Sound Zone Control inside Spatially Confined Regions in Acoustic Enclosures

Møller, Martin Bo

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

SOUND ZONE CONTROL INSIDE SPATIALLY CONFINED REGIONS IN ACOUSTIC ENCLOSURES

BY
MARTIN BO MØLLER

DISSERTATION SUBMITTED 2019

AALBORG UNIVERSITY
DENMARK

Sound Zone Control inside Spatially Confined Regions in Acoustic Enclosures

PhD Dissertation

Martin Bo Møller

Dissertation submitted September 30, 2019

Dissertation submitted:	September 30, 2019
PhD supervisor:	Prof. Jan Østergaard Aalborg University
Assistant PhD supervisor:	Assoc. Prof. Efren Fernandez-Grande Technical University of Denmark
Industrial PhD Supervisor:	Jesper Kjær Nielsen Bang & Olufsen a/s
Former PhD Supervisor:	Assoc. Prof. Søren Krarup Olesen Aalborg University
Former Industrial PhD Supervisor:	Jan Abildgaard Pedersen formerly at Bang & Olufsen a/s
PhD committee:	Professor Rodrigo Ordoñez (chairman) Aalborg University Professor Dr.-Ing. Sascha Spors University of Rostock Assoc Associate Professor Jordan Cheer University of Southampton
PhD Series:	Technical Faculty of IT and Design, Aalborg University
Department:	Department of Electronic Systems

ISSN (online): 2446-1628
ISBN (online): 978-87-7210-512-3

Published by:
Aalborg University Press
Langagervej 2
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2019

Curriculum Vitae

Martin Bo Møller

Martin Bo Møller is currently a research specialist at Bang & Olufsen a/s, Denmark. He holds a BSc in electrical engineering and an MSc in engineering acoustics, both from the Technical University of Denmark.

Abstract

The central topic of this thesis is sound zones. The concept of sound zones relates to a scenario where multiple people would like to listen to individual playback signals in the same acoustic enclosure e.g. a car cabin or a domestic room. It is assumed that the listeners do not want to wear headphones due to discomfort associated with prolonged use or safety considerations e.g. while driving. The proposed solution, sound zones, is based on controlling the resulting sound field produced by a number of loudspeakers. Control methods are applied to reproduce the desired playback signals in different spatial regions of the room with minimal leakage of sound between the zones.

Results from perceptual listening tests, involving a target and an interfering playback signal, indicate that the interfering sound should be significantly reduced in order for the scenario to be acceptable. Very high separation is possible in simulations. However, simulation studies tend to predict much higher separation compared to experimental results. The focus of the presented work is to experimentally investigate the topic of sound zones and determine factors which limit the acoustical separation.

Two prominent factors were investigated as part of the work: The impact of measurement noise and the influence of the reproduction scenario changing after being measured e.g. due to variations in the ambient temperature. Experimental results revealed that high acoustical separation is associated with high sensitivity to both measurement noise and changes in the reproduction system. To alleviate the measurement noise, a method was proposed for estimating the measurement uncertainty and automatically adjusting the control method. The time-dependent change of the reproduction scenario was accommodated by adopting a method from control theory.

Overall, the work described in this thesis contributes to 1) understanding the limitations of the creation of sound zones in practical settings, and 2) introducing a control method for sound zones which can adapt to changes in the reproduction scenario.

The thesis consists of two parts. In the first the background and motivation for the work is presented. The second part contains seven scientific papers which describe the conducted research in detail.

Resumé

Det centrale emne for denne afhandling er lydzoner. Emnet lydzoner omhandler et scenarie, hvor adskillige personer i samme rum (fx en bilkabine eller en dagligstue) ønsker at lytte til forskellige kildematerialer. Det antages, at anvendelse af hovedtelefoner ikke er ønskværdigt grundet enten ubehag ved at bære dem gennem længere perioder eller af sikkerhedsmæssige årsager, eksempelvis under bilkørsel. Den foreslåede løsning, lydzoner, realiseres ved at styre det resulterende lydfelt udsendt af et antal højttalere. Dermed bliver det muliggjort at afspille det ønskede kildemateriale i specifikke områder af rummet med minimal akustisk lækage mellem zonerne. Resultater fra perceptuelle studier viser, at oplevelsen af at blive forstyrret af ét kildemateriale, mens man forsøger at lytte til et andet, kræver kraftig reduktion af den uønskede lyd, før situationen vurderes som acceptabel. Dæmpning af lyd i denne størrelsesorden kan opnås i simuleringer. Der forudsiges dog typisk langt højere akustisk separation mellem lydzonerne i simuleringsstudierne sammenlignet med eksperimentielle resultater. Hovedformålet med det præsenterede arbejde er at anvende eksperimentielle opstillinger til at undersøge hvilke faktorer, der begrænser separationen.

To primære faktorer blev undersøgt i dette arbejde: indflydelsen af målestøj og effekten af ændringer i reproduktionsscenariet fx grundet temperaturvariationer i omgivelserne. De eksperimentielle resultater viste, at høj akustisk separation er forbundet med en høj følsomhed over for både støj og ændringer i rummet samt højttalerne. En metode til at begrænse indflydelsen af målestøj blev foreslået. Fremgangsmåden er baseret på at estimere usikkerheden i målingerne og automatisk inkludere denne i kontrolmetoden. For at adaptere lydzonekontrollen til ændringer i reproduktionsscenariet, blev en metode fra reguleringsteknik modificeret hertil.

Generelt set bidrager denne afhandling til 1) forståelsen af hvilke faktorer, der begrænser muligheden for at skabe lydzoner og 2) at introducere en kontrolmetode, der kan adaptere til ændringer i reproduktionscenariet.

Afhandlingen består af to dele. I den første del introduceres baggrunden og motivationen for det udførte arbejde. Den anden del består af syv artikler, der i detaljer beskriver det videnskabelige arbejde.

Contents

Preface

This thesis was submitted to the Doctoral School of IT and Design at Aalborg University in partial fulfillment of the requirements for the degree Doctor of Philosophy (PhD). The work was carried out over the period of October 1^{st} *2013 to September* 30^{th} *2019 on part-time (50% working hours). The work was carried out at Aalborg University and at Bang & Olufsen a/s, who sponsored the work. During the project, a three-month research visit was carried out with Filippo Maria Fazi in the "Virtual Acoustics and Audio Engineering" group at the Institute of Sound and Vibration Research, University of Southampton, UK.*

The topic of sound field control applied to the generation of sound zones has been on my mind for some time now. The journey started when a fellow student at the Technical University of Denmark, Martin Olsen, and I in 2010 undertook conducting a pilot study and our Masters' thesis work on the topic of sound zones in collaboration with Bang & Olufsen. During the thesis work and after my subsequent employment at Bang & Olufsen, I participated in the Perceptually Optimized Sound Zones (POSZ)[1] project (running from October 1st 2010 to April 30th 2014) was equally funded by the University of Surrey, United Kingdom, and Bang & Olufsen. POSZ was centered around four PhD projects running at Surrey, where two students were investigating perceptual attributes describing sound zones, while the other two focused on the control methods for creating sound zones. From October 1st 2013, I initialized my own sound zones related PhD project on part-time at Aalborg University, while continuing my employment at Bang & Olufsen. Six years, and many cups of coffee, later it is time to wrap up the PhD work.

In connection with this work a number of people deserves my deepest gratitude. First of all, I would like to thank Jan Abildgaard Pedersen, Søren Bech, and Bang & Olufsen for providing me with this opportunity.

My thanks to the team of supervisors: Jan Østergaard, Jan Abildgaard Pedersen, Søren Krarup Olesen, Jesper Kjær Nielsen, and Efren Fernandez-Grande. Your collected width of knowledge has been an enormous source of inspiration

[1] www.posz.org

throughout the project and continuously reminded me that there is always more to learn.

I am thankful to the colleagues I have had throughout the years. Morten Lydolf for our many late evenings tinkering with measurement procedures and equipment, Jakob Dyreby for your insights into active loudspeaker design, Jussi Rämö for our shared hours in the sound zones setup, and Neo Kaplanis for being my go-to reference for sound quality. To the Acoustic Research and Development departments at Bang & Olufsen, I am happy to have such wonderful people around me and apologize for the times where I have hoarded more than my fair share of signal and power cables for my setups.

A special thank you to Martin Olsen, for the discussions, the collaboration, and especially the coffee we have shared since entering the world of acoustics at DTU.

To Filippo Maria Fazi, I am grateful for our detailed discussions usually touching a wealth of topics in a short amount of time. Thank you for hosting me during my stay in Southampton. It was truly inspiring to visit VAAE and being a part of the group.

Throughout the years, I have had the pleasure of supervising and interacting with students, both at Bang & Olufsen and at universities. Asger, Daan, Lloyd, Mario, Kostas, Pierre, Anders, Oliver, Poul, Vincenzo, Francesc, Shu Ning, Cornelius, and all the rest, it has been my delight to be involved in your projects and try to answer your endless questions.

Finally, my deepest gratitude to my family and friends who have kept me sane throughout the years.

Martin Bo Møller
Aalborg University and Bang & Olufsen a/s, September 30, 2019

List of publications

This thesis is based on the following publications:

(A) **M. B. Møller**, M. Olsen, "Sound Zones: On Performance Prediction of Contrast Control Methods", *Proc. Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control*, Guildford, United Kingdom, July 2016.

(B) M. Olsen, **M. B. Møller**, "Sound Zones: On the Effect of Ambient Temperature Variations in Feed-forward Systems", *Proc. Audio Engineering Society Convention 145*, Berlin, Germany, May 2017.

(C) **M. B. Møller**, M. Olsen, "Sound Zones: On Envelope Shaping of FIR Filters", *Proc.* 24^{th} *International Congress on Sound and Vibration (ICSV 24)*, London, United Kingdom, July 2017.

(D) **M. B. Møller**, J. K. Nielsen, E. Fernandez-Grande, S. K. Olesen, "On the Influence of Transfer Function Noise on Low Frequency Pressure Matching for Sound Zones", *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 331–335, Sheffield, United Kingdom, July 2018.

(E) **M. B. Møller**, J. K. Nielsen, E. Fernandez-Grande, S. K. Olesen, "On the Influence of Transfer Function Noise on Sound Zone Control in a Room", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, pp. 1405–1418, vol. 27, no. 9, 2019.

(F) **M. B. Møller**, M. Olsen, "On In Situ Beamforming in an Automotive Cabin using a Planar Loudspeaker Array", *Proc.* 23^{rd} *International Congress on Acoustics*, pp. 1109–1116, Aachen, Germany, September 2019.

(G) **M. B. Møller**, J. Østergaard, "A Moving Horizon Framework for Sound Zones", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, pp. , vol. , no. , In peer-review.

Part I

Extended Summary

Sound Zones in Reflective Environments

1 Introduction

1.1 Context of the study

The focus of this dissertation is to use a sound *reproduction system*[2] to reproduce different *playback signals*[3] as sound in different *control regions*[4] of a single reflective environment (e.g. a domestic room or an automotive cabin). This application is called *sound zones*[5] and will be introduced in more detail after being related to sound reproduction in general.

Reproducing sound with a single loudspeaker creates a sound event (it radiates sound relative to the radiation capabilities of the loudspeaker and the playback signal). In sound reproduction for entertainment purposes, multiple loudspeakers are often used to create the perception of a specific sound event. An example is a stereo pair of loudspeakers used to create the perception of a sound event originating from between the loudspeakers [102]. Knowledge of the human auditory system and perception of sound is important to understand the perceived characterization of the created sound event. However, it is also possible to characterize the created sound event physically (e.g. through microphone measurements) and compare it to a target reference (the intended sound event). The work presented in this dissertation focuses exclusively on the physical aspects and limitations in creating a target sound event.

In the following, two adjacent scientific fields are briefly introduced to provide context for sound zones.

[2]Reproduction system is used to denote a set of loudspeakers with controllable input signals reproducing sound in an anechoic or reflective environment e.g. a room. The reproduction system includes both the loudspeakers and the environment.

[3]The playback signal is considered to be a monophonic signal stored in a digital format which is desired to be reproduced as sound.

[4]Control region is used to denote a spatial region where it is desired to control the sound field.

[5]Sound zones are two or more control regions in which different playback signals are desired.

1.1.1 Sound field reproduction

Creating a target sound event with a single loudspeaker is generally not possible as the loudspeaker radiation characteristics are typically different from the target event. This discrepancy can be characterized from the sound field the loudspeaker can generate relative to the characteristics of the target sound event. If the sound event is generalized to be independent of its associated playback signal, it can be characterized in terms of the radiated sound field in response to a wide band impulse. While a single loudspeaker is typically unable to match the target sound field, adding more loudspeakers increases the possibility of recreating the target. An example of a target sound field is sound being reproduced independently to each ear of a listener in a crosstalk cancellation system [3]. To cancel the undesired acoustic path to each ear of a listener requires in general at least two loudspeakers and knowledge of the acoustic path from each loudspeaker to each ear. If more loudspeakers are available, more elaborate methods for controlling sound fields can be implemented. This introduces the possibility to recreate target sound fields using methods like wave field synthesis and higher order ambisonics [107]. In both cases, the size of the region where the target sound field can be reproduced generally increases with the number of loudspeakers and decreases with increasing frequency [107].

1.1.2 Room correction

Sound radiated by a loudspeaker in a room is reflected by the room boundaries. The sound field is, therefore, dissimilar to radiation in free-field and depends on the specific loudspeaker and room. Mitigating the effects of the room when reproducing sound is known under different names e.g. room compensation, room equalization, and room correction. Many different approaches have been proposed to compensate for room effects as seen in the overview work by Cecchi et al. [21]. The general challenge in room correction is that it is not possible to perfectly invert the response of a loudspeaker in a room [81]. Therefore, the scientific field of room correction often relies on approximating the response of a loudspeaker in the room with a simplified and invertible model. The intended application for the room correction spans the areas of correcting the sound field at a single point in space [79], over correcting a region in the room [11, 104, 108], to global correction of the sound field globally in the room [22, 96].

For both wave field synthesis and higher order ambisonics, it is generally assumed that the loudspeakers are radiating sound in an anechoic environment [107]. One application of room correction is thus to reduce the room influence such that the result of these adapted methods is closer to the assumed anechoic performance [104, 108].

1.1.3 Sound zones

The goal of sound zones is to control the reproduction of sound in specific regions within a room. This directly relates to sound field reproduction and

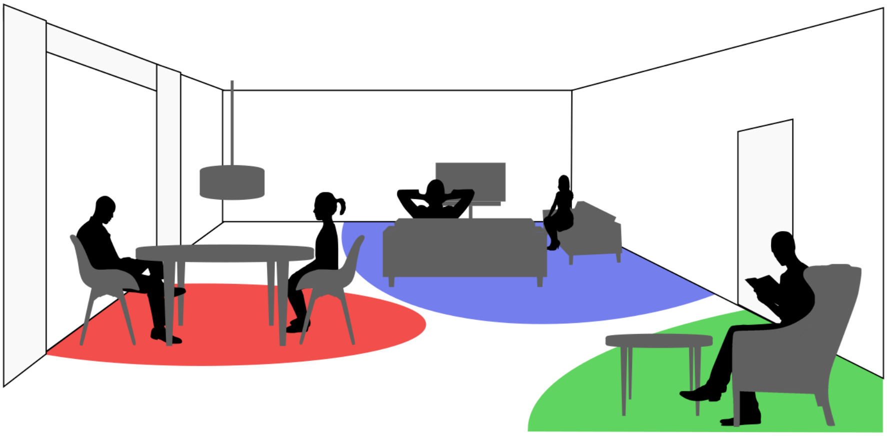


Fig. 1: Sketch of sound zones in a domestic environment.

room correction, with the important distinction that different playback signals are desired in each control region of the room.

When a loudspeaker reproduces sound in a room, the sound propagates everywhere in the room. This is not ideal if multiple people are occupied with different activities in the same room and each desires sound reproduction matching their activity. As an example, scenario, consider the scenario depicted in Fig. 1 where a group of people is gathered in a living room but are occupied by different activities. In this scenario, the activities are divided into three color-coded zones. In the red zone, a couple plays a board game at the dining table. In the blue zone, a different couple watches the news on the TV. In the green zone, the last member of the group is sitting, quietly reading a book. It is assumed that the people in the different zones desire different playback signals: in the red zone pop-music is desired as background for the game, in the blue zone the speech from the news reader should be audible, and in the green zone silence is the preferred acoustic atmosphere for being immersed in the book. In the case where a single loudspeaker is used to reproduce all the desired playback signals, a combination of all signals would be audible at every point of the room. Hence, the desired scenario of reproducing individual playback signals within specific regions of the room is not naturally occurring but requires control of multiple loudspeakers to fulfill the conflicting demands of the listeners.

