new; much work has been done in the past, but the difficulties in conducting a laboratory-controlled study, for direct comparisons between halls while maintaining realism is entirely

non-trivial. The current work aims to generate a highly realistic auralization in concert hall acoustics using the concert hall orchestral research database, or CHORDatabase, a spherical microphone and loudspeaker array measurement database of a wide variety of existing concert halls. Using the measurements made in these halls, highly realistic auralizations can be generated, and these auralizations can be used as of the basis of studying subjective preference. Additionally, auralizations with high spatial accuracy can be generated using higher-order spherical microphone array measurements. The access to increased spatial resolution in measured room impulse responses (RIRs) provides new outlooks and potential for identifying future metrics to predict concert hall acoustics perceptions.

6.2 Previous Studies of Concert Hall Perception

The first, and highly famous attempt to address the problem of what creates a good acoustic environment was done by Sabine.[1] To solve speech intelligibility concerns in a lecture hall, Sabine conducted many studies, and provided the first objective metric in room acoustics: reverberation time (RT). Although Sabine is most known for this metric, he also studied and understood the importance of loudness, individual strong reflections, and much more. Much time passed without any additional proposals of new objective metrics, and questions arose as to the sufficiency of RT alone in explaining perception. Through the 1950s and 1960s, researchers and practitioners began to search for other impressions and objective factors to explain the subjective differences that RT did not capture, analyzing the effects of variation in the character of a room's decay curve[3-4] and the impact of individual early reflections on spaciousness.[8] From that time to today, much work has been done to study perception in concert halls, including interview-based, laboratory, architectural / acoustic measurement-based, measured auralization, and simulated auralization studies.

6.2.1 Interview and Survey-based Approaches

The first category of studies includes interview or survey-based approaches. The initial and most notable of these was done by Beranek in 1962, where he interviewed musicians, conductors, and critiques regarding their experiences in different concert halls. The findings of these studies, along with a subsequent compilation of objective measurements, have taken the form of multiple books.[5,15-16] Other survey-based approaches have been implemented during live concerts. Hawkes and Douglass conducted such a study in four different rooms,[23] and Barron conducted such a survey of 11 major British concert halls.[24] For his dissertation work at IRCAM, Kahle also conducted a similar study with a 29 question survey, but he used a group of around ten assessors that traveled to different halls.[22] Objective measurements were

also collected in each of the halls. Kahle identified a list of eight separate perceptions of relevance. The most recent work using a survey-based approach was done by Skålevik, extending Beranek's study through an online survey.[97] For all of these studies, the main drawback lies in the lack of control and comparability during subjective data collection. In a live performance, or even more so for remembering back months (or even years) or a live performance, the musical passage, quality of orchestra, conductor, and many other non-acoustic factors change between hall experiences. The influence of all of these factors made direct comparison between halls very difficult to accurately obtain.

6.2.2 Simplified Laboratory Auralization Approaches

A few studies attempted to solve the repeatability problem using highly controlled and simplified room auralizations. The goal of these laboratory studies was to simulate a room-like effects with individual loudspeakers in an anechoic chamber. These researchers investigated the subjective effects of individual early reflections, reproduced from spatially separated loudspeakers with delay units. Often, uncorrelated artificial reverberation signals were played out of multiple loudspeakers to create a room-like reverberation. The Dresden group studied the perceptions of spaciousness and clarity, and proposed a metric, the room impression index (R) to predict the overall perception of concert hall quality.[20] Similar work was conducted by Lavandier in 1989 out of IRCAM in France.[21] Her work involved many different listening studies, aimed at perceptually validating the common objective acoustic metrics. This initial laboratory work inspired the later work by Kahle.[22] Both of these studies provide a high degree of control and repeatability, but the realism of the hall-like auralizations could be called into question. The simulation of discrete, specular early reflections might be perceptually valid, but the recreation of the reverberant energy in a room was highly simplified. Although this simplification was most likely necessary with available technology at the time, it is essential to ensure a high degree of realism in laboratory auralizations when drawing conclusions that are to be extended to realistic concert halls.

6.2.3 Measured and Simulated Auralization Approaches

The invention of binaural techniques and the binaural mannequin, or *dummy head*, provided new outlooks for merging realistic auralizations with laboratory-based control.[27] By making a recordings or binaural room impulse response (BRIR) measurements with a dummy head, multiple concert halls could be reproduced side-by-side over headphones or with a pair of loudspeakers. The dummy head captures the entire character of a sound field, including spatial effects, represented as the binaural signals associated with the acoustic effects of the size, geometry, and pinnae shapes of the mannequin. Yamaguchi performed such a

measurement-based auralization study in the Yamaha music hall, but the capture technique used a stereo microphone pair, spaced at an ear separation without including a physical model of the human head.[29] Such a limited binaural measurement technique could call the study into question. Kimura made recordings with a dummy head while playing anechoic music out of an omnidirectional loudspeaker in 13 different multipurpose halls.[31] Both of these studies were done using headphone-based auralizations. Schroeder et al. conducted recordings using a dummy head in 22 different European concert halls from two loudspeakers placed at either end of the stage.[30] Additionally, BRIRs were captured for later analysis. Reproductions for their study implemented crosstalk cancellation over two loudspeakers. Soulodre and Bradley compared the subjective differences in 10 different BRIRs from North American concert halls, reproduced over a loudspeakers with mechanical barriers and cross-talk cancellation filters to provide high channel separation.[34]

The most recent work using binaural techniques has been performed by Weinzierl et al. using simulation-based binaural auralizations to study a wide range of different room environments for both musical performance and unamplified speech.[43] Room acoustic models for 35 rooms at two listening positions were generated for three different anechoic passages. Each combination of room, receiver, and anechoic passage was rated on 46 different subjective terms. These terms were developed by a focus group of room acoustical experts from various European countries. With such a large number of potential stimuli, subjects only saw 14 randomly selected stimuli from the 190 total stimuli. Each subject rated all 46 words, but they were allowed to skip any terms they felt were unsuitable for a given stimulus. As each subject only saw a limited range of the stimuli, a large number of 190 subjects were required for data collection. This study extended traditional statistical analysis techniques to a large-scale study, resulting in a subset of perceptual terms known as the Room Acoustical Quality Inventory (RAQI). Despite the scale of the study, it was based upon auralizations created in one simulation program, RAVEN.[98-99] Although simulation enabled the generation of a large variety of auralizations with reasonable time and effort, the accuracy of simulations compared to real room environments can be called into question.[44] Additionally, all previous studies were conducted using a dummy head or the head-related transfer function (HRTF) of a dummy head. Although plausible, the use of non-individualized HRTFs has been shown to cause localization errors and internal locations problems, another potential limitation regarding the realism of auralizations.[58]

Finally, the most recent work to received much attention involved the loudspeaker orchestra studies from the Finnish group led by Tapio Lokki out of Aalto University.[38-41] The

team developed an orchestra made up of commercial loudspeakers, and they conducted RIR measurements using a 6-channel microphone array. Auralizations were made using an in-house technique known as spatial decomposition method (SDM). The study did not use pre-defined room acoustical vocabulary, but rather, subjects provided their own subjective descriptors and definitions. Then, they rated each concert hall stimulus using their own terms. Since no standardized vocabulary was used, statistical clustering techniques grouped words that were rated similarly into multiple categories. By use of the distributed array of loudspeakers across the stage, a large increase in source realism was possible compared to other works using one to two loudspeakers on stage. Despite this advance in source representation, source directivity was limited to only that of commercially available loudspeakers, oriented to best match the radiation of each instrument.[36] As instruments exhibit a highly complex, frequency-dependent radiation pattern, such properties cannot be accurately represented with one or two commercial loudspeakers.[72] This limitation may cause perceptual deviations from the realistic scenario, but it was still a variable that was held-constant in the study, and it was a large improvement from the prior studies. Additionally, no standardized omnidirectional loudspeaker was used during the measurements, preventing accurate calculation of multiple room acoustic metrics that are quite sensitive to source directivity.[49]

6.2.4 Summary of Significant Subjective Terms

To build from the large context of previous work, a summary of the subjective terms found to be most important in the studies mentioned in sections 6.2.1 through 6.2.3 has been generated in Table 6.1. For each of the studies, words were identified in the broad categories defined as follows: reverberance (RT), clarity I, strength (G), intimacy (Int), proximity (Prx), source width (SW), envelopment (Env.), diffusion (Diff), brilliance (Bril), warmth (Wrm), tonal color (TClr), balance (Bal), and ensemble (Ens). This summary contains only the studies from Table 6.1 that identified subjective attributes to best represent subjective concert hall preference ratings. This criterion does not include individual experience-based assessments, such as those performed by Jordan.[10-11] Beranek's textbook editions from 1996 and 2003 are not included, as he does not suggest how all of the attributes related to overall quality, as he did in his 1962 book.[5,15-16] Beranek does however provide a paper in 2003 in which he suggests a list of the most important attributes, which is included in Table 6.1.[17] Studies containing only geometric analyses were also not included in Table 6.1, as it is a very different task to connect architectural measurements with subjective impressions. For a more visual representation, all words that were used in these concert hall studies, not only indicating words of importance, are displayed as a word cloud in Figure 6-1. The size of each

Table 6.1: Summary table containing the subjective attributes deemed to be best related to overall preference in previous literature. Studies focused on only architectural measurements or individual opinions are not listed. Also, studies that did not identify subjectively important factors were not included.

First Author	Year	Attributes Identified as Most Perceptually Important													Factors / Dims.
		RT	C	G	Int	Prx	Spatial Attr.		Diff	Tonal Attr.			Bal	Ens	
							SW	Env		Brill	Wrm	TClr			
Sabine	1922	X		X											–
Somerville	1953-57	X													–
Beranek	1962	X		X	X				X		X		X	X	–
Marshall	1967			X			X								–
Hawkes	1971	X	X	X	X	X				X			X		4 – 6
Yamaguchi	1972	X	X	X											3
Schroeder	1974	X		n/a				X							2 – 4
Kimura	1976	X	X			X	X			X					2
Wilkens	1977		X	X						X					3
Reichardt	1978		X				X								–
Barron	1988	X		X	X			X							–
Lavandier*	1989	X		X			X								14
Kahle	1995	X	X	X					X	X	X		X		8
Sotiropoulou	1995		X			X				X					4
Soulodre	1995		X	X						X					–
Beranek	2003	X		X	X			X	X						
Lokki	2010-16	X	X	X		X	X	X			X				2 – 4
Kuusinen	2017	X	X	X	X										8
Skålevik	2017	X	X	X							X				
Weinzierl	2018	X	X	X	X			X		X		X			4 / 6 / 9
Total Counts & Range:		**15**	**12**	**15**	**6**	**5**	**5**	**9**	**3**	**7**	**6**	**3**	**3**	**1**	**2 – 14**

The n/a indicates that level differences were removed in study by Schroeder.

*the authors did not have access to a full translation so *temporal factors* and *effects of early reflections* could not be fully categorized.

Figure 6-1: A word cloud of all subjective terms used in concert hall studies. Larger words are words that are used most commonly across all studies from Table 6.1.

word corresponds to number of times it was used in previous studies. Even though words are not selected for importance, similar words tend to appear as most often selected as important in Table 6.1 and are most often included in Figure 6-1.

Analyzing the table, some terms occur quite consistently across all studies. These attributes include reverberance, clarity, and strength. Some other attributes that occur fairly often include envelopment and brilliance. Warmth, intimacy, proximity, and source width occur somewhat often, and diffusion, ensemble, and balance occur only a few times (mainly in Beranek's works). This ranking is not intended to suggest an ordering of subjective importance, as studies can have vastly different experimental conditions. Rather, it provides a good indication of the primary words that are repeatedly used across many different studies. Using this summary, a set of 10 individual attributes were selected to cover the total range of terms found to be related to concert hall perception. The list of the selected attributes will be further described in section 6.4.1. Along with specific terms, disagreement exists regarding the number of factors needed to explain differences in overall preference. From the studies that attempt to identify factors of importance, the size of such a list ranges from 2 to 14 factors. It is important to note that some studies aim to identify a set of orthogonal factors, while other attempt to develop a comprehensive list of impressions. By nature, studies identifying a set of orthogonal factors will likely generate a smaller list, where studies aiming at completeness will generate a large list of factors that are still highly correlated with one another.

Although much work has been performed, confusion still exists as to what makes a hall most preferred. Many studies aimed to identify a comprehensive set of words, and the authors feel that now, well-defined terms are needed to provide more consistency within the field of architectural acoustics, across many studies. This effort has been started by Weinzierl et al. (2018), but it is important to continue this work with measurement-based auralization using accurate source directivity representation and spatially accurate auralization techniques. Additionally, for individual preference, repeatability of an individual's preference has been little studied. If an individual's preference can be determined in a repeatable manner, then techniques to predict this preference, such as a quick listening test, could enable more efficient studies regarding how individual preference can vary between music genres, listener demographics, and even cultural or geographic regions.

6.3 Realistic Measurement-based Auralizations

To allow for both realistic auralization and subsequent objective analysis, the concert hall orchestral research database, or CHORDatabase was used. This database is a set of spherical microphone array RIR measurements that were captured in two source conditions. Full details about this measurement database can be found in chapter 5. First, a three-part omnidirectional sound source was used to provide a standardized, repeatable condition for objective metric calculation and spatial sound field analysis. Secondly, a 20-channel compact spherical loudspeaker array (CSLA) was used to provide built-in frequency dependent source radiation patterns for realistic auralization of a full orchestra. A 20-position source measurement grid was consistently placed on the stage of 21 different concert hall environments, and separate RIR measurements for 20 different orchestral instruments were captured at a single seat using the spherical microphone array.

6.3.1 Realistic Full-orchestral Auralizations

Full-orchestral auralizations were generated using the CSLA described in chapter 4. The CSLA was used in 21 different hall environments listed in Table 5.1 For each orchestral source location, filters were designed to build accurate frequency-dependent radiation patterns for each instrument into each RIR. A separate measurement was captured for 20 different source locations in an orchestral measurement grid shown in Figure 4.14. After these instrument RIRs were captured, each measurement was diffuse-field equalized, depending upon the instrument radiation pattern that was selected, to correct for the non-flat response of the array and sensitivity differences between instrument radiation patterns. Then, each measurement was crossed over to a low frequency measurement at the same source location

made with the subwoofer component of the omnidirectional sound source at 200 Hz. This measurement provided better low frequency signal-to-noise ratios (SNRs) for auralization.

The anechoic recordings that were used to generate full-orchestral auralizations were generated by TU Berlin for Beethoven's 8th Symphony.[100] These recordings were made in an anechoic chamber with a conductor present. Recordings were made with multiple instrumentalists in the chamber at a time, groups by strings (2 violins, 2 violas, 1 cello, 1 double bass), woodwinds (2 flutes, 2 oboes, 2 bassoons, and 2 clarinets), brass (2 French horns and 2 trumpets), and percussion (1 timpani). This set of 21 musicians also took multiple recording takes to generate a larger set of recordings to represent a 61-piece full orchestra. To help isolate each instrument recording, large barriers were placed between players. Also, recording microphones were strategically located to maximize signal and minimize crosstalk between instruments. Although some crosstalk exists in each individual recording, the effects are inaudible in the context of a full-orchestral auralization. To extend the 20 source position measurements in each hall, a larger grid of 61 source positions was developed based upon the sparse 20 source grid. This larger grid is shown in Figure 6-2. Once each instrument's spherical microphone array RIR (MicRIR) was encoded into a spherical harmonic RIR (ShRIR), the direct sound was isolated in each ShRIR. For source positions in the 61-source grid that did not align with the 20 actual measurement locations, a copy of the ShRIR was made, and the direct sound was slightly rotated to the new source location. Next, the amplitude was modified to adjust for differences in spherical spreading at the new location. All of this processing was done relative to the measurement location for each of the 20 sources, and the closest source with a matching directivity pattern was used as the base ShRIR. This technique generated a new set of 61 ShRIRs, each corresponding to a full 61-piece orchestra for Beethoven's 8th Symphony.

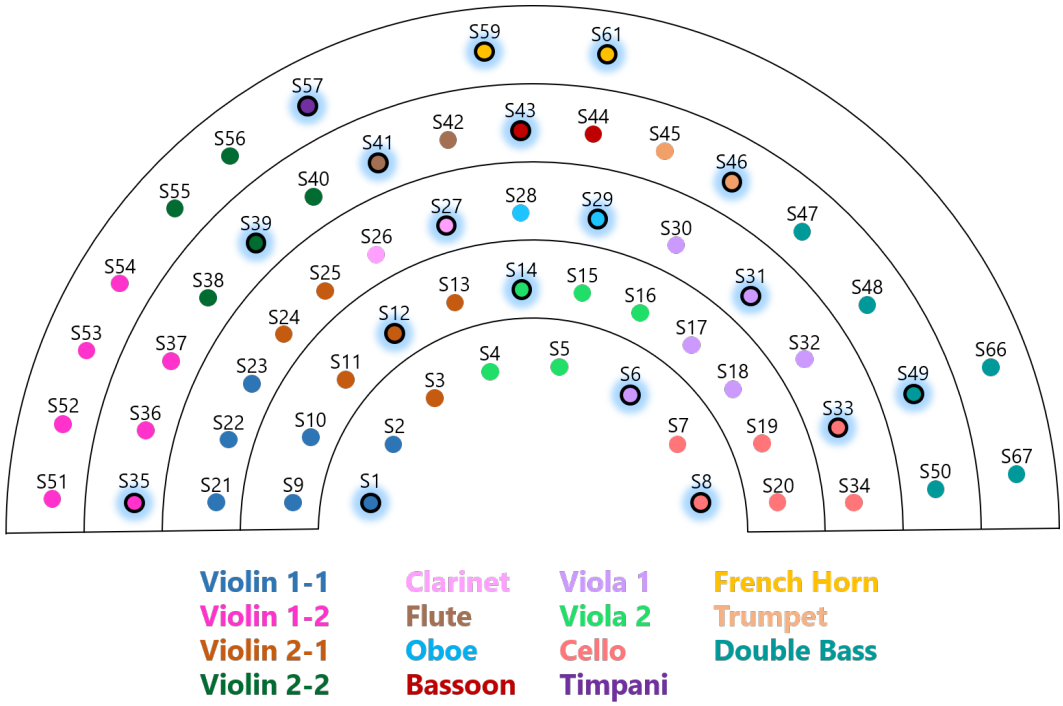


Figure 6-2: A 61-piece orchestral grid, compatible with the 18-source measurement grid made in each concert hall. Actual measurement locations are highlighted with a blue glow. Since Beethoven's 8th symphony did not contain trombones or tubas, they are excluded from this setup (reduced from 20).

Each ShRIR was convolved with one of the 61 anechoic recording tracks. To help balance the section sound without biasing the balance to any particular hall, an anechoic auralization was generated for the 61-piece orchestral setup. For each source, a plane wave was generated for a listener seated 15 m from the conductor location, and the relative level of the plane wave was attenuated for spherical spreading and air absorption at standard room conditions.[101-102] All of the instruments were superimposed to generate a full orchestral auralization, and overall section gains (strings, woodwinds, brass, percussion) along with individual instrument gains (violin, viola, etc.) were determined which provided a proper subjective balance between instruments. Effectively, a simulated auralization of the orchestra setup in a large anechoic chamber was generated for a receiver seated 15 m from the orchestra. This adjustment ensured a proper instrumental balance, without biasing the balance towards any hall. The final anechoic motif selected for the study was from the second movement, from a starting time of 50.0 s to an ending time of 67.35 s in the provided tracks. This process was repeated for all full-orchestral measurements, generating 21 unique auralizations of a full orchestra in different room environments. These auralizations were used as the basis for a subjective experiment regarding the overall perception of concert halls. Experimental design for this study is provided in section 6.4.

6.3.2 Spherical Microphone Array Beamforming Analysis

Along with measurements for realistic auralization, the CHORDatabase includes an omnidirectional source MicRIR measured using a three-part omnidirectional sound source. This source has a low-, mid-, and high-frequency component, and when combined into a single broadband RIR, omnidirectional source radiation is maintained up to the 5 kHz one-third octave band, described in detail in section 5.9.1. Sensitivity differences, amplifier gain differences, and diffuse field equalization were all properly equalized for each source, and a 3-band linear-phase crossover filter was designed to ensure in-phase summation between loudspeakers. The direct sound of the three measurements were also checked for time-alignment during post-processing, to compensate for differences in measurement latency. This measurement is standardized and repeatable, so other researchers or consultants could theoretically take similar measurements and produce comparable results for objective sound field analyses. It is important to ensure compatibility with existing measurement setups when defining new objective metrics and analyses. If others do not have access to an adequate setup for a specific metric, it will not find widespread use.

For the current study, a spherical microphone array was first used to extract a standard omnidirectional microphone response and a lateral dipole response, with the null oriented at the source location. Dick and Vigeant showed that these responses can be directly calculated from the MicRiR output of the array.[94] From these responses, standard metrics from ISO 3382 were calculated for all of the different auralization seat locations.[47] Spherical array beamforming or plane-wave decomposition (PWD) was also used to provide advanced spatial analysis of room sound fields. Using a spherical microphone array, the directional response of the microphone can be arbitrarily controlled in post-processing to have a beam-like response pattern. This beam pattern can be flexibly rotated around the full-sphere, and at each orientation, a directional RIR (DirRIR) was generated. Finally, energy in specific frequency and time ranges of the DirRIRs was analyzed to produce spatial energy maps of a RIR. An example of such a map of the early energy in a RIR from 10 – 100 ms is shown in Figure 6-3.

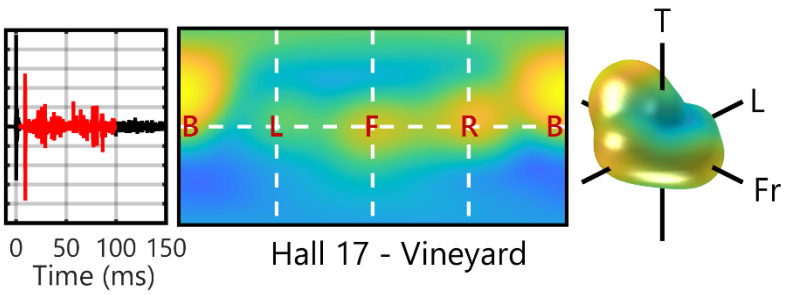


Figure 6-3: An example of a spatial energy map created using plane-wave decomposition of the early energy (10 – 100 ms) in a vineyard-style hall in the CHORDatabase from the 1 – 4 kHz bands.

These maps were generated for different time regions in the RIR of each hall used in the present subjective study. These time and spatial regions are correlated with subjective attributes and factors in the present study were then correlated with subjective attributes and factors resulting from a principal components and factor analysis described in section 6.5. With full access to the time and spatial regions in the RIR, the CHORDatabase can be used to propose new metrics, with increased spatial resolution and potentially better correlation with the fundamental subjective aspects of concert hall acoustics. Correlations between subjective data and spatial energy beamforming analyses will be presented in section 6.6.

6.4 Subjective Study Experimental Design

The first goal of the present study was to map out overall subjective impression in concert halls. The second goal was to then connect both overall average preference and individual preference onto this *perceptual space* or map. Some previous studies focused on preference alone, presenting an A-B forced-choice rating task. Others have focused on studying a wide variety of factors (20, 30, 40, …) and even others have used individually provided and defined factors to map out this preference space. The ultimate goal of these works is to determine the number of dimensions required to explain perceptual difference between halls. This study was targeted to meet both goals, which drove key decisions on experimental design including the number of subjective terms, the number of halls, the rating task, and the randomization design. All of these will be described in sections 6.4.1 through 6.4.5.

6.4.1 Subjective Attribute Selection

Looking at the summary of previous terminology used in concert hall acoustics studies, found in Figure 6-1 and Table 6.1, a subset of ten words was selected that covered almost the entire range of terms found to be important in previous literature: brilliance, envelopment, intimacy, proximity, reverberance, source width, spatial clarity, strength, temporal clarity, and warmth. This set is not fully exhaustive, but rather, it is focused on terms that have been

found to be subjectively significant in explaining differences between concert halls. Despite this reduction, words deemed most important in previous studies explain the majority of the variance of perceptual ratings, so selection of these words will still ensure that most of the perceptual space is well-sampled. The list, along with definitions of each term and the high and low anchors provided on the rating scale, are all given in Table 6.2. It should be noted that terms relating to balance, blend, and ensemble of the orchestra were not included. As the orchestra in this study was an unchanging, constant variable, such terms would not be appropriate to draw overall conclusions on using the current fixed-arrangement orchestra auralizations. These terms are likely related to orchestra arrangement, and most likely musicians adapt to each hall in a non-linear way.

The high and low anchor selection and numerical scale range (-50 to 50 or 0 to 100) and most of the definitions and anchors were based upon the room acoustic quality inventory (RAQI) focus group of experts led by Weinzierl et al. with some modifications.[43] The terms were translated to English, but for a few of the terms, the authors felt the translation may cause misinterpretation by the native English-speaking subjects. Also, the definition and anchors for the term envelopment was taken from previous studies on this specific perception by Dick and Vigeant.[52]

6.4.2 Subjective Rating Task and Interface

These tests are traditionally done using a paired-comparison or a single-stimulus rating task. The paired-comparison task is straightforward in nature and allows subjects to make fine judgements between stimuli through direct comparison. Despite its inherent accuracy, it is difficult to implement using large numbers of stimuli, as the number of possible comparisons grows exponentially with total number of stimuli. The single-stimulus method allows for subjective ratings given to be elicited efficiently, as only one rating is needed for each hall, opposed to comparing it with every other stimulus. Although efficient, the rating provided with this method typically requires a larger sample size to determine significant conclusions. It is also quite difficult for subjects to compare subtle aspects of concert halls using this method, such as source width or envelopment.

Table 6.2: The ten selected subjective attributes included in the experimental design of the subjective study. The high and low anchors, along with definitions provided to the subjects are listed below. Most of the anchors and many of the definitions are from the RAQI work by Weinzierl et al.[43]

Attribute	High Anchor	Low Anchor	Definition
Brilliance	Very brilliant	Not brilliant	Brilliance is the perception of emphasized treble (higher-pitched) sound energy within a room.
Envelopment*	Completely surrounded*	Not at all surrounded*	Envelopment is the perception of being fully immersed or surrounded by a room's sound energy.*
Intimacy	Intimate	Remote	Intimacy is the perception that you are listening in a smaller, more intimate space as opposed to a larger, more remote space.*
Proximity*	Close	Far*	Proximity refers to the perceived closeness of the orchestra from where you, the listener, are seated in the hall.
Reverberance	Reverberant	Dry	Reverberance is created when sound lingers in a space, even after the orchestra has stopped playing.*
Source Width	Very wide*	Not wide (narrow)*	Source Width is the perceived size of the orchestra in the horizontal (left to right) dimension when looking at the stage.
Spatial Clarity*	Clear	Blurred	Spatial Clarity is the ability to distinguish individual musicians from one another. When musicians are distinct and easy to separate spatially, a hall is considered clear. Conversely, when musicians are hard to spatially separate, a hall is considered blurred.*
Strength*	Loud	Soft	Strength is the apparent loudness of the orchestra in the hall.
Temporal Clarity	Clear	Blurred	Temporal Clarity is the ability distinctly hear successive notes over time in a musical passage.*
Warmth	Warm	Cool	Warmth is the perception of emphasized bass (lower-pitched) sound energy or bass-mid sound energy within a room. Conversely, Cool rooms are lacking in bass or bass-mid sound energy.
Preference	I like it	I don't like it	n / a

*denotes terms that have significant modifications from those proposed in the RAQI study.[43]

The test type selected for this study was a multiple-stimulus comparison technique, based upon the multiple-stimulus with hidden reference and anchor (MUSHRA) test used in audio quality studies.[103] This study allows for direct comparison of multiple stimuli, which provides subjects the ability to directly compare stimuli. Even with this comparison, the test is also time efficient, as subjects can compare all stimuli side-by-side, instead of using many separate pairings. Studies have shown that this technique produces the same subjective results as the single stimulus case in a study on reverberance, but subjective answers were found to be more repeatable and reliable with the multiple-stimulus test.[104] This increase in reliability allows for statistical conclusions to be reached more quickly, or with greater

statistical power. The testing interface was developed in the real-time audio processing and visual coding language Max7.[88] A screen capture of the testing interface is provided in Figure 6-4. This interface was implemented to work for all ten of the subjective terms.

