
The Creation of Perceptually Optimized Sound Zones Using Variable Span Trade-Off Filters

Ph.D. Dissertation
Taewoong Lee



Department of Architecture, Design, and Media Technology
Aalborg University
Rendsburggade 14, 9000 Aalborg, Denmark
Dissertation submitted October 15, 2021

Dissertation submitted: October 15, 2021

PhD Supervisor: Prof. Mads Græsbøll Christensen
Aalborg University

Assistant PhD Supervisor: Signal Processing Specialist Jesper Kjær Nielsen
Siemens Gamesa Renewable Energy A/S
formerly Assoc. Prof. at Aalborg University

PhD Committee: Professor Stefania Serafin (Chairman)
Aalborg University

Professor Thushara D. Abhayapala
The Australian National University

Associate Professor Filippo Maria Fazi
University of Southampton

PhD Series: Technical Faculty of IT and Design,
Aalborg University

Department:: Department of Architecture, Design,
and Media Technology

ISSN: xxxx-xxxx
ISBN: xxx-xx-xxxx-xxx-x

Published by:
Aalborg University Press
Skjernvej 4A, 2nd floor
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

© Copyright by Taewoong Lee

Printed in Denmark by Rosendahls, 2021

Normalsider: XXX sider (á 2.400 anslag inkl. mellemrum).
Standard pages: XXX pages (2,400 characters incl. spaces).

Curriculum Vitae

Taewoong Lee



Taewoong Lee received the M.Sc. degree in mechanical engineering from Korea Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2011. From 2011 to 2015, he was with SM Instruments, Daejeon, South Korea, working on analysis, visualization, and evaluation of sound and vibration as a senior research engineer. He is currently working toward the Ph.D. degree at the Department of Architecture, Design, and Media Technology with the Audio Analysis Lab, CREATE, Aalborg University, Aalborg, Denmark. He has been a visiting researcher at the Department of Electrical Engineering, ESAT-STADIUS, KU Leuven, Leuven, Belgium, in 2019. His research interests include audio signal processing, array signal processing, spatial audio, sound field reproduction, and applications.

Curriculum Vitae

Abstract

Acoustical isolation in a shared space, e.g., a living room, could be naturally achieved if the users are located in different rooms or use headphones at the cost of limited social interaction between the people. A personal sound system aims to create sound zones that provide such acoustical isolation for different audio contents by using a set of loudspeakers.

Generally, two different types of sound zones are considered: a bright zone and a dark zone. The bright (or listening) zone denotes an area in which the desired audio content is reproduced as faithfully as possible, or the acoustic potential energy is maximized. On the other hand, the dark (silent or quiet) zone indicates the area whose acoustic potential energy is minimized as much as possible. Tackling the problem of creating sound zones is usually done by either maximizing an acoustic contrast, i.e., the acoustic potential energy ratio between the bright and dark zones, or minimizing a reproduction error, i.e., the difference between the reproduced and desired sound fields. As traditional sound zone control methods optimize such physical quantities, the human auditory system, i.e., how we perceive sound, is not directly related to them.

This thesis focuses on proposing a framework that generates sound zones in a perceptually optimized manner. A fundamental foundation based on a subspace method, i.e., a generalized eigenvalue decomposition (GEVD), is proposed to provide such a framework, which controls the trade-off between acoustic contrast and reproduction error by tuning user parameters. On top of it, the human auditory system is integrated into the framework first in a nonadaptive manner later in an adaptive manner. The proposed framework is compared with the well-known sound zone control methods and evaluated via performance metrics, including acoustic contrast and reproduction error; furthermore, formal listening tests are also conducted. Apart from this, we have investigated practical aspects to understand the proposed method better, e.g., the computational complexity and the performance analyses for the tuning parameters. The frequency domain approach is investigated to reduce computational complexity while pertaining similar performance to the time domain approach. Besides, another subspace-based method, i.e., the conjugate gradient (CG), is also proposed to reduce the computational complexity and provide fast convergence compared to the GEVD-based approach. Lastly, we investigated a variety of precise control strategies for the proposed frameworks.

Abstract

Resumé

At opnå akustisk isolation i et fælles område som f.eks. en dagligstue kan opnås ved at bruge høretelefoner eller ved at bruge forskellige rum. Begge løsninger har dog den ulempe, at social interaktion mellem lytterne besværliggøres betydeligt. Målet med et personligt lydzonesystem er at muliggøre social interaktion samtidig med, at lyttere kan nyde forskellige lydmaterialer uden at forstyrre hinanden, og dette gøres ved at kontrollere en række højtalere.

Et personligt lydzonesystem kan grundlæggende set opbygges ved hjælp af to typer af lydzoner: en lys zone og en mørk zone. Den lyse zone (også kaldet lyttestonen) refererer til et område, hvori det ønskede lydmateriale bliver reproduceret bedst muligt. Med bedst muligt menes f.eks., at lydfeltet bliver genskabt eller at den akustiske energi maksimeres. I modsætning til dette referer den mørke zone (eller stillezonen) til et område, hvori den akustiske energi forårsaget af det ønskede lydmateriale dæmpes mest muligt. Et lydzonesystem bliver oftest designet ved enten at maksimere den akustiske kontrast mellem zonerne (det vil sige forholdet mellem akustiske energi i den lyse zone og den mørke zone) eller ved at forsøge at minimere reproduktionsfejlen (det vil sige forskellen mellem det reproducerede og ønskede lydfelt). Begge disse to tilgange designer lydzonesystemet udelukkende på baggrund af fysiske parametre og tager ikke hensyn til den menneskelige lydopfattelse.

Denne afhandling fokuserer på at foreslå en generel struktur for hvordan den menneskelige lydopfattelse kan indarbejdes i et lydzonedesign. Strukturer tager udgangspunkt i en underrumsmetode (nærmere bestemt en generaliseret egenværdiopløsning (GEVD)), som kan bruges til at vægte akustisk kontrast mod reproduktionsfejl ved at ændre brugerparametre. Modeller for den menneskelige lydopfattelse bliver indarbejdet i dette på en måde, der også kan udnytte det ønskede lydmaterials karakteristika. De foreslåede metoder sammenlignes med velkendte lydzonemetoder gennem forskellige fysiske metrikker som akustisk kontrast og reproduktionsfejl samt gennem formelle lyttetest. Praktiske problemer såsom beregningskompleksitet og parametertuning med de foreslåede metoder undersøges også. En metode i frekvensdomænet foreslås som en beregningseffektiv metode, der løser det oprindelige problem i tidsdomænet med god nøjagtighed. Derudover foreslås også en alternativ, beregningseffektiv underrumsmetode (conjugate gradient (CG)) til GEVDen. Endeligt undersøges en række præcise kontrolstrategier til de foreslåede metoder.

Resumé

List of Papers

The main body of this thesis consists of the following papers:

- [A] **T. Lee**, J. K. Nielsen, J. R. Jensen, and M. G. Christensen, “A Unified Approach to Generating Sound Zones Using Variable Span Linear Filters,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Calgary, AL, Canada, Apr. 2018, pp. 491–495.
- [B] J. K. Nielsen, **T. Lee**, J. R. Jensen, and M. G. Christensen, “Sound Zones as an Optimal Filtering Problem,” in *Proc. 52th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Oct. 2018, pp. 1075–1079.
- [C] **T. Lee**, J. K. Nielsen, and M. G. Christensen, “Towards Perceptually Optimized Sound Zones: A Proof-of-Concept Study,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Brighton, UK, May 2019, pp. 136–140.
- [D] **T. Lee**, J. K. Nielsen, and M. G. Christensen, “Signal-Adaptive and Perceptually Optimized Sound Zones with Variable Span Trade-Off Filters,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 2412–2426, 2020.
- [E] **T. Lee**, L. Shi, J. K. Nielsen, and M. G. Christensen, “Fast Generation of Sound Zones Using Variable Span Trade-Off Filters in the DFT-Domain,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 363–378, 2021.
- [F] L. Shi, **T. Lee**, L. Zhang, J. K. Nielsen, and M. G. Christensen, “A Fast Reduced-Rank Sound Zone Control Algorithm Using the Conjugate Gradient Method,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May, 2020, pp. 436–440.
- [G] L. Shi, **T. Lee**, J. K. Nielsen, and M. G. Christensen, “Generation of Personal Sound Zones with Physical Meaningful Constraints and Conjugate Gradient Method,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 823 – 837, 2021.

List of Papers

The following patent application has been filed in relation with the project:

- [1] **T. Lee**, J. K. Nielsen, J. R. Jensen, and M. G. Christensen, “Generating Sound Zones Using Variable Span Filters”. *WIPO Publication No.*, WO2019-197002, 2019.

Contents

Curriculum Vitae	iii
Abstract	v
Resumé	vii
List of Papers	ix
Preface	xiii
I Introduction	1
1 Introduction	3
1.1 Objectives and hypothesis	3
1.2 Structure	6
2 Creation of sound zones	6
2.1 Sound field modeling	9
2.2 Sound zone control methods	14
2.3 Evaluation of sound zones	17
3 Contributions	18
4 Conclusion and directions for future research	23
References	24

Contents

Preface

This thesis is submitted to the Technical Faculty of IT and Design at Aalborg University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy. The work was carried out from March 2017 to July 2020 in the Audio Analysis Lab at the Department of Architecture, Design, and Media Technology (CREATE) at Aalborg University.

The thesis concerns the generation of perceptually optimized sound zones using variable span trade-off filters and is divided into two parts. In the first part, an overview is given to the generation of sound zones, and previously proposed methods are reviewed. The main body of the thesis is its second part, which consists of a number of papers that have been published in peer-reviewed conferences or journals. The papers have been organized not in chronological order but according to both their significance and relevance.

First of all, I would like to send my deepest gratitude to my supervisor, Mads Græsbøll Christensen for this research opportunity, his detailed guidance, patience, and encouragement throughout the entire period of my PhD study. His endless support helped me from all the aspects of research. I also would like to express my sincere appreciation to my co-supervisor, Jesper Kjær Nielsen for his support, countless technical opinions, and discussions. They provided the foundations and ideas for the research. They never closed the doors to me every time I asked them whether they have time for discussion. I am also thankful to all the CREATE colleagues, especially the Audio Analysis Lab members, for their valuable time on discussions and listening tests.

I also would like to thank Toon van Waterschoot from KU Leuven in Belgium for hosting me and for discussions and support. I was grateful for being part of the DSP group during my stay in Belgium. The time that I had participated in measurements and experiments at the Library of Voices is one of the memorable moments of my life.

My heartfelt appreciation should go to my beloved wife, Eunsil, my son, Lohan, and my family. Nothing would have been possible without their unfaltering love, support, and encouragement for all these years.

Last but not least, a special thanks to Jungmin, who, although no longer with us, is my life mentor from both professional and personal perspectives. This thesis is dedicated to him.

Taewoong Lee

Preface

Leuven, Belgium, August 2021

Part I

Introduction

1 Introduction

The thesis concerns the creation of perceptually optimized sound zones, specifically regarding the methods of sound zone control and its thorough investigation. In the following sections, the objectives, the hypothesis, the structure are introduced.

1.1 Objectives and hypothesis

People often encounter a situation in which they expect to have different audio contents while in an acoustically shared space. Typically, acoustically separated regions in such a situation can be simply obtained by using a pair of headphones, or spatially separated regions can be obtained by being in an isolated space; however, social interaction between people might be prohibited in both of the cases. Alternatively, loudspeakers can be used for achieving the acoustically and spatially separated regions, which are referred to as *sound zones*. The main idea behind using loudspeakers is to exploit constructive and destructive interference by designing filters for each loudspeaker. Using a single loudspeaker, which could be seen as the simplest example, is insufficient to generate acoustically separated regions for different audio programs as it cannot generate such interference. Therefore, a system of two loudspeakers, which is referred to as *stereo*, can be considered; however, a set of more than two loudspeakers, which is referred to as a *personal sound system*¹, is typically exploited to create such sound zones [36, 38, 74]. In general, a bright (or listening) zone in which the listener can experience the desired audio content and a dark (or silent, quiet) zone in which no audio content is present are created simultaneously, as depicted in Fig. 1 (a). *Sound zone control* creates these zones by designing control filters for each of the loudspeakers to control the audio content² according to a particular design criterion. Scenarios for multiple bright zones also could be realized, as conceptually illustrated in Fig. 1 (b) for different audio contents with two bright zones and Fig. 1 (c) for the same

¹It should be noted that different names have been used in the literature, e.g., personal audio system [20, 21, 25, 41], multizone system [131], and private audio system [100].