1.2 Motivation and aim of the study

The interest in sound zones using active control of loudspeakers has been a topic of active research since the 90's [36]. While a wealth of literature has been published in the field, a large proportion of the work assumes loudspeakers simulated as point sources radiating sound into free-field. In relation to experimental implementations, there is a lack of investigations considering sound zones as a system designed for reproducing individual playback signals in small enclosures like automotive cabins and domestic rooms. In order to enable future field tests with such systems, it is of interest to investigate which scientific constraints that might affect these and limit their performance.

The aim of this study is to determine and evaluate factors limiting the potential for creating sound zones in reflective environments.

1.3 Scope and limitations of the study

The work presented in this dissertation is centered around specific implicit assumptions regarding the reproduction scenario. Choices have been made throughout the work relative to these assumptions regarding the experimental designs included in the published work. The premise for the conducted work is that sound zones are utilized for recreational use in domestic environments. Furthermore, parts of the work are conducted in automotive cabins which accentuates some of the challenges found in domestic rooms e.g. due to the irregular geometry of the cabin.

It should be noted that the purpose of the work is not to design a sound zones system, but to investigate the factors which would limit the performance of such a system. In the following a few additional limitations on the scope of the conducted work are introduced.

1.3.1 Reproduction environment

Given the domestic application, it is assumed that the loudspeaker positions are selected by listener (and not the designer of the sound zones system). It can therefore not be assumed that the separate loudspeakers are placed strictly following e.g. a circular geometry or close to the ears of the listeners. Consequently, it is necessary to characterize the available loudspeakers and control regions in situ. The approach adopted here is to measure the impulse responses from each loudspeaker to microphones in the control regions.

1.3.2 Number of control regions

For simplicity, two control regions are considered in the investigated scenarios. However, actual sound zones systems might include additional regions or regions of different sizes. Such scenarios would be subject to the same restrictions in terms of the number of loudspeakers relative to the total size of the regions in which the sound field should be controlled, as in the regular two zone scenario [92].

1.3.3 Spatial audio

Throughout this work, it is assumed that the playback signal is monophonic. This is to simplify the investigations. While spatial audio is feasible in sound zones, as seen in [32], the desired solution is subject to additional constraints. Reproduction of a stereo-phonic signal in a sound zone would require a sound zone solution for the left signal and another for the right signal. Such rendering scenarios would reduce the acoustical separation between sound zones and are not considered throughout this work.

1.3.4 Frequency range of concern

The goal of sound zones for entertainment purposes is to reproduce audio content in the entire audible frequency range, 20 Hz - 20 kHz. The main focus in this work is low frequency reproduction of sound zones. The reason for this is related to the desire to explicitly control the sound field in the zones without controlling the loudspeaker location. As it is known from the literature on the reproduction of plane wave fields in three-dimensions using spherical loudspeaker arrays, such control rapidly becomes prohibitive as frequency increases due to the required number of loudspeakers [111]. The interest has been to thoroughly investigate and understand the behavior and limitations of low frequency sound field control. The applicable frequency range of these methods should ideally cover a sufficiently broad frequency range to facilitate appropriate cross-over to loudspeaker beamforming techniques.

1.4 Research questions

The main motivation for the work presented in this dissertation is investigating the controllability of the sound field in a room, related to the application of sound zones. Given the chosen application, the level of control is often evaluated in terms of the acoustical separation between the playback signals reproduced for different listeners. For entertainment applications, the playback signals reproduced in the control regions should retain a minimum of sound quality. These goals can be summarized as the following two research questions:

- How much acoustical separation is attainable and what is causing the limitation?
- How can constraints on the physical reproduction system and parameters related to sound quality be introduced in sound zones methods?

1.5 Significance of the study

The work presented in this study is focusing towards implementations of sound zones. Experimental results indicate that multiple factors in combination limit the acoustical separation. The factors are all different types of mismatches between the reproduction system and the model approximating the reproduction system. The resulting performance degradation can be reduced but will in general lead to reduced performance relative to having an accurate model. The proposed solution is in part to identify the potential mismatch and include the knowledge in the control method. The other part is to apply a control framework which enables adaptation to changes and modelling physical systems. Such a framework is adopted from control theory to the generation of sound zones.

1.6 Structure of the thesis

The dissertation is structured as a collection of papers, meaning that it initially consists of an extended summary followed by a series of papers published in relation to the work. The work presented in the papers are a number of investigations with limited scope, but all of them related back to sound zones as a system. To provide the reader with an overview of the system related to the published work, the intention of the background section is to provide a broad, rather than deep, view of the literature.

In the following section 2, general considerations regarding sound zones are introduced with a brief overview of state-of-the-art in sound zones. In section 3, the contributions presented in the papers are summarized, before the results are discussed in section 4. Conclusions and considerations for future work are presented in section 5. This is followed by the second part of the dissertation which consists of the seven papers published as part of the work.

2 Background and state-of-the-art

The purpose of a sound zones system is to reproduce different playback signals in different control regions inside the reproduction environment. The general approach relies on superposition. For simplicity, it is assumed that the target is to reproduce different playback signals in two distinct regions of the same room, as depicted in the right-most solution in Fig. 2. The two control regions are denoted A and B and playback signal A and B are the desired signals in the corresponding regions. If it is possible to reproduce signal A in control region A while suppressing it in region B, and vice versa, it is possible to reproduce the desired playback signals in both control regions with minimal interference, as illustrated in Fig. 2. Typically, the control region where the playback signal is desired is termed the *bright zone*, while the *dark zone* is the region where the sound is suppressed [27]. Thus, with respect to content A, control region A is the bright zone and region B is the dark zone.

Throughout this work, the investigated scenario concerns the generation of one bright zone and one dark zone. It is assumed that a similar performance would be attainable if the definition of bright and dark zone was interchanged between the control regions.

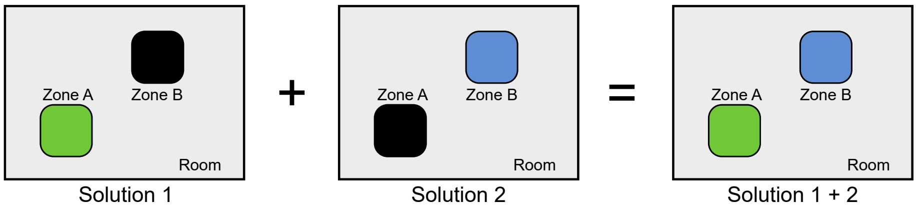


Fig. 2: Principle of superposition for sound zones. Solution 1 reproduces playback signal **A** in control region A together with silence in region B. Solution 2 reproduces playback signal **B** in region B together with silence in region A. Adding solution 1 and 2 provides playback signal **A** in region A together with signal **B** in region B.

One of the challenges related to controlling sound fields in the entire audible frequency range (20 Hz - 20 kHz) is the corresponding wavelength, which ranges from around 17 m at 20 Hz to 1.7 cm at 20 kHz. The challenge arises because sound fields in rooms change behavior between low and high frequencies. Objects like listeners can be insignificant at low frequencies and scatter the sound field at high frequencies. Furthermore, loudspeaker directivity depends on the diaphragm size of the loudspeaker driver relative to the wavelength.

The typical approach to deal with these frequency dependent changes is to apply different control strategies in combination, each covering part of the frequency range, as suggested in [36]. This composite application of control methods across the audible frequencies is conceptually illustrated in Fig. 3. At low frequencies, the sound can be controlled in the regions using *in situ measurements*, while mid frequencies can be reproduced using *beamforming* to focus the radiated sound towards the bright zone. At the highest frequencies, the inherent *directivity of loudspeakers* can be used, in combination with the expected increased absorption properties of the surfaces in the room, to focus the sound to the desired region. Correspondingly, most publications on sound zones deal with a limited frequency range where the proposed method is effective, and it is assumed that the remaining frequency range is unimportant or covered by a different method.

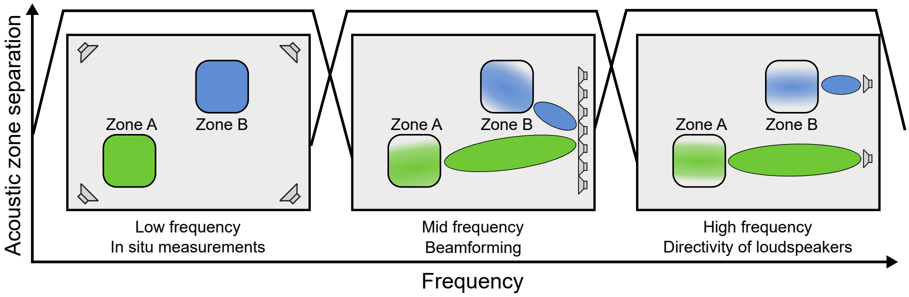


Fig. 3: Illustration of different methods for creating sound zones applied in different frequency ranges. At low frequencies the loudspeakers are distributed throughout the room and the control relies on constructive and destructive interference in the regions. At mid frequencies, beamforming is used to focus sound towards each specific region. At high frequencies, the inherent high directivity of the loudspeakers can be utilized to focus the sound without beamforming.

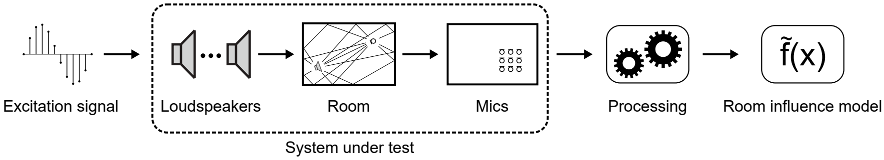
Given that the playback signals for each zone are available to the control system, sound zones are usually treated as a feed-forward problem. Creating sound zones can then be split into three stages as listed below and sketched in Fig. 4.

1. **Identification of the room influence model:** The radiation of sound from the loudspeakers to the control regions is determined, e.g. by measuring room impulse responses (RIRs). Sketched in Fig. 4a

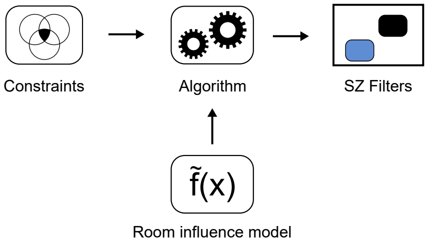
2. **Calculation of control filters:** Given the room influence model and constraints specifying the desired sound field, a digital filter is determined for each loudspeaker to pre-process the playback signal to be reproduced in the bright zone and suppressed in the dark zone. Sketched in Fig. 4b

3. **Rendering and evaluation of sound zones:** The playback signal is filtered by the determined loudspeaker control filters to reproduce the signal in situ. The effect can then be evaluated using microphone recordings or by performing listening tests. Sketched in Fig. 4c.

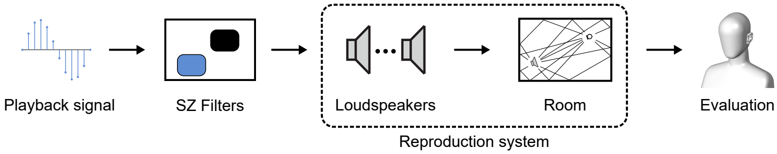
In the following, these three stages are examined in more detail to establish current results from the literature and form the background for the contributions in Sec. 3.



(a) Identification of the room influence model. An excitation signal is reproduced through each of the loudspeakers individually and the resulting sound pressure is recorded at the microphone positions in the control regions. These recordings are then post-processed to determine the room influence model as e.g. a collection of impulse responses.



(b) Calculation of control filters. The estimated room influence model is divided into parts describing a bright and dark zone and used in combination with a number of application-specific constraints in an algorithm generating a set of control filters for creating sound zones.



(c) Rendering and evaluation of sound zones. The playback signal is convolved with the determined control filters and reproduced through the loudspeakers. The performance of the sound zones system can then be evaluated either through listening tests or by measuring the resulting sound field using microphones.

Fig. 4: Diagrams of illustrating the three stages for creating sound zones with feed-forward control.

2.1 Room influence models

To control the sound field generated by multiple loudspeakers, it is necessary to know the response of each loudspeaker in the control regions. In Fig. 5 a room with four loudspeakers and two control regions is sketched. Although the sketch depicts the room in a single height plane, the control regions are in general three-dimensional enclosed regions of space. The goal is to control the sound field within the regions, but due to practical considerations, they are often sampled with microphones at discrete positions, which is also illustrated

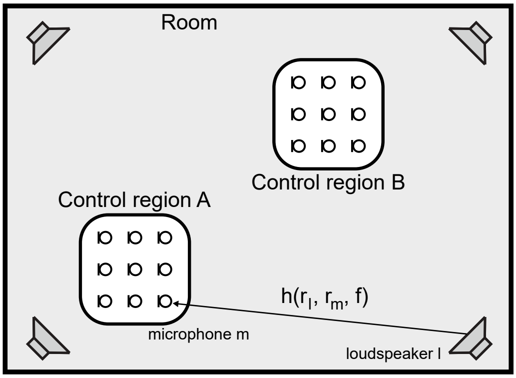


Fig. 5: Conceptual illustration of a room where it is desired to create sound zones within control region A and B. The control regions are defined as enclosed regions in the room. The transfer functions from the loudspeakers to the control regions are sampled at discrete microphone positions with the transfer function $h(\mathbf{r}_\mathrm{l}, \mathbf{r}_\mathrm{m}, f)$ from loudspeaker l to microphone m at frequency f. Here $\mathbf{r}_\mathrm{l}$ and $\mathbf{r}_\mathrm{m}$ denote the position of the loudspeaker and microphone, respectively.

in the figure.

The response from a loudspeaker to a microphone depends on multiple factors. These include loudspeaker and microphone position as well as the frequency content of the playback signal. Characterizing the sound radiated from each loudspeaker through the room to each of the control regions can be summarized in different *room influence models*. This model is used to express the sound field in the control regions which results from a specific input signal to the loudspeakers.[6] One example of a room influence model is a matrix of room transfer functions (RTFs) from each loudspeaker to each microphone in the control region at a single harmonic excitation frequency [62]. A different example is a model which consists of the impulse response from all loudspeakers to all microphones, also expressed as a matrix. This matrix has a block structure where each block is a Toeplitz matrix describing the convolution of a loudspeaker input signal with a RIR [63]. In the following, different room influence models used in the sound zones literature are introduced.

In Fig. 4a, the room influence model is depicted as being identified through measurements. It is also possible to assume a room influence model through e.g. the analytic expression for a point source radiating sound into free-field, or into a rectangular room using either Green's function [78] or the image source model [2].[7] In some articles, continuous distribution of point sources such as a continuous line or a circle is assumed. In [84, 85] such source distributions were considered in free-field, where the transfer function of the distribution is

[6]The choice of words "room influence model" is used as a general term which also includes radiation of sound under anechoic conditions.

[7]In that case, there is no identification stage to create sound zones, the room influence model is just assumed.

expressed in the wavenumber domain by applying the spatial Fourier transform. This room influence model can be used to determine control filters for a discretized loudspeaker array. Such filters are determined by dividing the transfer function of the source by the desired pressure function in the wavenumber domain and sampling the result. This approach was seen to be effective for radiating sound in specific directions using a line array of loudspeakers in an anechoic chamber in [86].

A similar room influence model, but used differently, is based on locally expanding the sound field in the control regions in basis functions such as circular or spherical harmonics. Such an approach was applied in [113] to express the target sound field in each control region by expansions around local coordinate systems for each region. The expansion coefficients were then translated to a global coordinate system corresponding to a loudspeaker array surrounding the control regions. The translation to a global coordinate system implies controlling the sound field in one large region encompassing all the smaller control regions. Thereby, the sound field is also subject to the limitations of controlling the field in the global region. In [114] the approach was modified by surrounding each control region by loudspeakers and describing the response from all loudspeakers in the local basis expansions of each control region. Solving the sound field control problem in all the local basis expansions simultaneously was seen to provide better possibility to position the control regions close to each other.

The expansion into local basis functions can also be applied to loudspeakers in a reverberant scenario. In [6, 8], it was suggested to surround the control region by a double layer microphone array and characterize the reverberant response of each loudspeaker. From image source simulations, the method was seen to compensate for the room reflections and reproduce the desired sound field in a single control region. This was applied to multiple control regions in [113]. The downside of this approach is the number of microphones required to accurately characterize the response due to each loudspeaker. In [57], the loudspeaker responses were modelled as a combination of their free-field radiation and a sparse number of plane waves. It was seen that compressive sensing can be applied to attain more accurate estimates of the transfer functions in the regions compared to the dual layer microphone array in [6].