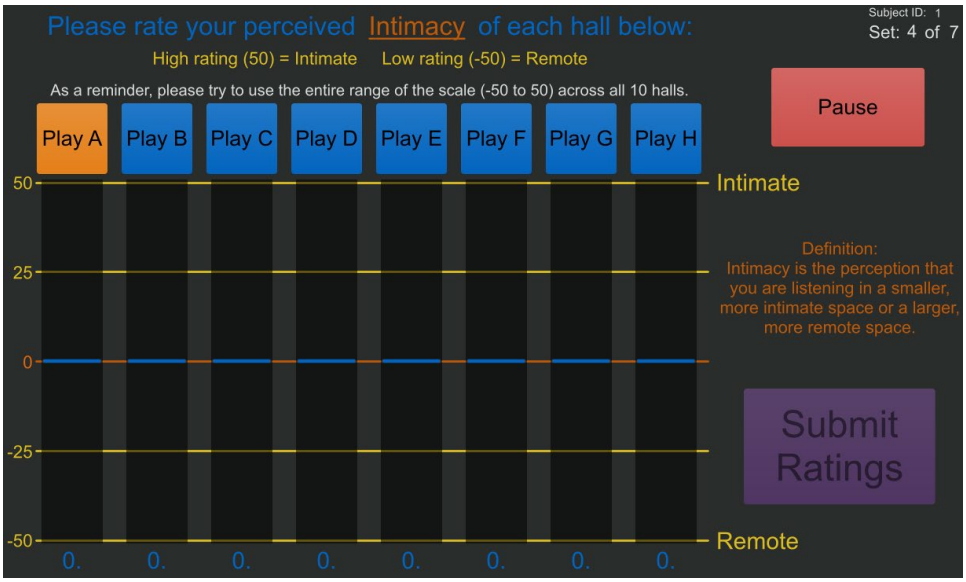


Figure 6-4: Testing interface for the multiple-stimulus comparative rating task used in the study. Subjects were able to switch freely and compare all eight halls side-by-side.

Using the interface, subjects could switch freely between eight hall stimuli while the orchestral passage looped. They could switch back-and-forth as many times as they would like and adjust the vertical slider bars relative to one another. Once every hall had been played at least once, and every slider had been adjusted from zero, the submit ratings button activated for final answer submission. After submission, another set of eight halls was loaded. The eight halls were presented in a random order using the interface. On each screen, the subject was prompted and reminded of which attribute was currently being rated. As subjects rated multiple attributes, much effort was taken to clearly define and remind the subject of the current word, its rating scale, and its definition. Finally, subjects were reminded to use the full range of the rating scale.

For the interface, eight was selected as an appropriate maximum number of stimuli. The selection procedure for halls included in the study will be described in more detail in section 6.4.3. This number provided a balance between increasing hall variety, without generating an overwhelming task for subjects. Using in-house piloting, ten stimuli were originally considered, but that number was found to be too tiring, increasing testing time and potentially biasing results due to fatigue.

6.4.3 Hall Selection using k-means Clustering

With only eight halls for comparison, the 21 possible full-orchestral auralizations available from the CHORDatabase needed to be reduced to a smaller subset. Additionally, a set of eight halls was deemed too small, possibly under sampling the full perceptual range of concert halls and limiting the extensions of results to the overall population. To balance these tradeoffs, the 21 halls were reduced to a representative set of 14 halls by removing 6 halls that were perceptually very similar. To identify halls of similarity, the broadband-averaged EDT, C80, and G was calculated for each of the 21 halls at the receiver location 15 m from stage. Each metric was mean-centered and normalized to the standard deviation across the 21 halls, so that each hall could be represented as a point in a three-dimensional metric space. K-means clustering was performed for $2 - 21$ clusters, where k is the user-selected number of clusters. This clustering algorithm uses randomly generated initial conditions, so it was run 1000 times for each value of k, and the result with the minimum within-cluster error was retained. By analyzing the hall groupings that occurred at each value of k, the following eight groups were generated of halls with similar metric values, plotted their raw (un-normalized) metric values in Figure 6-5. Clear groups emerge, and some groups have more members than others.

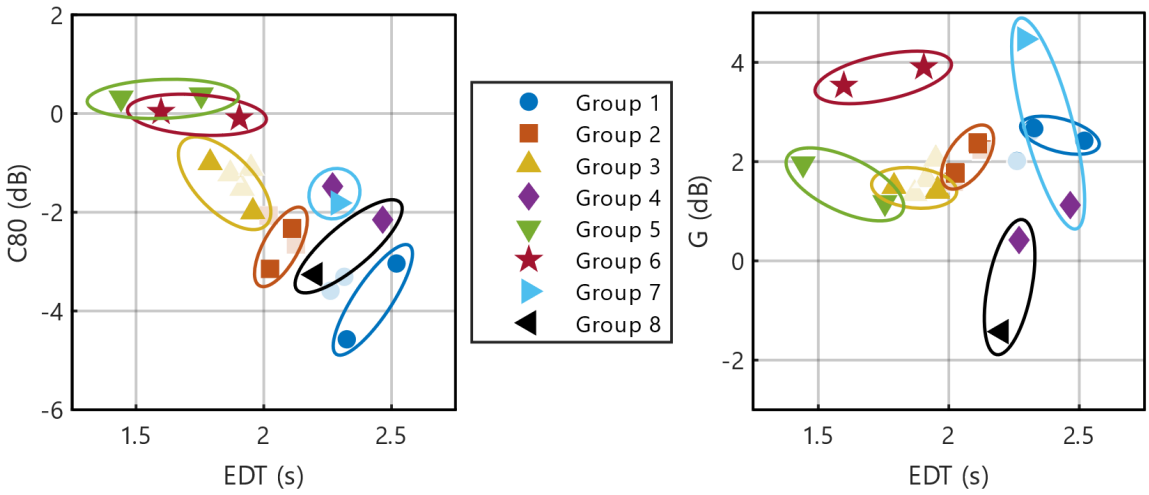


Figure 6-5: Halls were placed on a three-dimensional space, defined using broadband averages of three room acoustic metrics, EDT, C80, and G. A clustering analysis was used to group the halls into similar sets or groups. Groups with more than two halls were reduced to a representative set of two (halls removed are shown as slightly grayed out) and halls were paired within each groups to seven pairs of different, but similar halls. This technique reduced the set to a smaller, representative sample.[***]

To reduce the 21 halls to a more manageable set, two representative halls were selected from each of the groups with more than two members. For the singleton-clusters, each of these

[***] An animated version of this figure generated by the author for use in presentations. can be found online at: https://sites.psu.edu/spral/files/2019/07/clustering_gif_fixed.gif.

halls were retained. This process was completed using informal subjective comparisons of halls, and the two halls that were mostly similar, but added some additional variety into the overall set were selected. Two halls were removed from both groups 1 and 2, and three were removed from group 3. This reduction created a smaller subset of 14 halls for the study. A controlled randomization technique, explained further in section 6.4.4, was used so that each subject received a randomized set of 7 of the 14 halls along with one anchor stimulus on the interface. This generated the set of eight stimuli the subjects were presented for the test. The anchor stimulus was a modified version of one hall, adjusted for each subjective term since different perceptual words can have competing low-end perceptual qualities (e.g. clarity and reverberance). The anchor stimulus was consistent for the same word for all subjects, allowing the ratings to be comparable across all subjects to a consistent baseline. Since the anchor changed between subjective terms, it was not included in statistical correlations and factor analyses.

6.4.4 Incomplete Block Randomization Considerations

A controlled randomization scheme was implemented to enable the inclusion of 14 halls and 10 subjective terms without fatiguing subjects. Each subject randomly received seven halls from the subset of 14, and this set of halls was consistent throughout the test. To ensure that each subject saw a stimuli set with a wide perceptual range, each hall in groups 1 – 5 was paired with the other hall from its group displayed in Figure 6-5. For the singleton clusters, since they did not have a pairing (and were not suitable to pair together), group 6 was split, and one of the two were each paired with the stimuli in groups 7 and 8. This pairing was motivated through perceptual listening, as cluster 6 appeared to contain subjective differences than were not sufficiently predicted by EDT, C80, and G alone. Since only three metrics were used in the grouping, any perceptual aspects not explained by EDT, C80, and G, such as spaciousness, could not be captured. These final groupings were all perceptually validated again by in-house listening and piloting within the laboratory group.

With these pairings, each subject received one random stimulus from each pair, ensuring an adequate perceptual range while still randomizing any hall-related effects. The anchor stimulus provided a consistent low baseline, allowing for compatibility across subjects. Additionally, each subject first rated preference twice for their assigned eight halls, and they then rated a subset of five of the ten total subjective attributes. To ensure equal sampling across hall and subjective attribute, an incomplete-block randomization design was used, shown by the diagram in Figure 6-6. Subjects 1 and 2 and subjects 3 and 4 would be paired to rate the same set of words, but each subject would receive the complement of the randomized

hall selection received by the other. In the same way, subjects 1 and 3 and subjects 2 and 4 were paired to rate the same set of seven halls, but with the complement of the randomly selected attributes. Thus, after every four subjects who completed the study, each hall will have been rated for each of the subjective terms one time, and each hall would be rated on preference four times (by two subjects with two repetitions). This scheme ensured that mean values estimates were unbiased by the randomization scheme while limiting testing time to prevent subject fatigue.

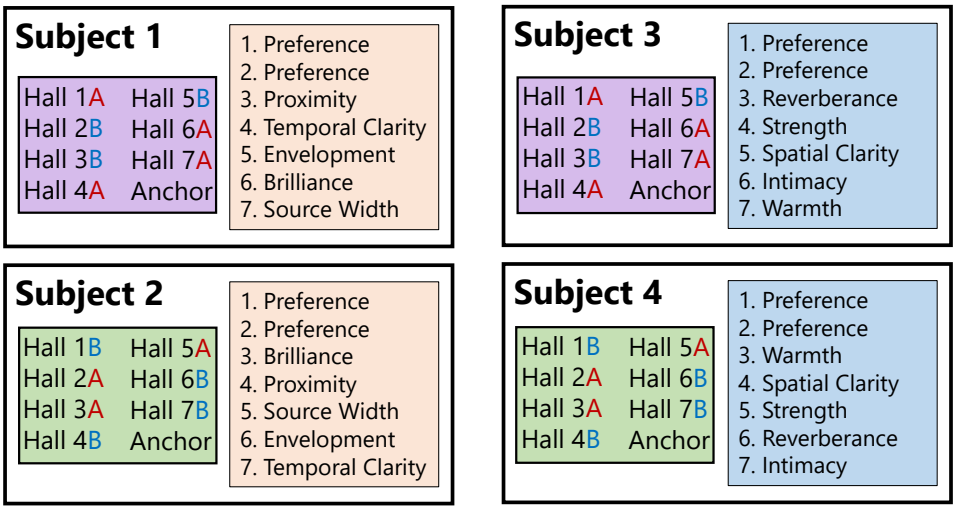


Figure 6-6: A diagram of the incomplete-block controlled randomization used in this study.

6.4.5 Final Study Format

For the final setup of the test, subjects began by filling out an informed consent form and a participant information survey. Next, a standard audiogram was administered, to ensure subjects had a minimum of 15 dB HL hearing thresholds from 250 – 8000 Hz. If they did not meet this criterion, they were excluded from the study and were given a $5 gift card. Additionally, subjects were required to have a minimum of 5 years formal musical training, and they were required to be currently active in a musical ensemble or private study. If these criteria were met, subjects were given a tutorial explaining the testing interface, and they were instructed to rate their preference of the eight different halls. To ensure that the attributes randomly selected for the subjects to rate did not influence their preference ratings, preference was rated first for each subject, and no mention of the attributes was made until after the preference rating portion of the test was complete. After the tutorial, subjects were presented with three different sets of the seven halls plus the anchor using the testing interface in Figure 6-4. Subjects rated preference three times for an identical set of stimuli (the subjects did not know it was the same stimuli), but the first set was used as a hidden practice, and these data was removed for additional training and to allow adequate time to adapt to the rating task.

After subjects completed the preference sets, they were given a 5-minute break, and then a new tutorial was administered for the subjective attribute ratings. The same style interface was used, but subjects were specifically instructed to now rate attributes with defined meanings and interpretations, as opposed to preference. After the tutorial, the test administrator presented the subject with definitions for the first three subjective attributes they would rate, and the subject was asked to explain back these terms to the administrator. The test administrator ensured that the subject was interpreting the terms properly, and if they noticed any problems in interpretation, they helped to re-orient the subject's interpretation of the word. After the first three terms were rated, subjects had another 5-minute break. They were then presented with the definitions of the fourth and fifth attributes, necessary corrections to interpretation were made, and subjects rated the final two attributes. To finish up, subjects completed a follow-up survey on their experience during testing and were compensated with a $15 gift card.

6.5 Results I: Correlation and Factor Analysis

First, the statistical techniques of correlation, principal components analysis (PCA), and multiple factor analysis (MFA) will be used to study the perceptual space, the multicollinearity existing in the perception of concert hall acoustics, and the factors that are most important in explaining the differences in perceptual ratings between halls.[105] A total of 16 subjects participated in the study, meaning a total of four ratings for each perception across all halls has been recorded. For preference, eight subjects provided two preference ratings for each hall, creating 16 total ratings of preference per hall. Each subjective attribute can be correlated with overall average preference to investigate which factors most closely related to overall preference. Additionally, overall average preference and each subjective impression can be correlated with existing objective metrics from ISO 3382, to accurately the current objective metrics predict their underlying perceptions using realistic full-orchestral auralizations. This comparison was not directly possible in previous work studying measurement-based full-orchestral auralizations, as no standard measurement source was used.[40]

To understand the size of the perceptual space in concert hall perception, first the correlations between all ten subjective attributes were analyzed, to observe the multicollinearity existing in these common perceptions. Then, a PCA was used to identify the number of factors required to explain most of the variation in perceptual ratings. Next, Varimax rotation was used to provide a rotation of the principal components that are easier to interpret with the original subjective attributes. The newly defined factors were then correlated with both the original subjective attributes and objective metrics to aid in

interpretation. Finally, with proper interpretations, a correlation analysis of individual (not average) preference ratings was conducted to investigate how preference in concert hall acoustics can vary at the individual level.

6.5.1 Perceptual Factor Space and Average Preference Results

First, the average subjective rating for each of the ten attributes and preference were calculated for each hall. Since the 10 attributes were clearly defined for all subjects, this averaging provides a more accurate subjective estimate of each term for the given hall. Since concert hall preference is not a defined attribute, averaging will remove the inter-subject variation of the data, but the average preference rating provides a general idea of what is preferred in concert halls, across all subjects. The information lost in averaging will be revisited in section 6.5.2.

6.5.1.1 Correlations between Subjective Attributes and Average Preference

First, the Pearson correlation coefficients were estimated between overall average preference and each of the ten attributes. The numbers shown in bold are significant with $p < 0.05$. The correlation was run between the average ratings for each hall, so the effective sample size in this calculation is the number of halls, $n = 14$. The first finding to note is the high correlation between overall average preference and proximity. This factor has been noted in previous literature, and most recently, Lokki et al. found that proximity was the single factor that best correlated with overall consensus preference in his work.[38] The agreement between his work and this currently work in connecting average preference connection with proximity is important, as both studies used different approaches to study a similar problem and still converge on this finding. Additionally, the factors of spatial and temporal clarity show high correlation with average preference, along with warmth. Interestingly, reverberance, strength, and intimacy, all which have been thought to be important in previous work show little connection with overall *average* preference.

Table 6.3: Correlations between average overall preference (Avg. Pref.) ratings across all halls and brilliance (Brill), envelopment (Env), intimacy (Int), proximity (Prox), reverberance (Rev), source width (SW), spatial clarity (SC), strength (Str), temporal clarity (TC), and warmth (Wrm). The bolded correlations significantly different from zero ($p < 0.05$) are those that exceed a 0.5 threshold.

	Individual Attributes									
	Brill	Env	Int	Prox	Rev	SW	SC	Str	TC	Wrm
Avg. Pref.	0.06	0.40	0.31	**0.81**	0.03	0.44	**0.60**	0.29	**0.68**	**0.58**

6.5.1.2 Correlations between Subjective Attributes and Metrics

All of the hall-average subjective impressions, including preference, were also correlated against existing objective metrics from ISO 3382, including reverberation time (T30, T20, T10), early decay time (EDT), clarity index for music (C80) and for speech (C50), center time (Ts), strength (G), early strength before 80 ms (G_E), late strength after 80 ms (G_L), lateral energy fraction (J_{LF}), and the late lateral energy level (L_J). The overall preference ratings exhibit highest correlation with G_E and C80, both highly related to early energy in a RIR. Very little correlation is found with any decay-based metrics or the lateral energy metrics. Looking at individual metrics, no clear correlations exist between the perception of brilliance and existing standard metrics. For envelopment, the metric to predict envelopment, L_J, is far worse at predicting the impression than either early strength or late strength. Intimacy shows no metrics with extremely high correlation, but it does have a very strong negative correlation with center time, Ts. Proximity shows promising correlation with G_E, despite no known metric existing for this perception.

Table 6.4: Correlations between the hall-averaged attribute rating, preference, and existing metrics. Values in bold are significantly different from zero ($p < 0.05$), meeting a magnitude threshold of 0.54

Att.	Metric											
	T30	T20	T10	EDT	C80	C50	Ts	G	G_E	G_L	J_{LF}	L_J
Avg. Pref	0.14	0.11	0.17	-0.22	**0.68**	**0.59**	-0.42	**0.65**	**0.79**	0.40	-0.08	0.11
Brill	0.30	0.36	0.41	0.41	-0.30	-0.07	0.32	0.27	0.07	0.45	0.08	0.42
Env	0.33	0.32	0.43	0.36	-0.13	0.09	0.36	**0.66**	0.42	**0.78**	0.07	0.50
Int	**-0.62**	**-0.64**	**-0.62**	**-0.75**	**0.56**	0.26	**-0.80**	0.01	0.25	-0.27	-0.34	-0.39
Prox	0.01	-0.02	0.04	-0.27	**0.68**	**0.76**	-0.49	**0.76**	**0.86**	0.51	-0.21	0.11
Rev	**0.69**	**0.69**	**0.72**	**0.60**	-0.31	0.01	**0.57**	0.43	0.18	**0.64**	0.41	**0.70**
SW	0.53	**0.55**	**0.63**	0.52	-0.12	0.18	0.36	**0.65**	0.44	**0.79**	0.15	0.52
SC	**-0.56**	**-0.56**	-0.49	**-0.71**	**0.73**	**0.56**	**-0.79**	0.34	**0.56**	0.00	-0.50	-0.27
Str	0.43	0.44	**0.57**	0.52	-0.15	0.21	0.39	**0.64**	0.41	**0.79**	0.15	**0.54**
TC	-0.33	-0.34	-0.32	**-0.56**	**0.69**	0.34	**-0.66**	0.10	0.38	-0.22	-0.32	-0.41
Wrm	-0.44	-0.48	-0.45	-0.50	0.51	0.40	-0.51	**0.54**	**0.61**	0.32	-0.09	-0.12

Reverberance shows strong correlation with all decay-based metrics, but surprisingly, EDT is not found to have the highest correlation, but rather, T10 shows the highest correlation. This finding is not in line with other literature on the perception of reverberance and its perceptual correlation with EDT.[10] Source width shows extremely weak correlation with the metric stated to specifically predict the perception, J_{LF}, and is much more strongly correlated with G or G_L. Strength shows high correlation with G and G_L, and both clarity factors are highly correlated with C80. Finally, warmth appears to be somewhat related to G

and G_E. All of these suggestions are preliminary, as more subjects are needed for stronger conclusions. Despite low sample sizes, trends emerge, showing better performance for clarity and decay-based metrics and rather poor performance is seen for any of the spatial energy metrics. Additionally, it appears some perceptions such as proximity might be highly related to early energy in the RIR.

6.5.1.3 Principal components and Factor Analysis

A correlation analysis was run between all of the ten perceptual attributes in the study. This correlation was run on the hall average. With a higher sample size, the significance test of the correlations coefficient not equal to zero is less helpful for identifying correlations (as most all of the numbers in the table are different from zero at a 0.05 significance level). Instead, the bold numbers below indicate correlations that exceed a 0.5 threshold. Clear multicollinearity exists, which indicates that PCA and MFA should be effective in reducing and simplifying the perceptual space. Strength shows high correlation with envelopment, reverberance, and source width. The clarity factors, proximity, intimacy, and warmth also all appear to be correlated.

Table 6.5: Correlations between the hall-average ratings for each of the ten subjective attribute ratings (n = 14). Large amounts of multicollinearity exist in the perceptual space.

	Brill	Env	Int	Prox	Rev	SW	SC	Str	TC
Env	0.20	–	–	–	–	–	–	–	–
Int	-0.16	-0.47	–	–	–	–	–	–	–
Prox	0.16	0.41	0.31	–	–	–	–	–	–
Rev	0.26	**0.67**	**-0.73**	0.00	–	–	–	–	–
SW	**0.65**	**0.68**	-0.25	0.47	**0.66**	–	–	–	–
SC	-0.14	0.06	**0.70**	**0.70**	-0.53	-0.06	–	–	–
Str	0.51	**0.77**	-0.37	0.37	**0.72**	**0.87**	-0.03	–	–
TC	-0.16	-0.21	**0.66**	0.50	**-0.61**	-0.16	**0.74**	-0.32	–
Wrm	-0.37	0.30	0.49	**0.63**	-0.25	0.05	**0.70**	0.07	0.50

To reduce the dimensionality of the perceptual space, principal component analysis (PCA) was used to calculate the eigenvectors of the ten perceptual attributes. The PCA was done on the raw perceptual ratings (un-averaged), excluding the anchor stimulus ratings and excluding the preference ratings. Since the anchor stimulus changed for each perception, it is not valid to view these ratings as consistent across subjective terms. The error variance remaining in the model after including each additional principal component (PC) is shown in Figure 6-7. The blue bars represent the variance explained by each individual PC, and the red

line represents the cumulative total variance explained after the addition of each new PC into the model. The Eigenvalues associated with these plots can be found in Table 6.6.

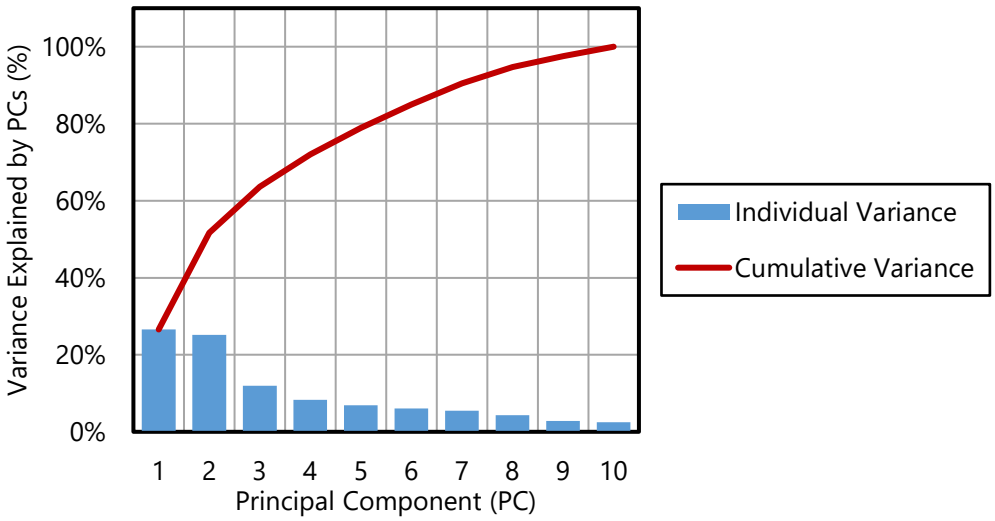


Figure 6-7: Results from the PCA of the perceptual space, showing the error remaining in the data set after the addition of each PC in (a) and the variance explained by each PC in (b).

The goal of the PCA is to find the number of factors that explains much of the variance in the model, while still trying to maintain as simple a model as possible. Often, the total number of dimensions is determined from the *elbow* in the scree plot, which can be an arbitrary choice.[105] For this study, the elbow appears to be somewhere between 2 – 4 factors. From inspecting the proportion of the variance explained by each PC from Table 6.6, it is clear that diminished benefit is found from adding five or more dimensions to the problem. Clear benefits are found with the addition of PC1 (27%) and PC2 (25%), and a fairly large benefit is found with the addition of PC3 (12%). The inclusion of PC4 (8%) in the model is somewhat debatable, as it has a larger benefit than PC5 or greater, but it is also much more marginal of a benefit compared to PC1 – PC3. For this study, both the three-dimensional and four-dimensional model will be considered, so that the benefits and costs of either choice can be investigated in final factor interpretations.

Table 6.6: Summary data from the PCA, showing the eigenvalue, portion of the total variance explained by each PC, and the cumulative variance as each new PC is added to the model.

Principal component	Eigenvalue	Proportion of Variance	Cumulative Variance
1	**2.657**	**26.6%**	**26.6%**
2	**2.514**	**25.1%**	**51.7%**
3	**1.195**	**12.0%**	**63.7%**
4	**0.834**	**8.3%**	**72.0%**
5	0.691	6.9%	78.9%
6	0.606	6.1%	85.0%
7	0.548	5.5%	90.5%
8	0.429	4.3%	94.7%
9	0.281	2.8%	97.5%
10	0.246	2.5%	100.0%
Total:	10.000	100.0%	

To help with interpretation, new factor scores were calculated for all 14 halls. Using the eigenvectors for each PC (which act as linear weighting factors), PC values were calculated for each hall by properly weighting and summing the average ratings for each of the ten attributes. These factor loadings generate factor scores for each hall, mapping each hall to the new perceptual factor space. Next, to aid in interpretation, a varimax rotation was performed on the PCs to rotate the PCs into new factors that were more easily interpretable to the original perceptions.[105] Varimax rotation attempts to maximize the variance in correlation coefficient estimates between each individual new factor and all of the original attributes. Effectively, it ensures that factors are highly correlated with certain attributes, while being mostly uncorrelated with other attributes. This optimization generates factors that were specifically identified with single or groups of attributes, as opposed to factors that somewhat related to each attribute, which is often the result of a PCA. The correlation with the original attributes is provided for the varimax rotated factors in Table 6.7. The varimax rotation solution is dependent upon the number of PCs that is selected to be maintained, so a separate solution is given for the four dimensions solution (factors 4.1, 4.2, 4.3, and 4.4) compared to the three-dimensional solution (factors 3.1, 3.2, and 3.3).

Table 6.7: Correlations between the varimax factors, both the set of retaining four factors and the set of retaining three factors, with the hall-averaged subjective attribute ratings and average preference. Percentages of the total variance explained by each factor are provided in parenthesis. The ordering of factors 3.1 and 3.2 has been swapped to match interpretation with factors 4.1 and 4.2 – 4.3.

Factor (% Var.)	Attribute										Avg Pref
	Brill	Env	Int	Prox	Rev	SW	SC	Str	TC	Wrm	
4.1 (25%)	-0.15	0.00	**0.78**	**0.43**	**-0.54**	-0.05	**0.77**	-0.04	**0.59**	**0.69**	**0.58**
4.2 (19%)	0.13	**0.84**	-0.13	**0.31**	**0.49**	**0.37**	0.19	**0.82**	**-0.35**	**0.44**	0.38
4.3 (16%)	0.18	**0.33**	0.02	**0.66**	**0.56**	**0.89**	0.13	**0.59**	0.08	0.04	**0.59**
4.4 (12%)	**0.9**	0.00	-0.12	**-0.31**	0.15	**0.29**	-0.24	0.26	**0.29**	**-0.33**	-0.06
3.2 (25%)	-0.19	0.07	**0.76**	**0.53**	**-0.46**	0.05	**0.78**	0.03	**0.55**	**0.71**	**0.64**
3.1 (26%)	0.23	**0.73**	-0.19	**0.46**	**0.68**	**0.7**	0.08	**0.85**	-0.26	0.21	0.40
3.3 (12%)	**0.77**	-0.18	0.03	-0.07	0.12	**0.46**	-0.14	0.14	**0.47**	**-0.36**	0.12

Clear patterns emerge for interpretation. First, in the four-factor solution, 4.1 appears to be related to clarity, having high correlations with spatial clarity, intimacy, temporal clarity, and warmth. The inclusion of warmth with the perception of clarity is an interesting result, as warmth is more often grouped with loudness and strength in other studies.[38] The second and third factors, 4.2 and 4.3, both have similar correlations with reverberance and strength. The factors differ in terms of their correlations with the spatial perceptions. Factor 4.2 is highly correlated with envelopment and strength, while factor 4.3 is also highly correlated with strength, but instead of envelopment, it is correlated with source width. This finding is interesting, for the perception of spaciousness has previously been reported as separable into two components: envelopment and apparent source with.[32,106] The separation of these two factors in the four-dimensional model (very loosely) supports the claim of two separable and independent spaciousness factors. It is important to note that source width has clear correlation with strength, but the metric to predict source width, J_{LF}, has no strength-based component. The 4.2 factor also groups the potential correlation between envelopment and warmth in the same factor. Both factors show a strong correlation with reverberance, opposite the negative correlation between reverberance and factor 4.1. Finally, the fourth factor, 4.4, appears to be related to timbre, with a very high correlation with brilliance and a slight negative correlation with warmth.