²It should be noted that a variety of terms also has been used in the literature, e.g., audio program [32, 49], input signal. In this thesis, either audio content or input signal is used, according to the context.

audio content in different languages for two bright zones. In principle, individual sound zone control for each audio content is required to achieve multiple bright zones.

The concept of sound zones was first proposed more than two decades ago [37, 38]. Since then, extensive research for creating sound zones has been accomplished. Broadly, the sound zone control methods can be divided into two categories: energy-based approaches [14–16, 28, 29, 40, 84, 108, 115] and field matching approaches [5, 8, 19, 52, 70–72, 90, 102, 118, 133, 135]. The approaches in the former category seek the control filters to maximize the acoustic potential energy ratio between the bright and dark zones, which is defined as acoustic contrast. On the other hand, the approaches in the latter category seek the control filters to minimize a reproduction error, which is defined as the difference between the desired and reproduced sound fields generated by the set of loudspeakers.

Most of the existing sound zone control methods optimize physical metrics, e.g., the reproduction error in the bright zone, the residual energy in the dark zone. The reproduced sound fields can be typically evaluated by calculating the performance of acoustic contrast or reproduction error. In [37], the listening tests found that acoustic contrast needs to be above 11 dB if sound and image such as a TV screen are present together but ideally around 20 dB in the case of sound only³. In ideal conditions, e.g., an anechoic environment [20], approximately 19 dB of acoustic contrast was reported. In [103, 104], the listening tests found that at least 25 dB of target-to-interferer ratio (TIR), which is defined as the acoustic energy or loudness ratio between the reproduced and interfering sound field in a single zone [48], is preferable in the case of multiple bright zones. However, in a real environment, e.g., in a car cabin, about 15 dB of acoustic contrast was reported in [26, 87]. Therefore, in a realistic scenario that provides around 15 dB of acoustic contrast, the reproduced sound field might not be the one that the user expects to experience.

Unfortunately, most sound zone control methods assume that the input signal has a flat spectrum to have acceptable performance on average regardless of the input signal statistics, as in [15, 70, 108, 118]. The main advantage of this approach is that the control filters can be computed offline and be independent on input signals, which makes the optimization problem simpler; however, the disadvantage of this approach is that the control effort (or the array effort), which is defined as the sum of mean squared control filters [29, 41], could be consumed at the frequencies in which the energy is barely present. Furthermore, those frequency components could be even amplified in the reproduced sound field in some cases.

When the reproduction error is minimized by using one of the field matching approaches, it is able to take a weighting filter into account. The weighting filter is typically applied for reducing a pre- or post-ringing artifact from the control filters, e.g., [13, 89, 109]. However, the weighting filter can in-

³These values were B-weighted [1]. Therefore, in most of the cases, it would be preferable to have more than these values if sound pressure level (SPL), which is unweighted, is considered.

1. Introduction

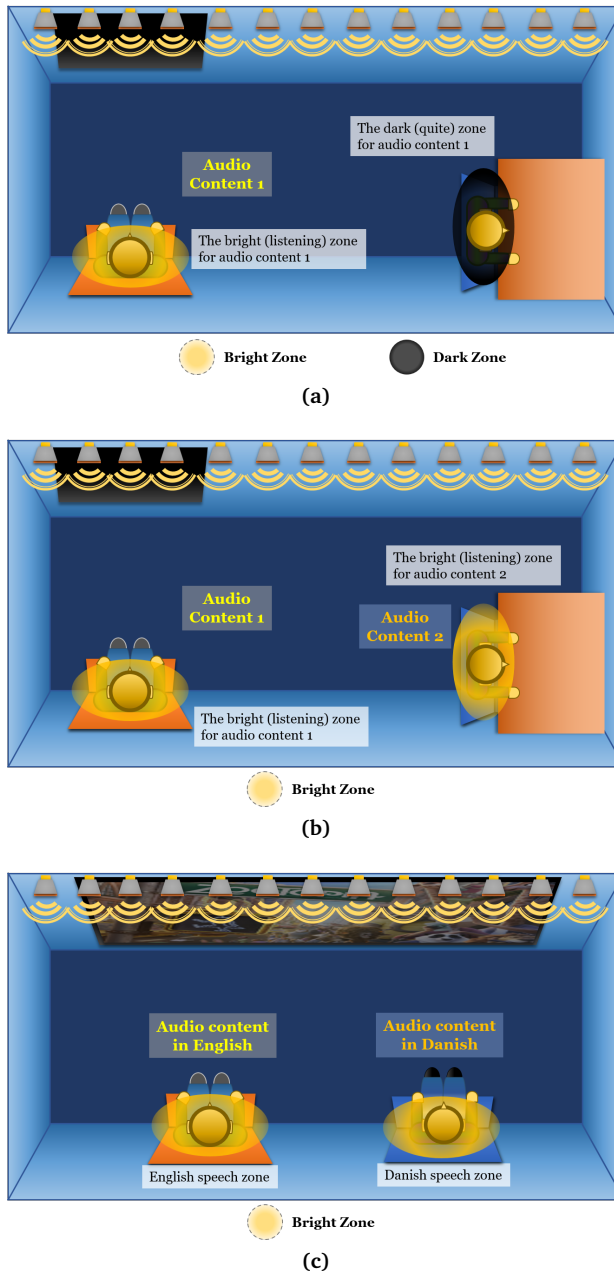


Fig. 1: Possible scenarios of personal sound system, (a) One bright zone and one dark zone, (b) Two bright zones for different audio contents, (c) Two bright zones for the same content in different languages.

deed account for the relative importance at different frequency components of the input signal so that the control effort can be frugally utilized for different frequencies, according to a specific criterion. This concept was successfully applied in perceptual audio coding in which quantization errors were perceptually masked by taking the human auditory system into account [10, 11]. Inspired by this concept in perceptual audio coding, the weighting filter can also be utilized in the context of sound zones. In this case, we no longer have quantization errors; rather, the reproduction errors in both of the bright and dark zones are the ones that should be modified, according to the human auditory system, which will be modelled as the weighting filter. To this end, in a given zone, the bright zone of the desired audio content plays a role as a masker, whereas the dark zone of other audio contents, which will be treated as interference, is considered as a maskee. One in the given zone then will hear the interference less audible, eventually and ideally, inaudible.

1.2 Structure

This thesis is in the form of a collection of papers. The first part covers the introduction of the work during the Ph.D. period, including a literature review. The second part of the thesis consists of three journal papers and four conference papers that have been published to peer-reviewed journals as part of the work.

In the following in Part I, the design and evaluation of sound zones are theoretically introduced in Section 2, including the literature review of the relevant state-of-the-art in sound zone control when necessary. Section 3 is devoted to expounding the contributions in the field of sound zone control made by the published work. The conclusion and directions for future research of the work are given in Section 4.

2 Creation of sound zones

After the invention of the loudspeaker [110, 116], the sound generated by loudspeakers is pervasive in everyday life. A single loudspeaker can reproduce an audio program in a space, but it is difficult to make any spatial impression. By using more than two loudspeakers, such an impression can be achieved. This spatial impression, e.g., a singer is moving from one location to another on the stage, was demonstrated more than 100 years ago, which delivered the performances on the stage to listeners using a binaural system, as described in [39, 57]. The stereophonic sound system (also known as *stereo*) was invented in the early 20th century [9], and stereo is still the most popular and well-known sound system, although extended sound systems were developed, e.g. 5.1 surround sound system [62] and 22.2 surround sound system [54].

In sound field reproduction⁴, typically, the desired sound field is considered,

⁴Because sound zone control can fall into the field of sound field reproduction, it is worth

2. Creation of sound zones

and the control methods compute the control filters to match the reproduced sound field to the desired sound field as closely as possible in a particular manner, which naturally falls into the field matching approaches, as mentioned in Sec. 1.1.

After the first demonstration by du Moncel in [39], reproducing the sound field of a concert hall in a telephone booth was studied by Camras in [17]. This attempt was based on Huygens's principle to recreate the concert hall impression, as shown in Fig. 2. The sound waves generated from the performance on the stage were recorded by the microphones facing outwards from volume V . The recorded signals were then played back through a set of loudspeakers inside volume V' . To this end, the listener in V' could experience the impression that one perceives in V . This concept was later theoretically expressed and denoted as *ambisonics* by Gerzon [52]. The desired sound field, assumed to be a plane wave field, was recorded by a specific type of microphone array and represented by using up to the first order of spherical harmonics expansion. When the input signals are fed into the loudspeakers, the coefficients in the reproduced sound field are calculated to match them to those in the desired sound field. Because the reproduction was on a single control point, it was later expanded to the higher order ambisonics (HOA) using high order spherical harmonics to have a higher accuracy in a wider area of reproduction rather than in a single control point, for example, as in [101, 128, 132]. Another very well-known method is the so-called *wave field synthesis* (WFS) proposed by Berkhout *et al.* in [5]. This approach is based on the Kirchhoff-Helmholtz integral, which describes the mathematical expression of Huygens's principle [101]. According to the Kirchhoff-Helmholtz integral, the pressure at any positions inside an arbitrary source-free region can be analytically expressed in case the pressure and the velocity on the surface of the region are known [68, 130]. In other words, if the surface is covered with a large number of equally spaced loudspeakers, the pressure at any positions inside the region is known. We refer the interested reader to [2, 120, 121, 136, 138] for more on Ambisonics and WFS.

The mode matching approach [101, 128, 133, 135], which is often categorized as HOA [132, 136], is also one of the well-known methods for sound field reproduction. Ambisonics and WFS are based on the physical phenomena according to the Kirchhoff-Helmholtz integral, whereas the mode matching approach is based on a numerical approach using a least-squares optimization [128, 135] or continuous loudspeaker concept [132, 133]. However, the mode matching approach can still be categorized in the field matching category because the desired sound field must be defined. In the earlier work of the mode matching approach, such as [101, 128, 132], a single zone was considered as in WFS or Ambisonics, and it was typically located at the center of the array. Later, this approach was expanded to account for multiple sound zones [133, 135]. In mode matching, any sound field is expressed by

reviewing the sound field reproduction first.

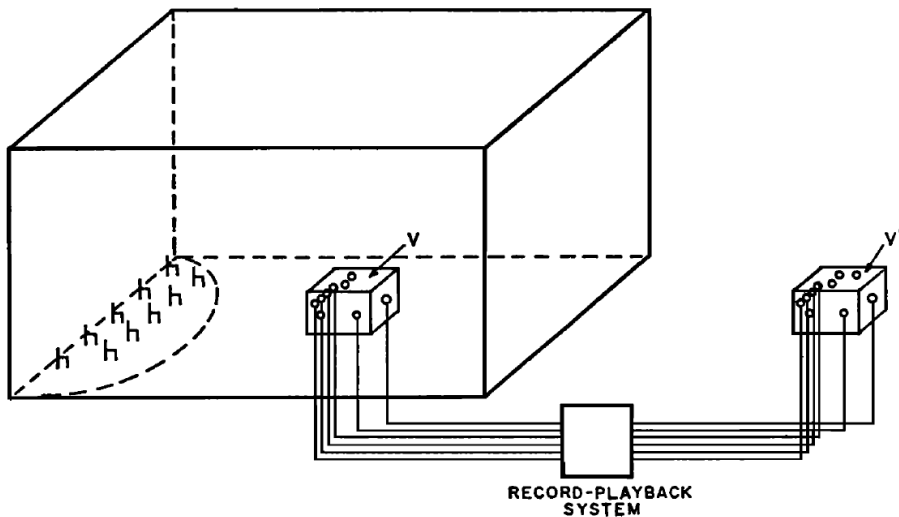


Fig. 2: Method for recreating a sound field. Note that the figure is excerpted from [17].

spatial harmonics expansion, e.g., spherical harmonics expansion. The exact reproduction of the desired sound field is possible using infinite number of coefficients of spherical harmonics, which requires an infinite number of loudspeakers. However, the truncated version of reproduced sound field is typically generated by a set of loudspeakers. The zone was normally located at the center of a circular loudspeaker array, e.g., [132], later it is located anywhere inside of the array using the so-called translation theorem in [133]. The relation between the size of the zone, the number of loudspeakers, and the frequency was theoretically revealed, for example, in [128, 132].