A different room influence model designed for utilizing a low number of microphones was suggested for room correction in [12]. The model is based on several observations regarding room impulse responses from a loudspeaker to a control region: 1) The direct sound from a loudspeaker to the microphones is generally independent of the room. 2) At low frequencies the wavelength is typically larger than the control region, hence, it is possible to interpolate between the measured responses at discrete positions in the region. 3) At high frequencies the late part of the room impulse responses tends to vary unpredictably between the microphone positions. It was therefore suggested to model the direct and low frequency sound deterministically from measured room impulse responses, while the high frequency reverberation was modelled stochastically. Filters for room correction were then determined by optimizing a cost-function including the expectation over the stochastic reverberation components. In [112], sound zones were formulated in terms of a comparable room influence model which can be extended to the probabilistic model from [12].

In the above examples, it is assumed that loudspeakers reproducing sound in a room are accurately modelled by a linear time-invariant (LTI) system. This enables the response from a loudspeaker at a particular microphone to be characterized as a linear and time-invariant filter. Typically, the room impulse responses are modelled as finite impulse response (FIR) filters. One approach to estimate such a FIR filter is reproducing an exponential sweep as the excitation signal with one loudspeaker and deconvolve the recorded microphone signal with the excitation signal [41, 42]. The consequences of the reproduction system not being an LTI system are considered in relation to the rendering and evaluation of sound zones in Section 2.3.

2.1.1 Outcome

A variety of room influence models can be used to express sound field control in enclosed regions. Some of them leverage specific geometry in loudspeaker placement in e.g. circular and spherical arrays, and some offer more efficient use of a given number of microphones to characterize the sound field in the control regions.

In this work, the aim is to investigate the attainable performance in reflective reproduction environments. Reducing the required number of microphones is important for practical applications. However, it is possible to avoid the additional complexity of translating measurements into e.g. spherical harmonics or plane waves by sampling the control regions with a fine spatial resolution. The remainder of this work is, therefore, utilizing room influence models which directly contain either room transfer functions or room impulse responses.

2.2 Sound zones control methods

With an estimate of a room influence model (depicted in Fig. 4a), the next stage for creating sound zones is to determine control filters for the loudspeakers (depicted in Fig. 4b). The filters are determined by formulating and solving an optimization problem, which requires the definition of an appropriate cost-function. The cost-function should describe the desired properties of the sound field reproduced in the control regions as well as the capability of the reproduction system. In the following, common parameters to include in optimization problems for sound zones are introduced. This is followed by a brief presentation of different control methods applied in different frequency ranges. Finally, typical control frameworks are described.

2.2.1 Controlled and constrained parameters

A key point in creating sound zones is determining which aspects of the sound field to control. A main concern is the acoustic separation between the bright and dark zone, which is evaluated as the contrast. Assuming harmonic excitation, the contrast at frequency f is defined as the ratio of mean square pressure reproduced in the bright and dark zone. The contrast can then be written as

$$\mathrm{contrast}(f) = \frac{V_{\mathrm{bright}}^{-1} \int_{Z_{\mathrm{bright}}} |p(\mathbf{r}, f)|^2 d\mathbf{r}}{V_{\mathrm{dark}}^{-1} \int_{Z_{\mathrm{dark}}} |p(\mathbf{r}, f)|^2 d\mathbf{r}} \approx \frac{M_{\mathrm{bright}}^{-1} \sum_{m_{\mathrm{B}}=1}^{M_{\mathrm{bright}}} |p(\mathbf{r}_{m_{\mathrm{B}}}, f)|^2}{M_{\mathrm{dark}}^{-1} \sum_{m_{\mathrm{D}}=1}^{M_{\mathrm{dark}}} |p(\mathbf{r}_{m_{\mathrm{D}}}, f)|^2}. \quad (1)$$

In the above equation, the pressure $p(\mathbf{r}, f)$ is the complex pressure at point $\mathbf{r}$, $Z_\bullet$ and $V_\bullet$ denote the domain and volume of the zone where $\bullet$ can be substituted by either bright or dark. As mentioned in Sec. 2.1, the control regions are commonly characterized at discrete microphone positions. This is reflected in the equation by approximating the integral with a discrete sum of microphone observations.[8] The pressure at the microphone positions is denoted $p(\mathbf{r}_{m_\bullet}, f)$. As the contrast directly describes the acoustic separation between two zones, it can be used in the cost-function for determining the control filters. Maximizing the contrast at a single frequency, corresponding to the room influence model, is known as acoustic contrast control (ACC) [27]. This is typically formulated as a matrix eigenvalue problem.

Besides the mean square pressure, it is also of interest to control the accuracy at each point in the reproduced sound field p_{R}. This can be evaluated as the mean square difference between the reproduced sound field and the target sound field p_{T}. The mean square error (also called the reproduction error) is expressed as

$$\mathrm{MSE}(f) = V^{-1} \int_{Z} |p_{\mathrm{T}}(\mathbf{r}, f) - p_{\mathrm{R}}(\mathbf{r}, f)|^2 d\mathbf{r} \approx M^{-1} \sum_{m=1}^{M} |p_{\mathrm{T}}(\mathbf{r}_m, f) - p_{\mathrm{R}}(\mathbf{r}_m, f)|^2, \quad (2)$$

where Z now denotes the domain of both control regions, while V is the sum of their volumes. Again, the integral is discretized to represent that the control regions are sampled by microphones. Determining the control filter responses at frequency f by minimizing the discretized mean square error is often referred to as pressure matching (PM), and formulated as a least-squares problem [97]. The choice of target sound field is the only thing that sets PM apart from typical single region sound reproduction as in [61]. A target example for sound zones is a plane wave in the bright zone and low or even zero pressure in the dark zone.

Generally, it is reported that ACC provides a higher contrast than PM at the expense of not specifically controlling the sound field in the bright zone [31, 56].

[8] The relation between the continuous integral and the corresponding discretization is rarely addressed in the sound zones literature and it is uncommon to assign weights to the microphone observations reflecting the implied numerical integration even when the sampling grid is not regular.

This effect is illustrated in Fig. 6, where it is seen that the ACC solution is not controlled inside the bright zone. In [4], it was reported that this lack of control using ACC led to reduced sound quality (relative to PM). To combine the benefits of both ACC and PM, a number of additional methods have been proposed to provide at tradeoff between reproduction accuracy in the bright zone and contrast. One variant is a weighted version of pressure matching where different emphasis is assigned to attain the target sound field in either the bright or the dark zone [23]. Planarity control is introduced as ACC with the added constraint that sound propagation in the bright zone impinges from a limited range of angles [31, 32]. Further formulations suggest to control both the sound pressure and particle velocity in the zones [13, 14, 28]. Lately, the tradeoff between contrast and mean square error has been generalized using variable span linear filters [66, 67, 83].

To protect the loudspeakers, it is of interest to limit the allowable filter gain in the optimization problem. This is also done to reduce *self-cancellation* (a large fraction of the sound radiated by each loudspeaker is cancelled by the other loudspeakers leading to low sound pressure level in the bright zone [39]. The degree of self-cancellation can be evaluated in terms of the *array effort*. The array effort is introduced as the ratio of the sum of signal power driving the array relative to the signal power required by a reference source to produce equal sound pressure level in the bright zone. Limiting the filter gains also reduces the amount of nonlinear distortion from the loudspeakers [26, 68, 71]. Frequency dependent limitations can be introduced to match the capabilities of a given loudspeaker driver [106].

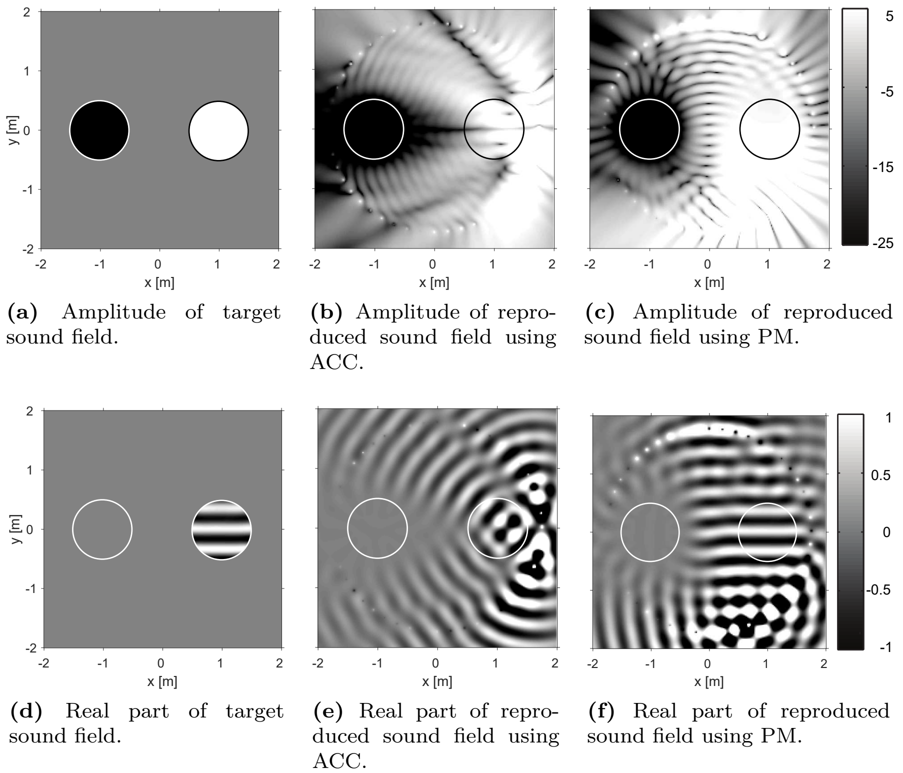
Besides only controlling physically motivated parameters, it might be of interest to include aspects in the optimization related to human perception of sound. In [45], the layout of a loudspeaker array was optimized using a model predicting distraction[9] between interfering playback signals. The results showed improved distraction ratings in a listening experiment, relative to a loudspeaker layout determined based on the contrast. In [33], sound zones for speech privacy were simulated using perceptual masking filters to predict and weight the audibility of errors in bright and dark zone. Finally, in [66] it was proposed to use masking curves to weight the effort of generating contrast across frequency relative to the desired audio content in the other zone.

Depending on the application several of these parameters should be included in the cost-function e.g. as a weighted sum. The number of parameters should be kept low to reduce the task of assigning priority to each of them. If desired bounds on the resulting parameters are know, this task can be simplified by formulating a constrained optimization problem as in [7, 112].

2.2.2 Solution strategies at different frequency ranges

The response of a loudspeaker in a room changes with frequency due to a variety of factors. The absorption of sound at the boundaries of the room change

[9]See Sec. 2.3.3 for a definition of distraction.



(a) Amplitude of target sound field.

(b) Amplitude of reproduced sound field using ACC.

(c) Amplitude of reproduced sound field using PM.

(d) Real part of target sound field.

(e) Real part of reproduced sound field using ACC.

(f) Real part of reproduced sound field using PM.

Fig. 6: Example illustrating the difference between optimizing the acoustic contrast (ACC) and the mean square error (PM) at 1000 Hz. The sound field is generated by a circular array of 48 point sources in free-field. The top row shows the amplitude in dB of the resulting sound field relative to the average level in the bright zone. The bottom row shows the real part of the sound field normalized to the average amplitude in the bright zone. The target sound field in the bright zone is a plane wave travelling parallel to the y-axis, zero pressure in the dark zone, and no constraints are placed outside the two zones.

according to the material properties at different frequencies. With decreasing wavelength, obstacles such as people and furniture tend to scatter impinging sound waves. The loudspeakers start to radiate sound directively above the frequency where the circumference of the driver is equal to the wavelength [110]. All in all, the sound field in a room becomes more complicated as the frequency increases. The change in behavior also influences the control of sound fields as discussed in [82]. As mentioned previously, it was suggested in [36] to cover the audible frequency range by utilizing different sets of loudspeakers and control methods in the low, mid, and high frequency ranges. This idea is illustrated in Fig. 3 and is adopted for the work conducted in this thesis and most of the literature on sound zones.

At low frequencies, where the dimensions of the reproduction environment are between 1/3 and 3 times the wavelength, the sound field is generally well described by an expansion of room eigenfunctions (also known as room modes and standing wave patterns) [78]. In this frequency range, the control regions are usually small relative to the wavelength. The effect is that the room influence model can be represented by a few components (either expressed as singular vectors, active room modes, or spherical harmonics) at low frequencies as discussed in [1, 11]. Consequently, it is possible to control the sound field in the regions using a low number of loudspeakers in this frequency range [1, 11]. This can be utilized to control the sound field with loudspeakers which are positioned irregularly throughout the reproduction environment. The room impulse responses can then be measured to microphone positions in each control region and used to control the sound field within the regions.

As the frequency increases and the size of the control regions become larger than the wavelength, the sound field becomes more complicated within the regions. This is reflected in the increased number of overlapping room modes in a narrow frequency band [55, 65, 78]. The increased complexity also increases the number of basis expansion terms required to accurately describe the sound field. Correspondingly, the required number of loudspeakers to accurately control the sound field in the control regions increases and rapidly becomes prohibitive for reproducing high frequency content [111]. The solution suggested in [36] and adopted in many subsequent publications [24, 29, 106] is to employ compact loudspeaker arrays to achieve the separation through beamforming. This has led to multiple studies on super directive beamforming using e.g. line arrays of loudspeakers [29, 37, 50, 89]. One approach in designing beamforming filters for a compact loudspeaker array is to formulate a least squares problem in free-field. This is done by constraining the resulting sound field at control points measured around the array. The problem is often ill-conditioned at low frequencies, which leads to excessive self-cancellation. The typical solution is to reduce the allowed array effort in the optimization.[10] Frequency dependent limits on the array effort has been used in e.g. [90] to reduce self-cancellation at low frequencies. This approach allows a tradeoff between self-cancellation and directivity across frequency.

The challenge with beamforming in rooms, is that boundaries of the room are reflective. One suggested solution is to avoid first order reflections from the array towards the control regions as proposed in e.g. [19, 51, 88]. While this might reduce the initial reflections from reaching the control regions, late reflections and reverberation will eventually reach the regions and limit the contrast. A different approach to reduce the reverberant sound is through reducing the acoustic power radiated into the room. This approach has been investigated in [48–50, 105] where the radiated power was reduced through the use of inherently directive loudspeakers (phase shift loudspeakers approximating hyper-cardioid directivity). To further reduce the radiated power, planar array configurations have been introduced to limit both vertical and horizontal

[10]Reducing the allowed array effort is equivalent to increasing Tikhonov regularization in the least squares problem [54].

directivity. In [105], it was seen that a planar array of phase shift loudspeakers was more directive and thus less influenced by the reverberant sound field in a room than a corresponding line array. This suggests that if possible, loudspeakers with inherent directivity can be a benefit to reduce the amount of acoustic power radiated in the room.

Another approach to reduce the power radiated to the reflective environments, is through loudspeakers placed close to the listeners. This approach has been proposed in transportation scenarios where loudspeakers can be included in the headrest of the seats in airplanes, trains, cars, etc. [26, 38, 58]. In [38], headrest loudspeakers were used to effectively reduce the sound leaked between adjacent airplane style seats in a small room. In [26], phase shift loudspeakers were mounted in the headrests in a car cabin. The target was to separate front and rear seats in sound zones. The control was seen to be less effective when the rear seats were selected as the bright zone. This was attributed to the directive headrest loudspeakers facing towards the front seats and thus not benefitting from the inherent directivity. This result highlights that the geometrical layout of loudspeakers and control regions significantly influences the attainable contrast.

An alternative approach for reproduction of sound locally is parametric loudspeakers. Modulating ultrasound at high sound pressure levels can produce audible sound reproduced in limited spatial regions. In [34], it is proposed to create sound zones for speech signals by combining regular loudspeaker arrays at low frequencies with parametric loudspeakers at higher frequencies. The choice of frequency range and application is based on the potential health risks and harmonic distortion associated with parametric loudspeakers, as discussed in [52]. While the directional properties of the technology are appealing, care should be taken in ensuring its suitability to a given application.

A circular piston moving in an infinite rigid baffle will start to become directive when the circumference is larger than the wavelength [110]. Thus, a typical loudspeaker driver becomes directive at high frequencies. However, a regular moving-coil loudspeaker driver only behave as moving piston at low frequencies. As the frequency increases, the loudspeaker diaphragm exhibit mechanical resonances due to bending waves in the diaphragm [10]. These vibrations will change the directivity pattern of the driver. Hence, if the application of the driver is to be directional at high frequencies, the driver design should be optimized for such an application.

As seen here, creating sound zones in reflective environments is highly dependent on the room. At low frequencies the room response can be utilized in controlling the sound field. At higher frequencies, the strategy is to avoid the effects of the room as much as possible. This is attained by careful positioning of the loudspeakers or beamforming to maximize sound pressure in the bright zone relative to the radiated acoustic power.

2.2.3 Control Systems

In Section 2.1, it was introduced that measured room impulse responses or transfer functions can be represented in different room influence models. Similarly, different control systems can be applied to generate sound zones from these models. While the control structure is generally closely related to the room influence model, it also inherently includes assumptions regarding the playback signal.