Comparing this result to the three-factor solution, factor 3.2 appears to be quite similar to factor 4.1, both relating to clarity, intimacy, and warmth. The factors for the 3-dimensional space have been re-ordered to align with the four-factor solution. Factor 3.1 appears to be mostly a combination of factors 4.2 and 4.3, now grouping strength, reverberance, envelopment, and source width together into a single factor. Although correlations are still strong with envelopment and source width, the correlations are diminished from when these

were treated as two separate factors. Finally, the brilliance factor appears to surface again in 3.3, related directly to factor 4.4, with a slight reduction in correlation. Some of the source width factor, 4.3, appears to also have been regrouped into factor 3.3 (the timbral factor) and not just 3.2 (the strength / spaciousness factor). This indicates a possible connection between source width and brilliance in the simpler, lower-dimensional model. As a final note, proximity does not have the highest correlations with any of the factors, but rather, it has a somewhat strong correlation with factors 4.1 – 4.3 and 3.1 – 3.2, all factors except the timbral factors. This overall trend indicates its connection to both clarity and strength / spaciousness. On the other hand, reverberance appears to be correlated with both main dimensions, but strongly negative to clarity and somewhat strongly positive to strength, envelopment, and source width. Clarity is again correlated with average preference, and in the four-factor model, source width is correlated with average preference.

As a final step, each factor can be correlated with existing objective metrics, to determine any possible metrics that might predict some of the identified factors. First, clarity index, C80, shows promise in predicting the first factor, either 4.1 or 3.2. C80 tends to outperform C50, but it should also be noted that Ts shows a strong negative correlation, even stronger than C80. Factors 4.2 – 4.3 and 3.1, the strength, spaciousness, reverberance factors, appear to be well correlated with strength and some of the decay time metrics. The later energy in the RIR appears to show promising results, even more so than total energy in the RIR. Additionally, T10 shows better correlation with these factors than EDT. When separated out by envelopment and source width, the envelopment factor (4.2) relates highly with late strength, while source width, factor 4.3, relates to late strength and early strength. Again, it should be noted that J_{LF}, the traditional source width metric, is not calculated with a strength-related dependence. Finally, the brilliance factors are hard to place, but it shows some limited promise related to EDT. All of the metrics in this table are also the ISO 3382 recommended broadband averaged results, typically the arithmetic mean of the 500 and 1000 Hz bands, except for the energy averaged results from 125 – 1000 Hz for L_J. Consideration of the frequency dependence of metrics could prove more promising for predicting factors 4.4 and 3.3. In either case, even less clear connection to existing metrics is found for factor 3.3.

Table 6.8: Correlations between the varimax factors and existing room acoustic metrics. Values in bold are significantly different from zero ($p < 0.05$), exceeding a magnitude threshold of 0.54.

Factor	Metric											
	T30	T20	T10	EDT	C80	C50	Ts	G	G_E	G_L	J_{LF}	L_J
4.1	**-0.61**	**-0.63**	**-0.60**	**-0.75**	**0.71**	0.45	**-0.81**	0.25	0.49	-0.08	-0.39	-0.39
4.2	0.29	0.29	0.39	0.34	-0.08	0.22	0.30	**0.75**	0.51	**0.86**	0.15	**0.54**
4.3	0.52	0.52	**0.61**	0.37	0.14	0.43	0.15	**0.77**	**0.63**	**0.80**	0.09	0.53
4.4	0.40	0.47	0.53	**0.56**	-0.44	-0.25	0.49	0.07	-0.14	0.31	0.14	0.38
3.2	**-0.56**	**-0.58**	**-0.54**	**-0.72**	**0.73**	0.50	**-0.80**	0.35	**0.57**	0.01	-0.38	-0.34
3.1	0.49	0.49	**0.59**	0.47	-0.09	0.25	0.36	**0.75**	0.51	**0.88**	0.18	**0.61**
3.3	0.37	0.43	0.48	0.39	-0.18	-0.05	0.24	0.10	0.00	0.23	0.04	0.27

6.5.2 Individual Preference Results

In previous sections, overall preference was averaged across all subjects for each hall. This method of analysis demonstrated an overall, average preference of concertgoers, but with averaging, the individual taste or variety was removed. For factors that all subjects tend to agree upon, they will average coherently, and such factors can be thought of a consensus dimensions of preference, with less inter-individual differences. Such factors that emerged in Table 6.3 include proximity, warmth, temporal clarity, and spatial clarity. Envelopment and source width did show somewhat high correlations, but brilliance, reverberance, and strength showed relatively low correlations with preference. To investigate whether these correlations were largely consistent or variable at the individual level, a separate correlation analysis between the preference ratings for each subject, the hall-averaged subjective attribute ratings, and the factor scores calculated from these hall-averaged ratings was run. These correlations are provided in Table 6.9.

Table 6.9: Correlations between the individually averaged preference ratings of each subject and all of the perceptual attributes, along with the final varimax rotated factor space (both 3 and 4 dimensions). Correlations that are stronger have been highlighted in color, red for a positive correlation and blue for a negative correlation. Correlations significantly different from zero ($p < 0.05$) are shown in bold ($n = 7$ for subjects, $n = 14$ for average preference rating).

Attribute / Factor	Subjects																
	Avg	6	3	16	15	5	2	12	4	10	11	7	8	9	13	1	14
Str	0.29	**0.89**	**0.78**	**0.77**	**0.72**	**0.83**	0.38	0.17	-0.13	-0.13	-0.22	-0.16	-0.50	**-0.92**	0.47	0.57	-0.20
Env	0.40	**0.81**	0.66	**0.68**	**0.75**	**0.78**	**0.82**	0.19	0.38	0.06	-0.04	-0.44	-0.42	**-0.91**	0.31	-0.07	-0.21
SW	0.44	**0.84**	**0.92**	**0.74**	0.56	**0.70**	0.30	0.24	-0.18	0.01	-0.13	-0.38	-0.12	**-0.78**	0.56	0.18	0.00
Rev	0.03	**0.84**	0.62	0.51	**0.85**	0.56	-0.23	-0.07	-0.61	-0.30	**-0.72**	**-0.69**	**-0.78**	**-0.76**	0.58	0.05	-0.08
Brill	0.06	0.56	0.36	0.63	0.09	0.07	0.04	-0.33	-0.23	**-0.68**	0.17	-0.17	0.00	-0.22	0.57	0.12	-0.28
Prox	**0.81**	0.51	**0.76**	0.45	0.38	0.31	0.63	0.51	**0.75**	**0.69**	**0.79**	0.64	0.21	-0.17	0.08	0.25	0.45
TC	**0.68**	-0.11	0.05	-0.29	0.03	-0.18	**0.68**	**0.78**	**0.85**	**0.87**	**0.93**	0.61	0.50	0.50	0.17	-0.37	0.39
SC	**0.60**	-0.46	-0.12	0.39	-0.19	0.16	0.66	**0.69**	**0.78**	**0.89**	**0.81**	0.61	0.55	-0.05	-0.37	-0.26	0.15
Wrm	**0.58**	-0.24	0.18	0.33	0.03	0.54	**0.76**	**0.82**	**0.92**	**0.91**	0.54	0.36	0.33	-0.39	-0.33	0.08	0.37
Int	0.31	-0.67	-0.29	0.09	-0.50	-0.23	0.23	0.58	**0.68**	0.55	**0.75**	**0.75**	0.62	0.37	-0.37	0.06	0.30
4.1 (25%)	**0.58**	-0.52	-0.12	0.13	-0.31	0.05	**0.67**	**0.68**	**0.94**	**0.82**	**0.95**	**0.72**	0.65	0.15	-0.37	0.08	0.34
4.2 (19%)	0.38	**0.80**	0.65	**0.78**	**0.77**	**0.88**	**0.71**	0.31	0.29	0.11	-0.12	-0.29	-0.47	**-0.96**	0.29	-0.30	-0.14
4.3 (16%)	**0.59**	**0.88**	**0.95**	**0.77**	**0.69**	**0.71**	0.36	0.27	0.02	0.12	0.14	0.00	-0.24	**-0.68**	0.56	0.20	0.18
4.4 (12%)	-0.06	0.57	0.34	0.40	0.16	0.03	-0.14	-0.34	-0.54	**-0.75**	-0.09	-0.40	-0.11	-0.12	0.66	0.24	-0.44
3.2 (25%)	**0.64**	-0.42	-0.02	0.21	-0.21	0.14	**0.70**	**0.74**	**0.93**	**0.87**	**0.93**	**0.70**	0.60	0.06	-0.34	0.08	0.36
3.1 (26%)	0.40	**0.92**	**0.81**	**0.81**	**0.77**	**0.85**	0.50	0.18	0.05	-0.02	-0.16	-0.31	-0.45	**-0.96**	0.48	-0.10	-0.08
3.3 (12%)	0.12	0.57	0.39	0.21	0.15	-0.20	-0.23	-0.21	-0.54	-0.59	0.18	-0.03	0.02	0.18	**0.68**	0.42	-0.10
MSS / MSE:		4.37	4.55	3.26	2.35	2.56	3.05	2.73	3.33	6.53	3.67	1.62	3.09	2.08	1.70	2.82	1.66

As each subject did not rate all 14 halls, these correlations are calculated for the seven halls that were rated, excluding the anchor. With the paired randomization, all subjects saw a large variety in halls, ensuring representative calculation of correlations across subjects, even with a reduced sample. Table 6.9 was reorganized in terms of subjective attribute and subject number to group together individual subject correlations that appear similar and group attributes with similar individual preference correlations. Table cells with a colored highlight indicate a correlation such that $|r| \geq 0.5$; red cells indicate a positive correlation and blue cells indicate a negative correlation. Interesting patterns emerge, showing large inter-individual variation in terms of preference. Previous literature suggests that often two groups of preference emerge: one group preferring loud, enveloping sound and another preferring clear, intimate sound. This overall trend is also supported in the results of this study. Subjects 3, 5, 6, 15, and 16 all appear to prefer loudness and envelopment, while subjects 2, 4, 7, 8, 10, 11, and 12 fall into the category associated with clarity, intimacy, and warmth.

These trends appear to connect with previous literature, with the simple two-group model of preference. With more investigation of Table 6.9, more subtle but seemingly individual trends take shape. For example, subjects in the strength and spaciousness group all have significant agreement with one another, but some subjects (5 and 15) show a slight tendency towards envelopment over source width, while subject 3 shows a stronger emphasis between source width and preference. This connection is further represented in the four-factor model, where envelopment-related and source width-related correlations appear to be split between factors 4.2 and 4.3. The remaining subjects in the group show no tendency towards either of the spaciousness attributes but appreciate both (6 and 16). Other individual differences include a higher correlation with warmth for subject 5. The perception of warmth appeared to be correlated to both factors 4.1, the clarity factor, and factor 4.2, the strength and envelopment factor in Table 6.7. This subject appears to connect with this perception, separate from all other subjects in their group. This might indicate they fall close to the group line, and don't fully associate with either group. Subjects 6 and 15 also show a negative correlation with intimacy, while others in the group did not have as strong preference against intimacy. This finding is clearer for subject 6 in their strong negative correlation with the clarity factor, 4.1 and 3.2. The majority of subjects had a positive correlation with factors 4.1 and 3.2, explaining why clarity is correlated with overall average preference. Despite this overall trend, certain subjects do exhibit a negative correlation with the clarity factor, mostly in terms of intimacy.

A final set of ungrouped subjects, 1, 9, 13, 14 all exhibit uniquely individual preference trends. Most surprisingly, subject 9 showed little positive correlation with many of the attributes, but rather, they exhibited a strong negative correlation opposite to the strength and spaciousness factors, 4.2, 4.3, and 3.1. The strength of this negative correlation is quite astonishing, as they show little correlation with any of the clarity group of attributes except temporal clarity as well. This subject goes against much conventional hall design wisdom, and although a hall will not likely be designed to this individual, it is important to know that such individuals do exist. With a small sample size, no complete conclusion can be drawn from a single subject, but to the testing of more individuals would demonstrate if similar trends remain.

Subjects 1, 13, and 14 show overall much weaker correlations with their preference and most all of the subjective attributes and factors. This could be due to an inconsistency in ratings, or this result could also be due to a preference that is not easily explained in the currently defined factor space. To determine the within-hall consistency of preference ratings by each subject, the ratio of the total mean sum of squared variance (MSS) to the mean sum of squared error between preference ratings of the same hall was computed. The equation for this calculation is provided in Eqn. 6.1:

$$\frac{MSS}{MSE} = \frac{\frac{1}{nk}\sum_{j=1}^{q}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}\right)^2}{\frac{1}{nk}\sum_{j=1}^{q}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2}, \qquad 6.1$$

where x_{ij} is the i^{th} perceptual rating of the j^{th} hall by a particular subject. The total number of halls rated by each subject is given by q and the total number of repetitions is given as n. The overall mean, $\bar{x}$, is used in the numerator as an estimate of total variance in a subject's preference ratings, and the individual average for each hall, $\bar{x}_j$, is used to calculate the variance in preference ratings of the same hall for each subject, effectively their error in providing consistent ratings for each hall between repetitions.

If this ratio takes on a value of 1, this result indicates an equal variation in preference ratings within the same hall and between different halls, showing no consistency in preference ratings. As this number increases, it indicates an increased consistency in differentiating preference ratings between halls. This statistic was calculated for each individual, and it provided in the last row of Table 6.9. For the subjects that provided no clear preference trends, two of the subjects, subjects 13 and 14, have two of the three lowest values of this ratio. This low ratio might indicate that these subjects provided less consistent ratings of their preference, which prevents clear correlations from emerging. One subject with less clear preference,

subject 1, had a reasonably comparable value of this ratio with other subjects with more easily defined preference. This might indicate that this subject's preference ratings were somewhat consistent, but the preference might not be clearly related to specific single sets of attributes or words. Even if consistent, preference may be less clearly defined or connected to common subjective terms for this subject. This number is normalized to the sample size, but in the current study, only two replicates of the preference ratings were possible within the same subject. With more repetitions within each subject, some of these subjects might have a more clearly defined preference.

Overall, these results suggest that conventional grouping of subjects into two groups reduces the complexity of preference in concert halls. This simplified representation is easier to identify at a large scale, but clear differences in individual preference are lost. Many previous studies have identified these groups to help explain differences in preference ratings, but the techniques were used more as a way of aiding in data interpretation, and it was not a clear methodology in the study. These results indicate that subjective groups might not be the best way to represent this data, removing the individual differences in preference seen in subjects 5, 6, and 9 or for differences between subjects in the same group, such as the possible envelopment emphasis of subjects 5 and 15 versus the source width emphasis of subject 3. Testing larger numbers of subjects will identify if these differences continue to be observed, and increased sample sizes of preference for each subject would help clarify if these differences remain statistically valid. Correlations between hall-averaged attribute ratings, that are well-defined, and individual preference ratings provide a methodological way to classify individual preference. Further correlations with the reduced perceptual factor space provide a continuous representation of preference, eliminating the need for subjective grouping.

6.6 Results II: Spatial Energy Map Subjective Correlations

Standard room acoustic metric analysis only has access to two domains: time and frequency. In the current study, the CHORDatabase includes RIR measurements with a 32-element spherical microphone array, so spherical beamforming analysis can be used for full spatial analysis of the RIR. For each room, spatial energy maps can be generated using spherical array beamforming over specific time ranges of a RIR. This type of analysis can provide complete access to the time and spatial behavior of a room. For the present study, the hall-averaged subjective ratings for envelopment, source width, proximity, and average preference have been correlated with specific time and spatial ranges of the RIR.

6.6.1 Time Energy Correlation Technique

First, the correlation of the energy in the room impulse response was varied with respect to a time integration window. The omnidirectional RIR was taken for each of the 14 halls used in the study, and a specific time range was isolated using a rectangular time window. Once a given time range was isolated, a Pearson correlation coefficient was calculated between the time-integrated energy on a decibel scale and the average subjective ratings for a particular hall. For example, the energy in the RIR, starting at 60 ms and ending at 240 ms, can be correlated with the average envelopment ratings of all 14 halls, averaged across subject. If little correlation is found, a coefficient close to 0 will result, indicating that no clear linear relationship exists between the subjective impression of envelopment and early energy in a concert hall. A value close to 1 indicates a positive linear relationship where a value close to -1 indicates a negative linear relationship.

As this analysis could be done over any time range, a procedure to generate a graphic map of these time correlations was developed. An example of such a map is shown in Figure 6-8. The horizontal axis indicates the start of the time integration window and the vertical axis indicates the end of the integration window. Each point indicates a pair of start and end integrations times, and the graph shows the correlation coefficient between the subjective rating data and the energy in that time range across all 14 halls. Since the starting time must be before the ending time, the plot is only valid in the upper left triangle where this condition is satisfied. By analyzing this plot, ranges of strong positive correlation (red), strong negative correlation (blue), or no correlations (white) can be identified. Once specific time ranges of interest are selected, these ranges can be isolated, and the correlation as a function of space can also be determined.

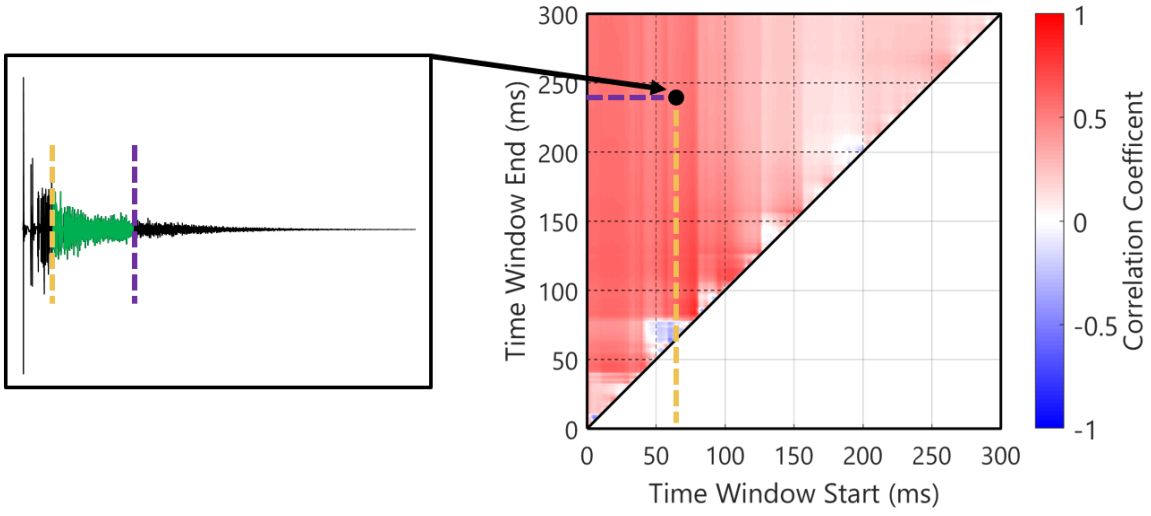


Figure 6-8: A time correlation map between hall-averaged subjective data and time energy integration ranges. Each point in the plot represented the correlation for a specific time range. The black dot would be the located where the correlation for energy in the time range from 60 ms to 240 ms, highlighted in the time-domain RIR to the left. Each point in the map represents a different time region in the RIR.

6.6.2 Spatial Energy Correlation Technique

For a fixed time range, the analysis process can also be done in the spatial domain. Using the spherical array beamforming techniques described in chapter 5, a set of DirRIRs were generated for each hall, calculating a RIR response for a beam-like microphone directivity in all directions around a listener. The desired time window was applied to each DirRIR, and the energy in that range was integrated separately for each direction. The energies in each direction were used to generate a spatial map for each hall, which illustrates the spatial distribution of energy in a given time range. Similar to the time correlation analysis, a spatial map of correlation coefficients can be generated, as shown in Figure 6-9. In this plot, the variable under study is direction, in terms of both azimuth and elevation. Each point represents the correlation between the hall-averaged subjective ratings, across all subjects, and the energy in that particular direction. Similar to the time maps, regions of high, low, or no correlation can be identified upon visual inspection.

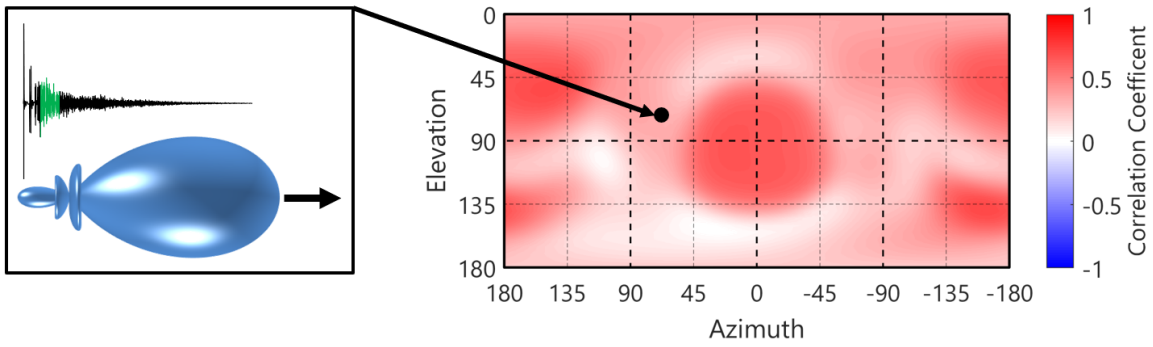


Figure 6-9: A spatial correlation map between hall-averaged subjective data and spatial energy beamforming maps. Each point in the plot represented the correlation for a DirRIR for a beam pattern oriented in that specific direction. This is generated by creating a beam-like microphone directional response (shown in blue) from the spherical microphone array, oriented in the direction of interest. This analysis is performed over a single, fixed time range, selected from the previous time-domain analysis.

These correlations were performed on the energy in each DirRIR generated using third-order Dolph-Chebyshev beams with -25 dB side lobe rejection from 1 – 4 kHz. In the following sections, correlations were analyzed in terms of time and space for the individual subjective perceptions of envelopment, source width, proximity, and average preference. These techniques can be highly useful for identifying regions of perceptual interest for subjective impressions without a clear metric, such as proximity, or for subjective impressions with metrics that are not yet well studied, such as envelopment and source width. This technique

was also applied to the four factors identified in the factor analysis described in section 6.5.1.3. Such analyses can help further understand the underlying perceptual factors that explain most of the variation in room acoustic perception.

6.6.3 Subjective Attribute Spatial Energy Correlation

This section will first focus on the temporal and spatial correlations of specific subjective attributes with each RIR. First, envelopment and source width were selected for analysis, as they are spatial perceptions that have less well-defined metrics and correlations. Then, correlations with proximity were investigated, due to its clear importance with overall preference and lack of any clearly defined metrics. To corroborate the results with proximity, the average preference ratings were correlated for comparison. Along with these individual attributes and average preference, the four dimensions of the perceptual space were correlated with similar time and spatial regions as these attributes. These correlations further demonstrate the interpretation of the individual factors, and how they relate to perception.

6.6.3.1 Envelopment

The first attribute under investigation was envelopment, and as a highly spatial perception, this type of analysis is highly intuitive. In his dissertation, Dick used correlation between different spatial and time ranges of the RIR, separately varying the starting integration time, the ending integration time, front azimuthal integration limits, rear azimuthal integration limits, and elevation integration limits.[52] Dick developed a proposed metric to predict envelopment using a time-integration of the RIR from 60 ms to 400 ms and a spatial-integration of the RIR from 30° to 130° elevation and from 20° to 120° and from -20° to -120° azimuth. In the present work, to visualize the starting and ending time integration limits at the same time, both the starting and ending time window ranges were adjusted in 1-ms increments, and the correlation on all possible integration limits is plotted in Figure 6-10. The lower right of the plot is not a valid region, as the starting integration limit cannot be greater than the ending integration limit.

Clear strong correlation occurs for energy in the RIR from 60 ms to later in the RIR, and a weaker but noticeable correlation exists up to 60 ms, shown in the bottom left corner of the triangle. First, the early range in the RIR from 0 to 60 ms was isolated for each hall, and the correlation analysis was performed over the spatial domain for this fixed time window; this is shown in Figure 6-11. The correlation in this range was not as prominent in the time-domain analysis, but this result is because the time-domain analysis was integrated over all spatial regions. Here, it is clear that in the early part of the RIR, envelopment is highly correlated

with a specific region of the RIR, and less correlated with other regions. For the later range from 60 ms to 500 ms, Figure 6-12 shows the spatial correlation with envelopment. It appears that late energy from all directions might be relevant to the perception of envelopment. This lack of clear correlation with a spatial region also indicates why this correlation was so strong in the time-domain analysis from Figure 6-10, which averaged over the spatial dimension.

This result slightly differs from Dick's findings. The spatial region identified in the early part of the RIR corresponds almost directly with Dick's, with a possibly narrower integration range from ±40° to ±100° azimuth and 40° to 130° elevation. The main difference is that Dick's metric started at 60 ms, where the current study shows clear correlation with this region, even before 60 ms. Additionally, this study shows less correlation with a specific spatial region following 60 ms. More subjects are needed in the present study to draw clear conclusions, but findings suggest that the spatial integration regions should be applied to the early part of the RIR as well, and the later part of the RIR should consider broader integration limits.

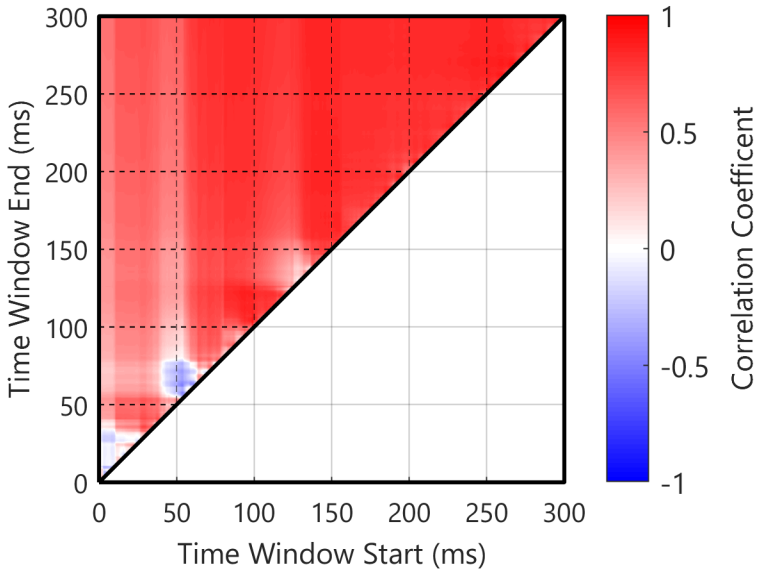


Figure 6-10: Correlations between subjective envelopment ratings as the start and end of a time integration window was varied, computed for the 14 halls in the study.

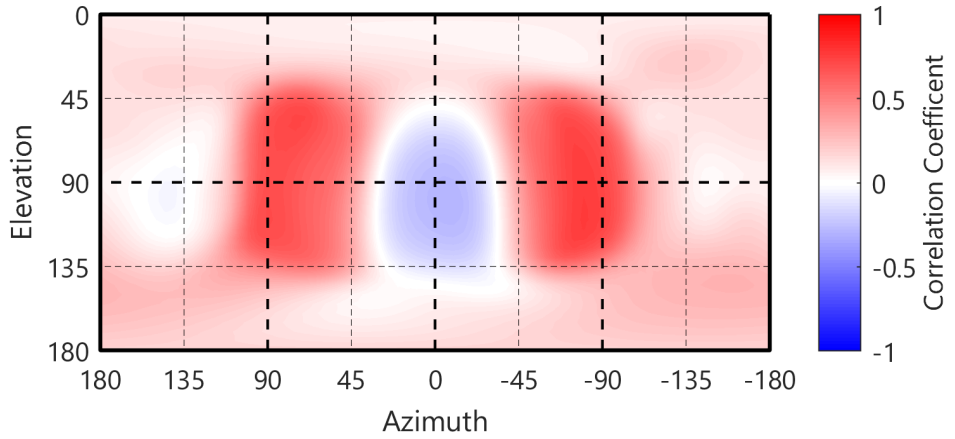


Figure 6-11: Correlation with envelopment as a functions of direction of arrival for the fixed time range from 0 – 60 ms in the RIR. Beamforming analysis was done using third-order Dolph-Chebyshev beamforming with a 20 dB side lobe rejection from the 1 – 4 kHz octave bands.