In the 1990s, the concept of the personal sound system was first introduced by Druyvesteyn *et al.* [37, 38]. Since then, diverse applications of the personal sound system have been studied, e.g., in car cabins [6, 18, 22–24, 26, 28, 30, 42, 79, 80, 98, 106, 119, 125, 129, 131, 134], for mobile devices [25, 43], for personal computers [20], at indoor environments [88, 95, 97, 117], and at outdoor concerts [12, 59–61]. The personal sound system is to provide spatially separated regions, i.e., sound zones, using a set of loudspeakers. Two different types of sound zones are typically considered: a bright zone and a dark zone⁵. The bright zone is defined as a control region in which the acoustic potential energy is focused or the reproduced sound field is generated as closely to the desired sound field as possible. In contrast, the dark zone is defined as a control region in which nothing is being played or the acoustic potential energy is as low as possible [29]. Note that the reproduced sound field is the field generated by the controlled loudspeakers. Because the reproduced sound fields

⁵These two zones were conceptually described in Sec. 1.1; however, these are now more precisely defined.

2. Creation of sound zones

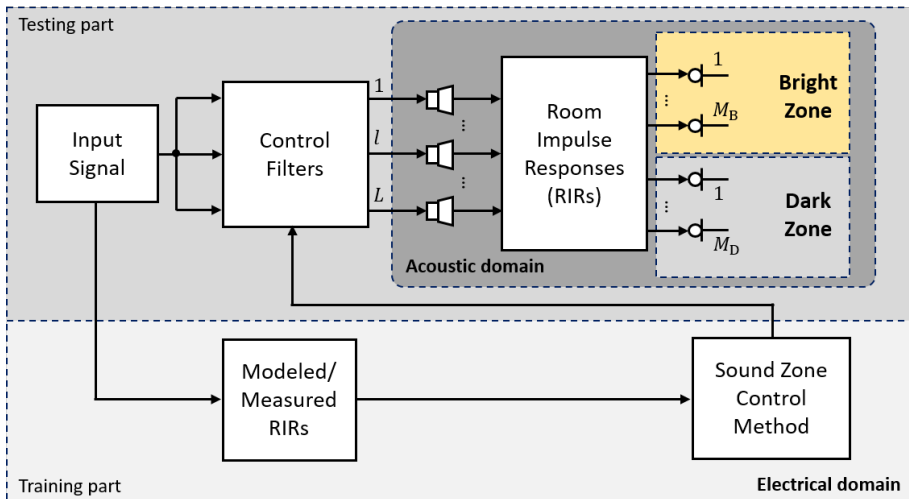


Fig. 3: A schematic block diagram of the generation of sound zones using a personal sound system.

in the sound zones are determined by several factors, e.g., room characteristics and the setup of zones as well as loudspeaker arrays, it is important to model the sound field in a mathematically sound manner to tackle the problem of creating sound zones. In the following, this modeling is more elaborated.

2.1 Sound field modeling

A schematic block diagram of a personal sound system is depicted in Fig. 3. The personal sound system aims at controlling over the sound zones in the space. To do so, as conceptually seen in the training part in Fig. 3, the control filters are first computed based on the measured or modeled RIRs. Then, as seen in the testing part in Fig. 3, the control filters convolved with the input signal are fed into the loudspeakers. Finally, the processed signals from each of the loudspeakers are delivered to and recorded at each of the sound zones. Often, the zone is discretized by spatially distributed M_B and M_D control points (or microphone positions) for the bright and dark zones, respectively. Sometimes, monitor points, typically located in between the control points, are considered to evaluate performance of such a sound zone system, e.g., in [33, 88, 126]. Throughout the thesis, subscripts $(\cdot)_B$ and $(\cdot)_D$ are used to denote the bright and dark zones, respectively, and a subscript a_C shows that quantity a belongs to either the bright zone or the dark zone, i.e., $C \in \{B, D\}$. Although no design limitation of the array is present, in practice, the locations of the loudspeakers are often fixed or given⁶, e.g., a car cabin in [26, 27].

⁶Recently, for a more realistic scenario for the loudspeaker locations, positioning loudspeakers in living rooms based on the user's choices and environmental constraints was studied [31].

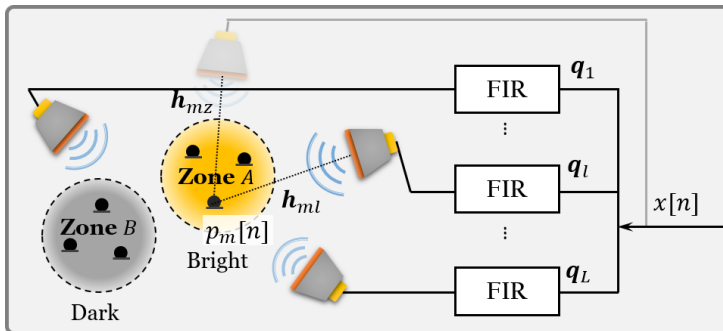


Fig. 4: An illustration of a system geometry of sound zones. Note that the figure is excerpted from [78].

As shown in Fig. 4, the reproduced sound pressure $p_m[n]$ at control point m can be modeled as linear convolution between the input signal $x[n]$ of length I , the room impulse response (RIR) $h_{ml}[n]$ of length K from loudspeaker l to control point m , and the control filter $q_l[n]$, which is a finite impulse response (FIR) filter, of length J at loudspeaker l for L loudspeakers, i.e.,

$$p_m[n] = \sum_{l=1}^L (h_{ml} * x * q_l)[n] = \sum_{l=1}^L \sum_{i=0}^{K-1} \sum_{j=0}^{J-1} h_{ml}[i] x[n-i-j] q_l[j], \quad (1)$$

and an index n and the symbol $*$ denote the discrete time sample index and the linear convolution operator, respectively. The reproduced sound pressure $p_m[n]$ at control point m will only be determined by the RIRs and the input signal when no control strategy is considered. In this case, the control filter $q_l[n]$ for loudspeaker l is simply the Kronecker delta function. The control filters will be computed in accordance with a certain design criterion, e.g., minimizing the difference between the reproduced sound pressure $p_m[n]$ and a desired sound pressure $d_m[n]$ for all control points in both the bright and dark zones. Various strategies for computing such control filters will be broadly but roughly reviewed later. Usually, different desired sound fields are defined for the bright and dark zones. A very small or zero amplitude is assigned for the desired sound field for the dark zone, whereas part of a sound field that is generated by virtual source z is defined as that for the bright zone [19, 102], i.e.,

$$d_m[n] = \begin{cases} (h_{mz} * x)[n] & m \in \mathcal{M}_B \\ 0 & m \in \mathcal{M}_D \end{cases}, \quad (2)$$

where \mathcal{M}_B and \mathcal{M}_D are the index sets of control points for the bright and dark zones, respectively. The impulse response $h_{mz}[n]$ from virtual source z to control point m can be modeled as in free field, which can be interpreted

2. Creation of sound zones

as the user desires to get a spatial impression that the reproduced sound is coming from the location of the virtual source without reverberations. In this case, the sound zone control methods need to perform dereverberation as well to minimize the difference between the desired and reproduced sound fields [51, 70, 92] (see [73, 86] for a similar concept in speech dereverberation).

The sound radiating characteristics from each loudspeaker in the space (or room) to each control point can be represented by the RIR between their locations, and the RIRs can be measured prior to calculating the control filters. The measurements can be done by using a specific type of excitation signals, e.g., an exponential sine sweep [45, 46, 91, 93, 123]. Alternatively, room acoustic models [3, 35, 53, 105, 107] also can be used to obtain such RIRs.

By taking these RIRs into account, a time-domain approach can be naturally considered as a general solution to tackle such a problem of creating sound zones. In this case, such a time-domain approach indeed provides an exact and general solution for the problem, including the crucial information of the input signals, i.e., the input signal statistics. However, often the time-domain approach suffers from its high computational complexity as it tackles typically one big optimization problem. Therefore, the input signal having a flat spectrum is assumed to provide acceptable performance on average regardless of the input signal statistics, which decreases the computational complexity significantly, as in [15, 70, 108, 118]. In other words, this assumption reduces the two linear convolutions in (1) to one linear convolution only between the RIR and the control filters over L loudspeakers. However, it should be noted that array effort can be used more efficiently for a specific frequency range present in the input signal if the input signal statistics are taken into account, as briefly claimed in [70, 118].

If we further assume that the input signal is periodic or infinitely long, then the time-domain approach, which tackles one big optimization problem, could be decoupled into a number of smaller independent problems in the frequency-domain (or often referred to as the subband-domain) [68, 88, 125]. Specifically, we can represent (1) for $N \geq I + J + K - 2$ samples of $p_m[n]$ as follows:

$$\mathbf{p}_m[n] = \sum_{l=1}^L \underline{\mathbf{h}}_{ml} \underline{\mathbf{x}}[n] \mathbf{q}_l^0, \quad (3)$$

where $\underline{\mathbf{h}}_{ml} = \text{circ}\{\mathbf{h}_{ml}^0\}$, $\underline{\mathbf{x}}[n] = \text{circ}\{\mathbf{x}^0[n]\}$, and $\text{circ}\{\mathbf{a}\}$ represents a circulant matrix that is fully defined by column vector \mathbf{a} , and each column is shifted by one element down related to the previous column. A zero-padded version of vector \mathbf{a} is represented by \mathbf{a}^0 . All the variables are summarized in Table 1.

As a circular matrix $\underline{\mathbf{A}}$ can be represented by using the discrete Fourier transform (DFT) matrix $\underline{\mathbf{F}}_N$, i.e.,

$$\underline{\mathbf{A}} = \underline{\mathbf{F}}_N^{-1} \text{diag}\{\underline{\mathbf{F}}_N \mathbf{a}\} \underline{\mathbf{F}}_N,$$

Table 1: Definitions of the variables

Variable	Description
$\mathbf{p}_m[n], \mathbf{h}_{ml}, \mathbf{x}[n], \mathbf{q}_l$	Vector version of $p_m[n], h_{ml}[n], x[n], q_l[n]$
$\mathbf{h}_{ml}^0, \mathbf{x}^0[n], \mathbf{q}_l^0$	Zero-padded version of $\mathbf{h}_{ml}, \mathbf{x}[n], \mathbf{q}_l$
$\underline{\mathbf{h}}_{ml}, \underline{\mathbf{x}}[n]$	Circulant matrix version of $\mathbf{h}_{ml}^0, \mathbf{x}^0[n]$

where

$$\underline{\mathbf{F}}_N = \left\{ \exp \left(-\frac{j2\pi(c-1)(r-1)}{N} \right) \right\}_{c,r=1}^N,$$

$$\underline{\mathbf{F}}_N^{-1} = \frac{1}{N} \underline{\mathbf{F}}_N^H,$$

and c and r represent the column and row indices, respectively, (3) can be rewritten as

$$\begin{aligned} \mathbf{p}_m[n] &= \sum_{l=1}^L \underline{\mathbf{F}}_N^{-1} \text{diag}\{\underline{\mathbf{F}}_N \mathbf{h}_{ml}^0\} \underline{\mathbf{F}}_N \underline{\mathbf{F}}_N^{-1} \text{diag}\{\underline{\mathbf{F}}_N \mathbf{x}^0[n]\} \underline{\mathbf{F}}_N \mathbf{q}_l^0, \\ &= \sum_{l=1}^L \underline{\mathbf{F}}_N^{-1} \text{diag}\{\underline{\mathbf{F}}_N \mathbf{h}_{ml}^0\} \text{diag}\{\underline{\mathbf{F}}_N \mathbf{x}^0[n]\} \underline{\mathbf{F}}_N \mathbf{q}_l^0. \end{aligned} \quad (4)$$