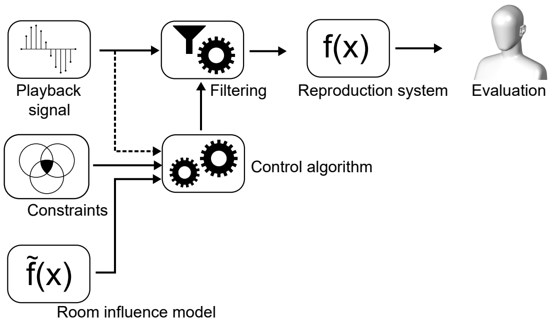
The first type of system considered is a feed-forward structure where the loudspeaker control filters are static i.e., they are time-invariant. This structure is sketched in Fig. 7a where the dashed line from the playback signal is assumed to be removed. If the playback signal is assumed to be periodic, the sound zones problem can be decomposed into a number of independent optimizations in the frequency domain [27, 58, 97]. If it is not desired to assume a periodic input signal, the feed-forward control system can be formulated in the time-domain. Here, the pressure at the microphones are modelled as linear convolutions between the loudspeaker signals[11] and the room impulse responses. These convolutions can be described in matrix form as a block matrix with Toeplitz blocks. If the playback signal is assumed to be white Gaussian noise the expected performance can be optimized. This reduces the room influence model to a linear convolution between the loudspeaker filters and the room impulse response [63, 106]. For implementation purposes it might be of interest to limit the length of the control filters to reduce the computational complexity of rendering sound zones. This is possible in the time-domain using contrast control [16–18, 103][12] or pressure matching [63, 106]. The downside of the time-domain formulation is that the dimensions of the problem quickly become very large compared to solving the independent optimizations in the frequency domain. Alternatively, the filters can be determined iteratively in the frequency domain knowing that they will be truncated in the time-domain [15].

A different formulation of the feed-forward structure arises by allowing the filters to be time-varying. This scenario is depicted in Fig. 7a by assuming the dashed line to be solid i.e. utilizing information regarding the playback signal. A variation of this approach is to update estimates of the correlation matrix of the playback signal as suggested for the variable span linear filter structure suggested in [66, 67, 83]. Otherwise, sound zones could be formulated in terms of time-frequency processing using the short-time Fourier transform as suggested in [35], where it is utilized to maximize speech privacy in a sound zones context. These structures enable updating the control filters depending on the playback signal.

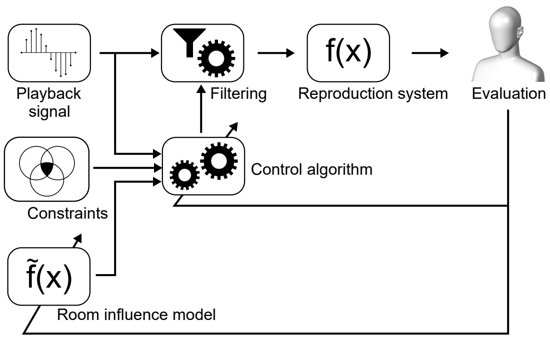
Lastly, one might consider creating sound zones using a feed-back structure where microphones are used to observe the reproduced sound field in the control

[11]The loudspeaker signal is the output of the linear convolution between the playback signal and the loudspeaker control filter.

[12]Note that defining the contrast control in the time-domain without constraining the average spectrum of the reproduced sound in the bright zone tends to produce solutions which strongly favors a single frequency as observed in [103].



(a) Feed-forward control. The room influence model, the constraints, and possibly the playback signal are fed into the control algorithm, where filters are determined assuming the room influence model to be accurate.



(b) Feed-back control. Like the feed-forward model, the room influence model, the constraints, and possibly the playback signal are fed into the control algorithm to determine control filters. However, here the result of the control is observed and fed back to update or correct both the control filters and / or the room influence model.

Fig. 7: Control structures

regions and adjust the filters accordingly. The feed-back structure requires representative microphone error signals to accurately control the sound field in the control regions. As the purpose of sound zones is to have listeners within the zones, this limits how close the microphones can be to the control region in a general domestic setting. General results from active control of sound suggest that a single loudspeaker can reduce a diffuse sound field by up to 10 dB within 1/10 wavelength of a microphone recording the error signal [59]. To control the sound field further away from the microphone positions requires predicting the field at a position away from the observations. Such an approach is suggested in [53, 60] where the predictions are based on measurements using both the permanent error microphones and microphones in a head and torso simulator, which is removed after initial calibration measurements.

2.2.4 Outcome

Various control methods exist to create sound zones, each of them is associated with different benefits and drawbacks. Typically, the tradeoffs are between the reducing computational complexity and controlling the relevant aspects of sound zones. In general, it is easy to conceptually describe the desired performance of a sound zones system e.g. low distraction due to the reproduced sound leaked between the zones, high sound quality in the bright zone, linear loudspeaker responses. However, it is less obvious to express such performance criteria directly as a cost-function with constraints that can be optimized. The cost-functions are, therefore, expressed using related parameters such as acoustic contrast, mean square error of the reproduced sound field, and the array effort.

The performance which can be attained for sound zones is related to the chosen loudspeakers and their location. At mid and high frequencies where beamforming solutions are feasible, it is generally of interest to radiate as little energy into the room as possible and avoid nearby reflecting surfaces. However, application constraints might make it impossible to avoid nearby boundaries, hence, it is of interest to know more about their influence on beamforming performance. With an estimate of a room influence model (depicted in Fig. 4a), the next stage for creating sound zones is to determine control filters for the loudspeakers (depicted in Fig. 4b). The filters are determined by formulating and solving an optimization problem, which requires the definition of an appropriate cost-function. The cost-function should describe the desired properties of the sound field reproduced in the control regions as well as the capability of the reproduction system. In the following, common parameters to include in optimization problems for sound zones are introduced. This is followed by a brief presentation of different control methods applied in different frequency ranges. Finally, typical control frameworks are described.

2.3 Rendering and evaluation of sound zones

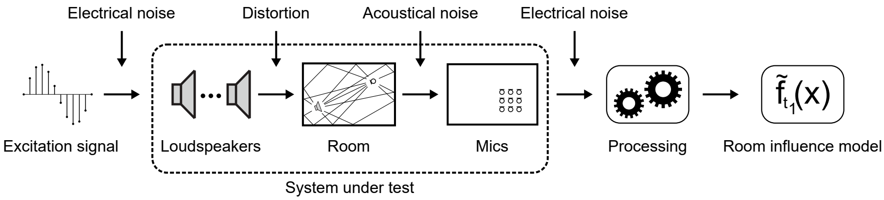
Having determined both the room influence model and the control filters for the loudspeakers in Secs. 2.1 and 2.2, the final step is to render sound zones and evaluate the results. There are in general two approaches to evaluate the performance of a sound zones system: using microphone measurements to calculate performance parameters or using human participants in designed listening tests.

In the following the consequences of a mismatch between the assumed room influence model and the state of the reproduction system are introduced before results from perceptual investigations on sound zones are presented.

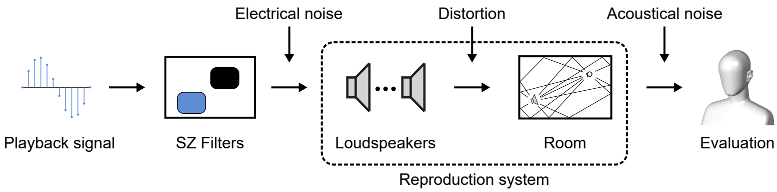
2.3.1 Physical evaluation considerations

When reporting results regarding the performance of sound zones the ecologically valid approach would be to design listening experiments and rate the performance of the implemented system. However, this is often impractical due to the time and resources required to run such tests and because the investigated method often only covers a limited frequency range as discussed in Sec. 2.2.2. A time-effective alternative is to evaluate the physical performance of sound zones by predicting the reproduced sound field. The sound field is predicted by applying the control filters to a room influence model. One concern when reporting such results is their robustness. A prediction of large acoustic separation is irrelevant if a tiny difference between the room influence model and the actual reproduction system makes it impossible to attain the reported result.

As stated in section 2.1, one of the assumptions applied when the room influence model is identified, is that the loudspeakers radiating sound into the room behaves as an LTI-system. While this is generally a good approximation, it is still an approximation with limited accuracy due to e.g. the speed of sound changing with temperature or nonlinear distortion in loudspeakers. Furthermore, measurements inherently include uncertainty. The mismatch between the identified room influence model and the reproduction system can thus be grouped into two categories: a time-dependent mismatch due to changes in the state of the reproduction system and a mismatch due to uncertainty in the



(a) Non-ideal room interaction model identification. The reproduction system is influenced by different disturbances. The input to the loudspeakers contains electrical noise, the loudspeakers will reproduce this noise and might also include nonlinear distortion if the level of the audio signal is large. There might be acoustic background noise recorded by the microphones and finally there might be additional electrical noise in the recording system.



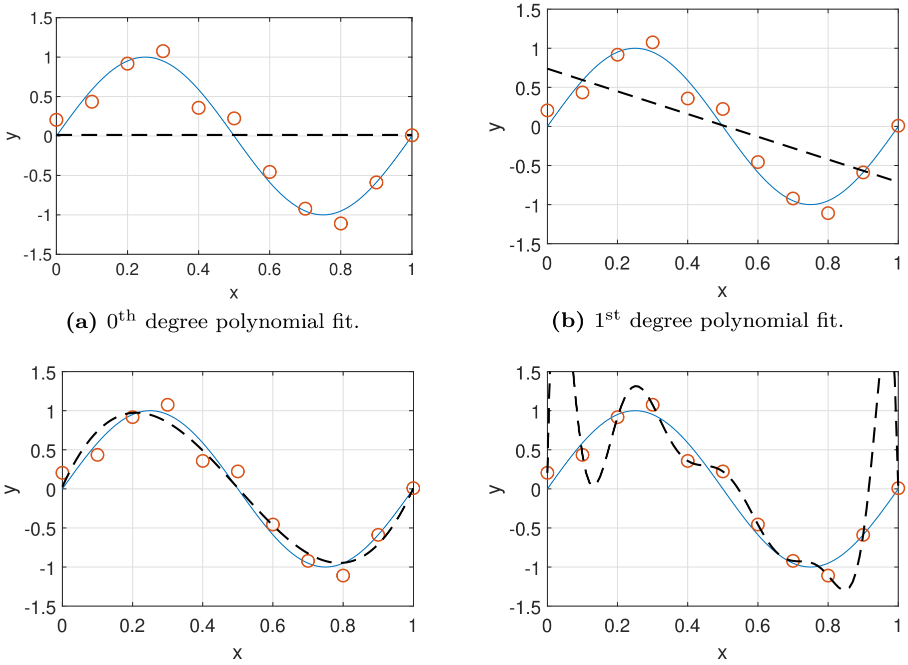
(b) Non-ideal Sound Field Reproduction. Identical types of disturbances can exist for the reproduction scenario. However, the realizations of the disturbances are different from the non-ideal room interaction model identification as the rendering happens at a different point in time.

Fig. 8: Mismatch between system model identified and reproduction system

estimated room influence model. This scenario is illustrated in Fig. 8, where noise and error sources have been indicated at different stages in in the system. The identified room influence model is an approximation of the reproduction system which is indicated by the superscript $\tilde{\cdot}$, while the sub-script t_1 denotes the time at which the model was determined.

When the reproduction system is used to render sound zones, it will by definition be at a later point in time than t_1. Hence, there will be at least a small mismatch between the room influence model and the reproduction system. Such effects should be considered when evaluating sound zones performance.

If the same room influence model is used to determine the control filters and to predict the resulting performance, the results might be highly sensitive to minor changes in the room interaction model without it being observed from the results. This behavior is known in the statistical model fitting literature as overfitting [9, 80]. The concept is illustrated for fitting polynomial models to noisy samples of a sine wave in Fig. 9. The example shows that adding more degrees of freedom (higher polynomial degree to fit the samples reduces the discrepancy between samples and model predictions. However, the 10$^{\text{th}}$ degree polynomial clearly does not match the sine wave from which the samples were generated. Furthermore, this behavior is not observed if the result is evaluated with the samples used to fit the model. From the discrete inverse problem literature [54] it is known that overfitting is generally related to a large norm of the determined solution (i.e. the filters). In the sound zones literature, this is generally associated with a high degree of self-cancellation.



(a) 0^{th} degree polynomial fit.

(b) 1^{st} degree polynomial fit.

(c) 3^{rd} degree polynomial fit.

(d) 10^{th} degree polynomial fit.

Fig. 9: Illustration of overfitting using different order polynomials to fit noisy samples of a sine wave (drawn with inspiration from [9]). (——): Sinusoid. (○): Noisy samples. (- - -): Fitted polynomial.

These observations beg the question of what a fair comparison between methods might be. Essentially, an evaluation should be application specific and indicate the suitability of a control method for a rendering scenario. Thus, if sound zones performance is predicted using a room influence model, the evaluation should incorporate inaccuracies in the room influence model.

2.3.2 Robustness of sound zones

The robustness to different sources of mismatch between room influence model and reproduction system has mainly been investigated in simulations. In the following, some of the results are introduced.

The robustness of an endfire array was investigated for free-field radiation in [37]. It was shown that high self-cancellation was associated with high sensitivity towards errors from position mismatch, source response mismatch, or changes due to discrepancies between free-field radiation and the addition of a reverberant sound field. Both the self-cancellation and the sensitivity to errors were seen to decrease with added regularization. Modelling the room reflections as a diffuse field was tested in [105], where it was used to predict the reduction in contrast from a compact array due to reverberation. Measurements showed a significant reduction in contrast relative anechoic results. Moreover, increasing the reverberation time of the room was seen to further reduce the contrast.

A different way to consider the mismatches is to evaluate their influence on the attained contrast. In [94, 95], the sensitivity to errors was investigated analytically. The errors were assumed to be small magnitude and phase errors in the loudspeaker and microphone responses as well as magnitude and phase errors due to uncertainty in their positions. The derivations showed the ACC method to be insensitive to small magnitude and phase errors in the loudspeaker responses as well as phase errors in the microphone responses. In [18], multiplicative errors were assumed on the transfer functions and including their expectation led to a robust formulation of the time-domain acoustic contrast control. This idea was adopted in [115–117] where the authors sought to reduce the number of in situ measurements required for sound zones. They measured a single transfer function from each loudspeaker to each control region and extrapolated to the rest of the region using the radiation model of a point source in free-field. The discrepancy between the extrapolated responses and the true transfer functions were approximated by an error distribution of identically distributed, independent amplitude and phase errors. Given rough estimates of the error distributions, the robust solution was seen to attain nearly identical contrast results as an ACC solution based on transfer functions measured at multiple control points in the control region.

Changes in the speed of sound and random displacement of the loudspeakers were investigated using free-field simulations in [30]. The authors investigated the influence of a 10 m/s change in the speed of sound change (corresponding to a 17°C change in temperature) and a random displacement of 10 mm. It was observed that ACC performance was generally insensitive to the speed of sound change, but highly sensitive to the misplacement of the loudspeakers. PM was seen to be sensitive to both error types and required regularization to reduce the effect of the changes. Changes in the ambient conditions are also a concern to sound zones applications where it is desired to control the low frequency radiation from open air concerts. A change in wind speed and temperature can reduce the control effectiveness by altering the relationship between the primary loudspeakers reproducing the concert sound and the secondary loudspeakers used to limit the propagation to the adjacent control areas [20].

The purpose of sound zones is to preproduce different audio to different listeners. This means that listeners will occupy the control regions, which scatters the impinging sound field at high frequencies. This effect was observed for a line array in a 17” monitor in [25], a dual layer circular array surrounding the listener in [23], and a 40 channel loudspeaker array surrounding two zones in [93]. In all cases, the contrast was seen to drop due to the scattering from the introduced listener and that the severity increased with frequency as the scatterer increased in size relative to the wavelength. While it might be possible to reduce this effect by assuming accurate knowledge of the scattered sound field, this approach was seen to be sensitive to small movements of the scatterer [23]. The severity of the contrast reduction due to the scattered sound field depends on the layout of the control regions relative to the loudspeakers. As an example, it was observed in [109] that a head-sized rigid sphere had negligible effect on the directivity pattern of an endfire loudspeaker array aimed towards the sphere.

Typically, the loudspeakers used for sound field control purposes are expected to act as linear transducers. However, it is well known that the moving coil loudspeaker driver is only approximately linear [64]. At large diaphragm displacements, the loudspeaker will start to generate harmonic and intermodulation distortion products. In [69–71], it was observed that while the 2nd and 3rd order harmonic distortion did not deteriorate the separation between sound zones significantly, high excursion levels also causes changes in the response at the fundamental frequency which was seen to reduce the resulting contrast. Suggestions for countering nonlinear distortion are generally to reduce the signal amplitudes of the loudspeakers through constraining either the array effort [70] or by constraining the maximal gain of each individual loudspeaker [26, 68]. It was seen that reducing the gain of the individual loudspeakers reduce the nonlinear artifacts, which leads to higher contrast at a given reproduction level in the bright zone [68].