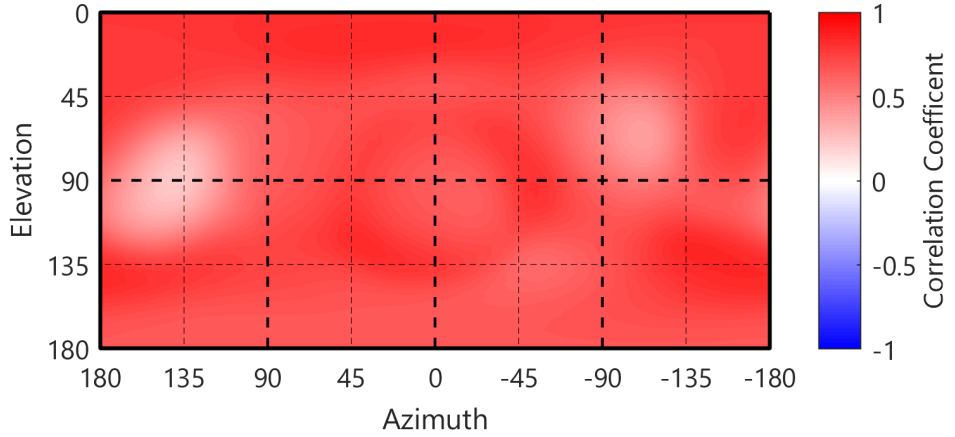


Figure 6-12: Correlation with envelopment as a functions of direction of arrival for the fixed time range from 60 – 500 ms in the RIR. Beamforming analysis was done using third-order Dolph-Chebyshev beamforming with a 20 dB side lobe rejection from the 1 – 4 kHz octave bands.

6.6.3.2 Source Width

Similarly for envelopment, the same procedure was used to investigate source width perceptual ratings. Source width showed less correlation in the time-domain analysis provided in Figure 6-13. A range of higher correlation exists from 80 ms to the end of the RIR and in the early part up to 60 ms in the RIR. The early energy from 0 to 60 ms correlated across space is shown in Figure 6-14. The correlations are less strong than envelopment, but a similar lateral region emerges for source width. Some correlation exists between envelopment and source width, so it is not surprising to see consistency between results. For the later energy, stronger correlations are seen from 80 ms to 500 ms shown in Figure 6-15. Source width appears to be related to energy in this time range, with a clear rejection of energy in front of a listener. A metric considering both early and late energy, rejecting most of the energy from the front of a listener, and accepting lateral early energy and all non-frontal late energy appears to suggest initial success.

The current metric for source width extracts spatial energy using a laterally oriented dipole microphone, rejecting energy from the front, above, behind, and below a listener, and it is integrated from 5 to 80 ms. Further, the metric is an energy ratio, normalized to the total early energy in the RIR, without any level-dependent component. Much more promising correlations are found with the spatial energy analysis in this section, as opposed to the 0.15 correlation coefficient with J_{LF} from Table 6.4. Much potential for a better-defined metric to predict source width appears possible with advanced spatial beamforming analysis over different time ranges of the RIR. The correlation region is quite similar to envelopment, as the two were also shown to be correlated to one another. Again, this result will increase in validity as more subjective ratings are recorded.

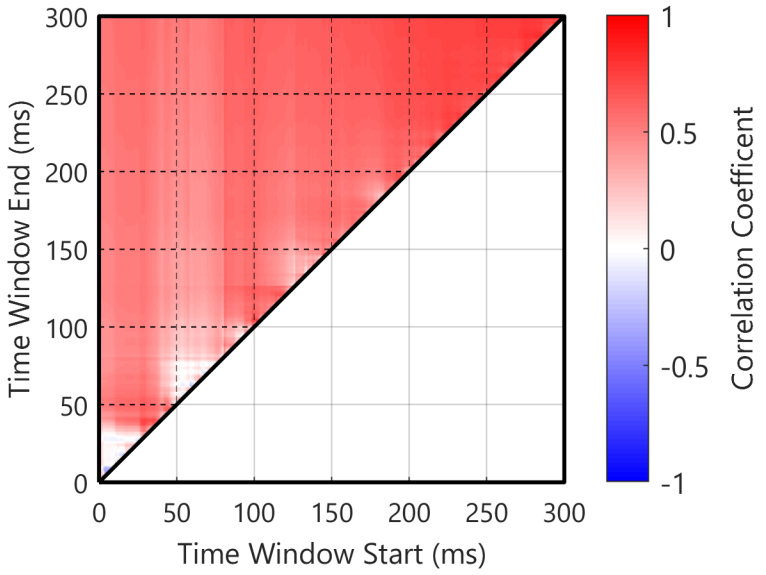


Figure 6-13: Correlations between subjective source width ratings as the start and end of a time integration window was varied, computed for the 14 halls in the study.

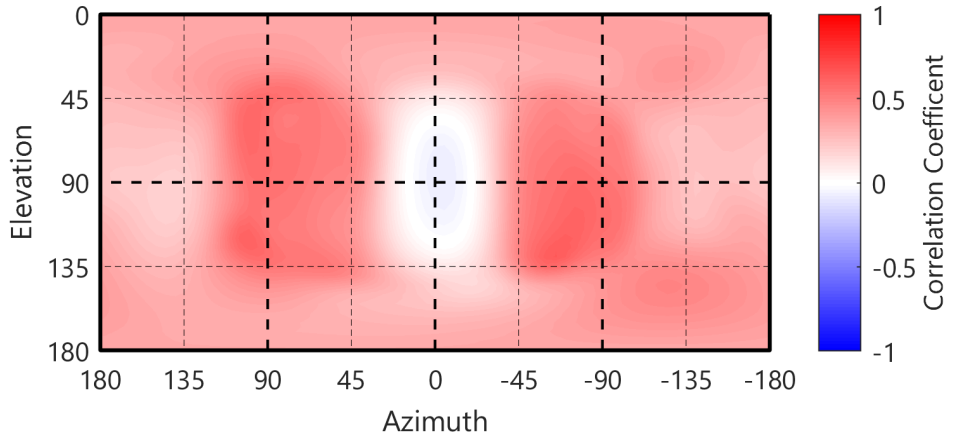


Figure 6-14: Correlation with source width as a functions of direction of arrival for the fixed time range from 0 – 60 ms in the RIR. Beamforming analysis was done using third-order Dolph-Chebyshev beamforming with a 20 dB side lobe rejection from the 1 – 4 kHz octave bands.

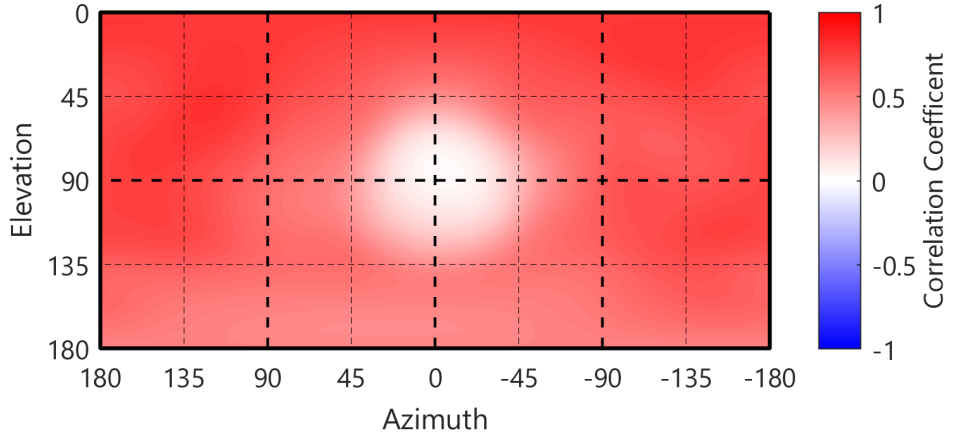


Figure 6-15: Correlation with source width as a functions of direction of arrival for the fixed time range from 80 – 500 ms in the RIR. Beamforming analysis was done using third-order Dolph-Chebyshev beamforming with a 20 dB side lobe rejection from the 1 – 4 kHz octave bands.

6.6.3.3 Proximity

Although it was found to correlation strongly with average preference, no metric exists for predicting the perception of proximity. As a first attempt, the level over different time ranges was first considered. From the correlation analysis shown in Figure 6-16, a range of interest appears to be from 70 ms to 150 ms in the RIR, with a weaker but clear correlation in the early energy, before 70 ms. Spatial correlations with both the early range and middle time range in the RIR are provided in Figures 6-17 and 6-18, respectively. Isolating the more prominent, later range, a spatial energy map was created to show the spatial character of this correlation in Figure 6-18. This map, shown over a similar starting time range to map for source width in Figure 6-15, appears somewhat complementary in correlation. Where source width was associated with later-arriving non-frontal energy, proximity seems to be related to frontal energy arriving in the mixing time range between early and late energy. The early energy, before 70 ms, was also investigated in terms of spatial correlation in Figure 6-17. Correlation is found in most all regions of the RIR, with a rejection of energy from the front. This result might seem counterintuitive, but it is important to note that all auralizations in the current study were made at a consistent source-receiver distance, isolating the effect of hall. As such, the current dataset cannot draw any conclusions on the strength of the direct sound energy; this would be mere speculation. As such, any results associated with the strength of the direct sound energy should be analyzed with caution.

While the time correlations remain strong for source width as the ending time integration limit increases, the correlation degrades for proximity past 150 ms. This finding suggests the importance of early energy, with a later cutoff than the traditional 80 ms. It should be noted that some correlation appears in the early part of the RIR, before 70 ms. Inspection of this energy appears to focus correlations on specific strong reflections that were present in halls rated with a high proximity. It is difficult to draw conclusions on this region alone, as reflections are not as densely occurring in the region. Due to the inclusion of only 14 halls with a single seat, results can artificially focus in on key events. With the inclusion of more auralization conditions (especially more seat locations within these halls), analysis of the early energy might demonstrate a spatial correlation that can be interpreted with perception.

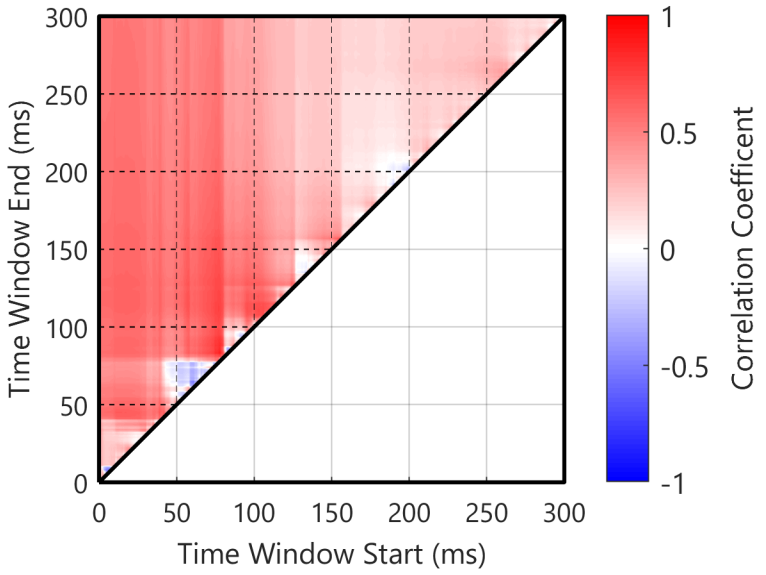


Figure 6-16: Correlations between subjective proximity ratings as the start and end of a time integration window was varied, computed for the 14 halls in the study.

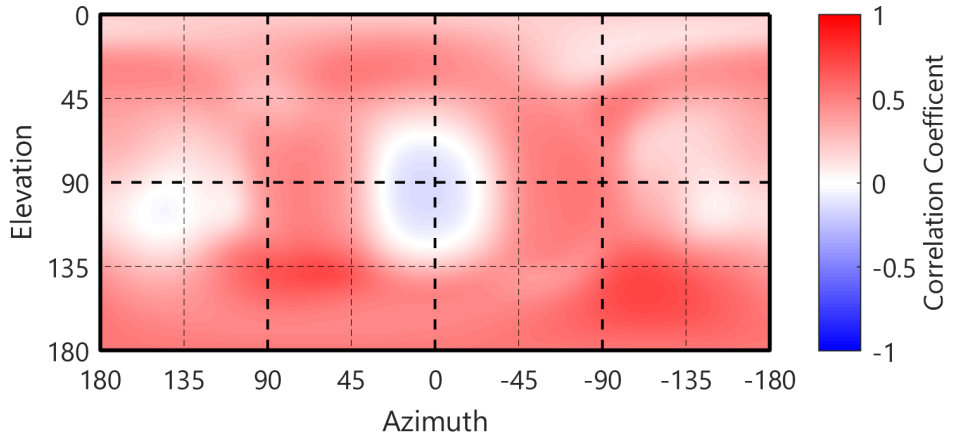


Figure 6-17: Correlation with proximity as a functions of direction of arrival for the fixed time range from 0 – 70 ms in the RIR. Beamforming analysis was done using third-order Dolph-Chebyshev beamforming with a 20 dB side lobe rejection from the 1 – 4 kHz octave bands.

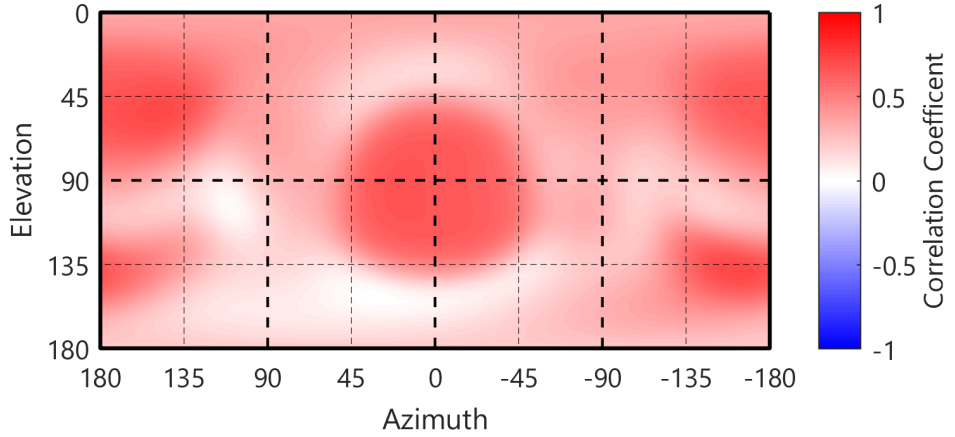


Figure 6-18: Correlation with proximity as a functions of direction of arrival for the fixed time range from 70 – 150 ms in the RIR. Beamforming analysis was done using third-order Dolph-Chebyshev beamforming with a 20 dB side lobe rejection from the 1 – 4 kHz octave bands.

6.6.3.4 Average Preference

Although averaging preference removes individual taste from the analysis, it is still beneficial to understand which aspects of the sound field are found to be agreeably positive across all listeners. First, time-domain correlations with different time ranges of the RIR are shown in Figure 6-19. The correlation has a clear similarity with proximity in Figure 6-16, with some additional emphasis on the early range in the RIR, before 70 ms. Looking at the spatial energy correlation in Figure 6-20, the early energy correlations are directly associated with the same spatial region that was found to connect with envelopment in Figure 6-11. Investigating the later energy in Figure 6-22, this energy appears to connect quite well with the proximity maps over the same region in Figure 6-18. Thus, average preference suggests a relationship with envelopment associated with the early sound energy in the RIR and proximity associated with the energy located in the mixing time range between early and late energy. As a final note, the overall magnitudes of the correlations with average preference are not as strong as they were with envelopment and proximity in this study. Again, this reduction is likely a result on the averaging across subject, which removes critical information from the subjective data analysis.

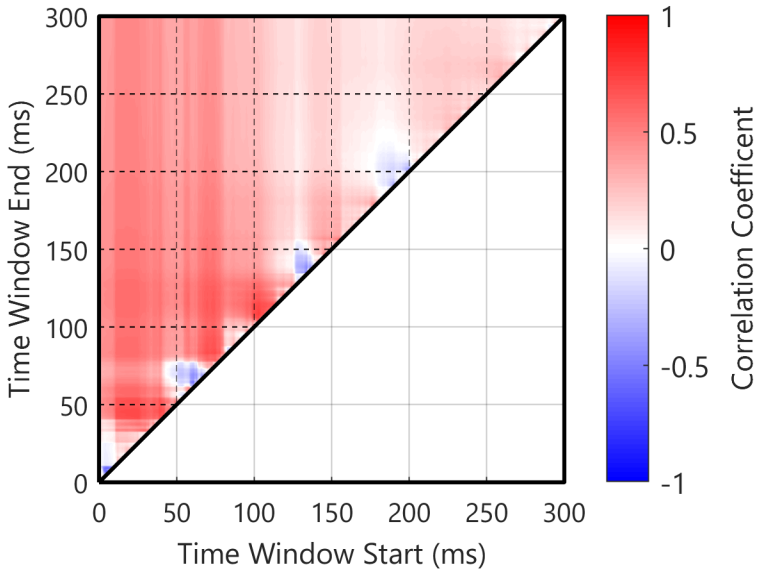


Figure 6-19: Correlations between average preference ratings as the start and end of a time integration window was varied, computed for the 14 halls in the study.

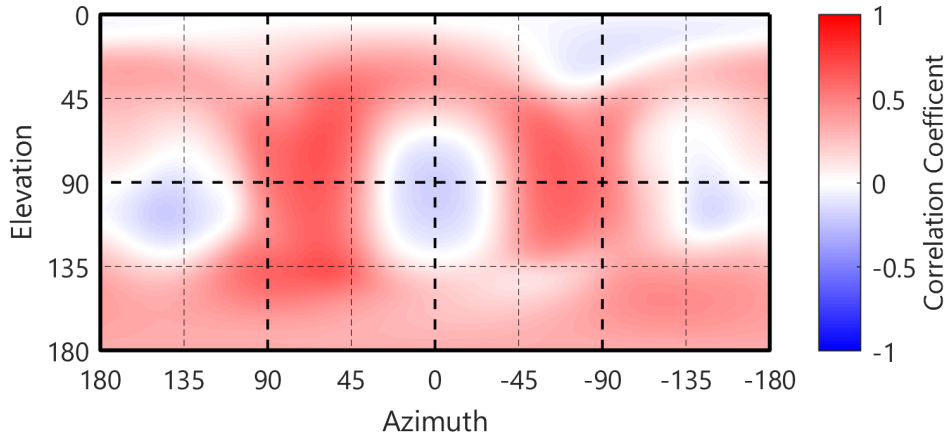


Figure 6-20: Correlation with average preference as a functions of direction of arrival for the fixed time range from 0 – 70 ms in the RIR. Beamforming analysis was done using third-order Dolph-Chebyshev beamforming with a 20 dB side lobe rejection from the 1 – 4 kHz octave bands.

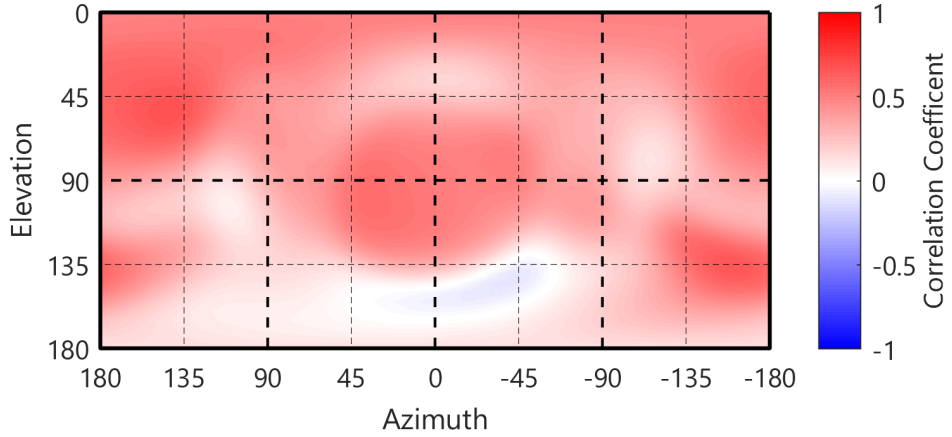


Figure 6-21: Correlation with average preference as a functions of direction of arrival for the fixed time range from 70 – 150 ms in the RIR. Beamforming analysis was done using third-order Dolph-Chebyshev beamforming with a 20 dB side lobe rejection from the 1 – 4 kHz octave bands.

6.6.4 Spatial Energy Correlation of Preference Factors

As a point of completeness, these time-domain and spatial-domain correlation techniques can also be implemented on the factors that emerged from the PCA analysis and varimax rotation. Taking the hall-averaged attribute ratings from the subjective study, factor scores were calculated using the provided factor loadings for factors 4.1 – 4.4 from the four-dimensional solution. The next four sections present the same series of correlation plots shown in 6.6.3, but results are now shown for correlations directly with the orthogonal factors identified to represent the perceptual space. These factors help to connect the fundamental perceptual aspects of the sound field with the temporal and spatial characters of the RIR.

6.6.4.1 Factor 4.1: The Clarity Factor

The first factor (4.1), primarily associated with clarity, intimacy, and warmth, has a time correlation shown below Figure 6-22. The traditional metric to predict clarity in concert halls is the clarity index for music, C80, which is inherently a ratio-based quantity, comparing the strength of early sound energy to later sound energy in a RIR. The cutoff for music is at 80 ms. Seen in Figure 6-22, a vertical line has been drawn at 80 ms and a horizontal line has been drawn at 140 ms, defining a visual transition between helpful and harmful energy for the perception of clarity. Visually, this appears as the separation between the positive (red) and negative (blue) correlation ranges. From the stark vertical line, an initial conclusion of 80 ms might be seen as an appropriate cutoff. On further investigation, when the time window starts at 80 ms and ends around 140 ms, still positive correlations can be found, although weak in nature. Tracing along the diagonal of the plot, where time windows tend to be smaller and *slide* through the RIR, clear benefits occur from energy before 60 ms. After that time, from 60 ms to 140 ms, energy still appears to be overall helpful towards clarity, but the benefit diminishes. Finally, after 100 – 140 ms, zero to negative correlations occur, indicating that energy after this time is detrimental towards clarity.

The current analysis technique is not performed on energy ratios, but rather, just energy levels in specific ranges of the RIR. This analysis technique could be adjusted to analyze a changing cutoff time between early and late energy, to generate a metric similar to C80 that accounts for both the positive and negative correlations in early and late energy. Such a method would help validate C80 as a metric, and this ratio quantity could also be correlated in the spatial domain, to identify if any spatial regions or frequency bands better represent the perception of clarity-related factors in concert halls. Spatial correlations were run on the early ranges of the RIR with this factor, but no clear trends of spatial interest emerged. A combined ratio-based analysis using this technique may still prove useful.

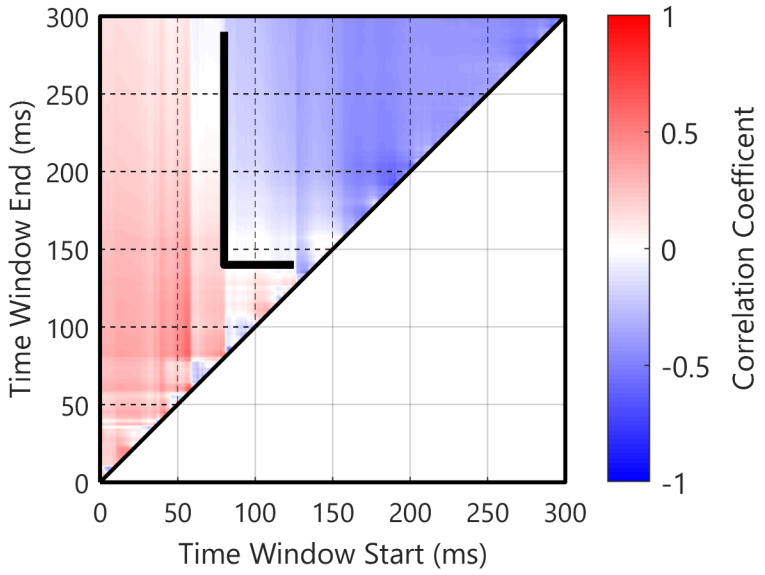


Figure 6-22: Correlations between factor 4.1 loadings as the start and end of a time integration window was varied, computed for the 14 halls in the study. A transition from helpful early energy and harmful late energy for the perception of clarity is observed between 80 and 140 ms.

6.6.4.2 Factor 4.2: The Strength and Envelopment Factor

The second factor (4.2) was associated with the strength and spaciousness factor in the three-factor solution, and in the four-factor solution, it better correlated with envelopment compared to source width in terms of spaciousness. The same time ranges used for envelopment with early sound (0 to 60 ms) in Figure 6-11 and late sound (60 to 500 ms) in Figure 6-12 are shown in Figures 6-23 and 6-24, respectively. The same key spatial region linked with envelopment is also important for this factor, further supporting the connection of factor 4.2 with envelopment. It also shows a somewhat negative correlation with frontally arriving energy, indicating the possible use of a ratio-like comparison of lateral energy to frontal energy. This type of metric would be the spatial equivalent of the time ratio-based metric for clarity, already used in the field. The later part of the RIR shows less clear spatial character and appears to be only associated with level.

As an important note, the current metric to predict envelopment, L_J, looks at the strength in the lateral portion of the late energy in a concert hall from 80 ms to the end of the RIR. The plot shown in Figure 6-24 is almost identical when the range is slightly adjusted to start at 80 ms, and no clear spatial correlations occur for later sound energy. Further investigation and refinement of a metric to predict envelopment has been started by Dick and Vigeant,[52,94] and this finding clearly supports the need for more work in this area, especially since it is associated with the second most prominent perception from the factor analysis. Currently, metrics that exist do not fully capture this important perception.

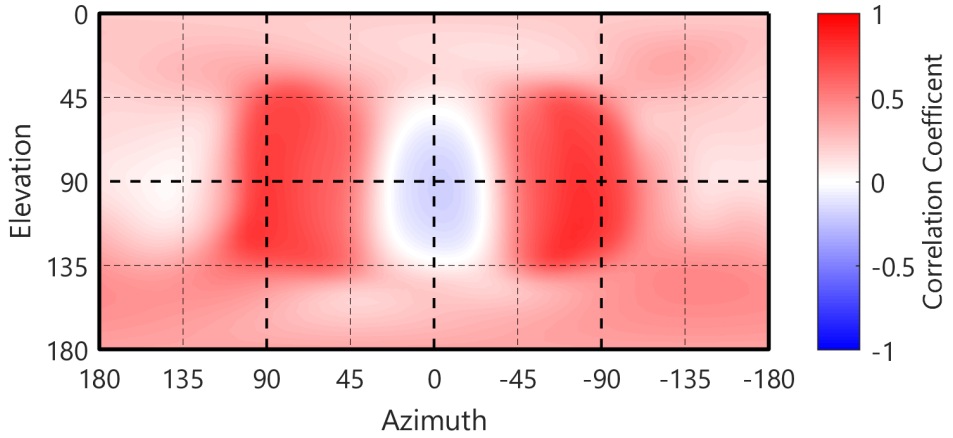


Figure 6-23: Correlation with factor 4.2 as a functions of direction of arrival for the fixed time range from 0 – 60 ms in the RIR. Beamforming analysis was done with a third-order Dolph-Chebyshev beam pattern with a 20 dB side lobe rejection over a frequency range from the 1 – 4 kHz octave bands.

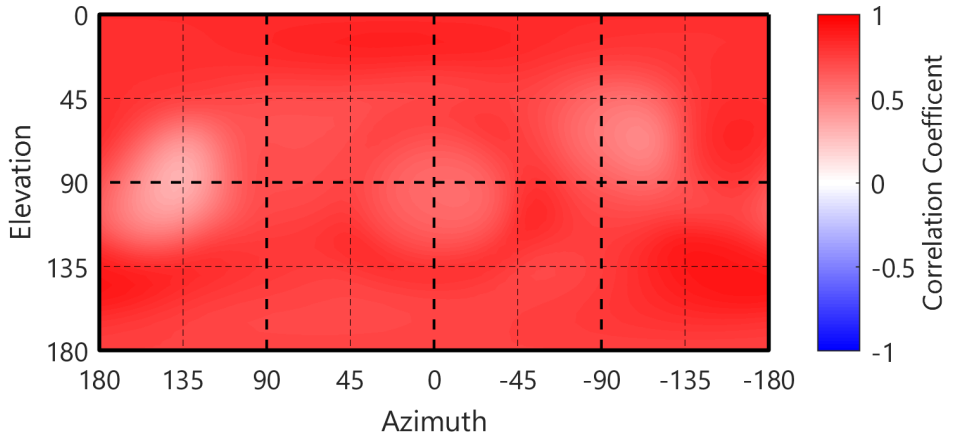


Figure 6-24: Correlation with factor 4.2 as a functions of direction of arrival for the fixed time range from 60 – 500 ms in the RIR. Beamforming analysis was done using third-order Dolph-Chebyshev beamforming with a 20 dB side lobe rejection over a frequency range from the 1 – 4 kHz octave bands.

6.6.4.3 Factor 4.3: The Strength and Source Width Factor

The strength and source width factor (4.3), despite its close relationship with the strength and envelopment factor (4.2), shows some unique differences when using this correlation procedure. The early energy shows a strong correlation with lateral energy in Figure 6-25, over much of the same spatial range as factor 4.2. The strength of the correlation with this region alone is less pronounced, and the importance of energy above, below, and even behind a listener suggest a more important role. When looking at the later energy in the RIR in Figure 6-26, more differences emerge between factors 4.2 and 4.3. While 4.2 associated with envelopment was highly correlated with most spatial regions, the possible source width factor in 4.3 seems to have higher correlation with specific spatial regions in the RIR, primarily associated with energy arriving above, behind, and to the sides of a listener. A possible rejection of energy from the front and late part of the response also appears to be suggested in Figure 6-26.