If $\underline{\mathbf{F}}_N$ is multiplied to the left side of (4), then eventually the following expression can be obtained:

$$\underline{\mathbf{F}}_N \mathbf{p}_m[n] = \sum_{l=1}^L \text{diag}\{\underline{\mathbf{F}}_N \mathbf{h}_{ml}^0\} \text{diag}\{\underline{\mathbf{F}}_N \mathbf{x}^0[n]\} \underline{\mathbf{F}}_N \mathbf{q}_l^0. \quad (5)$$

Finally, each and every frequency bin $k \in \{0, 1, \dots, N-1\}$ can now be treated independently, i.e.,

$$\begin{aligned} P_m[k] &= \sum_{l=1}^L H_{ml}[k] X[k] Q_l[k] \\ &= X[k] \mathbf{H}_m^T[k] \mathbf{Q}[k] \end{aligned} \quad (6)$$

where

$$\mathbf{H}_m[k] = [H_{m1}[k] \quad \dots \quad H_{mL}[k]]^T \in \mathbb{C}^{L \times 1}, \quad (7)$$

$$\mathbf{Q}[k] = [Q_1[k] \quad \dots \quad Q_L[k]]^T \in \mathbb{C}^{L \times 1}, \quad (8)$$

$P_m[k], H_{ml}[k], X[k]$, and $Q_l[k]$ are the k th element of the following vectors, respectively: $\underline{\mathbf{F}}_N \mathbf{p}_m[n]$, $\underline{\mathbf{F}}_N \mathbf{h}_{ml}^0$, $\underline{\mathbf{F}}_N \mathbf{x}^0[n]$, and $\underline{\mathbf{F}}_N \mathbf{q}_l^0$. It should be noted that (6)

2. Creation of sound zones

has the same form as (3) but in the different domain, and it is only valid when the aforementioned assumptions hold. Lastly, the input signal having a flat frequency spectrum can be seen as a special case, i.e., $\mathbf{x}^T[n] = [1 \ \mathbf{0}_{1 \times (I-1)}]$. The sound pressure distribution in a single zone $\{P_m[k]\}_{m=1}^{M_C}$ can then be represented by

$$\mathbf{P}_C[k] = [P_m[k] \ \cdots \ P_{M_C}[k]]^T, \quad (9)$$

$$= X[k] \underline{\mathbf{H}}_C[k] \mathbf{Q}[k], \quad (10)$$

where

$$\underline{\mathbf{H}}_C[k] = [\mathbf{H}_1[k] \ \cdots \ \mathbf{H}_{M_C}[k]]^T \in \mathbb{C}^{M_C \times L}. \quad (11)$$

The acoustic potential energy in each zone can be represented as

$$\begin{aligned} e_C[k] &= \sum_{m \in \mathcal{M}_C} \|\mathbf{P}_C[k]\|_2^2 \\ &= \mathbf{Q}^H[k] \underline{\mathbf{R}}_C[k] \mathbf{Q}[k], \end{aligned} \quad (12)$$

where

$$\underline{\mathbf{R}}_C[k] = \underline{\mathbf{H}}_C^H[k] \underline{\mathbf{H}}_C[k] \in \mathbb{C}^{L \times L}, \quad (13)$$

which is often called the spatial correlation matrix. The desired sound pressure and distribution in the frequency-domain $D_m[k]$ and $\mathbf{D}_C[k]$ also can be represented in the same manner, i.e.,

$$D_m[k] = X[k] H_{mz}[k], \quad (14)$$

$$\mathbf{D}_C[k] = [D_1[k] \ \cdots \ D_{M_C}[k]]^T, \quad (15)$$

where $H_{mz}[k]$ represents the transfer function between virtual source z and control point m . The location of the virtual source is highly user- and/or scenario-dependent.

To design the optimal control filters, a mean squared error (MSE) criterion is considered and defined for the bright and dark zones, respectively, i.e.,

$$\begin{aligned} \mathcal{S}_B[k] &= \|\mathbf{D}_B[k] - \mathbf{P}_B[k]\|_2^2 \\ &= |X[k]|^2 \left(\mathbf{Q}^H[k] \underline{\mathbf{R}}_B[k] \mathbf{Q}[k] - 2\mathbf{Q}^H[k] \underline{\mathbf{H}}_B[k] \mathbf{H}_z[k] + \|\mathbf{H}_z[k]\|_2^2 \right), \end{aligned} \quad (16)$$

$$\begin{aligned} \mathcal{S}_D[k] &= \|\mathbf{0}_{M_D} - \mathbf{P}_D[k]\|_2^2 \\ &= |X[k]|^2 \mathbf{Q}^H[k] \underline{\mathbf{R}}_D[k] \mathbf{Q}[k], \end{aligned} \quad (17)$$

where $\mathbf{0}_{M_D}$ is an all-zeros vector of length M_D , and the optimization strategies using these MSE criteria are reviewed in the following section.

Typically, designing control filters in the frequency-domain, as expressed in (6), is commonly used, for example, [19, 29, 32, 90, 102] and the references therein, the following sections will be explained based on the frequency-domain approaches⁷.

⁷The frequency bin index k is omitted for brevity, unless otherwise specified.

2.2 Sound zone control methods

In the following, the state-of-the-art methods for sound zone control are briefly reviewed. As alluded to earlier in Sec. 1, the methods are largely categorized into two: energy-based approaches and field-matching approaches.

Energy-based approach: acoustic contrast control

First, the control filters can be computed using energy-based methods. Choi and Kim introduced the concept of an acoustically bright zone and an acoustically dark zone [29, 68], and they proposed the method known as acoustic contrast control (ACC) that seeks the control filters to maximize the acoustic contrast γ_{AC} , i.e., the acoustic potential energy ratio between the bright and dark zones, which is defined as

$$\gamma_{AC} = \frac{e_B}{e_D} . \quad (18)$$

The corresponding constrained optimization problem, which often referred to as a direct acoustic contrast formulation [41], can be written as

$$\text{minimize } e_B \quad \text{subject to } e_D \leq \epsilon , \quad (19)$$

where e_B and e_D denote the acoustic potential energy in the bright and dark zones, respectively, and ϵ is the acoustic potential energy targeted in the dark zone. It should be mentioned that e_D is often called the residual energy. The solution of (19) is the eigenvector corresponding to the largest eigenvalue of the generalized eigenvalue problem [29], i.e.,

$$\mathbf{R}_D^{-1} \mathbf{R}_B \mathbf{U} = \mathbf{U} \mathbf{\Lambda} , \quad (20)$$

where $\mathbf{\Lambda}$ is a diagonal matrix whose diagonal elements are the eigenvalues sorted in descending order, and \mathbf{U} is a nonsingular matrix whose columns are the eigenvectors sorted in the same order as the corresponding eigenvalues on the diagonal of $\mathbf{\Lambda}$; therefore, often the first column in \mathbf{U} is chosen to be the solution, which is the ACC control filter. Alternatively, the solution can also be obtained by minimizing the residual energy with a constraint on the energy in the bright zone, and this approach is often called as an indirect acoustic contrast formulation, as described in [41]. It should be noted that the solution maximizing the energy difference (not the acoustic contrast) between the bright and dark zones [115] is the same solution as the indirect approach, as explained in [41].

The sound zone control methods in the frequency-domain compute the solution for a single frequency bin at a time, then the time-domain control filter is obtained after applying the inverse Fourier transform to the control filter in the frequency-domain. Because ACC only exploits the energy information, which does not take phase information into account, the spatial distribution

2. Creation of sound zones

of the reproduced sound field is often difficult to control, for example, as reported in [64]. Later, the direction of a wave propagation is considered in the planarity control method [32–34].

As can be seen from (13) and (20), the control filters are computed mainly based on the geometrical information of the personal sound system, which includes the locations of the loudspeaker array and the zones. Once the control filters are given, the input signals convolved with such the control filters will be fed into the loudspeakers, and the reproduced signals are emitting through the space to be delivered to the people inside the bright zones. Often, the RIRs or the transfer functions are assumed to be known by modelling or measured in advance to compute the control filters offline, which implies that such information is treated as time-invariant or controllable. However, as the modelling errors or the RIR measurement errors are inevitable in practice, the robustness against such errors were well studied. In [99, 100], the relation between the RIR measurement errors and the performance degradation of acoustic contrast were discussed. In addition to this, the influence to the performance of acoustic contrast due to the environmental change, e.g., temperature, wind speed, humidity, was studied in [58, 98]. Besides, the influence of the background noise on the RIRs to the acoustic contrast was also studied in [87].

Often, the generalized eigenvalue problems become ill-posed if the fields generated by each and every loudspeaker are similar to each other. As the eigenvalues represent the acoustic contrast, the more the spatial-correlation matrix of the dark zone is close to singular, the higher the acoustic contrast is. However, in this case, a massive amount of input power is required to the reproduced sound fields as well as such the acoustic contrast [33, 137]; hence, typically a regularization scheme, for example, the Tikhonov regularization, is added to provide robustness against such an issue [41, 68]. It should be noted that the array effort is directly related to the Tikhonov regularization⁸. In [137], error-based robust ACC methods were studied and compared to other regularization methods.

Although ACC guarantees the maximum acoustic contrast at the frequency bins on the discrete Fourier transform (DFT) grid, it cannot guarantee such performance on other frequencies that are not on the DFT grid [15, 88]. Therefore, the time-domain ACC [15] (often referred to as the broadband ACC or BACC) as well as the sub-band approach [28] was introduced. These methods aim at maximizing acoustic contrast by solving the same optimization problem as (19) but in the time-domain. However, as reported in [15, 40, 88], the BACC method results in a control filter that filters all the frequency components except for a single frequency and its vicinity where the maximum acoustic contrast is achieved, which fulfills the objective of such an optimization problem [15]. In this case, the reproduced sound field by BACC is severely distorted; therefore, additional constraints that mitigate such distortion were investigated [14, 15, 108]. Recently, the ACC method in the wave-number do-

⁸The array effort is defined as the sum of squared control filters [41].

main to ensure the maximum acoustic contrast over the control regions, including the control points, was proposed [55, 56, 135], and the ACC method using gradient descent algorithm [67] was also studied. Other aspects of ACC, for example, nonlinear distortion [81–84], pre-ringing artefacts [89, 129], were also studied.

Field matching approach: pressure matching

Field matching approaches also provide the sound zones. In this case, the desired sound field for the bright and dark zones should be defined in advance, i.e., (2) and (15), which allows one to control both of the magnitude and phase [19, 102]. To match the desired sound field and the reproduced sound field, the aforementioned MSE criteria are minimized via a ℓ_2 -norm manner⁹, i.e.,

$$\text{minimize } (\mathcal{S}_B + \mathcal{S}_D) \quad \text{subject to} \quad \|Q\|_2^2 \leq \epsilon, \quad (21)$$

where ϵ is the allowed amount of the input power. First, the reproduction of the plane wave field was proposed [69], later the plane wave field for the bright zone and the attenuated version of the field for the dark zone were reproduced [102], which is also known as the pressure matching (PM) method. Although the above mentioned approaches ensure the minimum reproduction error, including [70, 92], this degree of control is obtained at the cost of the degradation of the acoustic contrast [64]. To improve the performance of acoustic contrast while controlling the phase or the propagation direction of the sound field, combining the energy difference maximization method [115] and the pressure matching method [102] was considered in [90]. Followed by this method, another combination, which is often referred to as ACC-PM or PM-ACC, that controls the trade-off between the signal distortion in the bright zone \mathcal{S}_B and the residual energy \mathcal{S}_D was considered originally in the frequency-domain [19] and later in the time-domain [118]. Controlling the trade-off can be done by tuning a user parameter that can weigh between the \mathcal{S}_B minimization and the \mathcal{S}_D minimization, which cannot be obtained simultaneously. Variations of the PM-ACC method were also studied. For example, designing the control filters for the application of outdoor concerts by splitting the control filters into two parts: primary sources and secondary sources. The primary sources deliver the audio contents to the bright zone, whereas the secondary sources creates the dark zones and minimize the sound pressure emitted from the secondary sources to the bright zone [58, 60]. Also, a strategy that introduces a gray zone, which is part of the dark zone but not being occupied by users, was studied in [96, 114]. This strategy allows a more flexible control over not only the bright and dark zones but also the gray zone depending on the user parameters for each and every zone. Lastly, other aspects of PM,

⁹It should be mentioned that the Wiener type solution can be obtained if both \mathcal{S}_B and \mathcal{S}_D are minimized.

for example, 1) efficient algorithms to reduce the computational complexity while preserving the similar performance to the least-squares method by assuming some of control filters are already known [124, 125] or by performing the inverse computation iteratively in the discrete Fourier transform (DFT) domain [109] and 2) pre-ringing artefacts [13, 109] were also studied.