From the literature it is seen that several factors might influence the acoustical separation between sound zones. Examples are temperature variations, nonlinear distortion, and scattering from the listeners. These investigations are almost exclusively made under either anechoic conditions or conditions where the influence of the room can be considered a diffuse contribution. Thereby the robustness of sound zones in environments with strong room influence remains unknown.

2.3.3 Perceptual Evaluation

The performance of sound zones can be characterized through physically based parameters such as acoustic contrast and reproduction error. However, this does not necessarily match the performance perceived by listeners in the sound zones. To characterize the perception of listening to one playback signal while another is interfering was the topic of two PhD projects [5, 44]. Listening tests were conducted with two loudspeakers emulating a sound zones system. One loudspeaker reproduced the desired playback signal (the *target*) to the listener. Another loudspeaker was used to emulate the acoustic leakage from an adjacent sound zone by reproducing a different playback signal (the *interferer*). In an initial listening test, the participants were asked to adjust the level of the interfering signal until the scenario was acceptable. It was seen that the level reduction required for the scenario to be acceptable depended on the type of playback signals used (speech, classical music, pop music) [43].

To further characterize listening to both a target and an interfering signal, an elicitation experiment was conducted to determine the most prominent perceptual attributes describing such a scenario [46]. The most significant attributes were used in a subsequent rating experiment and principle component analysis revealed that around 99% of the variance could be explained by two dimensions. The first dimension accounting for 89% of the variance was labeled after the attribute *distraction* with the description “How much the alternate audio pulls your attention or distracts you from the target audio?” and end-points “not at all distracting” to “overpowered”. The second dimension, accounting for 10% of the variance was labeled after the attribute *balance and blend* with the description “How you judge the blend of sources to be?” and end-points “complementary” to “conflicting”.

Due to the dependence of the playback signal type, further listening tests were made with specific signal types. To predict distraction ratings for music used as both target and interfering signal, a model was presented in [47]. This model combines binaural and monaural recordings of both target and interfering sound to predict the distraction. The loudness ratio between target and interfering sound is a large part of the model, but significant improvement was seen by including overall loudness, frequency content of the interferer, and interference related perceptual score from the PEASS[18] toolbox [40]. While the model shows good predictions for the listening tests using one loudspeaker to reproduce the target and one to reproduce the interferer, this scenario hardly resembles reproduction systems for creating sound zones. To this purpose, the predictions of the model were validated using two implemented sound zones systems using both music and speech signals in [100] and [99], respectively. Furthermore, the distraction model in [47] is computationally expensive, taking approximately 13 minutes to calculate predictions of a 10 s sample. A real-time modification of the model was proposed in [98] reducing the computation time to 0.04% of the original model. The performance of the real-time model was validated against the original model in [101]. It was seen that the real-time model predictions are highly comparable to the original model. From the results, it was observed that individuals participating in the distraction rating experiments seemed to have different internal references regarding what is distracting which led to large variances between subjects. However, the distraction ratings were generally low when the target to interferer ratio (TIR) was above 25 dB and 29 dB in the two zones, respectively. TIR was defined as the difference in equivalent A-weighted sound pressure levels (dB $L_{A,eq}$) between target and interferer in one zone, calculated using 10 s integration time.

While the distraction model is useful for predicting the effect of the interfering audio in a sound zones scenario, it is not a full characterization of the overall sound quality of such a system. Such an evaluation would likely include considering the quality of the reproduced sound in the bright zone, which might

[18] Perceptual evaluation of audio source separation (PEASS).

be degraded due to the involved processing. In [4] the target sound quality, the distraction, and the overall sound quality were rated in a listening experiment for acoustic contrast control, pressure matching, and different variations of planarity control [31]. The main physical changes across the methods were differences in contrast (ACC attaining the highest contrast and degrading over the PC variations towards PM) and uniformity of the direction of impinging sound which followed the opposite trend (most uniform for PM and least uniform for ACC). The results of the rating experiments revealed that overall sound quality depended on both the quality of the reproduced sound in the bright zone and the distraction from interfering sound. However, overall sound quality was observed to have a significant interaction between control method and the combination of target and interferer content.

The results above provide an idea of the requirements for low distraction ratings in terms of contrast across the reproduced frequencies. It should be noted that while the distraction model appears to generalize well, the observed TIR values are specific for the tested sound zones system. Hence, although a constant contrast of 25 to 29 dB across frequency might be sufficient in comparable scenarios, using the distraction model for evaluating the expected performance of a full sound zones system is recommended. It is further noted, that this range of contrast values is based on the assumption that the playback signals are loudness matched and reproduced at identical levels in the sound zones. If that is not true, additional contrast would be required.

2.3.4 Outcome

The prediction of sound zones performance is potentially subject to overfitting, the effect of which is only seen if the evaluation procedure specifically is designed to account for it. Two types of overfitting might occur in the prediction of sound zones performance: one is overfitting the inaccuracies in the room influence model, while the other is overfitting the time-dependent state of the reproduction system. The choice of control method and evaluation of the results should reflect this.

The robustness of sound zones has been investigated in various simulations including different errors used to emulate mismatches between the room influence model and the reproduction system. However, few investigations regarding the robustness are performed using experiments in reflective environments such as rooms or automotive cabins.

Perceptual investigations have revealed that distraction is the most significant attribute used to describe listening to a target and interfering signal. Distraction ratings depend on the target and interfering playback signal types e.g. speech and pop music. Low distraction ratings have been observed in listening tests with 25 to 29 dB $L_{A,eq}$ difference between target and interferer.

2.4 Summary

To determine the control filters for the loudspeakers in the sound reproduction system, the capabilities of the system to reproduce sound in the control regions must be characterized. This characterization is included in a room influence model, which in the presented work is a matrix of room transfer functions or impulse responses estimated in situ.

In the literature, general physical parameters to be controlled are the acoustic contrast between a bright and a dark zone, the reproduction error in the bright zone relative to a target sound field, and the array effort. Besides a suitable choice of the control parameters, it has been seen that the arrangements of loudspeakers have a large effect on the attainable performance. The loudspeaker arrangements proposed in the literature depend on the target frequency range and assumptions regarding the location of the listeners.

The control filters are determined under the assumption that the room influence model is an accurate representation of the reproduction system. However, when estimated in situ the model is both subject to uncertainty in its estimation and time-varying behavior of the reproduction system due to e.g. changes in the ambient temperature. The robustness of sound zones to such changes have primarily been considered in simulations of point sources in free-field. It is therefore of interest to experimentally investigate the robustness in reflective environments to include and identify factors which are not included in the simulations.

It is possible to predict the performance of a sound zones system using a room influence model estimated in situ. However, overfitting the sound zones solution to the estimated model is a general concern. Most comparisons of sound zones methods are conducted in simulated or anechoic conditions. It is of interest to investigate whether the general results regarding the performance of the methods are significant under experimental conditions, or whether small inaccuracies in the experiments will suppress the difference between the methods.

The work presented in the literature regarding loudspeaker array beamforming generally considers the effect of the reproduction environment as being either a diffuse reverberation or a specular reflection of the far field directivity. These assumptions are reasonable when the loudspeaker array is far from reflecting surfaces. However, in many applications such a loudspeaker array would be positioned close to reflecting surfaces e.g. walls and furniture in a domestic room or the dashboard and windscreen in an automotive cabin. It is of interest to investigate whether the effects of nearby reflecting boundaries can inherently and robustly be included by the definition of the control regions for in situ estimation of the room influence model.

Compact loudspeaker arrays are used at mid and high frequencies to focus the sound towards the bright zone and away from the dark zone. However, after being reflected off the boundaries of the room, sound eventually reaches the dark zone, thereby, limiting the contrast. This is especially a problem towards lower frequencies where the compact array is not directive. It is, therefore, of interest to use distributed loudspeakers in as broad a frequency range as possible. Ideally, this should facilitate a transition to the compact array in the frequency range where it is effective. The majority of this work is directed towards limitations of controlling sound fields using distributed loudspeakers in acoustic enclosures.

3 Contributions

The contributions from the conducted research are a collection of papers. The papers are organized as a number of studies investigating various aspects concerning the generation of sound zones. The contributions are grouped according to the two research questions. The first research question is related to the contrast and limiting factors. In the previous section, it was reported that the mismatch between the room influence model and the reproduction system will limit the performance. The Papers A, B, D, E, and G provide different investigations on such mismatches. The second research question is related to the control methods for creating sound zones. Paper C introduces a modification of an existing control method, while a method from control theory is adapted to sound zones in Paper F. An overview of the papers is provided in Fig. 10.

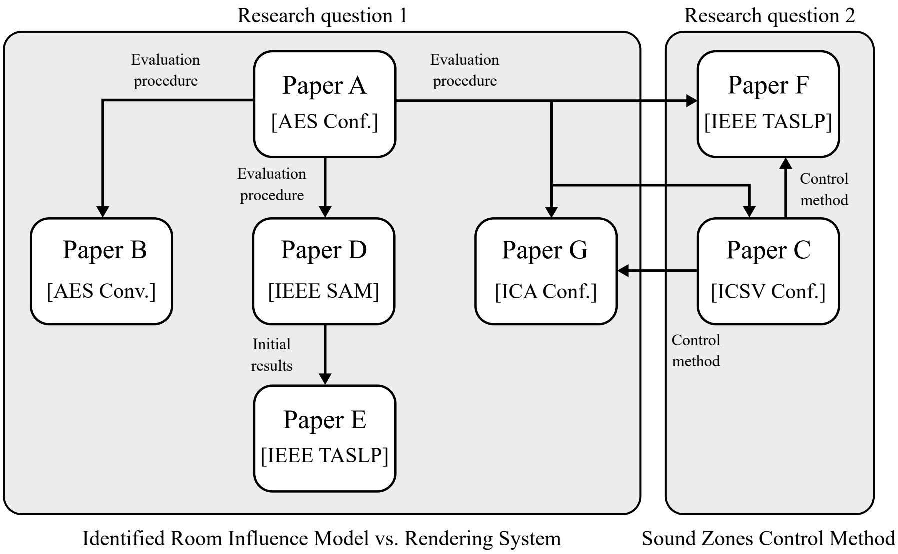
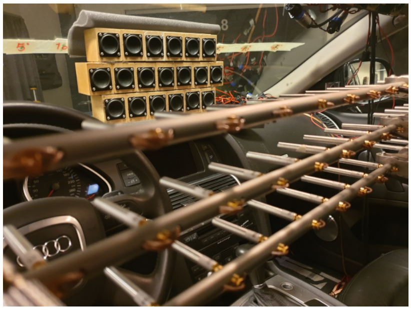


Fig. 10: Relationship between the papers. Paper A [74]: Evaluating sound zones performance. Paper B [91]: Influence of temperature variations. Paper C [75]: Reducing pre-and post-ringing in sound zones FIR filters. Paper D [72]: Estimation of in situ transfer functions using classical statistics and the influence of measurement noise. Paper E [73]: Bayesian inference of transfer functions and its use for sound zones. Paper F [77]: Sound zones formulated in a moving horizon framework. Paper G [76]: In situ beamforming using a planar loudspeaker array in an automotive cabin.

Throughout the work, experiments have been utilized to test hypotheses regarding the significance of factors limiting acoustical separation between sound zones. Pictures of the setups used for the experiments are provided in Fig. 11. As can be observed from the pictures, some of the setups were built for reproducing sound zones in the entire audible frequency range. The purpose of these implementations was to support listening tests validating the distraction model, as presented in [99, 100]. The full-bandwidth implementations utilized methods published in the literature and only contribute to the present work as inspiration.

(a) Setup used in paper A and C.

(b) Setup used in paper D, E, and F.



(c) Setup used in paper G.

Fig. 11: Examples of experimental setups built and used throughout the project.

In the rest of this section, the contributions are summarized within the two groups relating them to the research questions.

3.1 Identified room interaction model vs. reproduction system

The literature, presented in Sec. 2.3.2, shows that the robustness of a sound zones system is evaluated in terms of the sensitivity to a mismatch between the estimated room influence model and the reproduction system. The majority of the literature dealing with robustness of sound zones investigates the influence in free-field simulations. To test the behavior in more realistic environments, a number of investigations has been conducted as experimental studies. The focus of the studies was factors which might limit the performance of sound zones. In all of these investigations, the sound zones system was assumed to use time-invariant control filters. These filters were determined using a room influence model estimated at time t_1. When rendering sound zones at time t, the reproduction system might have changed (the system is sketched in Fig. 12). Such a scenario can be emulated by evaluating the performance using a room influence model estimated at time t_2. This approach was adopted in the contributions in this section to investigate factors limiting the performance of sound zones. Thus, the investigations are expressed in terms of the mismatch between independently estimated room influence models.

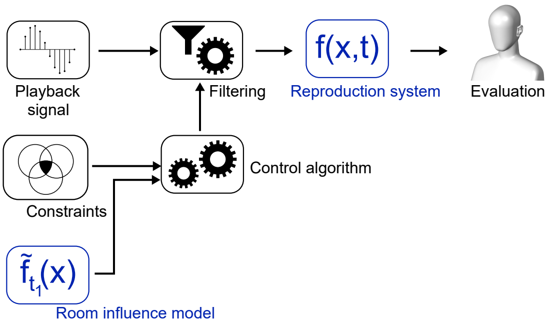


Fig. 12: Feed-forward control structure used to investigate room influence model mismatch and evaluation. The blue highlighted blocks indicate the primary blocks being investigated in this subsection which covers papers A, B, D, E, and F.

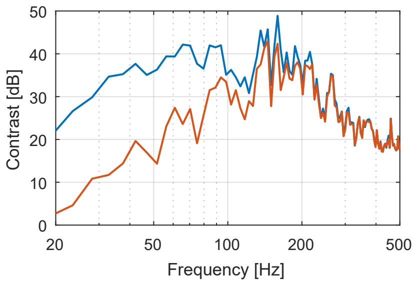
3.1.1 Evaluating the performance of sound zones - Paper A

Hypothesis 1: The differences in performance between alternative control methods are highly dependent on the chosen procedure for evaluating the results.

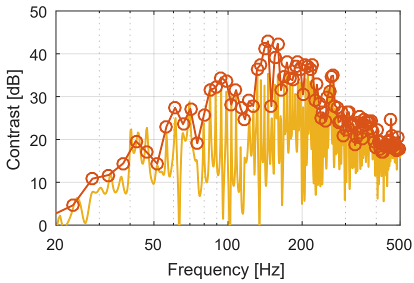
In the literature, a number of different ranges for the performance of sound zones is suggested. The performance of sound zones depends on the number and location of the loudspeakers, the acoustic enclosure, and the size and positions of the control regions [92]. It is, therefore, rarely possible to directly compare the performance of various control methods between multiple publications. Instead one can evaluate the relative performance between methods. Before the robustness of sound zones to various factors were investigated, a baseline experiment was conducted. The purpose of Paper A [74] was to evaluate a number of ways in which overfitting might influence the predicted contrast as mentioned in Sec. 2.3. This allowed several methods from the literature to be compared under similar conditions. For this investigation, two room influence models were estimated from separate measurements taken with approximately 10 second separation using the woofers depicted in Fig. 11a. The potential overfitting to the noise in the measured room influence model is illustrated in Fig. 13a. In the figure, a significant reduction in the predicted contrast is seen when using separately measured room influence models for determining the control filters and evaluating the results.

Another effect arises from the filter design process. If the sound zones problem is solved in the frequency domain, independently for each discrete

Fourier transform (DFT) bin, the solution to the independent optimization problems provide a complex scalar for each loudspeaker at each DFT bin. For an actual implementation, it would be necessary to design a digital filter for each loudspeaker matching these complex scalars across frequency. If the FIR filters are designed from the frequency sampling method, they might only be accurate at the designed DFT bins. This effect is displayed in Fig. 13b where the resulting contrast is evaluated at the design DFT grid and at a DFT grid with four times higher resolution. Clearly, the resulting filters do not yield the desired contrast across frequency, but this is only revealed by evaluating the result at a DFT grid which is different from the DFT grid used for designing the filters.