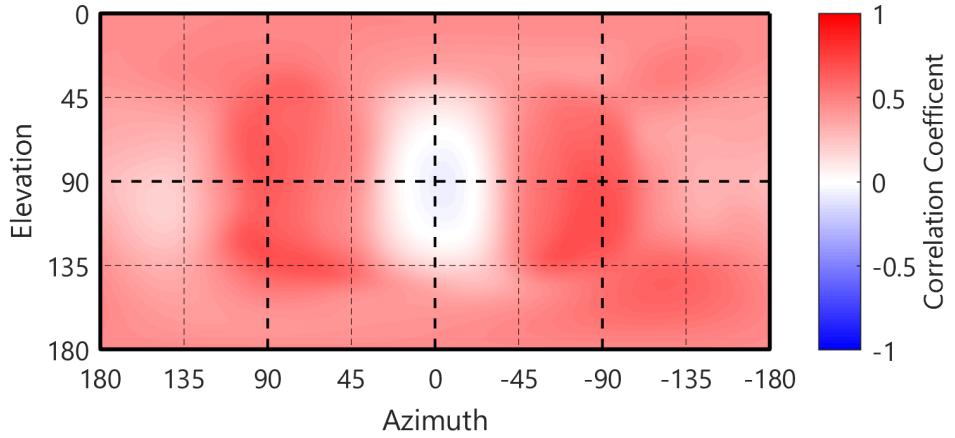


Figure 6-25: Correlation with factor 4.3 as a functions of direction of arrival for the fixed time range from 0 – 60 ms in the RIR. Beamforming analysis was done using third-order Dolph-Chebyshev beamforming with a 20 dB side lobe rejection over a frequency range from the 1 – 4 kHz octave bands.

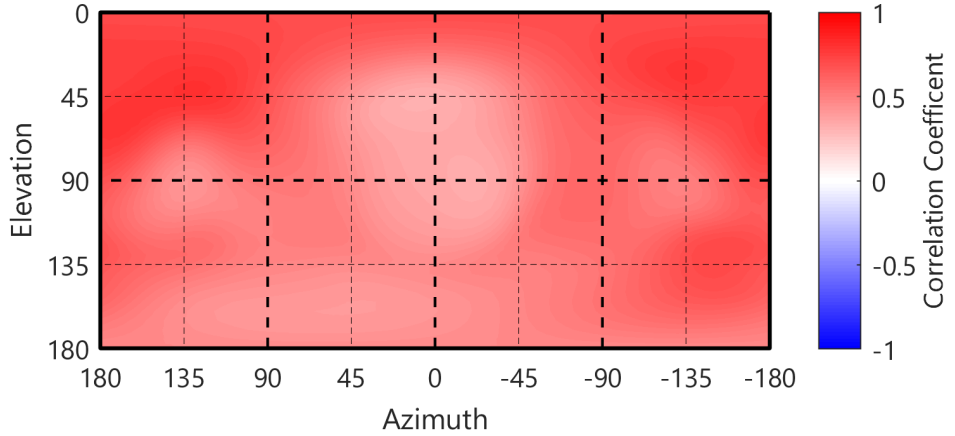


Figure 6-26: Correlation with factor 4.3 as a functions of direction of arrival for the fixed time range from 60 – 500 ms in the RIR. Beamforming analysis was done using third-order Dolph-Chebyshev beamforming with a 20 dB side lobe rejection over a frequency range from the 1 – 4 kHz octave bands.

6.6.4.4 Factor 4.4: Brilliance

For completeness, the time correlations with the brilliance factor, factor 4.4, are provided in Figure 6-27. This correlation is performed over the viable range of third-order beamforming possible with the Eigenmike, representing the total energy from 1 – 4 kHz in the RIR. As no frequency-band specific correlations are currently being used, it is not surprising that weak correlations are found in broadband time-domain analysis alone. Future analysis should look at frequency-band specific correlations, and possible ratios in the correlation with these quantifies will be investigated. Timbral perceptions, such as brilliance, have no widely accepted metrics for prediction, but it is interesting to note that a slightly positive correlation emerged with later sound energy in the RIR. This finding is not entirely new, as Soulodre and Bradley proposed a metric to predict the perception of treble as the ratio comparing the late arriving energy (after 80 ms) in the 4 kHz octave band to the late arriving energy in both the 1 – 2 kHz octave band.[34]

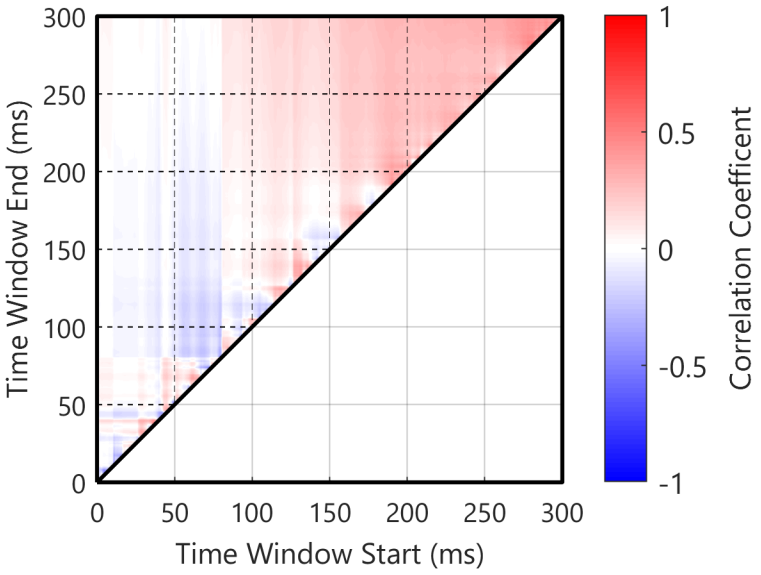


Figure 6-27: Correlations between factor 4.4 loadings as the start and end of a time integration window was varied, computed for the 14 halls in the study.

6.7 Conclusions

Using a state-of-the-art spherical microphone and loudspeaker array concert hall measurement database, auralizations were generated for a full orchestra in a repeatable manner between 21 different concert halls with a wide variety in shape, size, and reverberation characteristics (CHORDatabase, chapter 5). Using RIRs measured with a CSLA, the accurate frequency-dependent radiation patters of each instrument was reconstructed and directly built-into the RIR measurement protocol. Additionally, spatially accurate RIR auralizations were generated using RIRs measured with a 32-channel spherical microphone array. Using these auralizations, a subjective study was designed to sample the ten most common subjective attributes, along with preference, for 14 different concert halls.

A correlation analysis revealed high degrees of correlation between all ten attributes, so a principal components analysis and subsequent varimax factor rotation was used to reduce this space. In the three-dimensional solution, explaining 64% of the total variance, the three main factors were identified as a strength and envelopment factor (26% of variance), a clarity factor (25%), and a brilliance timbral factor (12%). For the four-dimensional solution, 72% of the total variance was explained and similar factor interpretations remained. The additional variance explained resulted from the separation of the strength and spaciousness factor into two factors, one of which appeared to be associated with envelopment (19%) and the other with source width (16%). Once this new factor space was generated, factor scores for each of the 14 halls were correlated with average ratings of preference and the ten attributes. Average preference was found to correlate strongly with the clarity factor (r = 0.64) and somewhat

strongly with the strength and spaciousness factor (0.40). When the four-factor solution is considered, correlation with the strength and source width factor and average preference increased (0.59) with a slight reduction in clarity's correlation (0.58). From the original attributes, the best correlation with average preference was found with proximity (0.81), temporal and spatial clarity (0.68 and 0.60, respectively), and warmth (0.58).

To consider how preference might vary at the individual level, the average attribute and factor scores for each of the 14 halls were correlated against individually averaged preference ratings of the seven-hall subset that was rated. Despite a reduction in sample size, quite strong correlations were found for the majority of subjects, and clear trends emerge. The current study supports the large-scale finding that most subjects can be sorted between two groups of preference, one preferring clarity and the other preferring loudness and spaciousness. The clarity group primarily associated with factor 4.1, and the loudness and spaciousness group primarily associated with factors 4.2 and 4.3. Inspecting the results on specific individuals, it becomes clear that not all individuals can be easily placed in one of these two groups. One specific individual associated their preference strongly in the opposite sense with loudness and envelopment, with a weak correlation with the clarity factor. This person enjoyed a softer, more direct sound, unrelated to either group and against classical wisdom regarding hall design. Other subjects fell on the borders between the clarity and spaciousness group, and even some subjects showed potential emphases between either source width or envelopment within the loudness and spaciousness group. Although grouping of subjects can assist with analysis, it does remove important information that seems to exist at the individual level.

Using spherical array beamforming techniques, a new method has been proposed to correlate subjective hall ratings of both individual attributes, preference, and the subjective factors with energy in each room's RIR over different time and spatial regions. A clear importance emerged regarding early lateral energy, from 0 to 60 ms in a range slightly frontal of strictly lateral, and this energy was correlated quite strongly with the perception of envelopment, the envelopment factor (4.2), and overall average preference. This region was also fairly correlated with source width. Traditional spaciousness metrics are only measured with a figure-of-eight or dipole pattern microphone, and these results indicate a clear need for more spatially informed metrics, considering spatial regions of the sound field measured with higher-order microphone arrays. When looking at proximity, correlations existed with early energy over most of the spatial region of the sound field, and correlation existed from 70 to 150 ms for more frontal energy. This region also was found to relate to overall average

preference, most likely causing the strong connection between proximity and average preference.

6.8 Acknowledgements

The authors would like to acknowledge Katie Krainc for assistance with subjective data collection and Nicholas Ortega for helpful discussion regarding statistical analysis and experimental design. This work was supported by the National Science Foundation (NSF) award #1302741.

Chapter 7
Overall Conclusions

Individual conclusions were drawn at the end of chapters four through six, specifically referenced to the compact spherical loudspeaker array (CSLA) processing, the CHORDatabase, and the subsequent perceptual study and spherical array beamforming analysis using the CHORDatabase. This chapter summarizes the overall work, highlighting the key findings from each previous section, tying all of the conclusions into a greater overall narrative. Additionally, the future work comprised from this research effort will be highlighted in detail. As much effort has gone into the accurate collection, processing, analysis, and subsequent subjective testing using the CHORDatabase, much potential for further work remains using this dataset.

7.1 Summary of Findings

7.1.1 The Concert Hall Orchestral Research Database (CHORDatabase)

The concert hall orchestral research database, or CHORDatabase (because every cool database needs a witty acronym) was generated using state-of-the-art RIR measurement techniques involving spherical array processing. A survey of researchers and consultants from around the US and Europe was used to gather suggestions of halls as measurement candidates which included a large variety in size, shape, and room acoustic perceptual qualities. Attempting to maintain variety in acoustic behavior and geometry, data as available was gathered for each hall. These data were used to select a representative short-list of halls to contact regarding measurements. In balancing travel cost, location, and hall variety, a total of 21 concert halls (15 in North America, 6 in Europe) were included in the database.

Standardized measurements were made in each hall using a three-part omnidirectional sound source for wide-bandwidth SNR and omnidirectional source behavior up to 5 kHz. This set of measurements was made at 242 unique seat locations across all halls and variable acoustic settings within the database. These measurements were comprised of both a standard receiver grid, preserving source receiver distance between halls, along with additional measurements to well-sample seating locations outside of the standard grid. Each seat

measurement took approximately 5 minutes. Using these measurements, the standard metrics used in the field were calculated from the spherical microphone array RIR, including lateral energy metrics. Overall distributions of metrics across the database revealed a substantial variety, covering a large portion of the range of variation of each metric listed in ISO 3382, Appendix A.[47] Some of these variations included an EDT range from 1.20 to 3.33 s, a C80 range from -5.7 to 1.8 dB, and a J_{LF} range from 0.01 to 0.47. The same equipment setup was used in all of the concert hall measurements, providing consistency across the entire database.

7.1.2 Spherical Array Room Impulse Response Beamforming Analysis

Along with standard metrics, streamlined techniques for spherical array beamforming analysis of each RIR in the database was generated. This processing scheme involved the integration of microphone capsule gain adjustments, SH encoding, radial filtering, cleaning of the RIR in the SH domain, Ambisonic format conversion, diffuse-field microphone equalization, spherical array beamforming, and higher-order Ambisonic auralization. Part of these processing techniques were done using existing toolboxes, including the SOFiA Toolbox,[91] the Ambisonic decoder toolbox,[84] and a SH toolbox by Politis,[92] all based out of the MATLAB environment. These processing techniques for both spherical microphone array beamforming and higher-order Ambisonics perform similar techniques in different conventions or formats.

These analysis techniques were used to generate spatial energy maps of discrete early reflections in a single RIR and average spatial energy maps for larger RIR time window. Comparisons of the early energy (0 – 100 ms) and late energy (100 – 1000 ms) were made for eight different halls of different shapes. These spatial energy maps showed clear connection to classical wisdom regarding the enveloping, spacious sound of shoebox halls and the frontal character of reverberant energy in fan-shaped halls. Finally, a full 3D spatial animation of the RIR in time was generated to demonstrate the full visualization capabilities of spherical array beamforming. This style of animation fully represents temporal and spatial information contained in the RIR. Clear potential for new spatial metrics exists from these data.

7.1.3 Radiation Control using a Compact Spherical Loudspeaker Array

To allow for a realistic auralizations, a custom 20-channel compact spherical loudspeaker array (CSLA) was built to reconstruct the frequency-dependent radiation patterns of different orchestral instruments. Although the previously mentioned setup allows for high spatial resolution analysis of the sound field, a realistic orchestra cannot be represented using a single point source. The CSLA provided a single sound source that could be flexibly controlled

to mimic each unique orchestral instrument. After superimposing 20 measurements across the stage. To minimize the impact of spatial aliasing, the array was designed to have a small overall diameter of 15.2 cm (6 inches). Twenty 40 mm diameter drivers with high linear excursion provided sufficient SNR down to a low frequency resonance of 200 Hz. To flexibly control the array, measured radiation patterns represented in the SH domain were encoded into a set of 20 driver-specific filters for each instrument. Using these filters, a single-channel measurement signal could be copied into a 20-channel signal, and after filtering this 20-channel signal with the 20-channel filter bank for a given instrument, the measurement signal would radiate from the source with the directivity characteristics matching that instrument. These filters included one-third octave band SH directivity information, radial filtering, decoding to loudspeaker driver signals, normalization to prevent individual driver distortion, and individual driver equalization.

Turntable measurements made in an anechoic chamber were used to validate the directional accuracy of the array for all instruments, up to 3 – 4 kHz. Above this frequency, the array exhibits spatial aliasing, and the accuracy of the directional radiation pattern reconstruction is reduced. A 20-position orchestral source measurement grid was designed to represent an orchestra that is compatible with stages of most concert halls. Measurements were made at a single seat 15 m from the stage using this consistent orchestral arrangement in each hall. Diffuse-field source equalization filters were designed to remove non-flat frequency characteristics and overall sensitivity differences from the measurements for each instrument. An additional measurement was taken at each source location with the omnidirectional source's subwoofer, and appropriate crossover filter and diffuse-field equalization filter were designed to match the sensitivity and response of the subwoofer to that of the CSLA. In total, all of the measurements for the full-orchestra took 1.5 – 2 hours in each hall for a single seat. This time is far reduced from making separate RIR measurements for each of the 20 drivers in the CSLA, which would allow for flexible source directivity control in post-processing.

SH rotation and direct sound modification was used to extend these 20 measurements to a 61-piece orchestral arrangement in each hall, compatible with a recently released high-quality set of instrument-separated anechoic recordings of Beethoven's 8th symphony.[100] With section and instrument balance adjustments, full-orchestral auralizations were generated in 21 different hall environments for real-time, realistic comparisons of concert halls with identical orchestras. The full-orchestral auralizations provide a realistic and repeatable auralization for comparing halls across the CHORDatabase.

7.1.4 Subjective Study of Individual Concert Hall Preference

Using these auralizations, a subjective study was designed to determine the overall number of dimensions needed to represent most of the perceptual variation between concert halls. Analysis from this study investigated how these primary perceptual dimensions related to both average and individual preference. The study consisted of subjective ratings of preference along with the ten most common subjective attributes found to be important in concert hall acoustic, including reverberance, warmth, intimacy, etc. For testing time limitations, a smaller representative set of 14 halls was used in the study.

Large amounts of correlation were found between each of the ten individual attributes, so a principal components analysis and varimax factor rotation was used to reduce the perceptual space. The space was reduced to a set of either three factors that explained 64% of the total perceptual variance or four factors that explained 72% of the total variance. The three main factors that emerged were interpreted as a clarity factor (explaining 25% of the variance), a strength and spaciousness factor (26%), and a brilliance factor (12%). When the four-dimension solution is retained, the additional explained variance is used to separate the strength and spaciousness factor into two factors, one associated with envelopment (19%) and the other associated with source width (16%), which is consistent with Morimoto's[106] and Bradley and Soulodre's finding.[32]

When considering the average preference across all individuals, the highest correlation was found with proximity, also deemed important to average listener preference in previous literature.[23,38] Additional aspects of higher correlations included clarity attributes, envelopment, warmth, and envelopment. When correlated against the perceptual factors, average preference correlated with the clarity dimension, and in the four-factor perceptual space, it had a high correlation with the strength and source width factor. To gain more insight into individual preference, individually averaged preference ratings for each subject were correlated against hall-average values for the perceptual factors and the individual subjective attributes. Results indicated large inter-individual preference differences and tastes, which at the highest level, could be explained in the two classical preference groups. One group preferred clarity while the other group preferred strength and spaciousness. The grouping of subjects removed quite significant variations in individual preference. While some subjects showed weaker and difficult to place preferences, one subject exhibited a highly unique preference, preferring softer, less enveloping halls, even somewhat disconnected from clarity. Others fell on the border between both groups, and a forced selection of a single, discrete group

seemed to harm the interpretation of their individual preference. The current study suggests the importance of considering a continuous preference space, beyond simple grouping methods.

7.1.5 Correlations between Beamforming Data and Subjective Ratings

To identify which aspects of a room's sound field correlate with different perceptual factors, the spatial and temporal spherical array beamforming data was correlated with envelopment, source width, proximity, average preference, and each of the factors from the four-dimensional perceptual space. Envelopment was found to be highly correlated with a specific directional range from ±40° to ±110° azimuth and 40° to 130° elevation for early energy before 60 ms. This directional range is highly related to previous studies of envelopment by Dick and Vigeant.[52,107] For later energy, correlations with envelopment were found over a most of the spatial range. Similar, but less pronounced spatial correlations with early energy were again found for source width but correlations were more prominent for later energy after 60 ms. Source width later correlations were strongest for directions above, behind, and to the side of the listener. Proximity showed strong correlations from 70 to 150 ms in directions oriented to the front of a listener, with other clear regions of correlations behind the listener as well. This finding indicates that energy in a mixing time between early and late ranges, especially from in front of a listener, might generate a suitable metric for proximity. Such a metric would prove useful in assessing the average, overall quality of a hall.

Along with individual attributes, each of the four factors were correlated with the spatial energy maps. The first clarity factor appeared to relate to the classical interpretations that early energy is beneficial and later energy in the RIR is not helpful. At closer inspection of the time-domain correlation, questions could be raised in regard to the cutoff range for clarity, suggesting the possibility of a gradual early-late transition from 60 to 130 ms. For the strength and envelopment factor, the same spatial and temporal correlations emerge, as were found with envelopment directly. For the strength and source width factor, similar spatial correlations with early energy were found as were found with envelopment, but the later sound energy, arriving after 60 ms, was found to best correlate in the directions above, behind, and possibly to the sides of a listener. Frontal late energy did not show as high correlations with source width. The final factor was associated with brilliance and did not show promising correlations over any of the broadband time-spatial integration analysis. With further consideration of the frequency-dependent nature of these maps, including ratios of different frequency ranges, a potential suitable metric may be found. No current widely accepted metric is known to predict this impression, despite its inclusion as an important perceptual factor.

Across all subjects, average preference was correlated with the spatial energy maps. As proximity strongly correlated with average preference, when analyzing the time-domain correlations, similar results were found. The range of energy from 70 ms to 150 ms was identified as range of interest, which showed higher spatial correlations with frontal energy and some rear energy above a listener. Surprisingly, lower correlation was found in this range with lateral energy. A highly untested explanation could be due to a masking phenomenon. If strong, early reflections, before 70 ms, occur in the lateral direction in highly preferred halls, reflections arriving after those strong reflections might be masked, and not contribute to perception. The reflections arriving from other directions may contribute to preference, as they are not masked due to their spatial separation. Spatial and temporal masking effects most likely contribute highly to some of these findings and predictions.

7.2 Future Work

Many other analyses and studies could be directly possible using the CHORDatabase measurements and auralizations. New analyses are even possible using the same subjective data from the current study. First, averages for all of the ten individual subjective attributes in this study could be individually correlated with existing metrics and with the time, spatial, and frequency domain data from the CHORDatabase. To limit the scope, the current study did not incorporate the frequency dependence of the energy into the time and spatial domain analysis. Also, the current analysis did not investigate the impact of calculating energy ranges or relationships, as are used for metrics like C80. The beamforming data could be used to calculate ratios with a sliding time range for correlation, or even ranges with smooth transitions. Further refinement for decay-based metrics could be done by individually backwards integrating each directional RIR from the spatial beamforming analysis, providing access to the spatial dependence of metrics like EDT and T30.

Another possibility for future metric development could center around treating reflections as individual events, determined directly from the spatial beamforming data. It is likely that the brain applies a more sophisticated processing scheme than simple discrete energy windowing in time. Incorporating masking-related concepts from the field of psychoacoustics might provide good direction for future metric development.[108] This was a concept explored somewhat by Seraphim,[9] and in a practical sense by Marshall.[8] As a final note regarding metrics, of the defined factors of importance, clarity is the perception which has a relatively clear metric of adoption, and strength has a metric, G, but limitations still exist in its integration time range and whether or not a specific loudness weighting should be applied. The factors of envelopment, source width, and brilliance either have metrics that appear to be

unsatisfactory, or do not have a standard metric to predict each effect. It is important that new metrics are developed to accurately predict these underlying perceptual factors.

Along with new metrics, a highly important and practically implementable result from this database would be a short listening test, taking 5 to 10 minutes, which would could be used to predict or elicit an individual's taste in concert halls. If this test could be generated, it could be repeated in an efficient manner for many subjects and also repeated across different musical types. A more large-scale investigation of the repeatability of individual preference, how it varies within across musical genres, and the relative preference makeup of different concertgoers would prove extremely useful for the consultant. If available, consultants could have boards of directors or a sample of concertgoers for a hall renovation or new construction project take this test. Hall designs could then be individually tailors to the users and needs of the space. This job is currently the role of the senior acoustic consultant, relying on their refined expertise and experience to communicate the project needs and set the project's acoustic goals. Having such a listening test to complement a consultant's expertise would prove invaluable in ensuring concert halls meet the needs of those who listen to music in the space.

In terms of auralization, much possible future work could be done to better understand how to accurately generate full-orchestral auralizations of a concert hall. In the current work, auralizations were made using static instrument radiation patterns and anechoic recordings from a single microphone placed in the near field of an instrument. Future investigations could first be done to determine how many source positions are needed to accurately represent an orchestra in a concert hall. It is clear that a distributed array of sources with accurate directivity representation is needed, but the number of sources and the level of spatial accuracy needed to represent instrument directivity is not known. Further, it is not known if the later portion of a room's decay requires accurate source distribution and directivity information, or if this information is only perceptually important in the initial, early part of the RIR. Such knowledge of perception would allow for computer or simulation programs to produce full-orchestral auralizations with reasonable computational efforts, especially if this information is only important in the first 100 – 200 ms of the RIR. This is quite important for real-time auralization, which will most likely be common in the future.

Further, the realism of auralizations could be improved with increased accuracy in the spatial capture of anechoic recordings. Instrument anechoic recordings ideally should be made using a surrounding spatial microphone array in an anechoic chamber. This setup would capture the spatially dependent character of the instrument's radiation, which could be subsequently auralized using a ShRIR. This representation of the anechoic recording,

potentially in the SH domain, would allow direct auralization of the source without assuming a particular or even static instrument directivity. This method would be difficult to implement for a full orchestra, but comparisons with more simplistic methods could be made against this case to justify certain assumptions or simulations. Additionally, the encoding of array channels to the SH domain would not be straightforward due to the source centering problem. At the least, a far-field measurement with a diffuse-field like anechoic recording equalization of a single array channel might prove more realistic.

The most prominent remaining measurement-based limitation regards the use of unoccupied concert halls. It is usually impractical to take measurements in occupied concert halls. A single occupied measurement can be possible at the intermission of a 'soft' opening of a concert hall, but an audience will not stay and sit still over the larger period of time required for full concert hall measurements. It is known that the addition of an audience will cause a significant effect of the decay time of a hall, adding absorption to the room. Although known, this effect varies quite uniquely between halls, depending upon the absorption difference between the seat itself and the occupied seat, and most likely other hall-geometry related factors. Beyond just reverberance, the addition of people and human heads located around a listener will scatter and attenuate mid- and high-frequency energy. This effect would likely impact the perception of early reflections quite significantly, altering perception. To adjust for the change in reverberation for the occupied condition, the later reverberant decay in the room could be controlled by modifying the slope of the measured RIR to a new target decay time, but the impact of the local scattering from other audience members remains unstudied. This limitation could be a very important and impactful consideration in generating realistic auralizations. In a quite similar way, these concerns are also important regarding the absorption, reflection, and scattering properties of musicians on stage, chairs, and stands.

Finally, with the increasing prevalence of binaural techniques in the growing fields of virtual and augmented reality, auralization using binaural techniques with real-time head tracking, allowing for head rotation, shows good promise for the future wide-spread use of auralization in many fields. If an efficient method can be generated to individually tailor a binaural rendering to a listener, without an individually measured HRTF, this technique would provide a highly realistic, yet cost effective setup for auralization. Loudspeaker arrays are large, expensive, and require dedicated space, while headphones are portable and could extend the usefulness of auralizations to many firms, and if streamlined, auralization could even be implemented on smaller projects with lower budgets and fees. Auralization is a powerful tool for scientific research and the future of real-world consulting projects.

References

[1] W. Sabine, Collected Papers on Acoustics, Caimbridge: Harvard University Press, 1922.

[2] A. Kuusinen, "Perception of Concert Hall Acoustics - Selection and Behaviour of Assessors in a Descriptive Analaysis Experiment," *MS Thesis, Aalto University,* 2011.

[3] T. Somerville, "An Empirical Acoustic Criterion," *Acustica,* vol. 3, no. 6, pp. 365-369, 1953.

[4] T. Somerville and J. Head, "Empiricial Acoustic Criterion (Second Paper)," *Acustica,* vol. 7, pp. 96-100, 1957.

[5] L. Beranek, Music, Acoustics and Architecture, Wiley & Sons, 1962.

[6] R. Muncey and A. Nickson, "The Listener and Room Acoustics," *J. Sound Vib.,* vol. 1, no. 2, pp. 141-147, 1964.

[7] A. Nickson and R. Muncey, "Critera for Room Acoustics," *J. Sound Vib.,* vol. 1, no. 3, pp. 292-297, 1964.

[8] A. Marshall, "A Note on the Importance of Room Cross-Section in Concert Halls," *J. Sound & Vib.,* vol. 5, no. 1, pp. 100-112, 1967.

[9] H. Seraphim, "Über die wahrnehmbarkeit mehrerer Rückwürfe von Sprachschall," *Acustica,* vol. 2, no. 81-82, 1961.

[10] V. Jordan, "Acoustical Criteria for Auditoriums and Their Relation to Model Techniques," *J. Acoust. Soc. Am.,* vol. 47, no. 2 (Part 1), pp. 408-412, 1970.

[11] V. Jordan, "A Group of Objective Acoustical Criteria for Concert Halls," *Appl. Acoust.,* vol. 14, pp. 253-266, 1981.

[12] "Concert Hall Research Group," Acoustical Society of America, [Online]. Available: https://chrgasa.org/. [Accessed 21 May 2019].

[13] W. Chiang, "Effects of Various Architectural Parameters on Six Room Acoustical Measures in Auditoria," *Dissertation, University of Floria,* 1994.

[14] A. Gade, "Room acoustic properties of concert halls: quantifying the influence of size, shape, and absorption area," *3rd ASA / ASJ meeting,* p. paper 5aAA1, Dec. 1996.

[15] L. Beranek, Concert and Opera halls: How they sound., Melville: Acoustical Society of America, 1996.

[16] L. Beranek, Concert Halls and Opera Houses, New York: Springer-Verlag, 2003.

[17] L. Beranek, "Subjective Rank-Orderings and Acoustical Measurements for Fifty-Eight Concert Halls," *Acta Acust united Ac,* vol. 89, pp. 494-508, 2003.