As alluded earlier to in Sec. 2, mode matching methods [7, 8, 132, 133] also can be used for reproducing sound zones. A relation between such methods to the least-squares problem was well studied in [47, 132], and the least-squares approach can reduce to the mode matching approach [47] if the loudspeakers are uniformly distributed and the control points are uniformly distributed on the surface of the bright and dark zones.

2.3 Evaluation of sound zones

To evaluate the sound zone control methods briefly and rather broadly reviewed in the previous Section, objective measures as well as listening tests are typically exploited.

As the sound zones are reproduced by solving either maximizing the acoustic contrast in (18) or minimizing the reproduction error in (16), it is natural to evaluate such the sound zones based on acoustic contrast and/or reproduction error. It should be noted that the two measures are functions of frequency and space. A set of SPL distribution maps at the selected frequencies is typically provided over space, including the bright and dark zones. In addition to it, a set of reproduction error maps is also provided in the same manner. It seems that there is a variation on the definition of the reproduction error; hence, a normalized reproduction error with respect to the desired sound field in the bright zone is also used [19].

To evaluate a scenario having more than two bright zones, a target-to-interferer ratio (TIR) is defined, as shown in Fig. 5, and used to evaluate such the scenario. It represents the ratio of acoustic potential energy or loudness between the bright zone of the desired signal and the dark zone of the rest signals, also known as the interfering signals, in a given zone. It is reported that at least 25 dB of TIR is desirable to provide a low distraction score in which the distraction is defined as how distracting the interfering signals are to the listener who pays attention to the desired signal [50, 103, 104].

Not only the physical measures above but also well-known perceptual metrics such as PESQ (Perceptual Evaluation of Speech Quality) [63] and STOI (Short-Time Objective Intelligibility) [122] are used for evaluating the sound zones in case the speech signals are desired [14–16]. Listening tests are also used to evaluate sound zone control methods. A MUSHRA listening test was conducted to evaluate the intelligibility of the desired signal in the bright zone by a hearing impaired person and in the dark zone by normal hearing people [85]. Speech intelligibility in the bright and dark zones was also evaluated in [127]. Lastly, a multiple-bright-zone scenario was evaluated by the TIR and comprehensive listening tests were performed [4, 50, 103, 104].

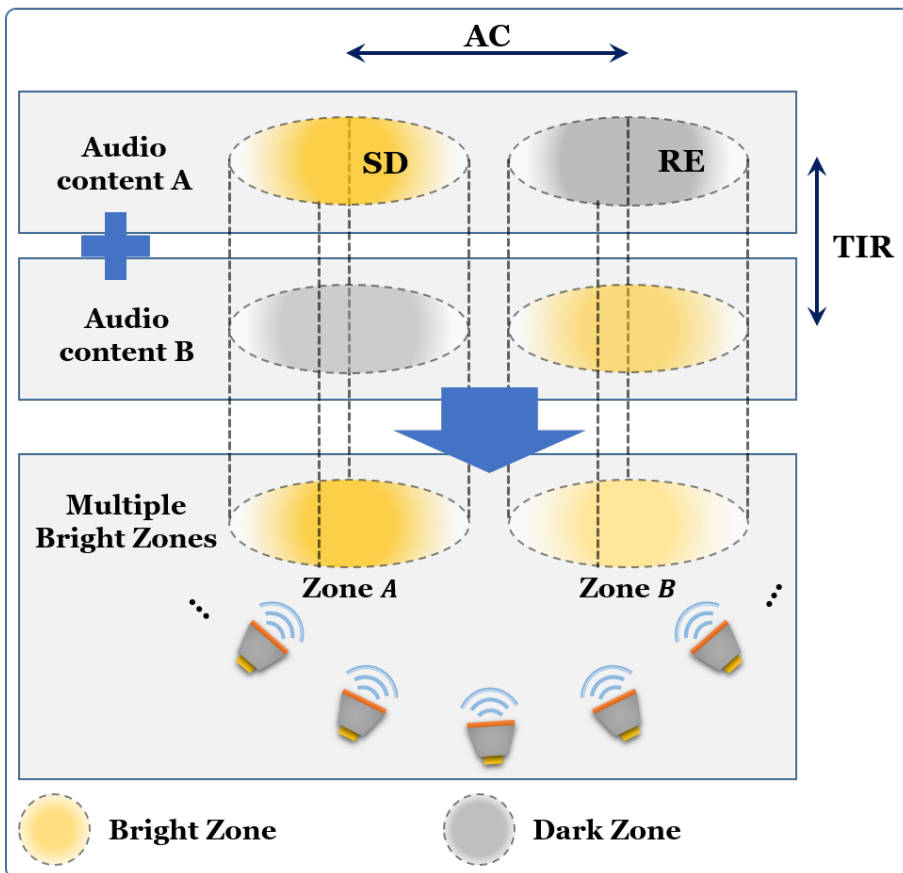


Fig. 5: An illustration of performance metrics: AC (acoustic contrast), SD (signal distortion), RE (Residual energy), and TIR (Target-to-Interferer Ratio).

3 Contributions

As alluded to in the previous section, we expound the creation of perceptually optimized sound zones by using a set of loudspeakers in this thesis. The problem of creating sound zones in a physically optimized manner is tackled to form the fundamental followed by applying perceptual weighting filters into the problem to account for the perceptually optimized sound zones.

The main body of this thesis consists of **Papers A to G**. It should be mentioned that the relation among the published papers and the key contributions are summarized in Fig. 6. In **Paper A**, we proposed the unified framework, the so-called variable span trade-off (VAST), in the time-domain, which establishes the theoretical foundation. Because the VAST framework was inspired by a subspace-based approach in speech enhancement in [44, 65, 66, 111],

3. Contributions

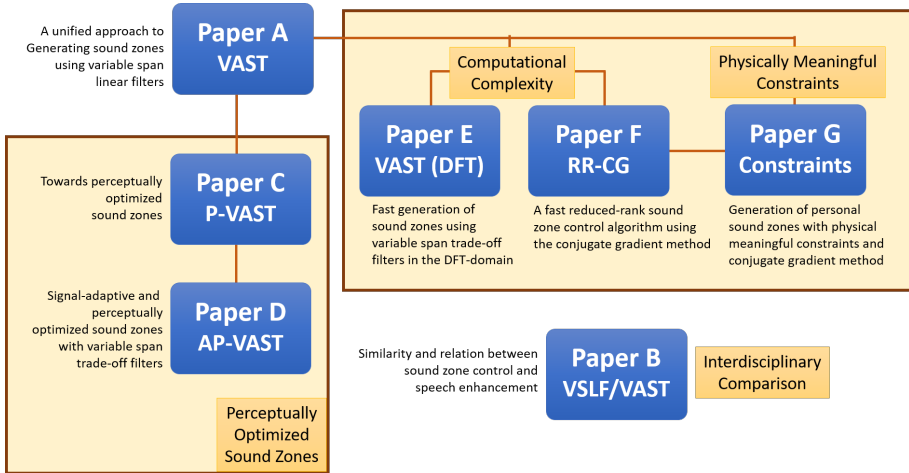


Fig. 6: Contributions and relationship between the papers.

we investigated on the relation between speech enhancement and sound zone control in **Paper B**. In **Papers C to D**, the input signal statistics and the human auditory system are accounted for creating perceptually optimized sound zones. The proof-of-concept study and an informal AB preference test were performed in **Paper C**. As an extension of this work, which is a generalized version of the work in **Paper C**, a method to create signal-adaptive and perceptually optimized sound zones was finally proposed in **Paper D**. On top of the method, more elaborated and thorough investigation in regard to the behavior of the proposed method was conducted, and a MUSHRA (Multiple Stimuli with Hidden Reference and Anchor) test to compare the proposed method with existing methods was performed. On top of it, the results were also thoroughly analyzed. As claimed in **Paper A**, **Paper C**, and **Paper D**, the time-domain approaches give the general and exact solution to generate sound zones across frequencies; however, it suffers from high computational complexity, which causes a substantial processing time. Therefore, a fast implementation in the DFT-domain to reduce the computational complexity was investigated and compared it with the time-domain approaches in terms of acoustic contrast and signal distortion in **Paper E**. Moreover, another subspace-based approach by using the conjugate gradient method to reduce the computational complexity was also proposed in **Paper F**, and this method was further extended in **Paper G**. The detailed descriptions regarding the contributions of each paper are as follows.

Paper A [77]: Time-domain framework for creating sound zones (general and exact) A unified framework that generates sound zones by using a GEVD-based subspace approach was proposed. This framework allows the user

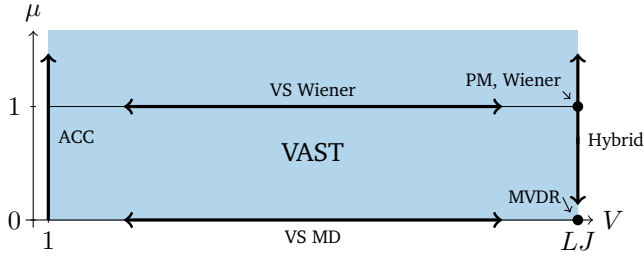


Fig. 7: An illustration that illustrates various special cases of the VAST framework depending on the two user parameters: the Lagrange multiplier μ and the subspace rank $1 \leq V \leq L \cdot J$ where L denotes the number of loudspeakers and J is the length of the control filters. Note that the figure is excerpted from [77].

to control a trade-off between acoustic contrast and signal distortion directly by adjusting the following two user parameters: the Lagrange multiplier and the subspace rank. Depending on these two parameters, the VAST framework reduces to the existing sound zone control methods, as illustrated in Fig. 7. This paper theoretically proved the relation between acoustic contrast and signal distortion, which was merely accepted empirically. The results showed that the maximum acoustic contrast only could be obtained with the largest signal distortion, and vice versa.

Paper B [94]: Relation between sound zone control and speech enhancement The VAST framework in **Paper A** was inspired by a GEVD-based subspace approach, the so-called variable span linear filters (VSLF) in speech enhancement, which controls the trade-off between signal distortion and noise reduction. In this paper, the relation in the context of optimal filtering problems between speech enhancement and sound zone control is discussed. Furthermore, we claimed that the subspace-based approach is more suitable to sound zone control as direct access to the input signals, the signal statistics, and the RIRs is available in sound zone control.

Paper C [75]: Perceptually optimized sound zones: A Proof-of-concept study The main hypothesis behind the method proposed in this paper is that integrating the input signal statistics and the mathematical expressions of the human auditory system into the VAST framework improves the perceived performance of sound zones. This method, which is referred to as perceptual VAST (P-VAST), then was evaluated objectively as well as subjectively. The results showed that the performance of physical metrics, such as acoustic contrast, signal distortion power in the bright zone, and target-to-interferer ratio (TIR) for P-VAST was not better than the existing methods, i.e., ACC and PM. Instead the perceptual metrics, i.e., STOI and PESQ, of P-VAST outperformed ACC and PM. Furthermore, an AB preference listening test supported the hypothesis because the results showed that the listeners preferred P-VAST to PM.