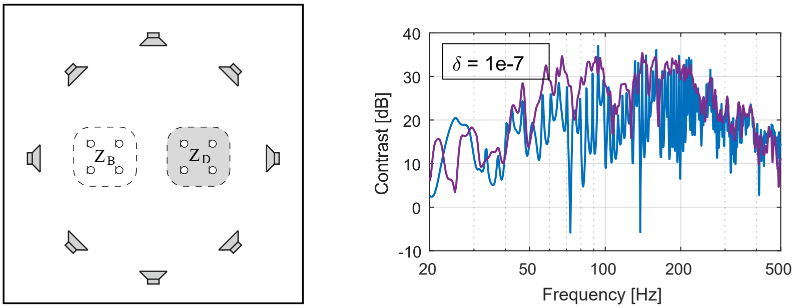
(a) Contrast results for two conditions: Loudspeaker control filters determined using room influence model 1 and evaluated using room influence model 1 (——). Loudspeaker control filters determined using room influence model 1 and evaluated using model 2, which is measured approximately 10 seconds after model 1 (——).



(b) Contrast results for two conditions: Loudspeaker control filters designed by frequency sampling and evaluated only at the sampled DFT bins (–⊖–). Loudspeaker control filters designed by frequency sampling and evaluated on a DFT grid having four times higher resolution than the sampled DFT bins (——).

Fig. 13: Results and plots repeated from Paper A [74]. Example of evaluation procedures potentially leading to wrong conclusions when evaluating contrast results.

With the suggested evaluation procedure, several methods for creating sound zones were compared in Paper A [74]. Two main observations were made from this comparison. First, that methods based on time-domain formulations were not subject to the artifacts related to FIR design by the frequency sampling method, as the FIR design was an inherent part of the formulation. Secondly, that a weighted pressure matching method formulated in the time-domain [106] provided contrast results equal to the acoustic contrast control at the design frequencies as shown in Fig. 14. This indicates that the superior contrast performance of ACC relative to other control methods might not be realized in a practical scenario.



(a) Setup used in the investigation. Eight 10” woofers placed in a circular arrangement surrounding two zones as seen in Fig. 11a. The sampled control regions were $0.4 \times 0.2 \times 0.23\ \mathrm{m}^3$.

(b) Contrast comparison between acoustic contrast control [27] (——) and weighted pressure matching [106] (——) using mismatched room influence models and DFT grids. The regularization level is $\delta = 10^{-7}$, see Paper A [74] for further details.

Fig. 14: Results and plots repeated from paper A [74]. Comparison between acoustic contrast control and time-domain pressure matching in experimental conditions.

3.1.2 Outcome of Paper A

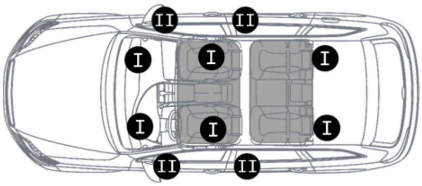
As highlighted by the experimental results presented in Paper A [74], the choice of evaluation procedure can significantly alter the conclusions regarding the performance. Even small variations between room impulse responses measured within 10 seconds of each other are enough to significantly reduce the predicted contrast. Thus, it is important to always consider the evaluation procedure when reporting results on sound zones. At the very least, separate measurements should always be used for determining the control filters and evaluating the performance.

3.1.3 Influence of temperature changes - Paper B

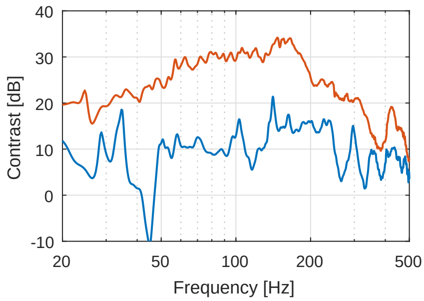
Hypothesis 2: Changes in the ambient temperature affect the sound zones control due to alterations in the transfer functions between loudspeakers and microphones.

The speed of sound depends on the temperature [55]. Hence, temperature changes will alter the state of the reproduction system and lead to a mismatch from a previously estimated room influence model. To gain insight in the significance of the sensitivity to temperature variations, an experiment was designed in an automotive setting. For this purpose, room influence models (for both filter design and evaluation) were estimated at both 23°C and at −2°C. Calculating filters using a 23°C room influence model and evaluating the result with a −2°C model yielded a large drop in the attained contrast as seen from the results in Fig. 15.

Further investigations indicated that the shift in temperature caused more than just a change in the speed of sound. The loudspeakers used in the investigation were seen to change resonance frequency due to the shift in temperature. This behavior was confirmed by measurements in a temperature-controlled climate chamber. Simple approaches to compensate for the changes in the loudspeakers and the speed of sound were suggested, but they failed to significantly reduce the effect of the temperature mismatch in terms of the acoustic contrast. It was suggested that the changes to the reproduction system might consist of more factors than a change in loudspeaker resonance frequency and a change in speed of sound e.g. that the boundary conditions of the car cabin might be temperature dependent.



(a) Setup in a car cabin with the front seats being the bright zone and the rear seats the dark zone. Ten woofers were used in the experiment, six of type I and four of type II distributed as indicated.



(b) Contrast when the filters are designed using a 23°C room influence model and evaluated using a second 23°C model (——). Contrast when the filters are designed using a 23°C system model and evaluated using a −2°C model (——).

Fig. 15: Results and plots repeated from paper B [91]. Setup and results from investigating the effect temperature changes has on generating sound zones in an automotive cabin.

3.1.4 Outcome of Paper B

Changes in temperature can significantly affect the contrast performance in an automotive environment. The experiment indicates that it is not straightforward to correct for the changes in temperature. Moreover, it might be necessary to separate the effects on the transducer, the medium, and the boundary conditions in the car cabin if it is desired to compensate the determined filters.

An additional outcome of the study is the importance of careful selection of loudspeaker drivers for a given application. By choosing a driver constructed of materials which maintain their mechanical properties across the investigated temperature range, the change in loudspeaker resonance frequency can be significantly reduced.

3.1.5 Influence of noise inherent in the measured transfer functions - Papers D and E

Hypothesis 3: The noise present in the measured transfer functions inherently limits the attainable separation between sound zones.

Even if the reproduction system does not change over time relative to the identified room influence model, there is still an effect which limits its accuracy. That effect is the noise present during the estimation of the room influence model. While background noise levels can be kept low in a laboratory environment, the same cannot be guaranteed in a general domestic or automotive setting. Following the results in Paper A [74], it is of interest to compare predictions of the sound zones performance to the actual performance. To investigate the effect of the measurement noise, simulations and experimen-tal investigations were conducted in Papers D and E [72, 73] using the setup sketched in Fig. 16a.

An initial experimental investigation (presented in Paper D [72] consisted of using repeated measurements to determine how different noise realizations influence the performance prediction. The results in Fig. 16b show the maximum and minimum predicted contrast using one measurement to determine the sound field control and one to evaluate the result, over all combinations of 30 repeated measurements.[19] The results indicate that there can be a large variation between the lowest and highest prediction of the contrast, when no regularization is added to the least-squares problem of pressure matching. When regularization is added, the contrast prediction decreases but so does the difference between maximum and minimum predicted contrast. This shows that the procedure of using different measurements for calculating filters and evaluating results might reveal some overfitting, but it is not a guarantee that the performance prediction is identical to the actual sound field.[20]

The investigation was continued in Paper E [73], with a slightly different outset. Instead of evaluating the estimated room influence model using another estimate of the model, a simulation study was presented where the effects of the noise could be evaluated with a noiseless model. It was observed that if the noise at each microphone is independent, identically distributed white Gaussian noise, both the regularized and the non-regularized sound zone methods provide consistent and similar results across 100 Monte Carlo simulation as seen from Fig. 16c. However, if the noise is perfectly coherent across the microphones, the Monte Carlo experiments show a large variance and the reg-

[19]Note that in the investigations for Paper D and E [72, 73], FIR filters are not designed, and the sound field control is only considered at the particular frequencies where the transfer functions were determined. This was a deliberate choice to highlight the direct influence of the estimation error at each frequency separately. The frequency independent investigation of the results is motivated by the framework used to estimate the variance in the transfer function measurements. Following that decision, the measurements were made as steady-state response measurements using multi-tone excitation signals.

[20]Of course, the quality of the prediction depends on the quality of the measurements. Measurements with higher signal to noise ratio would show less variation between the predictions.

ularization effectively reduces this variation, which is seen from Fig. 16d. The regularization added to the sound zones methods was automatically determined given Bayesian estimation of the uncertainty in the transfer functions. Thus, no manual selection of regularization parameter was required.

3.1.6 Outcome of Papers D and E

The results, presented in the papers, highlight that noise in the transfer functions can reduce the performance. However, noise which is independent between the microphones tends to automatically regularize the problem (which is known to reduce the performance if the regularization becomes sufficiently large). If the noise is coherent on the other hand, it was seen to have a much greater impact on the resulting performance. In that case, introducing regularization based on the estimated uncertainty in the transfer functions was seen to reduce the influence of the noise.

3.1.7 In situ beamforming at mid/high frequencies - Paper F

Hypothesis 4: Mid / high frequency sound field control is susceptible to changes in the reproduction scenario within complex acoustic enclosures.

As explained in Sec. 2.2.2, when the frequency increases it becomes necessary to change strategy for creating sound zones. One suggested approach to generate sound zones at these frequencies is to use beamforming to focus the sound towards the bright zone and away from the dark zone. In free-field, this approach works well but, in a room or a car, the radiated sound will be reflected by the boundaries of the environment. The common way to treat this effect is to consider the sound field to consist of the direct sound from a loudspeaker array and a diffuse contribution or specular reflections [19, 51, 88, 105]. However, such approaches assume that the boundaries of the environment are far from the loudspeaker array. In general, it is more common that loudspeaker arrays intended for beamforming are positioned close to the boundaries of a room or a car cabin. To evaluate the influence of complex acoustic boundaries, an experimental study was conducted by placing a planar loudspeaker array on the dashboard of a car (see Fig. 11c) and comparing different approaches to focus sound towards the driver's position and away from the front passenger seat.

Two of the methods presented in Paper F [76] are compared in Fig. 17. The first method is delay and sum beamforming, assuming the loudspeakers to be point sources radiating in free-field. The comparison is made against the method from Paper C [75]. This method relies on a room influence model, which for the displayed result was measured from the loudspeakers to microphone positions in front of the headrests, as sketched in Fig. 17a.[21] It is observed that if the contrast is evaluated at the measured microphone positions

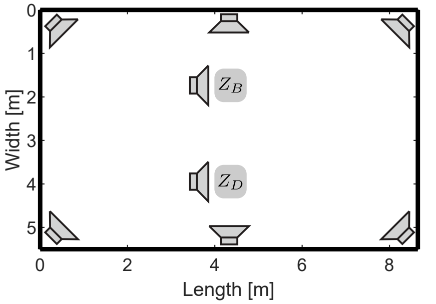
[21]In Paper F [76], the method is denoted *in situ zonal control* to specify the associated room influence model relative to other investigated choices.

(using separately measured impulse responses), the approach based on in situ measurements provides superior contrast relative to the delay and sum beamformer. A different performance evaluation was made using RIRs measured to the microphones in a head-and-torso-simulator (HATS) emulating occupied front seats. These measurements were made with the HATS in each of the front seats while the other seat was occupied by a mannequin. Furthermore, the measurements in each seat were conducted at different heights and head-rotations of the HATS. Evaluation with these RIRs reveal that the in situ control is sensitive to the introduction of the HATS and mannequin below approximately 2 kHz. The corresponding contrast results show similar performance using delay and sum beamforming and in situ control, given a mismatch due to the inclusion of “listeners” in the front seats. This indicates that the in situ control utilizes all reflection paths available to the loudspeakers from the measured room influence model. Thus, the solution can be considered over-fitted to the state of the reproduction system and suffers from changes in the reflection paths caused by introducing listeners in the front seats.

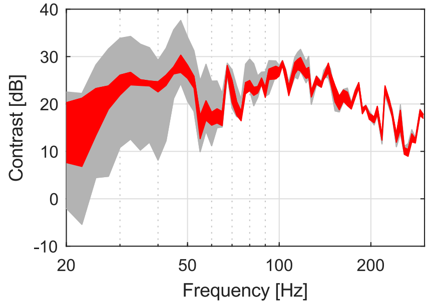
3.1.8 Outcome of Paper F

It appears that the discrepancy between in situ measurements and free-field beamforming is primarily a concern at lower frequencies (where the particular array only has limited inherent directivity). At higher frequencies where the array inherently is directive, it also becomes less sensitive to the nearby boundaries and attain similar performance using both delay and sum beamforming and in situ control. This also highlights the importance of the design of the array. It will only naturally become focused in presence of the boundaries if it is well designed for the intended frequency range.

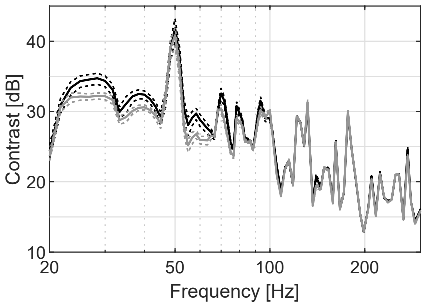
The results highlight that designing sound zones filters, for a compact array, based on in situ measurements can be problematic if the measurements are not a good representation of the rendering scenario. In such cases, it can be almost as effective to assume simplified radiation behavior using delay and sum beamforming.



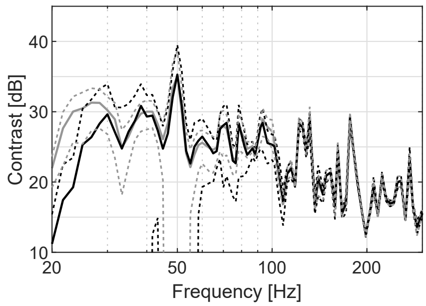
(a) Setup used in the contribution. Eight 10" woofers are placed around the room. Six are placed at the edges of the room and the last two are placed as close to the zones as possible.



(b) Evaluation of contrast using 30 measured room influence models. One model is used to determine the control and another for evaluating the performance. The shaded areas show the range between maximum and minimum predicted contrast across all combinations of the 30 measurements. The gray area depicts PM results with no regularization. The red area depicts PM results with regularization based on the sample variance across the 30 measurements. See Paper D [72] for further details.

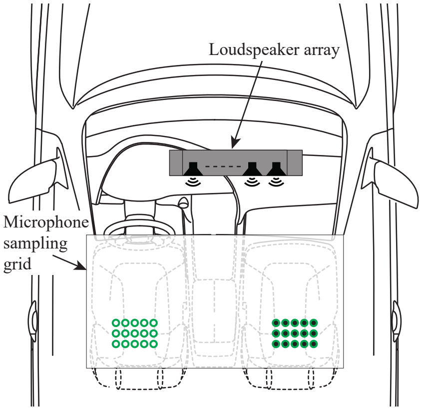


(c) Influence of noise during transfer function estimation. Rectangular room simulated using Green's function. Independent noise realizations at each microphone used to estimate the room influence model. The results are evaluated with a noiseless room influence model. Average performance (solid) ± 1 standard deviation (dashed) over 100 Monte Carlo iterations. (——): PM with no regularization. (——): PM with regularization based on the uncertainty in the room influence model.

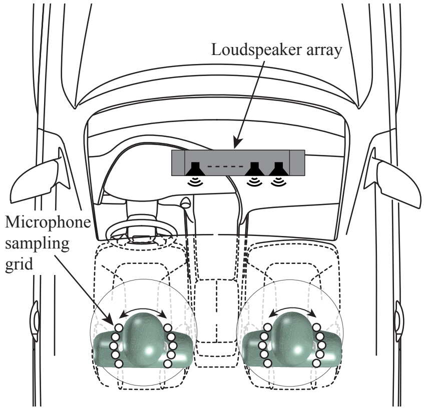


(d) Influence of noise during transfer function estimation. Rectangular room simulated using Green's function. Perfectly coherent noise at all microphones used to estimate the room influence model. The results are evaluated with a noiseless room influence model. Average performance (solid) ± 1 standard deviation (dashed) over 100 Monte Carlo iterations. (——): PM with no regularization. (——): PM with regularization based on the uncertainty in the room influence model.

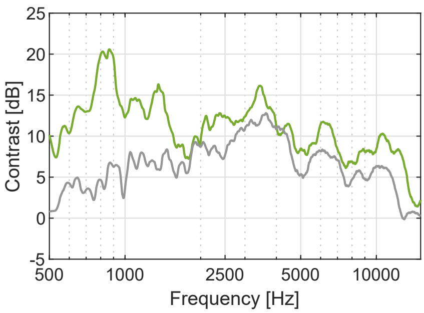
Fig. 16: Results and plots repeated from Papers D [72] and E [73]. Example showing the differences between e valuating the results using different room influence models or using a noiseless model (only available in simulations).