[18] W. Reichardt, O. Abdel Alim and W. Wchmidt, "Definition und Meßgrundlage eines objektiven Maßes zur Ermittlung der Grenze zwischen brauchbarer und unbrauchbarer Durchsichtigkeit bei Musikdarbietung," *Acustica,* vol. 32, pp. 126-137, 1975.

[19] V. Thiele, "Richtungsverteilung und zeitfolge der schallrückwürfe in räummen," *Acustica,* vol. 3, pp. 291-302, 1953.

[20] W. Reichardt and U. Lehmann, "Definition of the room impression index R by determining the room impression of the basis of subjective examination of musical performance (German)," *Appl. Acoust.,* vol. 11, no. 2, pp. 99-127, April 1978.

[21] C. Lavandier, "Perceptive validation of an objective model for characterization of room acoustic quality (Validation perceptive d'un modèle objectif de caractérisation de la qualité acoustique des salles)," *Dissertation, Université du Maine,* 1989.

[22] E. Kahle and M. Bruneau, "Validation of an Objective Model of the Perception of Room Acoustical Quality in an Ensemble of Concert Halls and Operas (Validation d'un modèle objectif de la perception de la qualité acoustique dans un ensemble de salles de concerts et d' opéras)," *Dissertation, Le Mans,* 1995.

[23] Hawkes, RJ and Douglas, H, "Subjective Acoustic Experience in Concert Auditoria," *Acustica,* vol. 2, pp. 235-250, 1971.

[24] M. Barron, "Subjective Study of British Symphony Concert Halls," *Acustica,* vol. 66, no. 1, pp. 1-14, 1988.

[25] A. Sotiropoulou and D. Fleming, "Concert Hall Acoustic Evaluations by Ordinary Concert-Goers: I, Multi-dimensional Description of Evaluations," *Acustica,* vol. 81, pp. 1-9, 1995.

[26] A. Sotiropoulou and D. Fleming, "Concert Hall Acoustic Evanluations by Orinary Concert-Goers: II, Physical Room Acoustic Critera Subjectively Significant," *Acustica,* vol. 81, pp. 10-19, 1995.

[27] P. Damaske, "Head-Related Two-channel Stereophonie with Loudspeaker Reproduction," *J. Acoust. Soc. Am.,* vol. 50, pp. 1109-1115, 1971.

[28] V. Mellert, "Construction of a Dummy Head After New Measurements of Thresholds of Hearing," *J. Acoust. Soc. Am.,* vol. 51, p. 1359, 1972.

[29] K. Yamaguchi, "Mutivariate Analysis of Subjective and Physical Measures of Hall Acoustics," *J. Acoust. Soc. Am.,* vol. 52, no. 5, pp. 1271-1279, 1972.

[30] Schroeder, MR, Gottlob, D and Siebrasse, KF, "Comparative study of European concert halls: correlation of subjective preference with geometric and acosutic parameters," *J. Acoust. Soc. Am,* vol. 56, no. 4, pp. 1195-1201, 1974.

[31] S. Kimura, "Study on criteria for acoustical design of rooms by subjective evaluation of room acosutics," *J. Acoust. Soc. Jpn.,* pp. 606-614, 1976.

[32] J. Bradley and G. Soulodre, "Objective measures of listener envelopment," *J. Acoust. Soc. Am.,* vol. 98, pp. 2590-2597, 1995.

[33] M. Neal, "Investigating the sense of listener envelopment in concert halls using third-order Ambisonic reproduction over a loudspeaker array and a hybrid room acoustics simulation method," *Master's Thesis, The Pennsylvania State University,* 2015.

[34] G. Soulodre and J. Bradley, "Subjective evaluation of new room acoustic measures," *J. Acoust. Soc. Am.,* vol. 98, no. 1, pp. 294-301, July 1995.

[35] K. Lorenz-Kierakiewitz and M. Vercammen, "Acoustical Survey of 25 Europeans Concert Halls," 2019.

[36] J. Pätynen, "A virtual symphony orchestra for studies on concert hall acoustics," *Dissertation, Aalto University,* 2011.

[37] J. Pätynen and T. Lokki, "Directivities of Symphony Orchestra instruments," *Acta Acust. united Ac,* vol. 96, pp. 138-167, 2010.

[38] T. Lokki, J. Pätynen, A. Kuusinen and S. Tervo, "Disentangling preference ratings of concert hall acoustics using subjective sensory profiles," *J. Acoust. Soc. Am.,* vol. 132, no. 5, pp. 3148-3161, November 2012.

[39] T. Lokki, J. Pätynen, A. Kuusinen, H. Vertanen and S. Tervo, "Concert hall acoustics with individually elicited attributes," *J. Acoust. Soc. Am.,* vol. 130, pp. 835-849, 2011.

[40] A. Kuusinen, J. Pätynen, S. Tervo and T. Lokki, "Relationships between preference ratings, sensory profiles, and acoustical measurements in concert halls," *J. Acoust. Soc. Am.,* vol. 135, no. 1, pp. 239-250, 2014.

[41] T. Lokki, J. Pätynen, A. Kuusinen and S. Tervo, "Concert hall acoustics: Reperoire listening position, and individual taste of the listeners influence the qualitative attributes and preferences," *J. Acoust. Soc. Am.,* vol. 140, no. 1, pp. 551-562, July 2016.

[42] S. Tervo, J. Pätynen, A. Kuusinen and T. Lokki, "Spatial Decomposition Method for Room Impulse Responses," *J. Aud. Eng. Soc.,* vol. 61, no. 1/2, pp. 17-28, 2013.

[43] S. Weinzierl, S. Lepa and D. Ackermann, "A measuring instrument for the auditory perception of rooms: The Room Acoustical Quality Inventory (RAQI)," *J. Acoust. Soc. Am.,* vol. 144, no. 3, pp. 1245-1257, September 2018.

[44] F. Brinkmann, A. Aspöck, D. Ackermann, S. Lepa, M. Vorländer and S. Weinzierl, "A round robin on room acoustical simulation and auralization," *J. Acoust. Soc. Am.,* vol. 145, no. 4, pp. 2746-2760, 2019.

[45] W. Ahnert and S. Feistel, "Advanced Measurements Techniques: Methods in Architectural Acoustics," *Architectural Acoustics Handbook,* pp. 75-118, 2017.

[46] A. Farina, "Simultaneous Measurement of Impulse Resposne and Distortion with a Swept-Sine Technique," *Proc. 108th Audio Eng. Soc. Conv.,* 19-22 February 2000.

[47] *ISO 3382:2009, "Acoustics – Measurements of room acoustics parameters – Part 1: Performance spaces".*

[48] D. Protheroe, "IRIS," Marshall Day Acoustics, [Online]. Available: http://www.iris.co.nz/. [Accessed 8th May 2019].

[49] T. Knüttel, I. Witew and M. Vorländer, "Influence of 'omnidirectional' loudspeaker directivity on measure room impulse responses," *J. Acoust. Soc. Am.,* vol. 134, no. 5, pp. 3654-3662, 2013.

[50] "High-Power Omnidirectional Loudspeaker - OmniPower Sound Source - Brüel and Kjær Sound & Vibration," Brüel and Kjær, [Online]. Available: https://www.bksv.com/en/products/transducers/acoustic/sound-sources/omni-power-light-4292. [Accessed 28th May 2019].

[51] G. Behler, "Uncertainties of Measured Parameters in Room Acoustics Caused by the Directivity of Source and/or Receiver," *Proc. of Forum Acust.,* December 2001.

[52] D. Dick, "A New Metric to Predict Listener Envelopment Based on Spherical Microphone Array Measurements and Higher Order Ambisonic Reproductions," *Dissertation, The Pennsylvania State University,* 2017.

[53] M. Gerzon, "Periphony: With-Height Sound Reproduction," *J. Audio Eng. Soc.,* vol. 21, no. 1, pp. 2-10, Januray/February 1973.

[54] A. Berkhout, "A Holographic Approach to Acoustic Control," *J. Audio Eng. Soc.,* vol. 36, no. 12, pp. 977-995, 1988.

[55] A. Berkhout, D. de Vries and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.,* vol. 93, no. 5, pp. 2764-2778, 1993.

[56] E. Williams, Fourier Acoustics: Sound Radiation and Nearfield Acoustic Holography, Academic Press, 1999.

[57] J. Blauert, Spatial Hearing – The Psychophysics of Human Sound Localization, revised edition., Cambridge: MIT Press, 1996.

[58] H. Möller, M. Sörensen, C. Jensen and D. Hammershöi, "Binaural Technique: Do We Need Invidual Recrodings?," *J. Audio Eng. Soc.,* vol. 44, no. 6, pp. 451-469, June 1996.

[59] E. Weisstein, "Moore-Penrose Matrix Inverse," MathWorld - A Wolfram Web Resource, [Online]. Available: http://mathworld.wolfram.com/Moore-PenroseMatrixInverse.html. [Accessed 27 May 2019].

[60] A. Heller, R. Lee and E. Benjamin, "Is My Decoder Ambisonic?," *Proc. of 125th Conv.of the Audio Eng. Soc.,* 1-5 October 2008.

[61] A. Heller and E. Benjamin, "The Ambisonic Decoder Toolbox: Extensions for Partial-Coverage Loudspeaker Arrays," in *Linux Audio Conference*, Karlsruhe, Germany, 2014.

[62] G. Romigh, D. Brungart, R. Stern and B. Simpson, "Efficient Real Spherical Harmonic Representation of Head-Related Transfer Functions," *IEEE J. Sel. Topics Signal Process.,* vol. 9, no. 5, pp. 921-930, 2015.

[63] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.,* vol. 45, no. 6, pp. 456-66, June 1997.

[64] V. Pulkki, "Spatial sound reproduction with directional audio coding," *J. Audio Eng. Soc.,* vol. 55, no. 6, pp. 503-516, June 2007.

[65] V. Pulkki, M. Laitinen, J. Vilkamo, J. Ahonen, T. Lokki and T. Pihlajamäki, "Directional audio coding - perception-based reproduction of spatial sound," *Proc. Int. Workshop on Spatial Hearing,* 11-13 November 2009.

[66] J. Merimaa and V. Pulkki, "Spatial Impulse Response Rendering I: Analysis and synthesis," *J. Audio Eng. Soc.,* vol. 53, no. 12, pp. 1115-1127, 2005.

[67] V. Pulkki and J. Merimaa, "Spatial impulse response rendering II: Reproduction of diffuse sound and listening tests," *J. Audio Eng. Soc.,* vol. 54, pp. 3-20, 2006.

[68] J. Pätynen, S. Tervo and T. Lokki, "Amplitude panning decreases spectral brightness with concert hall auralizations," *Proc. 55th Int. Conf. Audio Eng. Soc.,* 27-29 August 2014.

[69] M. Neal and M. Vigeant, "A measurement database of US and European concert halls for realistic auralization and study of individual preference," *Proc. IOA Conf. Aud. Acs.,* vol. 40, no. 3, October 2018.

[70] M. Vigeant, L. Wang and J. Rindel, "Investigations of Orchestra Auralizations Using the Multi-Channel Multi-Source Auralization Technique," *Acta Acust united Ac,* vol. 94, pp. 866-882, 2008.

[71] J. Meyer, "Acoustics and the Performance of Music, 5th Edition," 2009.

[72] N. Shabtai, G. Behler, M. Vorländer and S. Weinzierl, "Generation and analysis of an acoustic radiation pattern database for forty-one musical instruments," *J. Acoust. Soc. Am.,* vol. 141, no. 2, pp. 1246-56, February 2017.

[73] J. Bodon, "Development, Validation, and Evaluation of a High-Resolution Directivity Measurement System for Played Musical Instruments," *MS Thesis, Bringham Young University (BYU),* 2016.

[74] O. Warusfel and N. Misdariis, "Sound Source Radiation Synthesis: from Stage Performance to Domestic Rendering," *Proc. 116th Conv. Audio Eng. Soc.,* 8-11 May 2004.

[75] R. Avizienis, A. Freed, P. Kassakian and D. Wesel, "A Compact 120 Independent Element Spherical Loudspeaker Array with Programmable Radiation Patterns," *Proc. 120th Conv. Audio Eng. Soc.,* 20-23 May 2006.

[76] F. Zotter, "Analysis and Synthesis of Sound-Radiation with Spherical Arrays," *Dissertation, IEM, KU Graz,* 2009.

[77] M. Kerscher, "Compact Spherical Loudspeaker Array for Variable Sound-Radiation," *MS Thesis, IEM, KU Graz,* 2010.

[78] J. Klein and M. V. M. Pollow, "Optimized spherical sound source for auralization with arbitrary source directivity," *Proc. EAA Smp. Aura. and Amb.,* pp. 56-61, 3-5 April 2014.

[79] L. Kinsler, A. Frey, A. Coppens and J. Sander, "Fundamentals of Acoustics," 1999.

[80] C. Nachbar, F. Zotter, E. Deleflie and A. Sontacchi, "AmbiX - A Suggested Ambisonics Format," *Ambisonics Symposium,* 2011.

[81] B. Rafaely, Fundamentals of Spherical Array Processing, Berlin: Springer-Verlag, 2015.

[82] M. Gerzon, "General Metatheory of Auditory Localisation," *92nd Conv. Audio Eng. Soc.,* 24-27 March 1992.

[83] J. Daniel, "Représentation de champs acousticques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia," *Ph.D. Dissertation, University of Paris,* 2001.

[84] A. Heller, E. Benjamin and R. Lee, "A Toolkit for the Design of Ambisonic Decoders," *Linux Audio Conference,* 12-15 April 2012.

[85] J. Daniel, "Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format," *Audio Eng. Soc. 23rd Int. Conf.,* 2003.

[86] D. Blackstock, "Chapter 10: Spherical Waves," *Fundamentals of Physical Acoustics,* pp. 335-375, 2000.

[87] M. Pollow and G. Behler, "Variable Directivity for Platonic Sound Sources Based on Spherical Harmonics Optimization," *Acta Acust united Ac,* vol. 95, pp. 1082-1092, 2009.

[88] "Cycling '74 Max," [Online]. Available: www.cycling74.com/max7/. [Accessed 16 7 2015].

[89] H. Wilkens, "A Multidimensional Description of Subjective Judgement of Concert-Hall Acoustics [German Name: Mehrdimensionale Beschreibung subjektiver Beurteilungen der Akustik von Konzertstilen.]," *Acustica,* vol. 38, no. 1, pp. 10-23, 1977.

[90] H. Kuttruff, Room Acoustics, 4th Edition, Elsevier Science Publisher, 2000.

[91] B. Bernschütz, C. Porschman, S. Spors and S. Weinzierl, "SOFiA Sound Field Analysis Toolbox," *Int. Conf. on Spatial Audio,* December 2010.

[92] A. Politis, "Real/Complex Spherical Harmonic Transform, Gaunt Coefficients, and Rotations," [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/43856-real-complex-spherical-harmonic-transform-gaunt-coefficients-and-rotations?s_tid=prof_contriblnk. [Accessed 3 May 2019].

[93] A. Koretz and B. Rafaely, "Dolph-Chebyshev Beampattern Design for Spherical Arrays," *IEEE Trans. Signal Process.,* vol. 57, no. 6, pp. 2417-2420, June 2009.

[94] D. Dick and M. Vigeant, "A comparison of late lateral energy (GLL) and lateral energy fraction (LF) measurements using a spherical microphone array and conventional methods.," *Proc. of EAA Auralization & Ambisonics Symposium, Berlin,* 2014.

[95] M. Barron, "Late lateral energy fractions and the envelopment question in concert halls," *Appl. Acoust.,* vol. 62, pp. 185-202, 2001.

[96] N. Xiang, "Room-Acoustic Energy Decay Analysis," *Architectural Acoustics handbook,* pp. 119-136, 2017.

[97] M. Skålevik, "Concert Hall Acoustics, Online Rating and Beranek's Data Collection," *24th ICSV,* 23-27 July 2017.

[98] D. Schröder and M. Vorländer, "RAVEN: A real-time framework for the Auralization of interactive virtual environments.," *Proc of Forum Acusticum,* pp. 1541-1546, 2011.

[99] D. Schröder, "Physically Based Real-Time Auralization of Interactive Virtual Environments," *Dissertation, RWTH Aachen University,* 2011.

[100] C. Böhm, D. Ackermann and S. Weinzierl, "Eine mehrkanalige und nachhallfreie Aufnahme von Beethovens 8. Sinfonie," *DAGA,* 2018.

[101] *ISO 9613-1:1993, "Acoustics - Attenuation of sound during propagation outdoors - Part 1: Calculation of the absorption of sound by the atmosphere".*

[102] *ISO 9613-2:1996, "Acoustics - Attenuation of sound during propagation outdoors - Part 2: General method of calculation".*

[103] *ITU-R BS.1534-1 (2001-2003), "Method for the subjective assessmment of intermediate quality level of coding systems".*

[104] M. Lawless and M. Vigeant, "Effects of test method and participant musical training on preference ratings of stiuli with different reverberation times," *J. Acoust. Soc. Am.,* vol. 142, pp. 2258-2272, October 2017.

[105] R. Johnson and D. Wichern, "Applied Multivariate Statistical Analysis," 2013.

[106] M. Morimoto and Z. Maekawa, "Auditory spaciousness and envelopment," *Proc. of the 13th Int. Cong. on Acoust., Belgrade,* pp. 215-218, 1989.

[107] D. Dick and V. M, "An investigation of listener envelopment utilizing a spherical microphone array and third-order ambisonics reproduction," *J. Acoust. Soc. Am.,* vol. 145, no. 4, pp. 2795-2809, 2019.

[108] B. Moore, "An Introduction to the Psychology of Hearing," 2003.

This Page is Intentionally Left Blank

Appendix A
Instrument Radiation Patterns

This appendix contains the radiation reconstruction results of the compact loudspeaker array for all 13 instruments used in the concert hall measurements in Figures A.1 – A.13.

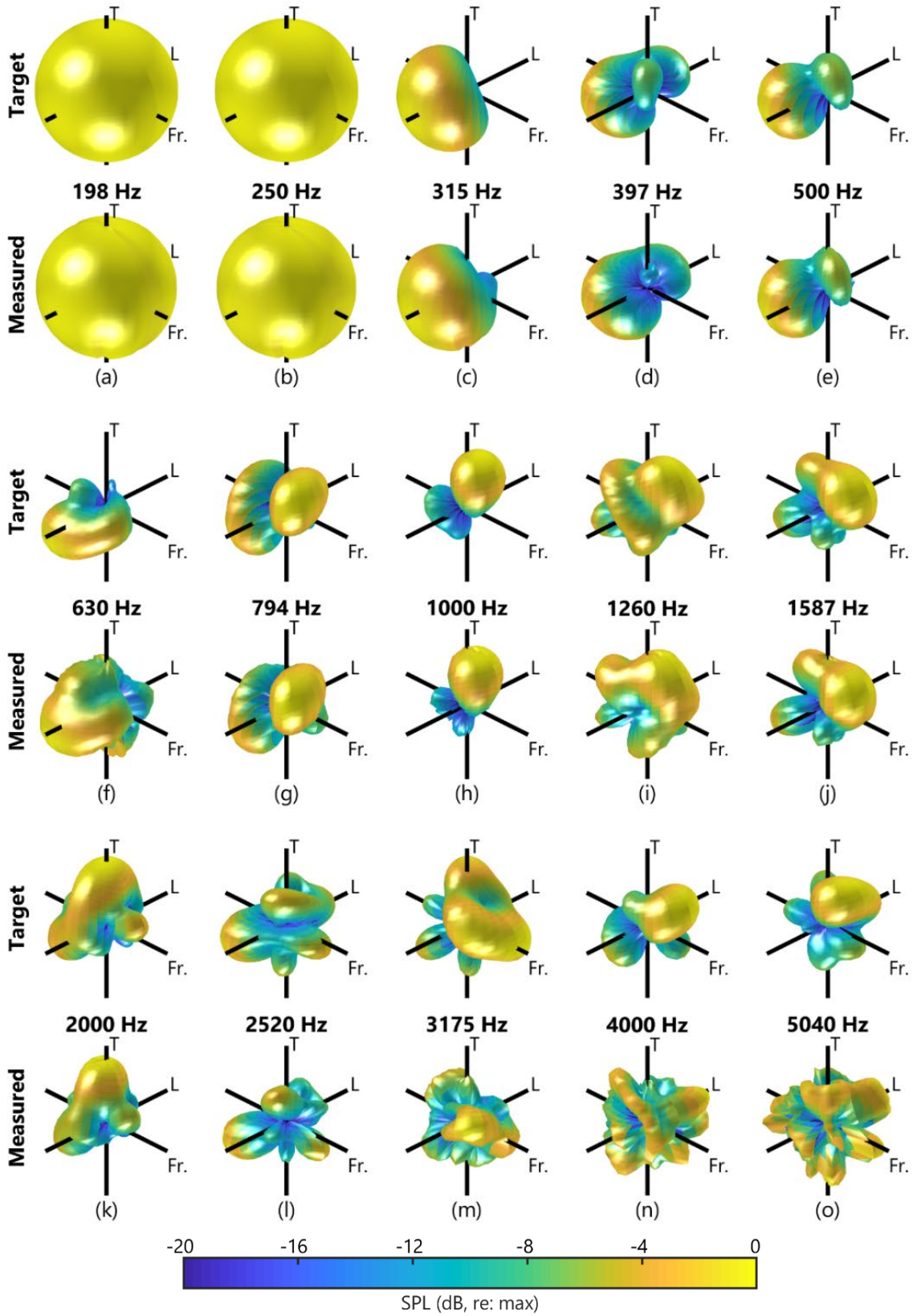


Figure A.1: Radiation reconstruction results for the bassoon. The SH order of the filters is increased from 0th to 1st, 2nd, and 3rd order for letters (a)-(b), (c), (d)-(e), and (f)-(o) respectively.

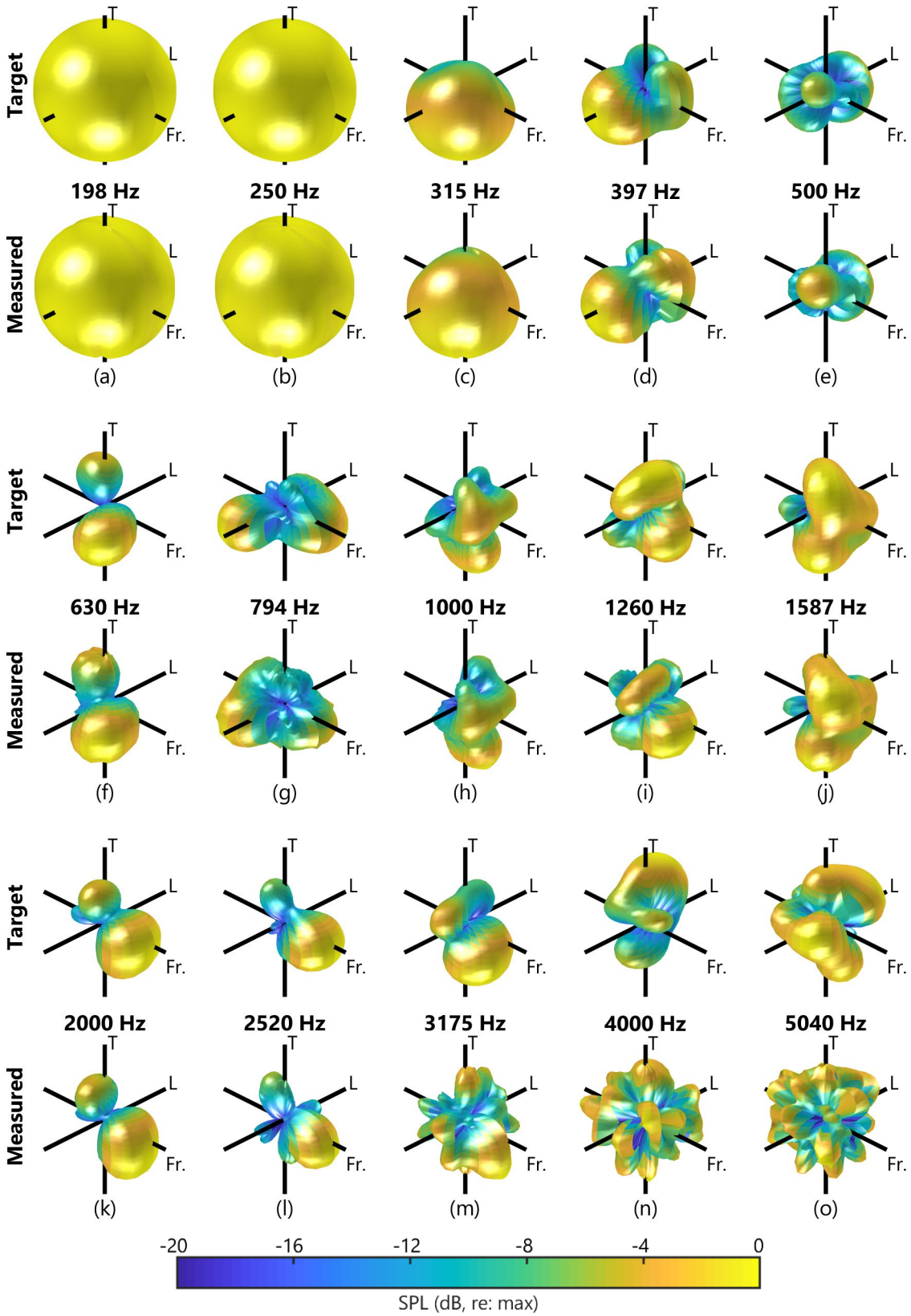


Figure A.2: Radiation reconstruction results for the cello. The SH order of the filters is increased from 0^{th} to 1^{st}, 2^{nd}, and 3^{rd} order for letters (a)-(b), (c), (d)-(e), and (f)-(o) respectively.

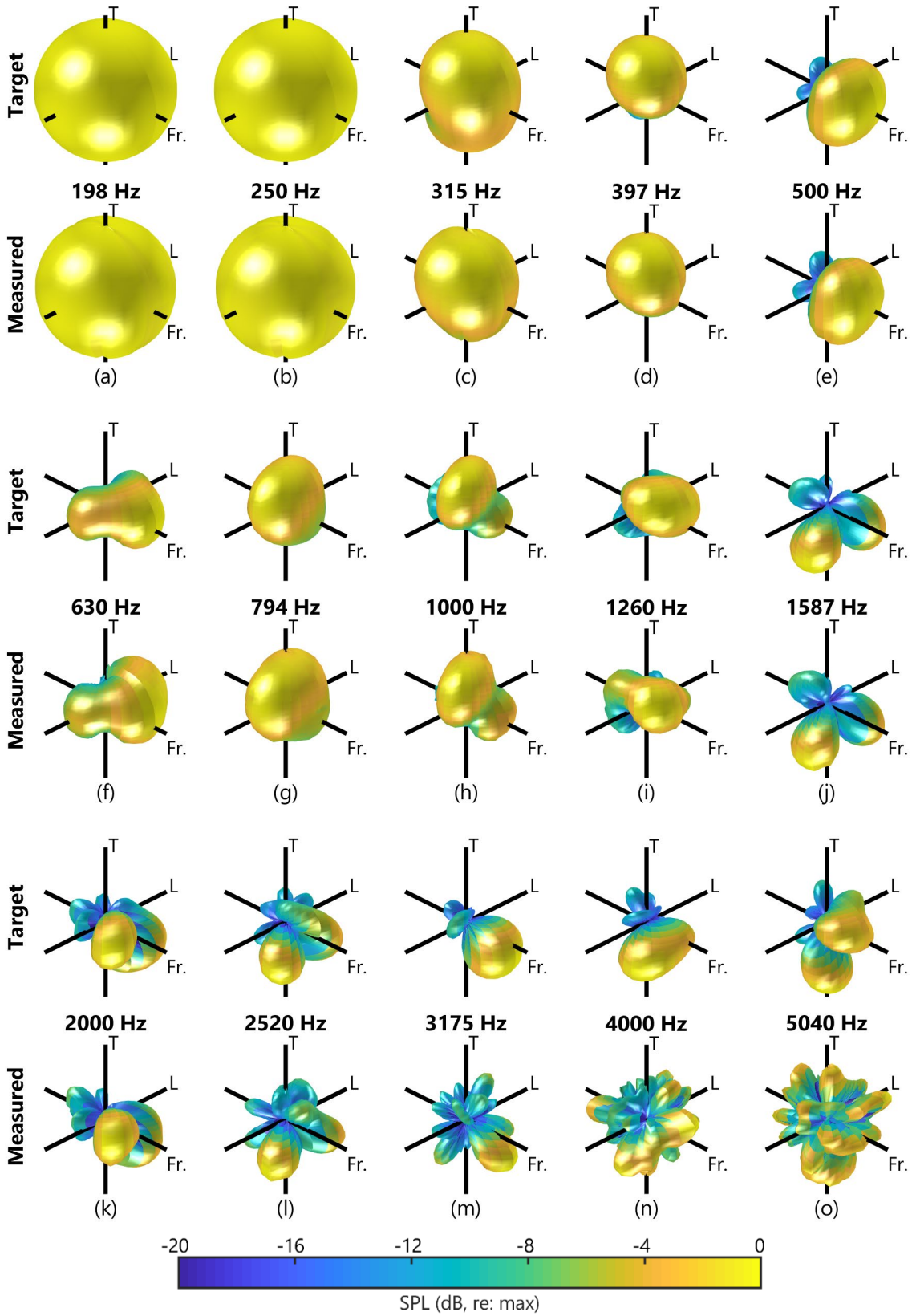


Figure A.3: Radiation reconstruction results for the clarinet. The SH order of the filters is increased from 0th to 1st, 2nd, and 3rd order for letters (a)-(b), (c), (d)-(e), and (f)-(o) respectively.