3. Contributions

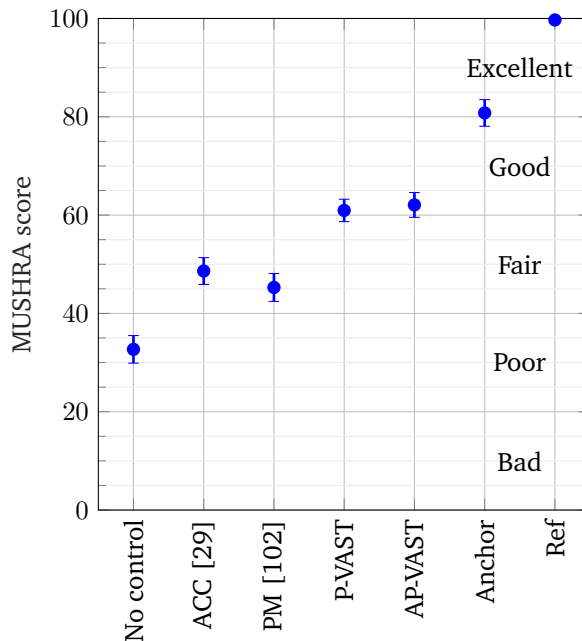


Fig. 8: The mean values and the 95% confidence intervals of all MUSHRA scores for four different methods and a hidden reference and the two anchors. In total, 1400 ratings, specifically, 200 ratings (20 participants for 10 different data sets) per method, were used. Note that the standard anchor and the hidden reference are denoted as Anchor and Ref, respectively. Note that this figure is excerpted from [76]

Paper D [76]: Signal-adaptive and perceptually optimized sound zones

In this paper, a method to create signal-adaptive and perceptually optimized sound zones was proposed as an extension of the previous work in **Paper A**, which is non-perceptual and non-adaptive, and **Paper C**, which is perceptual but non-adaptive. In **Paper D**, the control filters are computed by taking the input signal statistics and the human auditory system into account on a segment level. This integration is inspired by perceptual audio coding, which shapes quantization errors perceptually so that the signal could be compressed without causing noticeable perceived difference. A formal MUSHRA listening test was conducted for the two-bright-zone case in five different scenarios. The MUSHRA score by the proposed method was at least 20% improved compared to that of existing methods¹⁰, as shown in Fig. 8.

Paper E [78]: Fast implementation of the VAST framework Until now, **Papers A to D** deal with the problem of creating perceptually optimized sound zones using the GEVD-based subspace approach in the time-domain. This ap-

¹⁰All the audio examples used for the listening test can be accessed through the following link: <https://tinyurl.com/apvast2020>

proach provides a general and exact solution to such a problem; however, it demands a high degree of computational complexity, as explained in **Paper A**. Therefore, fast implementation of the VAST framework in the DFT-domain (an approximate solution compared to the solution in the time-domain) is proposed in **Paper E**. In this paper, we proposed a VAST frameworks in the frequency-domain to reduce the computational complexity from $\mathcal{O}(L^3 J^3)$ to $\mathcal{O}(L^3 J)$ where L is the number of loudspeakers and J is the length of the control filters by solving the same problem described in **Paper A** in an approximated manner. Two types of the framework in the frequency-domain were proposed: a narrowband VAST framework and a broadband VAST framework. The narrowband VAST framework, which is a typical frequency-domain approach, solves the problem of creating sound zones for each frequency bin in the frequency of interest. In contrast to the narrowband approach, the broadband VAST framework solves all the individual problems as one general problem while having the same order of the computational complexity as that of the narrowband approach, i.e., $\mathcal{O}(L^3 J)$. The broadband approach minimizes the overall MSE across the frequency of interest, whereas the narrowband approach minimizes the MSE at each frequency bin. These two approaches are compared to the existing sound zone control methods, including the VAST framework in **Paper A**.

Paper F [113]: Fast algorithm for generating sound zones Reducing the computational complexity of the VAST framework proposed in **Paper A** also can be done in the time-domain using the conjugate gradient (CG) method. We proposed an algorithm to generate sound zones using the CG method, which reduces the computational complexity from from $\mathcal{O}(L^3 J^3)$ to $\mathcal{O}(V_{CG} L^2 J^2)$ where V_{CG} is the number of iterations for search directions of the CG method. In the VAST framework, the eigenvectors obtained from a GEVD constitute the basis of the solution space, whereas the search directions in the CG method form the basis for the solution space in the proposed algorithm. The results showed that the proposed algorithm in this paper was approximately seven times faster compared to the VAST framework, although a 4 – 5 dB performance degradation in terms of acoustic contrast is observed.

Paper G [112]: Physically meaningful constraints for generating sound zones The VAST framework provides flexible performance in terms of acoustic contrast and signal distortion by tuning the two user parameters: the subspace rank and the Lagrange multiplier. In **Paper G**, we present a variety of strategies to obtain an accurate and precise control across the sound zones by introducing physically meaningful constraints, including signal distortion in the bright zone (SD), acoustic contrast (AC), energy reduction in the dark zone (ER) and by reformulating the optimization problem with them. Furthermore, a modified CG method was introduced to reduce the computational complexity and to provide similar performance to the VAST framework.

4 Conclusion and directions for future research

This thesis proposed a unified framework, the so-called AP-VAST, based on a GEVD and a joint diagonalization to create perceptually optimized sound zones. In order to obtain such a framework, we first tackled solving a problem to generate physically optimized sound zones, which minimizes the signal distortion in the bright zone with a constraint on the residual energy in the dark zone. We then proved that this framework allowed the user to control the trade-off between signal distortion in the bright zone and acoustic contrast directly by adjusting the two user parameters: the Lagrange multiplier and the subspace rank. We also proved that the framework could precisely compute the upper and lower bounds of acoustic contrast and signal distortion. Moreover, we have extended this framework to provide perceptually optimized sound zones by taking the input signal statistics and the human auditory system into account. Furthermore, as this framework is formulated in the time-domain, which is the exact and general solution to create perceptually optimized sound zones, we have also proposed a fast version of such a method formulated in the DFT-domain. Apart from these, we also have investigated an accurate and precise control of the performance metrics, i.e., acoustic contrast, signal distortion, residual energy, or target-to-interferer ratio, over the bright and dark zones.

What could make perceptually optimized sound zones more practical and even commercially viable? One area of further research would be implementing the AP-VAST to precisely control the performance metrics mentioned above. In **Paper D**, we have shown that the AP-VAST outperformed the existing methods in terms of the MUSHRA listening test score. We believe that an adaptive strategy to update the two user parameters as in **Paper G**, instead of using a fixed value over time and frequency, can help improve the performance of the AP-VAST. The two user parameters could be signal-independent; hence, various user cases could also be considered to improve performance further and better understand the method. On top of these, robustness against the RIR measurement and experiments under practical scenarios has to be considered to make the AP-VAST commercially applicable. One could investigate the performance metrics for the reverberation time under the given system geometry and the environment, providing an insight into the relationship between them. Typically, the control filters should be long enough to perform dereverberation if the desired sound field corresponds to free field conditions and deal with the given reverberant environment, which causes a delay (or latency) in the reproduced sound field. Therefore, an effort could be spent on reducing such delay and computational complexity while preserving performance. Lastly, a computational complexity analysis on such a method could also be one of the research topics to reduce computational complexity further.

References

- [1] R. M. Aarts, "A comparison of some loudness measures for loudspeaker listening tests," *J. Audio Eng. Soc.*, vol. 40, no. 3, pp. 142–146, Mar. 1992.
- [2] J. Ahrens, *Analytic Methods of Sound Field Synthesis*, 1st ed. Heidelberg, Germany: Springer, Jan. 2012.
- [3] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.
- [4] K. Baykaner, P. Coleman, R. Mason, P. J. B. Jackson, J. Francombe, M. Olik, and S. Bech, "The relationship between target quality and interference in sound zone," *J. Audio Eng. Soc.*, vol. 63, no. 1/2, pp. 78–89, Jan. 2015.
- [5] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, vol. 93, no. 5, pp. 2764–2778, May 1993.
- [6] S. Berthilsson, A. Barkefors, L.-J. Brännmark, and M. Sternad, "Acoustical zone reproduction for car interiors using a mimo mse framework," in *Proc. 52nd Int. Conf. Audio Eng. Soc.: Automotive Audio*, Munich, Germany, Sep. 2012.
- [7] T. Betlehem and P. D. Teal, "A constrained optimization approach for multi-zone surround sound," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Prague, Czech Republic, May 2011, pp. 437–440.
- [8] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 81–91, Mar. 2015.
- [9] A. D. Blumlein, "Improvements in and relating to sound-transmission, sound-recording and sound reproducing systems," GB Patent 394 325, Dec., 1931.
- [10] M. Bosi and R. E. Goldberg, *Introduction to digital audio coding and standards*, 1st ed. Dordrecht, The Netherlands: Kluwer, 2003.
- [11] K. Brandenburg and T. Sporer, "NMR and masking flag: Evaluation of quality using perceptual criteria," in *Proc. 11th Int. Conf. Audio Eng. Soc.*, Portland, OR, USA, May 1992, pp. 169–179.
- [12] J. Brunskog, F. M. Heuchel, D. C. Nozal, M. Song, F. T. Agerkvist, E. Fernandez-Grande, and E. Gallo, "Full-scale outdoor concert adaptive sound field control," in *Proc. 23rd Int. Congr. Acoust.*, Aachen, Germany, Sep. 2019, pp. 1170–1177.
- [13] M. Buerger, C. Hofmann, C. Frankenbach, and W. Kellermann, "Multizone sound reproduction in reverberant environments using an iterative least-squares filter design method with a spatiotemporal weighting function," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, New Paltz, NY, USA, Oct. 2017.
- [14] Y. Cai, M. Wu, L. Liu, and J. Yang, "Time-domain acoustic contrast control design with response differential constraint in personal audio systems," *J. Acoust. Soc. Am.*, vol. 135, no. 6, pp. EL252–EL257, Jun. 2014.
- [15] Y. Cai, M. Wu, and J. Yang, "Design of a time-domain acoustic contrast control for broadband input signals in personal audio systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 341–345.
- [16] —, "Sound reproduction in personal audio systems using the least-squares approach with acoustic contrast control constraint," *J. Acoust. Soc. Am.*, vol. 135, no. 2, pp. 734–741, Feb. 2014.

References

- [17] M. Camras, "Approach to recreating a sound field," *J. Acoust. Soc. Am.*, vol. 43, no. 6, pp. 1425–1431, Jun. 1968.
- [18] J.-H. Chang and W.-H. Cho, "Evaluation of independent sound zones in a car," in *Proc. 23rd Int. Congr. Acoust.*, Aachen, Germany, Sep. 2019, pp. 5174–5181.
- [19] J.-H. Chang and F. Jacobsen, "Sound field control with a circular double-layer array of loudspeakers," *J. Acoust. Soc. Am.*, vol. 131, no. 6, pp. 4518–4525, Jun. 2012.
- [20] J.-H. Chang, C.-H. Lee, J.-Y. Park, and Y.-H. Kim, "A realization of sound focused personal audio system using acoustic contrast control," *J. Acoust. Soc. Am.*, vol. 125, no. 4, pp. 2091–2097, Apr. 2009.
- [21] J.-H. Chang, J.-Y. Park, and Y.-H. Kim, "Scattering effect on the sound focused personal audio system," *J. Acoust. Soc. Am.*, vol. 125, no. 5, p. 3060, May 2009.
- [22] J. Cheer, "Active control of the acoustic environment in an automobile cabin," Ph.D. dissertation, University of Southampton, Dec. 2012.
- [23] J. Cheer, S. Elliott, and W. Jung, "Sound field control in the automotive environment," in *Proc. 3rd Int. ATZ Automotive Acoust. Conf.*, Zurich, Switzerland, Jun. 2015.
- [24] J. Cheer and S. J. Elliott, "Design and implementation of a personal audio system in a car cabin," in *Proc. 21st Int. Congr. Acoust.*, Montreal, QC, Canada, Aug. 2013.
- [25] J. Cheer, S. J. Elliott, Y. Kim, and J.-W. Choi, "Practical implementation of personal audio in a mobile device," *J. Audio Eng. Soc.*, vol. 61, no. 5, pp. 290–300, May 2013.
- [26] J. Cheer, S. J. Elliott, and M. F. Simón-Gálvez, "Design and implementation of a car cabin personal audio system," *J. Audio Eng. Soc.*, vol. 61, no. 6, pp. 412–424, Jun. 2013.
- [27] W.-H. Cho and J.-H. Chang, "Consideration on the design of multi-zone control system in a vehicle cabin," in *Proc. 146th Conv. Audio Eng. Soc.*, Dublin, Ireland, Mar. 2019, e-Brief 494.
- [28] J.-W. Choi, "Subband optimization for acoustic contrast control," in *Proc. 22nd Int. Congr. Sound Vib.*, Florence, Italy, Jul. 2015, pp. 849–856.
- [29] J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1695–1700, Apr. 2002.
- [30] M. Christoph and M. Kronlachner, "Improvement of personal sound zones by individual delay compensation," in *Proc. Audio Eng. Soc. Int. Conf. Sound Field Control*, Guildford, UK, Jul. 2016.
- [31] C. Cieciora, R. Mason, P. Coleman, and J. Francombe, "Understanding users' choices and constraints when positioning loudspeakers in living rooms," in *Proc. 148th Conv. Audio Eng. Soc.*, May 2020, e-Brief 596.
- [32] P. Coleman, "Loudspeaker array processing for personal sound zone reproduction," Ph.D. dissertation, University of Surrey, May 2014.