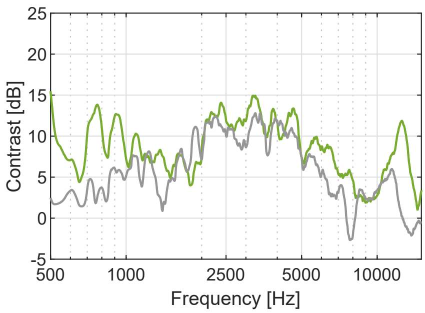
(a) Microphone evaluation in the car cabin. The target head positions were sampled in a 3D-grid of $5 \times 3 \times 3$ microphones with 5 cm between adjacent microphones.



(b) HATS evaluation using the microphones in the ear-canal. To emulate movement of a person the HATS was adjusted to three different heights (corresponding to the heights of the microphone evaluation) and five head rotations for each height.



(c) Microphone evaluation



(d) HATS evaluation

Fig. 17: Results and plots repeated from Paper F [76]. Contrast results attained with a planar loudspeaker array on the dashboard of a car, see Fig. 11c. The contrast is evaluated with the driver's seat as the bright zone and the front passenger seat as the dark zone. Delay and sum beamformer (——) and in situ control (——) in an automotive cabin, evaluated with microphones and with a HATS. Note that the contrast results have been smoothed with a 1/12$^{\text{th}}$-octave rectangular window to improve the readability of the results.

3.2 Sound zone control methods

The second part of the contributions consists of changes to the sound zones control methods. The focus of the contributions is controlling an aspect related to sound quality and adaptation to changes in the room influence m odel. The control methods in these contributions are demonstrated at low frequencies. However, the methods could just as well be applied at higher frequencies e.g. to determine filters for beamforming.

3.2.1 Pre- / post-ringing - Paper C

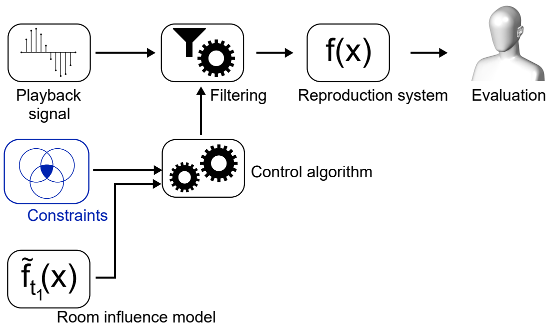


Fig. 18: Feed-forward control structure used for the investigation in paper C. The blue highlighted block indicate that the main focus is expressing the added constraint on pre- and post-ringing of the reproduced audio in the bright zone.

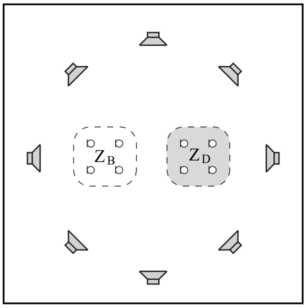
Hypothesis 5: The pre- / post-ringing in the control ilters deteriorates the sound quality. It is possible to reduce this ringing without reducing the separation between the zones.

One observation which arose during the experimental implementations leading up to the work in [100, 101] was that the sound zones processing introduced audible artifacts in the resulting bright zone. The main noticeable artifacts were related to pre- and post-ringing in the reproduced sound. This spread of energy was observed visually in the filters used for creating sound zones with various methods in Paper A [74].

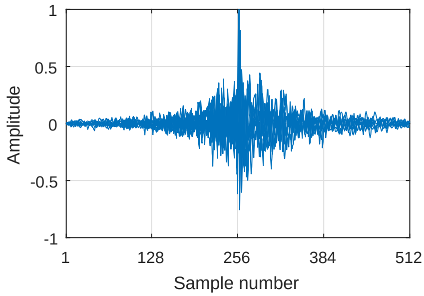
To reduce the pre- and post-ringing in the reproduced sound, the ringing in the control filters was targeted. The proposed solution is to add a weighted ℓ_2-norm penalty to the FIR filters when minimizing the cost-function for time-domain pressure matching [106]. The weight is chosen relative to the desired reduction of pre- and post-ringing in the filters.[22] From the results, shown in Fig. 19, it is seen that the proposed solution can reduce pre- and post-ringing

[22] A similar penalty could be imposed on the resulting pressure impulse responses in the bright zone. However, such an approach does not offer control of how the resulting sound field is attained. As seen from the results of Paper F [76], utilizing all the reflection paths in the RIRs can make the resulting sound field control sensitive to changes in the environment.

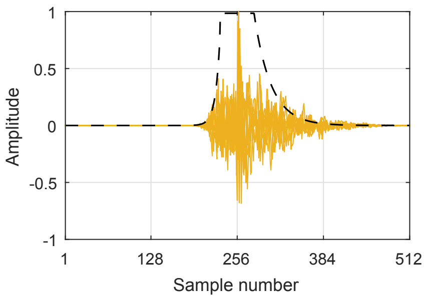
in the control filters. Furthermore, it is seen that for the particular case it is possible to do so without reducing the resulting contrast between the zones, relative to the regular solution with uniformly weighted ℓ_2-norm.



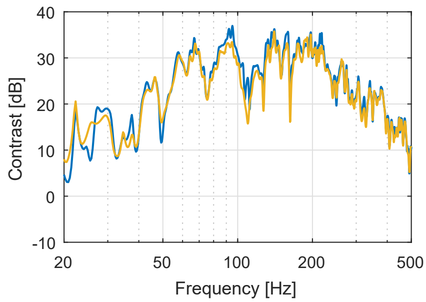
(a) Setup used in the investigation. Eight 10" woofers placed in a circular arrangement surrounding two zones as seen in Fig. 11a. The sampled control regions were $0.4 \times 0.2 \times 0.23$ m^3.



(b) Eight FIR filters plotted on top of each other, uniform weighting of the ℓ_2-norm (——).



(c) Eight FIR filters plotted on top of each other, specific FIR weighting of the ℓ_2-norm (——). The reciprocal values of the weighting function (- - -).



(d) Contrast from filters based on uniformly weighted ℓ_2-norm (——). Contrast from filters based on weighted ℓ_2-norm for reducing pre- and post-ringing in the filters (——).

Fig. 19: Results and plots repeated from Paper C [75]. Example of how FIR filters with different shapes (and thus pre- / post-ringing) properties can provide almost identical contrast results.

3.2.2 Outcome of Paper C

Introducing a weighted ℓ_2-norm penalty, it is possible to indirectly control the pre- and post-ringing of the resulting FIR filters. Note that controlling the pre- and post-ringing of the FIR filters does not guarantee similar properties of the reproduced sound within the bright zone. However, it will limit pre-ringing if the propagation delays from the loudspeakers to the bright zone are nearly equal.

3.2.3 Moving horizon framework - Paper G

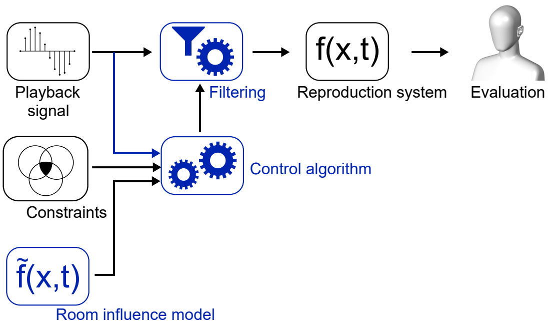


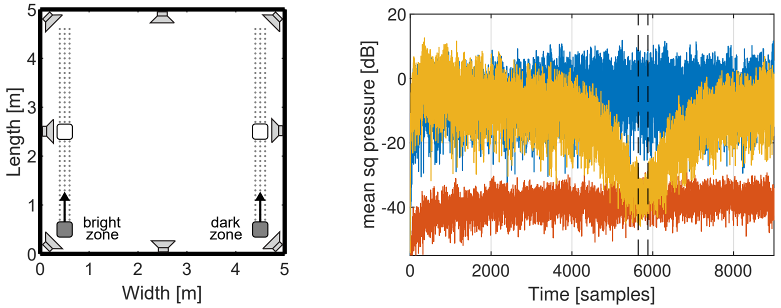
Fig. 20: Feed-forward control structure used for the moving horizon control in paper G. The blue highlighted blocks indicate that the desired audio content and a time-varying room interaction model are used as inputs to change the control filters over time.

Hypothesis 6: Including detailed information concerning the rendering scenario in the control structure can improve the resulting separation.

As observed in the previous contributions and the literature, using a room influence model which does not match the current state of the reproduction system leads to a loss of performance, relative to accurate knowledge of the reproduction system. Regularization can reduce the sensitivity to changes in the reproduction system. However, this robustness comes at the expense of reduced performance in the scenario where the room influence model is accurate. A different approach would be to utilize a control framework where changes in the reproduction system can be incorporated into the room influence model as they occur. The proposed solution is a framework, based on a state-space representation of the room influence model, known in control theory as e.g. moving horizon and model predictive control.

Moving horizon is a control method where a new solution is determined at each time-step (e.g. corresponding to the sampling frequency of the playback signal. Besides the determined solution, the framework also predicts how the solution applied at the current time-step will affect the control at future time-steps. This formulation allows the framework to be extended to model systems with inertia. An example of this could be including a loudspeaker diaphragm displacement model, which would be included for constraining the loudspeakers to operate in their linear range.

The work presented in Paper G, is used to show-case two properties of the moving horizon framework. First, the ability to incorporate updates of the room influence model is shown with an example using simulated room impulse responses and moving zones. The example in Fig. 21 is a scenario where two zones move throughout the simulated room following a trajectory at constant velocity. The control actions are based where the zones are now and where they will be in the future. The corresponding mean square sound pressure in the moving bright and dark zone are plotted over time in Fig. 21b with white noise as the playback signal. As a reference, the mean square pressure in the moving dark zone is shown for a scenario where the room influence model is not updated relative to the changing zone locations. Instead, the room influence model is kept static using RIRs corresponding to the location of the white-filled squares in Fig. 21a. From the pressure responses, it is observed that not updating the RIRs to match the zone locations clearly reduces the performance.



(a) Moving zone setup. The gray-filled squares represent zones which move with a constant velocity in the direction of the arrows. The dots indicate positions where the RIRs are known and the white-filled squares are the stationary reference positions for the zones. The mean square pressures in the moving zones are plotted against time in Fig. 21b.

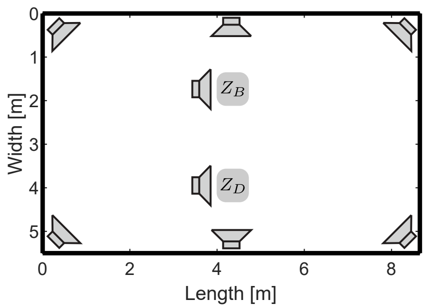
(b) Mean square pressure in the moving bright and dark zones from Fig. 21a plotted against time. (——): Pressure in the moving bright zone. (——): Pressure in the moving dark zone when the room influence model is updated relative to the positions of the moving zones. (——): Pressure in the moving dark zone when the room influence model is static, matching the reference positions (the white-filled squares in Fig. 21a). (- - -): Time interval where the updating room influence model matches the static model.

Fig. 21: Results and plots repeated from paper G [77]. Comparison of two zones moving with constant velocity along the trajectories indicated by the arrows and the control filters being updated along with the zone position or just being optimized for one fixed position.

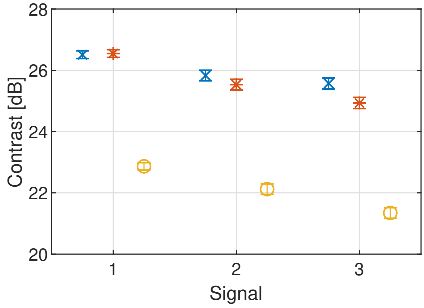
An additional benefit of the moving horizon framework is that it is applied directly to the audio content being reproduced, and not to the expected performance given a white noise playback signal. The effect of this is shown by comparing the performance of the moving horizon framework (with static zones) to the time-domain pressure matching solution as presented in Paper C [75]. This example is based on measured RIRs in the setup of Fig. 11b and utilizes separate measurements for calculating the filters and evaluating the results. The results are presented for three 10-second-long playback signals; signal 1 is white noise, signal 2 is electronic music, and signal 3 is rap / rnb music. The parameters of the moving horizon framework were adjusted to attain similar performance, relative to the static filters, in terms of reproduction error and loudspeaker input signal energy for signal 2. As seen from the results in Fig. 22, it is possible to attain a higher performance in terms of contrast while keeping the other two evaluation parameters almost identical.

3.2.4 Outcome of Paper G

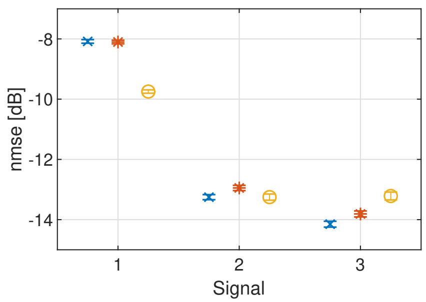
The moving horizon framework is a flexible formulation of the sound zones problem in state-space form which enables adaptation to time-varying effects. It offers improved performance at the expense of having to recompute the filters at each time-step. With the state-space formulation it can be extended to include physical models e.g. of a loudspeaker diaphragm displacement.



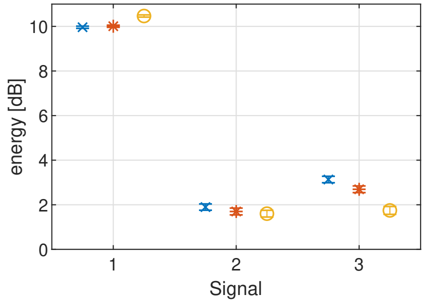
(a) Stationary setup. Eight 10" woofers are placed around the room. Six are placed at the edges of the room and the last two are placed as close to the zones as possible. The control regions are $0.2 \times 0.2 \times 0.2$ m^3.



(b) Contrast. Note that these results are determined in the time-domain, hence, the contrast is the ratio of spatially averaged squared pressures in the bright and dark zone at each time sample. For detailed description see Paper G [77].



(c) NMSE. The mean square error is here defined as the spatial average of squared deviation from the target sound pressure in the bright zone at every time sample. The normalization is the average of the squared target pressure in the bright zone across both the bright zone and the duration of the audio content. For a detailed description see Paper G [77].



(d) Loudspeaker input signal energy. The loudspeaker signal energy is defined as the sum of squared input sample levels to the loudspeakers at each time sample. For further details see Paper G [77].

Fig. 22: Results and plots repeated from Paper G [77]. Mean and 95% confidence interval over the last 11,000 samples of the three 10 s playback signals, down sampled to 1.2 kHz sampling frequency. The playback signals are: 1) white Gaussian noise, 2) electronic dance music, and 3) rap/rnb music. (×): Moving horizon knowing only the current audio sample. (∗): Moving horizon knowing the current and the coming audio samples. (○): Static filters as used in Paper C.

4 Discussion

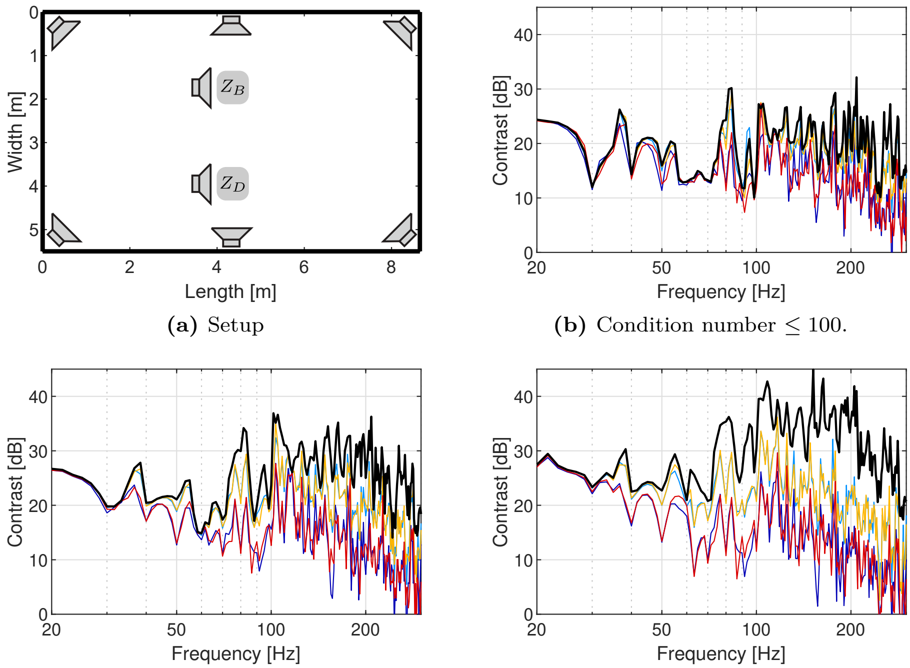
The contributions, summarized in the previous section, are a collection of confined studies focusing on limited aspects of sound zones. In this section, the results are collected and discussed in relation to a system for generating sound zones.