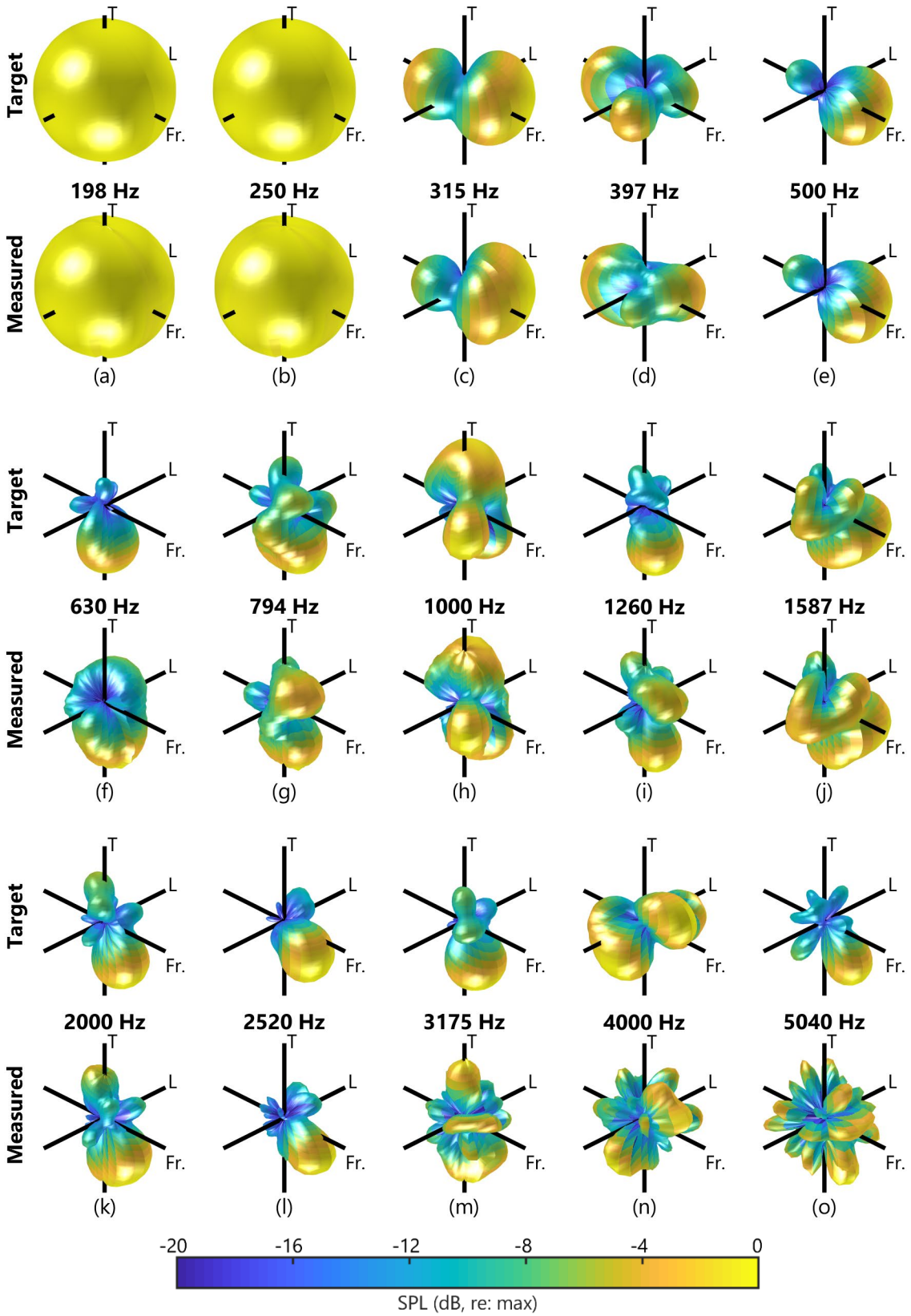


Figure A.4: Radiation reconstruction results for the double bass. The SH order of the filters is increased from 0th to 1st, 2nd, and 3rd order for letters (a)-(b), (c), (d)-(e), and (f)-(o) respectively.

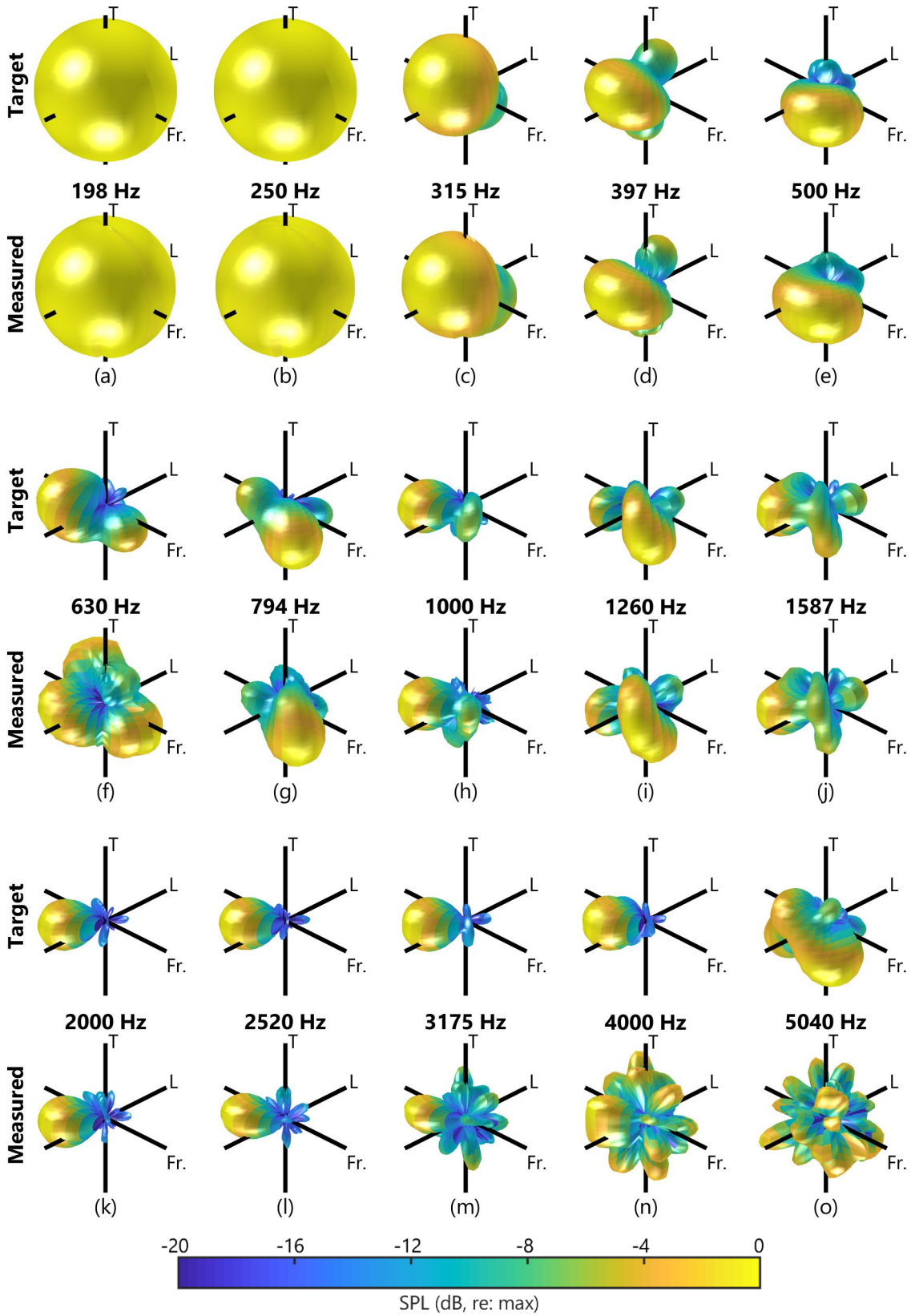


Figure A.5: Radiation reconstruction results for the French horn. The SH order of the filters is increased from 0th to 1st, 2nd, and 3rd order for letters (a)-(b), (c), (d)-(e), and (f)-(o) respectively.

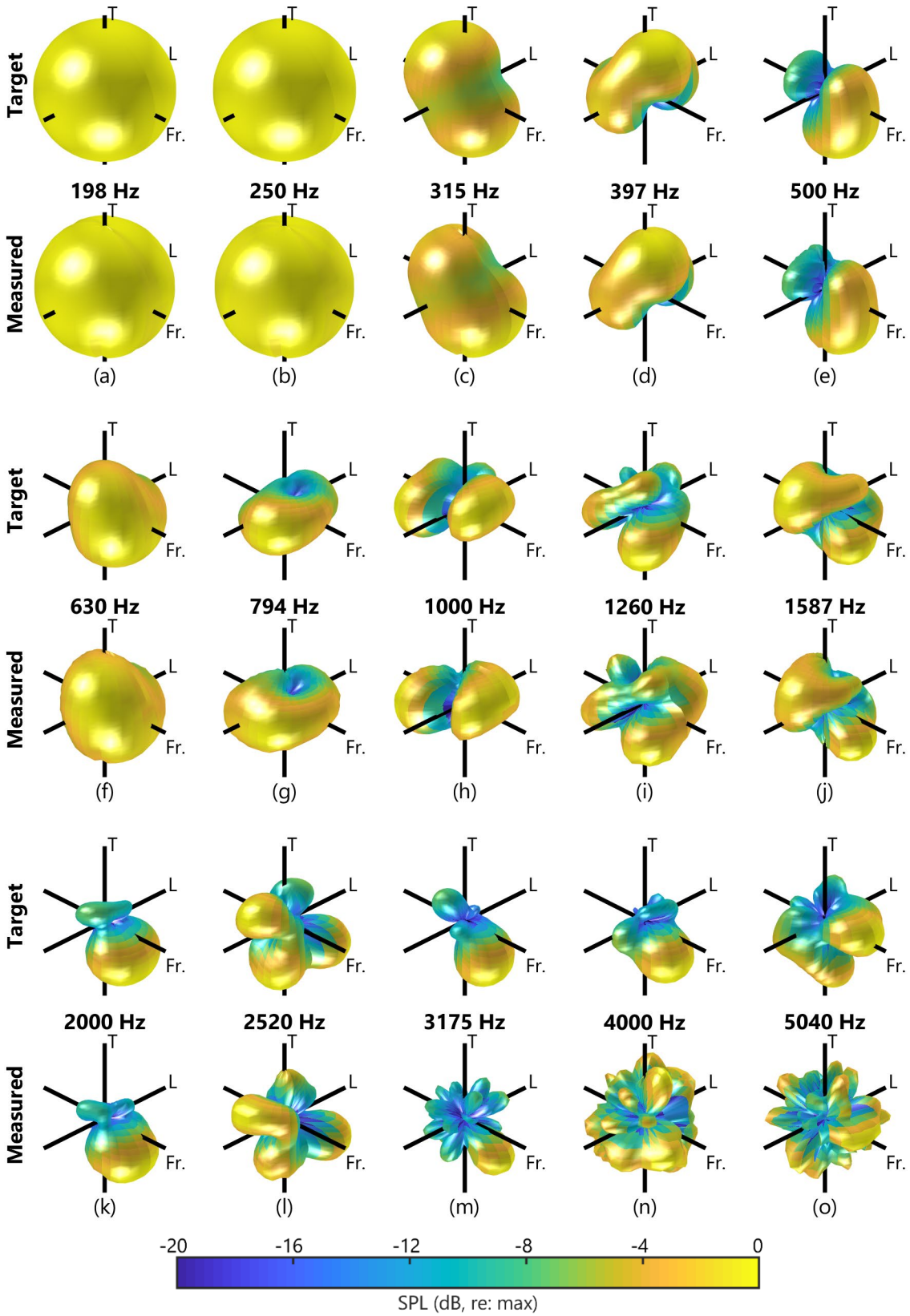


Figure A.6: Radiation reconstruction results for the oboe. The SH order of the filters is increased from 0^{th} to 1^{st}, 2^{nd}, and 3^{rd} order for letters (a)-(b), (c), (d)-(e), and (f)-(o) respectively.

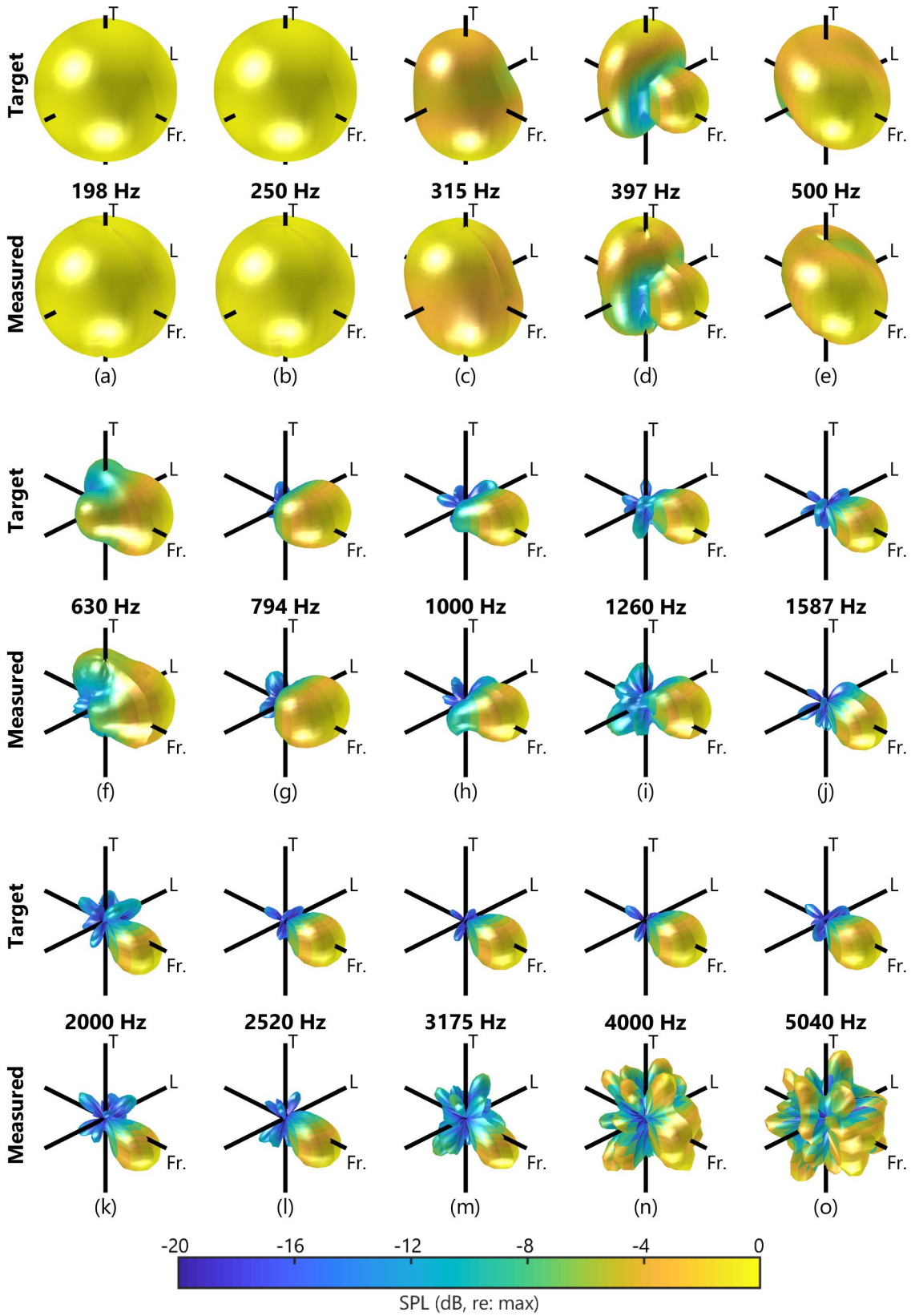


Figure A.7: Radiation reconstruction results for the tenor trombone. The SH order of the filters is increased from 0th to 1st, 2nd, and 3rd order for letters (a)-(b), (c), (d)-(e), and (f)-(o) respectively.

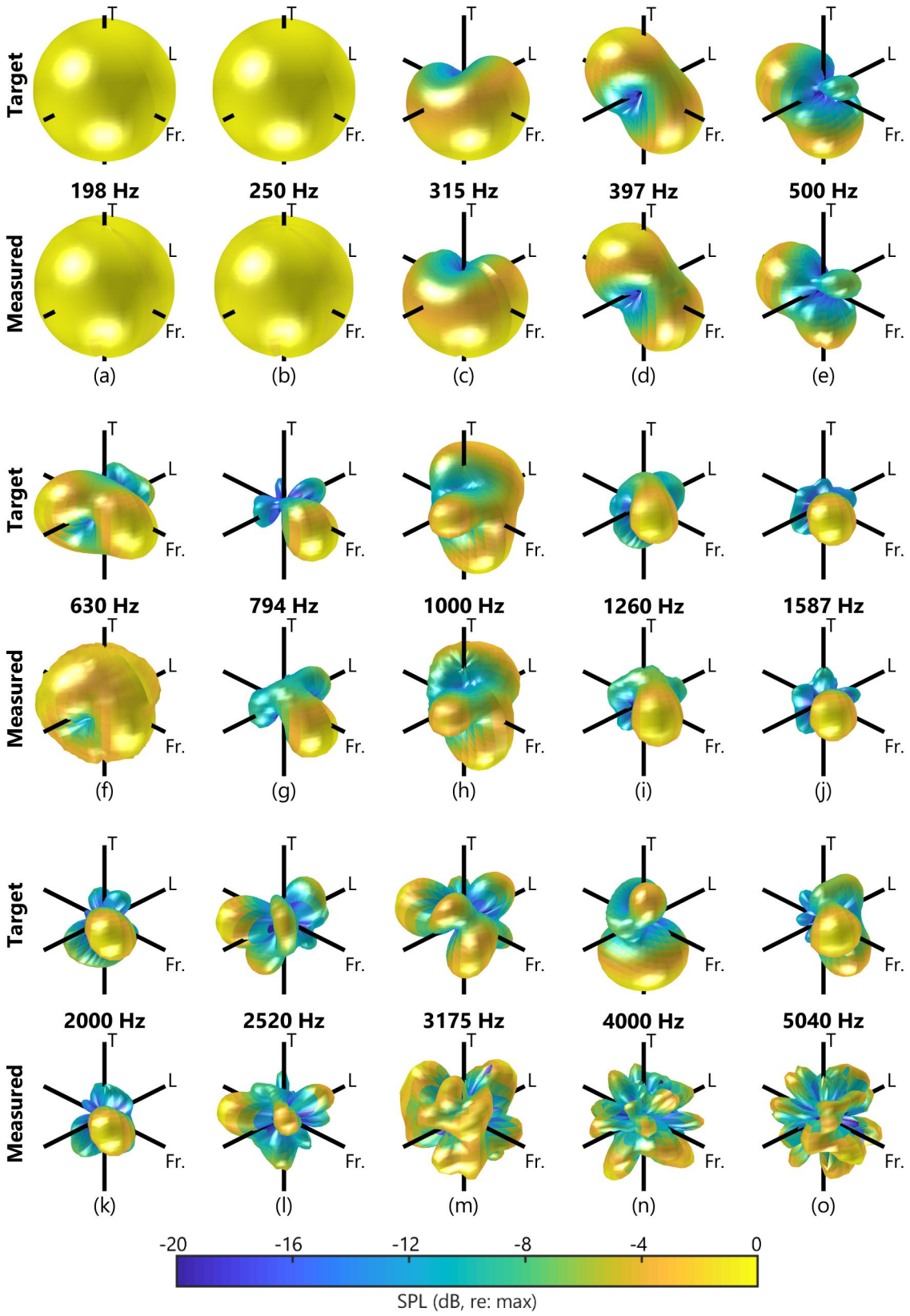


Figure A.8: Radiation reconstruction results for the transverse flute. The SH order of the filters is increased from 0th to 1st, 2nd, and 3rd order for letters (a)-(b), (c), (d)-(e), and (f)-(o) respectively.

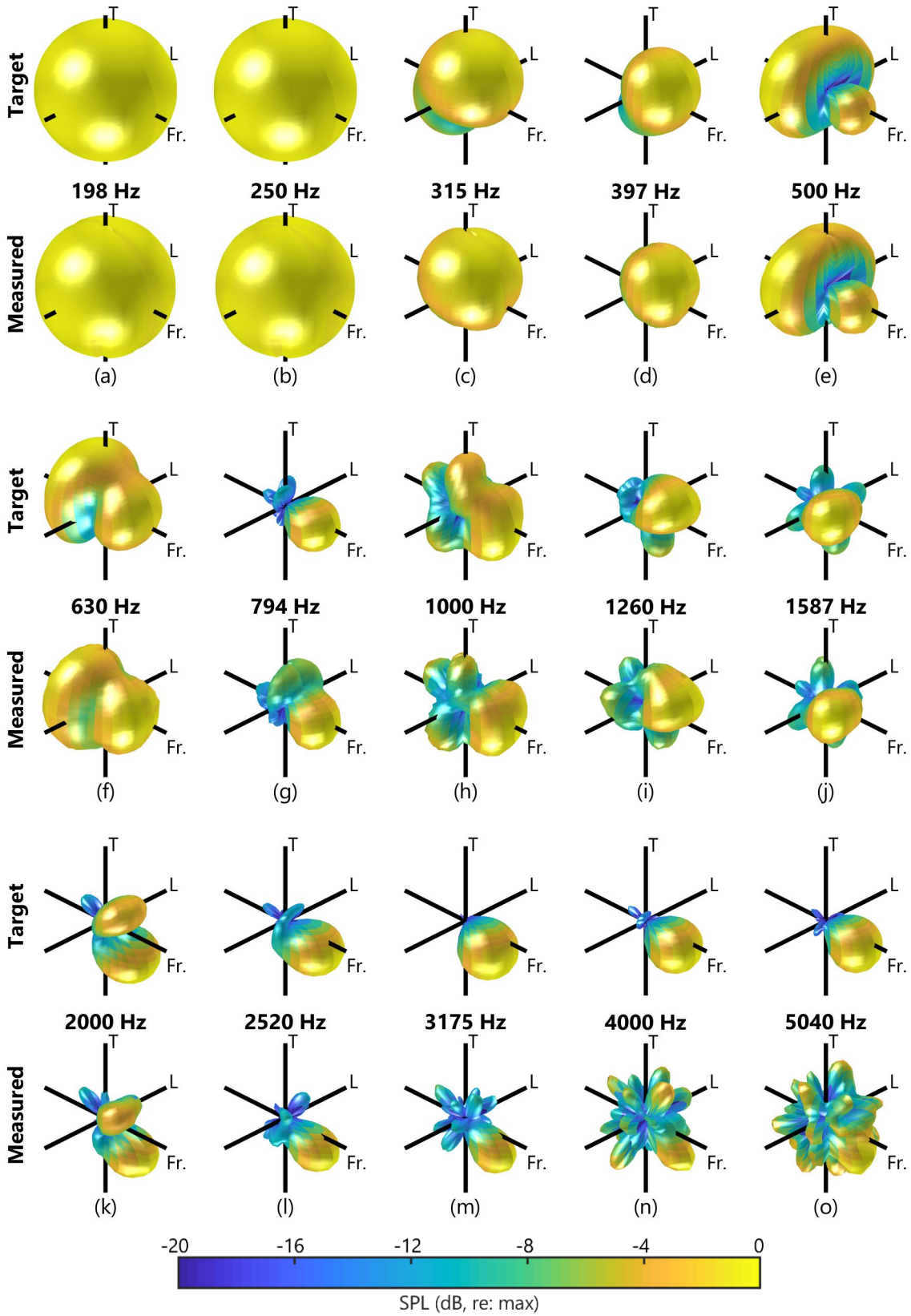


Figure A.9: Radiation reconstruction results for the trumpet. The SH order of the filters is increased from 0th to 1st, 2nd, and 3rd order for letters (a)-(b), (c), (d)-(e), and (f)-(o) respectively.

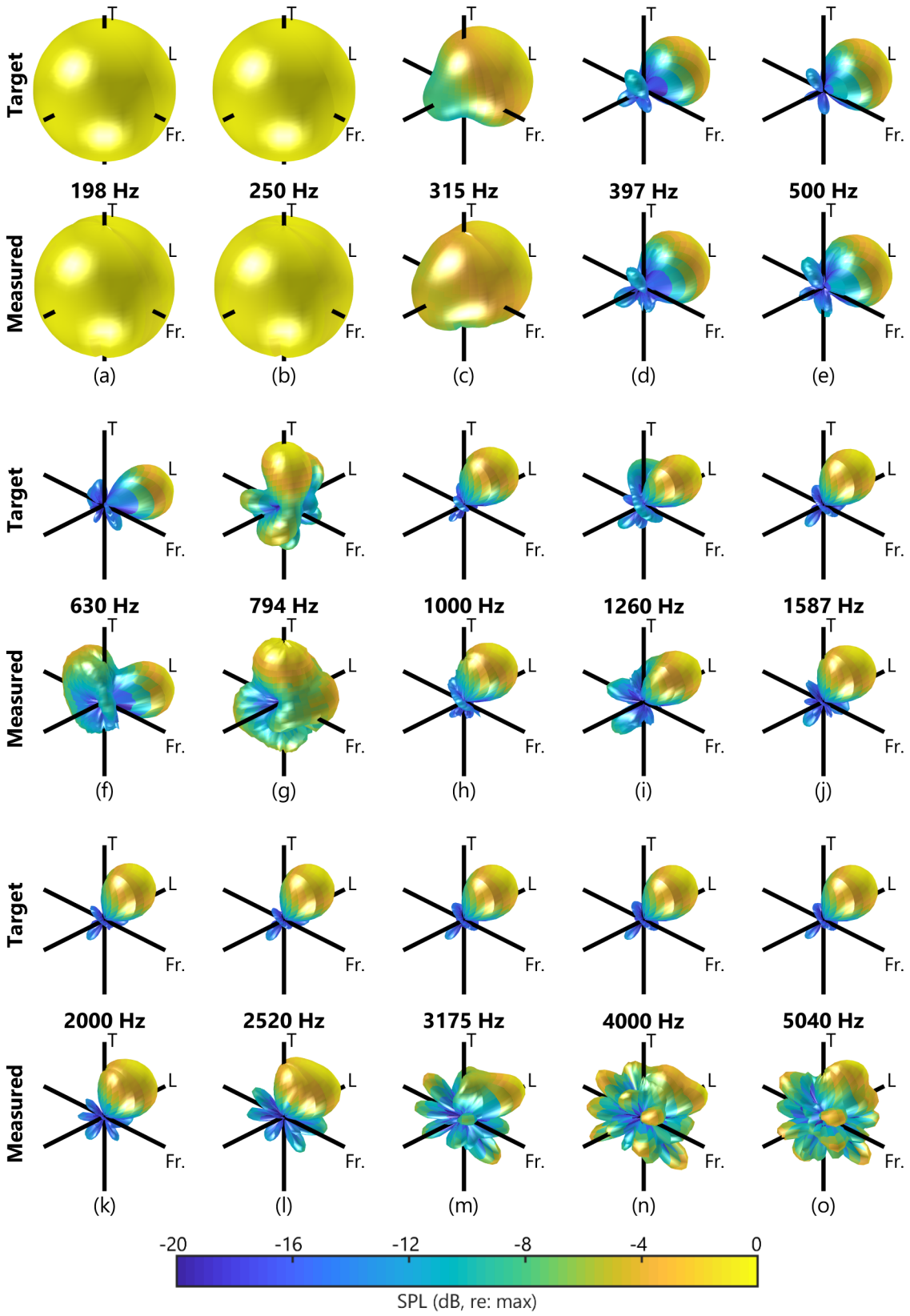


Figure A.10: Radiation reconstruction results for the tuba. The SH order of the filters is increased from 0^{th} to 1^{st}, 2^{nd}, and 3^{rd} order for letters (a)-(b), (c), (d)-(e), and (f)-(o) respectively.