References

- [33] P. Coleman, P. J. B. Jackson, M. Olik, M. B. Møller, M. Olsen, and J. A. Pedersen, “Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array,” *J. Acoust. Soc. Am.*, vol. 135, no. 4, pp. 1929–1940, Apr. 2014.
- [34] P. Coleman, P. J. B. Jackson, M. Olik, and J. A. Pedersen, “Personal audio with a planar bright zone,” *J. Acoust. Soc. Am.*, vol. 136, no. 4, pp. 1725–1735, Oct. 2014.
- [35] E. De Sena, N. Antonello, M. Moonen, and T. van Waterschoot, “On the modeling of rectangular geometries in room acoustic simulations,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 4, pp. 774–786, Apr. 2015.
- [36] J. Donley, “Reproduction of personal sound in shared environments,” Ph.D. dissertation, University of Wollongong, Jan. 2018.
- [37] W. F. Druyvesteyn, R. M. Aarts, A. Asbury, P. Gelat, and A. Ruxton, “Personal sound,” in *Proc. Inst. Acoust.*, vol. 16, no. 2, 1994, pp. 571–585.
- [38] W. F. Druyvesteyn and J. Garas, “Personal sound,” *J. Audio Eng. Soc.*, vol. 45, no. 9, pp. 685–701, Sep. 1997.
- [39] T. du Moncel, “The telephone at the Paris Opera,” *Sci. Am.*, pp. 422–423, Dec. 1881.
- [40] S. J. Elliott and J. Cheer, “Regularisation and robustness of personal audio systems,” ISVR Technical Memorandum 995, Tech. Rep., 2011.
- [41] S. J. Elliott, J. Cheer, J.-W. Choi, and Y. Kim, “Robustness and regularization of personal audio systems,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 7, pp. 2123–2133, Sep. 2012.
- [42] S. J. Elliott and M. Jones, “An active headrest for personal audio,” *J. Acoust. Soc. Am.*, vol. 119, no. 5, pp. 2702–2709, May 2006.
- [43] S. J. Elliott, H. Murfet, and K. R. Holland, “Minimally radiating arrays for mobile devices,” in *Proc. 16th Int. Congr. Sound Vib.*, Krakow, Poland, Jul. 2009, pp. 2811–2817.
- [44] Y. Ephraim and H. L. van Trees, “A signal subspace approach for speech enhancement,” *IEEE Trans. Speech Audio Process.*, vol. 3, no. 4, pp. 251–266, Jul. 1995.
- [45] A. Farina, “Simultaneous measurement of impulse response and distortion with a swept-sine technique,” in *Proc. 108th Conv. Audio Eng. Soc.*, Paris, France, Feb. 2000, Paper 5093.
- [46] —, “Advancements in impulse response measurements by sine sweeps,” in *Proc. 122nd Conv. Audio Eng. Soc.*, Vienna, Austria, May 2007, Paper 7121.
- [47] F. M. Fazi and P. A. Nelson, “A theoretical study of sound field reconstruction techniques,” in *Proc. 19th Int. Congr. Sound Vib.*, Madrid, Spain, Sep. 2015, pp. 1–6.
- [48] J. Francombe, P. Coleman, M. Olik, K. R. Baykaner, P. J. B. Jackson, R. Mason, S. Bech, M. Dewhirst, J. A. Pedersen, and M. Dewhirst, “Perceptually optimized loudspeaker selection for the creation of personal sound zones,” in *Proc. 52nd Int. Conf. Audio Eng. Soc.: Sound Field Control*, Guildford, UK, Sep. 2013.
- [49] J. Francombe, R. Mason, M. Dewhirst, and S. Bech, “Determining the threshold of acceptability for an interfering audio programme,” in *Proc. 132nd Conv. Audio Eng. Soc.*, Budapest, Hungary, Apr. 2012, pp. 1–17, Paper 8639.

References

- [50] —, “Elicitation of attributes for the evaluation of audio-on-audio interference,” *J. Acoust. Soc. Am.*, vol. 136, no. 5, pp. 2630–2641, Nov. 2014.
- [51] P.-A. Gauthier and A. Berry, “Adaptive wave field synthesis for sound field reproduction: Theory, experiments, and future perspectives,” *J. Audio Eng. Soc.*, vol. 55, no. 12, pp. 1107–1124, Dec. 2007.
- [52] M. A. Gerzon, “Periphony: With-height sound reproduction,” *J. Audio Eng. Soc.*, vol. 21, no. 1, pp. 2–10, Feb. 1973.
- [53] E. A. P. Habets, “Room impulse response generator,” Technische Universiteit Eindhoven, Tech. Rep., Sep. 2010, Ver. 2.1.20141124.
- [54] K. Hamasaki, K. Matsui, I. Sawaya, and H. Okubo, “The 22.2 multichannel sounds and its reproduction at home and personal environment,” in *Proc. 43rd Int. Conf. Audio Eng. Soc.: Audio for Wirelessly Networked Personal Devices*, Pohang, South Korea, Sep. 2011.
- [55] Z. Han, M. Wu, Q. Zhu, and J. Yang, “Two-dimensional multizone sound field reproduction using a wave-domain method,” *J. Acoust. Soc. Am.*, vol. 144, no. 3, pp. EL185–EL190, Sep. 2018.
- [56] —, “Three-dimensional wave-domain acoustic contrast control using a circular loudspeaker array,” *J. Acoust. Soc. Am.*, vol. 145, no. 6, pp. EL488–EL493, Jun. 2019.
- [57] B. F. Hertz, “100 years with stereo - The beginning,” *J. Audio Eng. Soc.*, vol. 29, no. 5, pp. 368–370, 372, May 1981.
- [58] F. M. Heuchel, D. Caviedes-Nozal, J. Brunskog, F. T. Agerkvist, and E. Fernandez-Grande, “Large-scale outdoor sound field control,” *J. Acoust. Soc. Am.*, vol. 148, no. 4, pp. 2392–2402, Oct. 2020.
- [59] F. M. Heuchel, D. C. Nozal, F. T. Agerkvist, and J. Brunskog, “Sound field control for reduction of noise from outdoor concerts,” in *Proc. 145th Conv. Audio Eng. Soc.*, New York, NY, USA, Oct. 2018, Paper 10107.
- [60] F. M. Heuchel, D. C. Nozal, J. Brunskog, E. Fernandez-Grande, and F. T. Agerkvist, “An adaptive, data driven sound field control strategy for outdoor concerts,” in *Proc. Audio Eng. Soc. Int. Conf.*, Struer, Denmark, Aug. 2017.
- [61] F. M. Heuchel, D. C. Nozal, F.-G. Efren, J. Brunskog, and F. T. Agerkvist, “Evaluation of independent sound zones in a car,” in *Proc. 23rd Int. Congr. Acoust.*, Aachen, Germany, Sep. 2019, pp. 1178–1183.
- [62] ITU-R BS.775-3, “Multichannel stereophonic sound system with and without accompanying picture,” International Telecommunication Union (ITU), Geneva, Switzerland, Aug. 2012.
- [63] ITU-T P.862, “Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” International Telecommunication Union (ITU), Geneva, Switzerland, Feb. 2001.
- [64] F. Jacobsen, M. Olsen, M. B. Møller, and F. Agerkvist, “A comparison of two strategies for generating sound zones in a room,” in *Proc. 18th Int. Congr. Sound Vib.*, Rio de Janeiro, Brazil, Jul. 2011.

References

- [65] J. R. Jensen, J. Benesty, and M. G. Christensen, “Noise reduction with optimal variable span linear filters,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 4, pp. 631–644, Apr. 2016.
- [66] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sørensen, “Reduction of broad-band noise in speech by truncated QSVD,” *IEEE Trans. Speech Audio Process.*, vol. 3, no. 6, pp. 439–448, Nov. 1995.
- [67] S. Kim and T.-S. Choi, “Design of multichannel fir filter using gradient descent optimizer for personal audio systems,” in *Proc. 148th Conv. Audio Eng. Soc.*, May 2020, Paper 10349.
- [68] Y.-H. Kim and J.-W. Choi, *Sound Visualization and Manipulation*. Singapore, Singapore: John Wiley & Sons, Aug. 2013.
- [69] O. Kirkeby and P. A. Nelson, “Reproduction of plane wave sound fields,” *J. Acoust. Soc. Am.*, vol. 94, no. 5, pp. 2992–3000, Nov. 1993.
- [70] —, “Digital filter design for inversion problems in sound reproduction,” *J. Audio Eng. Soc.*, vol. 47, no. 7/8, pp. 583–595, Jul. 1999.
- [71] O. Kirkeby, P. A. Nelson, and H. Hamada, “The ‘Stereo Dipole’: A virtual source imaging system using two closely spaced loudspeakers,” *J. Audio Eng. Soc.*, vol. 46, no. 5, pp. 387–395, May 1998.
- [72] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, “Fast deconvolution of multichannel systems using regularization,” *IEEE Trans. Speech Audio Process.*, vol. 6, no. 2, pp. 189–194, Mar. 1998.
- [73] I. Kodrasi, S. Goetze, and S. Doclo, “Regularization for partial multichannel equalization for speech dereverberation,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 9, pp. 1879–1890, Sep. 2013.
- [74] C.-H. Lee, J.-H. Chang, J. Y. Park, and Y.-H. Kim, “Personal sound system design for mobile phone, monitor, and television set; Feasibility study,” *J. Acoust. Soc. Am.*, vol. 122, no. 5, p. 3053, Nov. 2007.
- [75] T. Lee, J. K. Nielsen, and M. G. Christensen, “Towards perceptually optimized sound zones: A proof-of-concept study,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Brighton, UK, May 2019, pp. 136–140.
- [76] —, “Signal-adaptive and perceptually optimized sound zones with variable span trade-off filters,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 2412–2426, 2020.
- [77] T. Lee, J. K. Nielsen, J. R. Jensen, and M. G. Christensen, “A unified approach to generating sound zones using variable span linear filters,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Calgary, AB, Canada, Apr. 2018, pp. 491–495.
- [78] T. Lee, L. Shi, J. K. Nielsen, and M. G. Christensen, “Fast generation of sound zones using variable span trade-off filters in the DFT-domain,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 363–378, 2021.
- [79] X. Liao, J. Cheer, S. J. Elliott, and S. Zheng, “Design array of loudspeakers for personal audio system in a car cabin,” in *Proc. 23rd Int. Congr. Sound Vib.*, Athens, Greece, Jul. 2016, pp. 4434–4441.
- [80] —, “Design of a loudspeaker array for personal audio in a car cabin,” *J. Audio Eng. Soc.*, vol. 65, no. 3, pp. 226–238, Mar. 2017.