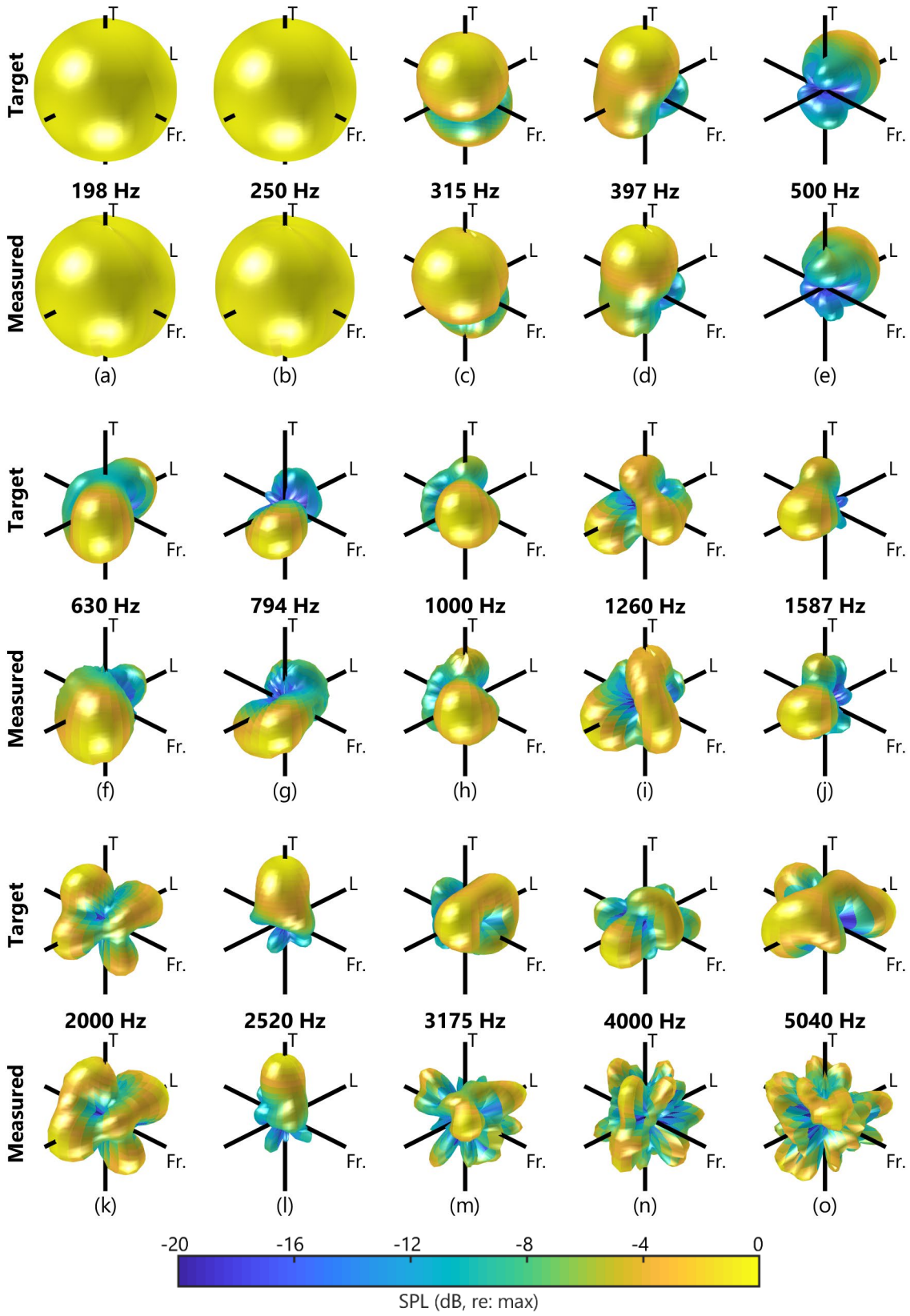


Figure A.11: Radiation reconstruction results for the viola. The SH order of the filters is increased from 0^{th} to 1^{st}, 2^{nd}, and 3^{rd} order for letters (a)-(b), (c), (d)-(e), and (f)-(o) respectively.

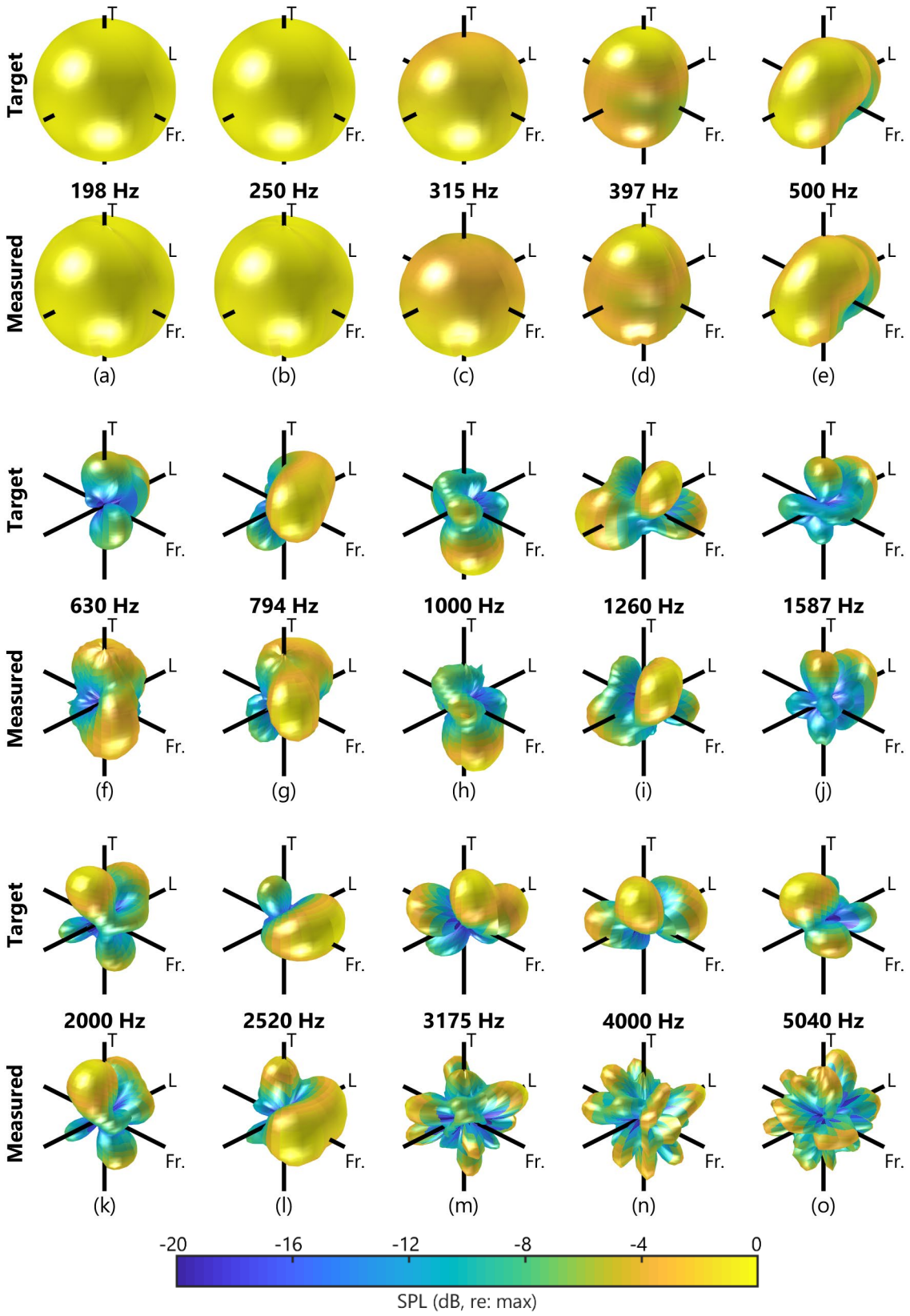


Figure A.12: Radiation reconstruction results for the violin. The SH order of the filters is increased from 0th to 1st, 2nd, and 3rd order for letters (a)-(b), (c), (d)-(e), and (f)-(o) respectively.

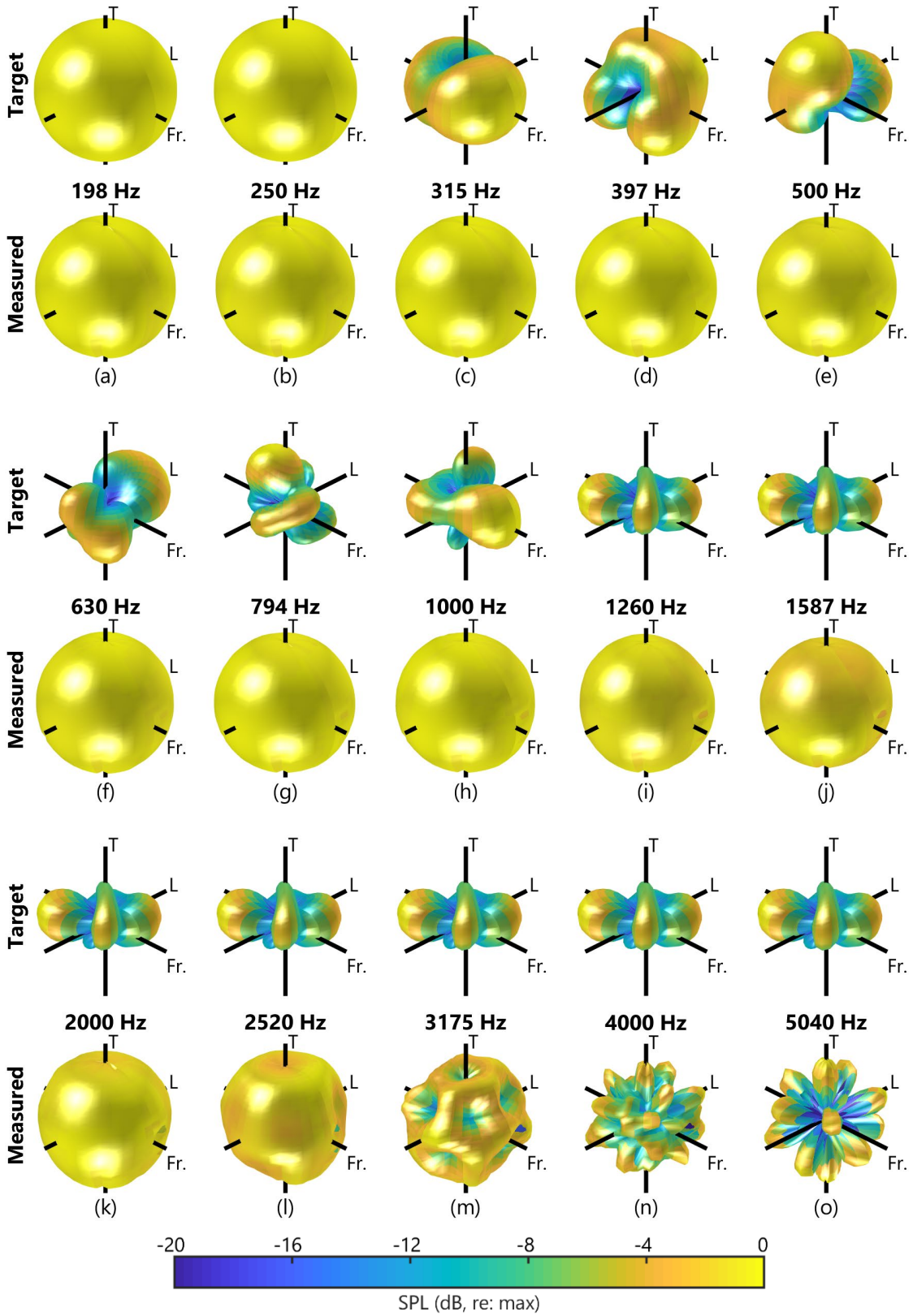
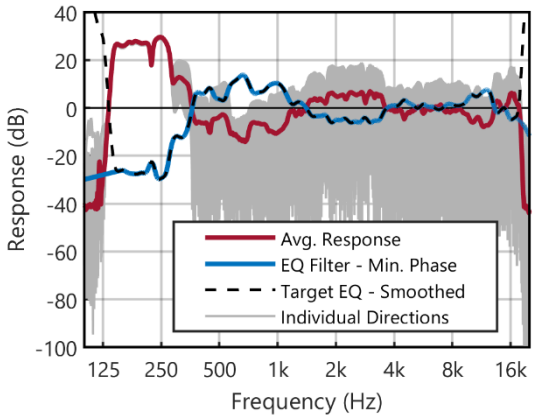


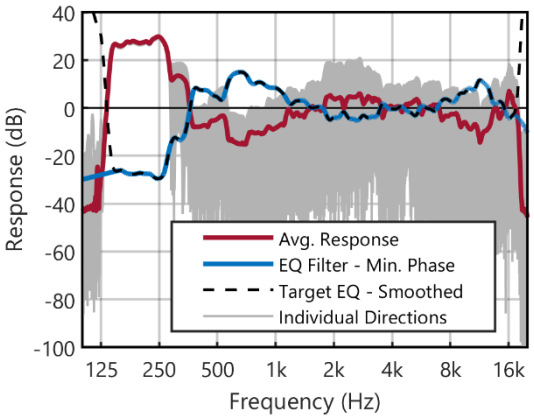
Figure A.13: Radiation reconstruction results for the timpani. To prevent any temporal artifacts that might be highly perceptible in a percussive instrument, the timpani source was measured with the omnidirectional radiation pattern. Comparisons are still provided above for completeness.

Appendix B
Instrument Directivity Filters

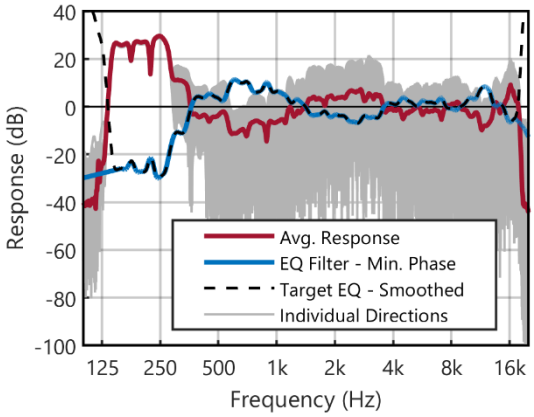
This appendix contains plots of the diffuse-field equalization filters for each of the 13 instruments used in the orchestral setup shown in Figure 4.14. The red line indicates the diffuse-field average response of each instrument, measured using a turntable setup in an anechoic chamber, described in Section 4.6. Each IR from each individual direction are shadowed being the plot in grey. Finally, the black dashed line indicated the inverted and frequency-smoothed target for the diffuse-field equalization filter. The blue line plots the final designed minimum-phase filter. Note: for the timpani, to prevent any temporal artifacts from the directional processing, an omnidirectional response was used. In the following plots the instruments shown are as follows: (a) bassoon, (b) cello, (c) clarinet, (d) double bass, (e) French horn, (f) oboe, (g) trombone, (h) transverse flute, (i) trumpet, (j) tuba, (k) viola, (l) violin, and (m) timpani (omnidirectional). The filters are each normalized or remove the overall gain differences that exist between each instrument radiation filter. This is most noticeably seen in the larger overall amplitude reduction built-in to the timpani (omnidirectional) diffuse-field equalization filter in (m).



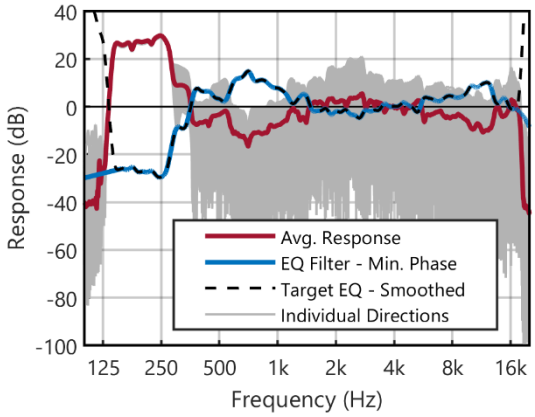
(a)



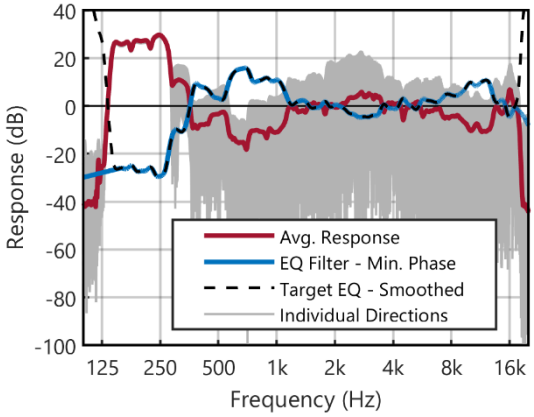
(b)


Response (dB)
Frequency (Hz)
Avg. Response
EQ Filter - Min. Phase
Target EQ - Smoothed
Individual Directions

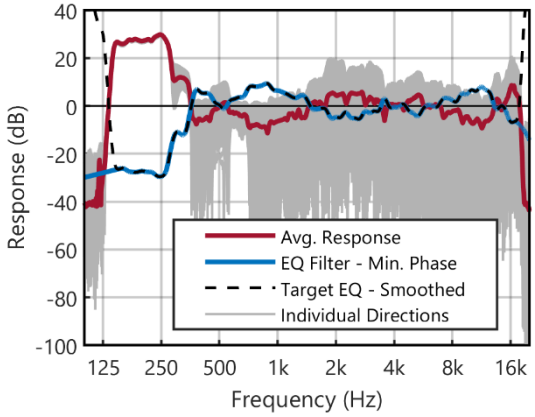
(c)


Response (dB)
Frequency (Hz)
Avg. Response
EQ Filter - Min. Phase
Target EQ - Smoothed
Individual Directions

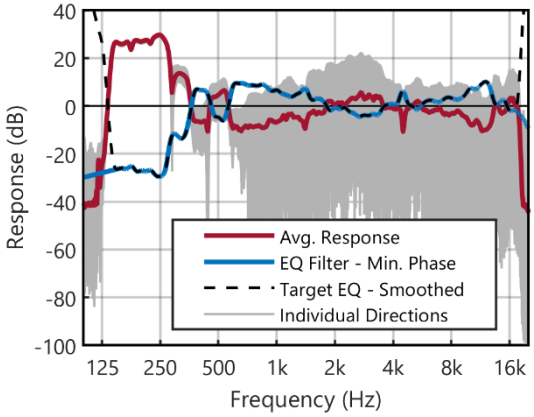
(d)


Response (dB)
Frequency (Hz)
Avg. Response
EQ Filter - Min. Phase
Target EQ - Smoothed
Individual Directions

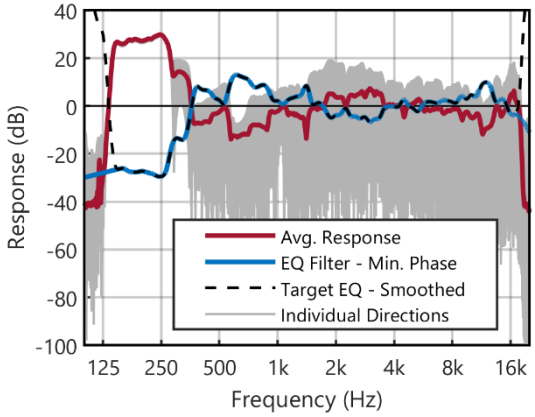
(e)


Response (dB)
Frequency (Hz)
Avg. Response
EQ Filter - Min. Phase
Target EQ - Smoothed
Individual Directions

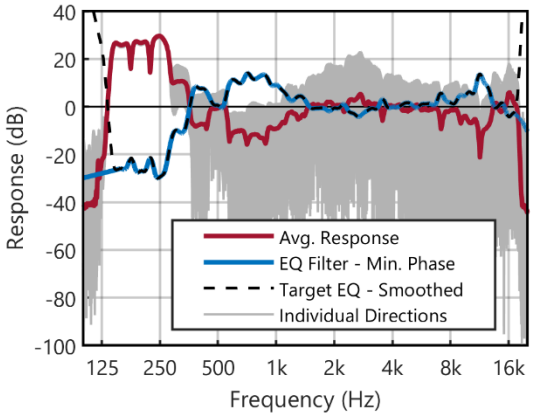
(f)


Response (dB)
Frequency (Hz)
Avg. Response
EQ Filter - Min. Phase
Target EQ - Smoothed
Individual Directions

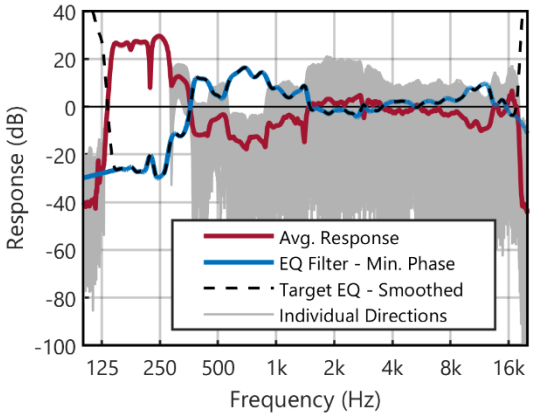
(g)


Response (dB)
Frequency (Hz)
Avg. Response
EQ Filter - Min. Phase
Target EQ - Smoothed
Individual Directions

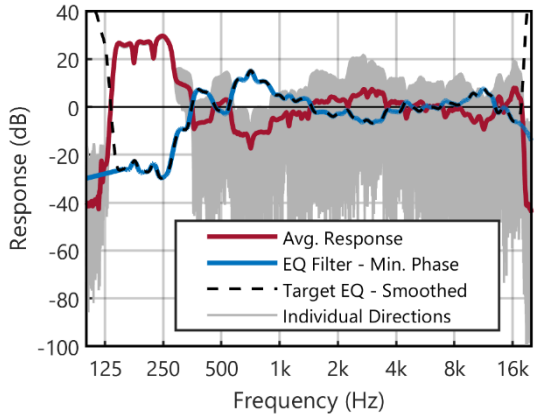
(h)


Response (dB)
Frequency (Hz)
Avg. Response
EQ Filter - Min. Phase
Target EQ - Smoothed
Individual Directions

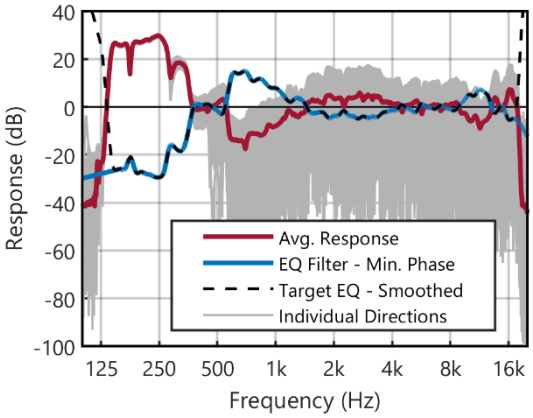
(i)


Response (dB)
Frequency (Hz)
Avg. Response
EQ Filter - Min. Phase
Target EQ - Smoothed
Individual Directions

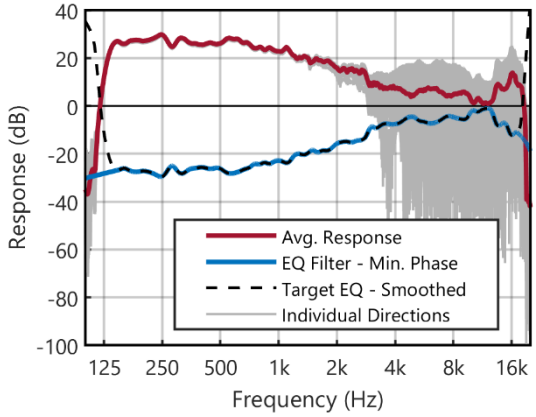
(j)


Response (dB)
Frequency (Hz)
Avg. Response
EQ Filter - Min. Phase
Target EQ - Smoothed
Individual Directions

(k)


Response (dB)
Frequency (Hz)
Avg. Response
EQ Filter - Min. Phase
Target EQ - Smoothed
Individual Directions

(l)


Response (dB)
Frequency (Hz)
Avg. Response
EQ Filter - Min. Phase
Target EQ - Smoothed
Individual Directions

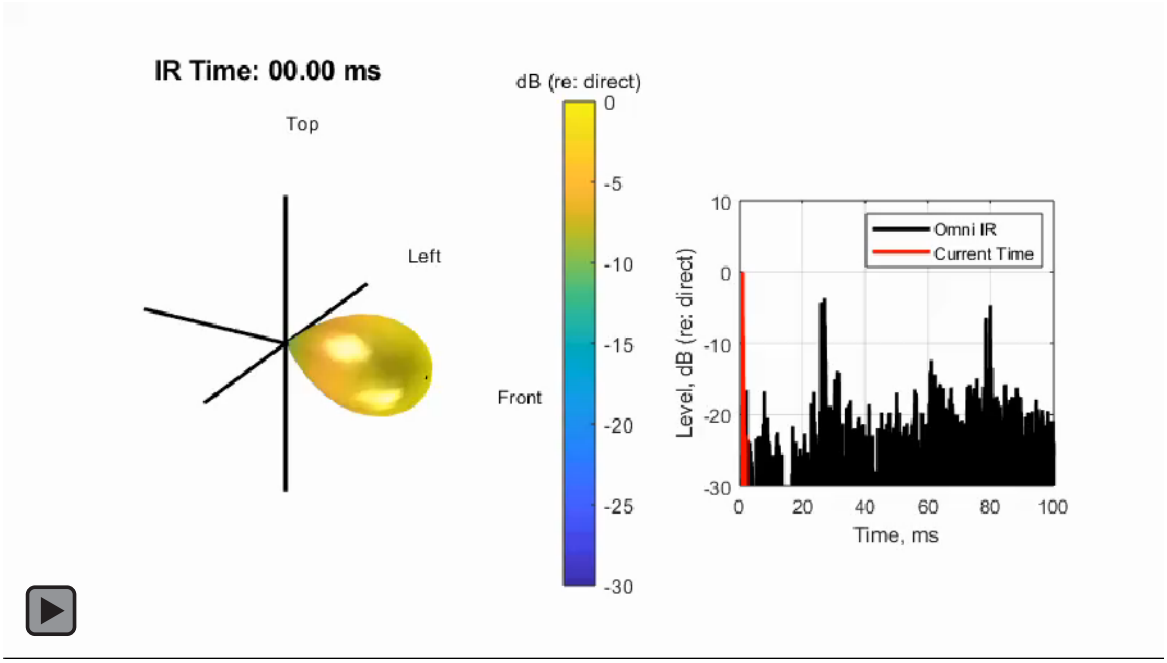
(m)

This Page is Intentionally Left Blank

Appendix C

RIR Beamforming Video Animation

This appendix contains the video animation of the beamforming analysis of a RIR, as described in section 5.6.2.3. This animation can also be found online at: https://sites.psu.edu/spral/animations-and-data/beamforming_video/



Vita

Matthew T. Neal

Matthew was originally born in Morgantown, WV, but grew up in Indiana, PA, graduating from Indiana Area Senior High School in 2009. Matthew was musically involved throughout high school, participating in three choral ensembles, concert band, jazz band, marching band, along with musicals and dramatics. He also participated in many music festivals in high school, most significantly being selected as a member of the MENC All-eastern honors chorus in 2009. After high school, Matthew entered Penn State to pursue a Bachelor of Architectural Engineering (BAE). Matthew graduated with his BAE in 2014, with a mechanical option emphasis, and a special interest in acoustics. Matthew also is a graduate of the Schreyer Honor's College at Penn State, and he received the award for outstanding record of study in the mechanical option. Matthew's undergraduate thesis focused on a geothermal retrofit of the Gaige Building on the campus of Penn State Berks, and his work was awarded the first-place senior thesis award for the mechanical option. While an undergraduate and M.S. student, Matthew was actively involved in the Essence of Joy choir at Penn State.

While finishing up his BAE, Matthew joined the Sound Perception and Room Acoustics Laboratory (SPRAL) led by Dr. Michelle C. Vigeant. Matthew participated in a National Science Foundation (NSF) sponsored research experience for undergraduates with Dr. Vigeant and began his MS research with her in the Fall of 2013. Matthew's M.S. focused on the design and construction of the Auralization and Reproduction of Acoustic Sound fields (AURAS) facility to virtually recreate concert hall sound fields, and he conducted a subjective study on the perception of listener envelopment in concert halls. After graduating with an MS in acoustics from Penn State's Graduate Program in Acoustics in August 2015, Matthew spent the fall as an architectural acoustics intern at Threshold Acoustics in Chicago, IL. Matthew rejoined SPRAL for his Ph.D. During graduate school, Matthew has been highly active in the Acoustical Society of America (ASA), giving nine lecture presentations at ASA national meetings, winning five best student paper awards. He also served as the ASA national student council representative for the Technical Committee on Architectural Acoustics (TCAA) from 2017 – 2019. Matthew was the recipient of the 2017 Leo and Gabriella Beranek Scholarship for Architectural Acoustics and Noise Control, and he also is recipient of both the INCA Leo Beranek Student Medal and the ASA Robert B. Newman Student Medal.