References

- [81] X. Ma, P. J. Hegarty, K. F. Jørgensen, and J. J. Larsen, “Nonlinear distortion reduction in sound zones by constraining individual loudspeaker control effort,” *J. Audio Eng. Soc.*, vol. 67, no. 9, pp. 641–654, Sep. 2019.
- [82] X. Ma, P. J. Hegarty, and J. J. Larsen, “Mitigation of nonlinear distortion in sound zone control by constraining individual loudspeaker driver amplitudes,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Calgary, AL, Canada, Apr. 2018, pp. 456–460.
- [83] X. Ma, P. J. Hegarty, J. A. Pedersen, L. G. Johansen, and J. J. Larsen, “Assessing the influence of loudspeaker driver nonlinear distortion on personal sound zones,” in *Proc. 142nd Conv. Audio Eng. Soc.*, Berlin, Germany, May 2017, Paper 9807.
- [84] X. Ma, P. J. Hegarty, J. A. Pedersen, and J. J. Larsen, “Impact of loudspeaker nonlinear distortion on personal sound zones,” *J. Acoust. Soc. Am.*, vol. 143, no. 1, pp. 51–59, Jan. 2018.
- [85] A. Marker, S. J. Elliott, M. F. Simón-Gálvez, and S. Bleeck, “Using listening tests to demonstrate the subjective performance of a superdirective TV loudspeaker array,” in *Proc. Forum Acusticum*, Krakow, Poland, Sep. 2014.
- [86] M. Miyoshi and Y. Kaneda, “Inverse filtering of room acoustics,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 2, pp. 145–152, Feb. 1988.
- [87] M. B. Møller, J. K. Nielsen, E. Fernandez-Grande, and S. K. Olesen, “On the influence of transfer function noise on sound zone control in a room,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 9, pp. 1405–1418, Sep. 2019.
- [88] M. B. Møller and M. Olsen, “Sound zones: On performance prediction of contrast control methods,” in *Proc. Audio Eng. Soc. Int. Conf. Sound Field Control*, Guildford, UK, Jul. 2016.
- [89] —, “Sound zones: On envelope shaping of FIR filters,” in *Proc. 24th Int. Congr. Sound Vib.*, London, UK, Jul. 2017, pp. 613–620.
- [90] M. B. Møller, M. Olsen, and F. Jacobsen, “A hybrid method combining synthesis of a sound field and control of acoustic contrast,” in *Proc. 132nd Conv. Audio Eng. Soc.*, Budapest, Hungary, Apr. 2012, Paper 8627.
- [91] S. Müller and P. Massarani, “Transfer-function measurement with sweeps,” *J. Audio Eng. Soc.*, vol. 49, no. 6, pp. 443–471, Jun. 2001.
- [92] P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, “Inverse filter design and equalization zones in multichannel sound reproduction,” *IEEE Trans. Speech Audio Process.*, vol. 3, no. 3, pp. 185–192, May 1995.
- [93] J. K. Nielsen, J. R. Jensen, S. H. Jensen, and M. G. Christensen, “The single- and multichannel audio recordings database (SMARD),” in *Proc. Int. Workshop Acoust. Signal Enhancement*, Juan-les-Pins, France, Sep. 2014, pp. 40–44.
- [94] J. K. Nielsen, T. Lee, J. R. Jensen, and M. G. Christensen, “Sound zones as an optimal filtering problem,” in *Proc. 52th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Oct. 2018, pp. 1075–1079.
- [95] M. Olik, J. Francombe, P. Coleman, P. J. B. Jackson, M. Olsen, M. Møller, R. Mason, and S. Bech, “A comparative performance study of sound zoning methods in a reflective environment,” in *Proc. 52nd Int. Conf. Audio Eng. Soc.: Sound Field Control*, Guildford, UK, Sep. 2013.

References

- [96] F. Olivieri, F. M. Fazi, S. Fontana, D. Menzies, and P. A. Nelson, "Generation of private sound with a circular loudspeaker array and the weighted pressure matching method," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 8, pp. 1579–1591, Aug. 2017.
- [97] F. Olivieri, M. Shin, F. M. Fazi, P. A. Nelson, and P. Otto, "Loudspeaker array processing for multi-zone audio reproduction based on analytical and measured electroacoustical transfer functions," in *Proc. 52nd Int. Conf. Audio Eng. Soc.: Sound Field Control*, Guildford, UK, Sep. 2013.
- [98] M. Olsen and M. B. Møller, "Sound zones: On the effect of ambient temperature variations in feed-forward systems," in *Proc. 142nd Conv. Audio Eng. Soc.*, Berlin, Germany, May 2017, Paper 9806.
- [99] J.-Y. Park, J.-W. Choi, and Y.-H. Kim, "Acoustic contrast sensitivity to transfer function errors in the design of a personal audio system," *J. Acoust. Soc. Am.*, vol. 134, no. 1, pp. EL112–EL118, Jul. 2013.
- [100] J.-Y. Park, M.-H. Song, J.-H. Chang, and Y.-H. Kim, "Performance degradation due to transfer function errors in acoustic brightness and contrast control: Sensitivity analysis," in *Proc. 20th Int. Congr. Acoust.*, Sydney, Australia, Aug. 2010.
- [101] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Eng. Soc.*, vol. 53, no. 11, pp. 1004–1025, Nov. 2005.
- [102] —, "An investigation of 2D multizone surround sound systems," in *Proc. 125th Conv. Audio Eng. Soc.*, San Francisco, CA, USA, Oct. 2008, Paper 7551.
- [103] J. Rämö, L. Christensen, S. Bech, and S. Jensen, "Validating a perceptual distraction model using a personal two-zone sound system," in *Proc. Meet. Acoust.*, vol. 30, no. 1, Boston, MA, USA, Jun. 2017, p. 050003.
- [104] J. Rämö, S. Marsh, S. Bech, R. Mason, and S. H. Jensen, "Validation of a perceptual distraction model in a complex personal sound zone system," in *Proc. 141st Conv. Audio Eng. Soc.*, Los Angeles, CA, USA, Sep. 2016, Paper 9665.
- [105] J. H. Rindel, "Room acoustic prediction modelling," in *Proc. Sociedade Brasileira de Acústica*, Salvador, Brazil, May 2010.
- [106] P. N. Samarasinghe, W. Zhang, and T. D. Abhayapala, "Recent advances in active noise control inside automobile cabins: Toward quieter cars," *IEEE Signal Process. Mag.*, vol. 33, no. 6, pp. 61–73, Nov. 2016.
- [107] L. Savioja and U. P. Svensson, "Overview of geometrical room acoustic modeling techniques," *J. Acoust. Soc. Am.*, vol. 138, no. 2, pp. 708–730, Aug. 2015.
- [108] D. H. Schellekens, M. B. Møller, and M. Olsen, "Time domain acoustic contrast control implementation of sound zones for low-frequency input signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Shanghai, China, Mar. 2016, pp. 365–369.
- [109] M. Schneider and E. A. P. Habets, "Iterative DFT-domain inverse filter optimization using a weighted least-squares criterion," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 12, pp. 1957–1969, Dec. 2019.
- [110] S. E. Schoenherr. Recording technology history. (accessed: 30.04.2020). [Online]. Available: <http://www.aes-media.org/historical/html/recording.technology.history/notes.html>

References

- [111] R. Serizel, M. Moonen, B. V. Dijk, and J. Wouters, “Low-rank approximation based multichannel wiener filter algorithms for noise reduction with application in cochlear implants,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 4, pp. 785–799, Apr. 2014.
- [112] L. Shi, T. Lee, J. K. Nielsen, and M. G. Christensen, “Generation of personal sound zones with physical meaningful constraints and conjugate gradient method,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 823–837, 2021.
- [113] L. Shi, T. Lee, L. Zhang, J. K. Nielsen, and M. G. Christensen, “A fast reduced-rank sound zone control algorithm using the conjugate gradient method,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2020, pp. 436–440.
- [114] M. Shin, F. M. Fazi, P. A. Nelson, and F. C. Hirono, “Controlled sound field with a dual layer loudspeaker array,” *J. Sound Vib.*, vol. 333, no. 16, pp. 3794–3817, Aug. 2014.
- [115] M. Shin, S. Q. Lee, F. M. Fazi, P. A. Nelson, D. Kim, S. Wang, K. Park, and J. Seo, “Maximization of acoustic energy difference between two spaces,” *J. Acoust. Soc. Am.*, vol. 128, no. 1, pp. 121–131, Jul. 2010.
- [116] E. W. Siemens, “Moving-coil transducer with a circular coil of wire in magnetic field,” U.S. Patent 149 797, Apr., 1874.
- [117] M. F. Simón-Gálvez, S. J. Elliott, and J. Cheer, “The effect of reverberation on personal audio devices,” *J. Acoust. Soc. Am.*, vol. 135, no. 5, pp. 2654–2663, May 2014.
- [118] M. F. Simón Gálvez, S. J. Elliott, and J. Cheer, “Time domain optimization of filters used in a loudspeaker array for personal audio,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 11, pp. 1869–1878, Nov. 2015.
- [119] H. So and J.-W. Choi, “Subband optimization and filtering technique for practical personal audio systems,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Brighton, UK, May 2019, pp. 8494–8498.
- [120] S. Spors and J. Ahrens, “A comparison of Wave Field Synthesis and Higher-Order Ambisonics with respect to physical properties and spatial sampling,” in *Proc. 125th Conv. Audio Eng. Soc.*, San Francisco, CA, USA, Oct. 2008, Paper 7556.
- [121] S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, and F. Zotter, “Spatial sound with loudspeakers and its perception: A review of the current state,” *Proc. IEEE*, vol. 101, no. 9, pp. 1920–1938, Sep. 2013.
- [122] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, “An algorithm for intelligibility prediction of time–frequency weighted noisy speech,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 7, pp. 2125–2136, Sep. 2011.
- [123] A. Torras-Rosell, “Methods of measuring impulse responses in architectural acoustics,” Master’s thesis, Technical University of Denmark, Oct. 2009.
- [124] L. Vindrola, M. Melon, J.-C. Chamard, and B. Gazengel, “Pressure matching with forced filters for personal sound zones application,” *J. Audio Eng. Soc.*, vol. 68, no. 11, pp. 832–842, 2020.
- [125] L. Vindrola, M. Melon, J.-C. Chamard, B. Gazengel, and G. Plantier, “Personal sound zones: A comparison between frequency and time domain formulations in a transportation context,” in *Proc. 147th Conv. Audio Eng. Soc.*, New York, NY, USA, Oct. 2019, Paper 10216.

References

- [126] D. Wallace and J. Cheer, “The design of personal audio systems for speech transmission using analytical and measured responses,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Brighton, UK, May 2019, pp. 8003–8007.
- [127] —, “Design and evaluation of personal audio systems based on speech privacy constraints,” *J. Acoust. Soc. Am.*, vol. 147, no. 4, pp. 2271–2282, Apr. 2020.
- [128] D. B. Ward and T. D. Abhayapala, “Reproduction of a plane-wave sound field using an array of loudspeakers,” *IEEE Trans. Speech Audio Process.*, vol. 9, no. 6, pp. 697–707, Sep. 2001.
- [129] S. Widmark, “Causal MSE-optimal filters for personal audio subject to constrained contrast,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 5, pp. 972–987, May 2019.
- [130] E. G. Williams, *Fourier acoustics: Sound radiation and nearfield acoustical holography*. Academic Press, 1999.
- [131] Y. J. Wu, “Spatial soundfield reproduction in complex environments,” Ph.D. dissertation, The Australian National University, Oct. 2010.
- [132] Y. J. Wu and T. D. Abhayapala, “Theory and design of soundfield reproduction using continuous loudspeaker concept,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 17, no. 1, pp. 107–116, Jan. 2009.
- [133] —, “Spatial multizone soundfield reproduction: Theory and design,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 6, pp. 1711–1720, 2011.
- [134] N. Yanagidate, J. Cheer, S. Elliott, and T. Toi, “Car cabin personal audio: Acoustic contrast with limited sound differences,” in *Proc. 55th Int. Conf. Audio Eng. Soc.*, Helsinki, Finland, Aug. 2014.
- [135] W. Zhang, T. D. Abhayapala, T. Betlehem, and F. M. Fazi, “Analysis and control of multi-zone sound field reproduction using modal-domain approach,” *J. Acoust. Soc. Am.*, vol. 140, no. 3, pp. 2134–2144, Sep. 2016.
- [136] W. Zhang, P. Samarasinghe, H. Chen, and T. D. Abhayapala, “Surround by sound: A review of spatial audio recording and reproduction,” *Appl. Sci.*, vol. 7, no. 6, May 2017, Art. no. 532.
- [137] Q. Zhu, P. Coleman, M. Wu, and J. Yang, “Robust acoustic contrast control with reduced in-situ measurement by acoustic modeling,” *J. Audio Eng. Soc.*, vol. 65, no. 6, pp. 460–473, Jun. 2017.
- [138] F. Zotter and M. Frank, *Ambisonics*, 1st ed. Cham, Switzerland: Springer, May 2019.