4.1 Significance of error sources in room influence model

The significance of the potential sources of mismatch between the room influence model and the reproduction system is application dependent. For instance, the resulting sound field from loudspeakers in a room is more complicated than a compact array in free-field. The sensitivity to changes in the room arises when the control method is able to utilize the room response to reproduce the tar-get sound field. In a room, a change in the speed of sound causes a change in the resonance frequencies of the room, whereas in free-field it only results in a phase-shift proportional to frequency and the change in the propagation speed. In Paper B [91], effects due to large changes in temperature were seen to severely decrease the contrast. To illustrate the sensitivity to small temper-ature variations a simple simulated example is included in Fig. 23. The results display the sensitivity to variations in the speed of sound associated with tem-perature changes up to 3°C. The results in Fig. 23 were simulated using Green's function in a rectangular room and applying the ACC method [27]. Tikhonov regularization has been adjusted at each frequency to ensure the condition number of the matrices involved in the ACC method stayed below 10^2, 10^3, and 10^4, respectively. From the plots it is seen that increasing the traditional regularization does not appear to improve the performance, it merely reduces the contrast of the scenario where the temperature has not changed. This indicates that high acoustic separation between sound zones is only feasible if the reproduction system does not deviate from the assumed room influence model.

The significance of measurement accuracy is highly dependent on the setting. While it is unlikely to be significant in a laboratory environment, it can present a problem in practical situations where the acoustic background noise is significant. It should also be noted that the measurement procedure is correlated with the resulting performance. If all the measurements are conducted simultaneously, acoustic background noise will introduce a coherent component between the microphones. On the other hand, if the measurements are conducted with a single microphone being moved between each measurement, the noise influence consists of independent observations. Thereby, the noise might be reduced due to averaging across the microphone positions. In Paper E [73] it was argued that the estimated uncertainty in the estimated room influence model should be included in the control method to automatically regularize the solution.

The above discussions indicate that temperature variations should be a



(a) Setup **(b)** Condition number ≤ 100.

(c) Condition number ≤ 1000. **(d)** Condition number ≤ 10000.

Fig. 23: Temperature sensitivity of acoustic contrast control simulated in the setup used in Paper E. The ACC loudspeaker weights are determined independently at each frequency at 20°C and evaluated using transfer functions where the temperature has been changed by −3°C (——), −1°C (——), 0°C (——), +1°C (——), and +3°C (——). The results are determined given diagonal loadings added to the transfer function matrices ensuring that the condition number does not exceed 100, 1000, and 10000, respectively.

concern when implementing sound zones in a laboratory environment. Outside the laboratory, measurement noise is an additional concern. Both effects are only significant if the control is based on in situ measurements. For compact loudspeaker arrays based on free-field simulations, the effects are unlikely to be significant, as discussed in Paper F [76].

4.2 Limitations of attainable contras

In Sec. 2.3.3, it was mentioned that 25 to 29 dB broadband contrast seems to be required for low distraction between the sound zones, assuming the reproduced playback signals to be loudness-matched in the zones. In order to achieve such level of separation requires a significant accuracy of the room influence model and careful decisions to be made during the design of the system to fit the room. The results presented in this dissertation indicate that such performance levels might be attained below approximately 300 Hz, assuming the reproduction system is perfectly known. From 300 Hz to around 1 kHz, it is unlikely that a compact loudspeaker array would be sufficiently directive to create significant contrast. This is related to the influence of the nearby boundaries affecting the radiated sound as mentioned in Paper F [76] and the effect of reverberation investigated in [105]. Hence, the most promising approach in this frequency range would be to violate the assumption of allowing the listeners to place the loudspeakers in their domestic room. By moving the loudspeakers close to the ears of the listener, the direct sound in the bright zone can be high relative to the direct and reflected sound reaching the dark zone. At higher frequencies beamforming solutions can be effective, but they rely heavily on the properties of the room. If sound is focused in the bright zone with a compact array, the sound pressure level in the dark zone depends on the acoustic power radiated by the array and the absorption of the walls reflecting the sound.

4.3 Sound field outside the controlled zones

In this work, the reproduced sound field is only evaluated in two control regions, namely the bright and dark zone. While this is sufficient to create and evaluate two sound zones it does not reveal what is happening outside the controlled zones. As the sound field is not controlled in this part of the room, it will generally contain a mixture of both playback signals. The severity of the leakage depends on the control method applied and the constraints introduced. For example, loudspeakers distributed around the room and characterized in the control regions leads to an uncontrolled sound field in the remainder of the room. While a mixture of playback signals is undesired for the listener, the uncontrolled sound field is only experienced outside the sound zones. This is part of the motivation for generating sound zones which are capable of following the listeners in the room, as introduced in Paper G [77]. The contrast performance depends on the size and location of the sound zones relative to the available loudspeakers. Hereby, a number of decisions must be made regarding the behavior of a dynamic sound zones system e.g. to determine what the system should do when two sound zones are very close to each other.

Approaches like beamforming seek to efficiently radiate sound to the target bright zone, hence, less energy is radiated into the room. Thereby, the effect of sound in the uncontrolled part of the room is reduced. However, this does not equate that the sound is imperceptible at other locations of the room. Beaming different playback signals to multiple control regions using a single loudspeaker array introduces an additional challenge. If the control regions should follow the listener in a room, the regions are not allowed to “block line-of-sight” between each other and the loudspeaker array. The immediate solution to this problem is to introduce additional loudspeaker arrays at different positions of the reflecting environment. This again poses questions regarding the system behavior of dynamic sound zones. A situation which should be considered is that it is likely confusing for the listeners, if the direction of sound impinging to them suddenly change as they move through the room.

4.4 General vs. specific sound zone systems

In Sec. 2.2.2, it was briefly discussed that the geometric layout of the control regions and the corresponding loudspeakers has great influence on the performance of a sound zones system. Therefore, the more assumptions which can be made regarding the position of listeners and their interaction with the sound zones system the better opportunity there is to optimize the loudspeaker locations and control methods to the particular scenario. In general, the fewer constraints which are enforced on the listeners' position and behavior, the more loudspeakers would be required to attain adequate control. As seen from the literature, many effective solutions use prior knowledge of the reproduction scenario to improve the performance. One example is to place loudspeakers in the headrest of an automotive style seat [26]. Other examples would be to optimize the geometric layout of loudspeakers to one given scenario. While this often includes placing loudspeakers close to the control regions, it could also include distributing the loudspeakers to have good coupling to the room, or to position the loudspeakers for being able to cancel the sound on the way from the bright zone towards the dark zone.

4.5 Three or more zones

The results for two control regions can be extended to three or more. With superposition, creating three sound zones is realized by choosing one control region as the bright zone and the other regions as a combined dark zone. This can be done for all three combinations of bright and dark zones. However, similarly to increasing the size of two control regions, the inclusion of additional regions will reduce the performance in each zone. Furthermore, if one set of loudspeakers is used to render all three playback signals, a higher strain is put on the loudspeakers with superposition of additional solutions. This is a concern related to nonlinear distortion in the loudspeakers.

5 Conclusion

5.1 Results

The goal of this study was two-fold: 1) To investigate the attainable separation between sound zones and factors limiting it. 2) To determine control methods to enabling constraints directly on the physical reproduction system and to control parameters related to sound quality.

Throughout this work, experimental investigations were conducted in order to determine factors limiting the attainable acoustical separation between sound zones. The initial experiment compared different control methods at low frequencies in a room using two subsequent measurement sets of room impulse responses. The first set of measurements was used to calculate loudspeaker control filters and the other was used to evaluate the result. The evaluation showed that the small differences between the measurement sets were sufficient to ensure that all the investigated control methods attained similar separation performance.

The initial experimental results indicated that additional constraints on the sound field in the bright zone could be added without deteriorating the acoustic separation. This was verified for a method reducing pre- and post-ringing artifacts in the reproduced sound field to improve the sound quality.

The potential influence of measurement noise was investigated in later experiments. While measurement noise is unlikely to be a concern in laboratory conditions where the background noise and the measurement duration can be controlled, this is not true outside laboratory conditions. Here, an approach to automatically assess and compensate for measurement noise is desired. The proposed solution relies on variational Bayesian inference to estimate both the transfer functions and the uncertainty in the estimate, which is included in the control method. Simulations were used to show that the type of noise had an impact on the resulting performance. Independent noise between microphones tended to act as regularization, whereas coherent noise had a pronounced effect on the resulting performance. The proposed control method was seen to have little effect in case of independent noise while improving the stability in case of coherent noise.

Time-varying changes in the room impulse responses were investigated for low frequency sound zones in a car cabin. The impulse responses were measured at $-2°$C and $+23°$C and it was seen that sound zones designed for one temperature performed poorly at the other. Besides a change in the speed of sound, the temperature was seen to alter the response of the loudspeakers. The temperature dependence of the loudspeakers was verified in a temperature-controlled climate chamber.

An investigation at mid to high frequencies was conducted with a compact loudspeaker array on the dashboard of the car. Control filters designed for free-field conditions and in situ measurements were used to evaluate the benefit of in situ measurements. The separation attained with the compact array was seen to increase when using in situ measurements in the control regions, but the results were sensitive to changes in the cabin due to the inclusion of two emulated passengers in the control regions.

The sensitivity to inaccurate measurements is generally treated by introducing regularization. While this typically increases the robustness of the control, it also decreases the performance relative to the scenario where the measurements are accurate. As an alternative, a method (known as moving horizon) was adapted from control theory to incorporate changes in the room impulse responses. An example application was sound zones moving to follow listeners at positions where the room impulse responses have been measured. Additionally, the proposed method was seen to outperform static time-domain filters in terms of separation.

The results of this work have indicated that several factors limit the attainable separation and need to be considered when creating sound zones. In order to work in real world conditions, it is important to target the mismatch between measurement and the state of the reproduction system. In this work 1) Measurement noise, 2) Temperature shifts, and 3) Surfaces close to a compact array, were shown to be of high importance. Robust and stable performance can be attained but generally leads to lowered performance relative to accurate measurements. The proposed solution is to incorporate knowledge of the uncertainty into the control method and adapt to changes in the reproduction system. Further work is required to infer and model the relevant changes for the adaptation.

5.2 Outlook

The perceptual investigations on sound zones indicate that around 25 to 29 dB of separation across the audible frequencies would be required to ensure most people would not be distracted by sound leaking between sound zones. The presented results indicate such performance is only attainable at low frequencies if the room impulse responses from loudspeakers to control regions are known with high precision relative to the time-dependent state of the loudspeakers and environment. At mid and high frequencies, beamforming seems to be inadequate to attain separation of this magnitude in an extended region of space. An implementation in a room with less than 0.33 s reverberation time (RT60) in the range 250 - 4000 Hz was seen to attain up to 20 dB separation between two planar sound zones of 0.53 m x 0.53 m using a 24-element line array [87]. In a room with longer reverberation time, this separation would likely be reduced. Alternative approaches in the literature utilize loudspeakers close to the ears of the listeners and only seek to control the sound field at the ears. This might be a suitable solution if the application allows such positioning of the loudspeakers.

5.3 Further Work

The temperature dependent changes of the room impulse responses measured in the automotive cabin indicated that multiple factors were changing with the ambient temperature. A future research direction could be to extend the experiments to identify and model the effects of loudspeakers, the behavior of air, and the acoustic impedance of the boundaries in the reproduction environment separately. The purpose would be to predict and compensate for temperature dependent changes in the environment.

Throughout the experimental investigations it has been seen how measurement noise and changes in the environmental conditions can change the performance of sound zones. While the measurement uncertainty can be estimated and introduced in the control of sound zones, the proposed procedure is based on steady state response measurements with multi-tone signals. It would be of interest to extend the procedure to deconvolution-based measurement techniques such as the exponential sweep. This would enable a formulation of the time-domain methods which inherently includes the uncertainty in the measurements.

The suggested moving horizon framework (Sec. 3.2.3) was proposed in order to incorporate time-varying models in the sound field control. While the formulation allows the update of the room impulse responses, it also predicts the consequences of current control decisions to future control scenarios. This extends naturally to e.g. constraining the loudspeaker diaphragm excursion to a linear range based on a short foresight of the audio content to be reproduced in the zones. Updating the transfer functions, as they change, requires knowledge of how they change. It is an active area of research to infer the sound pressure (or even the transfer function) at a position away from the physical microphone measurements.

A general challenge in creating sound zones is to determine the priority of the terms to be controlled. This challenge is two-fold as it is desired to define requirements for the performance of the system, but it might not be possible to fulfill the requirements. The problem arises when priority changes depending on the attainable performance. For example, it might be preferable for the listeners in two sound zones to agree on one common playback signal, rather than reproducing individual playback signals in the zones with clearly audible interference between the zones. Another aspect of this problem is the allowable deterioration in the sound quality of the reproduced playback signals relative to the attainable separation. The sound quality deterioration could e.g. be in terms of reduced uniformity of where the sound impinges from, spectral coloration of the reproduced sound, and time-smearing artifacts such as pre- and post-ringing. Future research could therefore investigate the primary factors for overall sound quality in sound zones and how to control the identified factors.

References

[1] *In-car noise field analysis and multi-zone noise cancellation quality estimation*, 2015.

[2] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, pp. 943–950, 1979.

[3] S. Atal and R. Schroeder, "Apparent sound source translator," Patent, 1966, uS Patent 3,236,949.

[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 zones," *J. Audio Eng. Soc.*, vol. 63, no. 1/2, pp. 78–89, 2015.

[5] K. R. Baykaner, "Predicting the perceptual acceptability of auditory interference for the optimisation of sound zones," Ph.D. dissertation, University of Surrey, 2014.

[6] T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *J. Acoust. Soc. Am.*, vol. 117, no. 4, pp. 2100–2111, April 2005.

[7] T. Betlehem and P. D. Teal, "A constrained optimization approach for multi-zone surround sound," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 437–440.

[8] T. Betlehem, "Acoustic signal processing algorithms for reverberant environments," Ph.D. dissertation, The Australian National University, 2005.

[9] C. M. Bishop, *Pattern Recognition and Machine Learning.* Springer, 2006.

[10] J. Borwick, *Loudspeaker and Headphone Handbook.* Butterworth, 1988.

[11] L. Brännmark, A. Bahne, and A. Ahlén, "Compensation of loudspeaker-room responses in a robust mimo control framework," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 6, pp. 1201–1216, October 2013.

[12] L.-J. Brännmark, "Robust audio precompensation with probabilistic modeling of transfer function variability," in *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct 2009, pp. 193–196.

[13] M. Buerger, C. Hofmann, and W. Kellermann, "Broadband multizone sound rendering by jointly optimizing the sound pressure and particle velocity," *J. Acoust. Soc. Am.*, vol. 143, no. 3, pp. 1477–1490, 2018.

[14] M. Buerger, R. Maas, H. W. Löllmann, and W. Kellermann, "Multizone sound field synthesis based on the joint optimization of the sound pressure and particle velocity vector on closed contours," in *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2015, pp. 1–5.

[15] M. Buerger, C. Hofmann, C. Frankenbach, and W. Kellermann, "Multizone sound reproduction in reverberant environments using an interactive least-squares filter design method with a spatiotemporal weighting function," in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 2017.

[16] 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, 2014.

[17] Y. Cai, M. Wu, and J. Yang, “Design of a time-domain acoustic contrast control for broadband input signals in personal audio systems,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 341–345.

[18] ——, “Robust time-domain acoustic contrast control design under uncertainties in the frequency response of the loudspeakers,” in *Proceedings of Inter-Noise 2014*, November 2014.

[19] A. Canclini, D. Markovic, M. Schneider, F. Antonacci, E. A. P. Habets, A. Walther, and A. Sarti, “A weighted least squares beam shaping technique for sound field control,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018, pp. 6812–6816.

[20] D. Cavides-Noval and J. Brunskog, “Parameter optimization of forward sound propagation models using bayesian inference for sound field control purposes,” in *Euronoise 2018 Proceedings*, 2018, pp. 2301–2308.

[21] S. Cecchi, A. Carini, and S. Spors, “Room response equalization - a review,” *Applied Sciences*, vol. 8, no. 1, 2018.

[22] A. Celestinos and S. B. Nielsen, “Controlled acoustic bass system (cabs) a method to achieve uniform sound field distribution at low frequencies in rectangular rooms,” *J. Audio Eng. Soc.*, vol. 56, no. 11, pp. 915–931, November 2008.

[23] J.-H. Chang and F. Jacobsen, “The effect of scattering on sound field control with a circular double-layer array of loudspeakers,” in *Audio Engineering Society Convention 132*, April 2012.

[24] 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, April 2009.

[25] 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, pp. 3060–3066, May 2009.

[26] J. Cheer, S. J. Elliott, and M. F. S. Gálvez, “Design and implementation of a car cabin personal audio system,” *J. Audio Eng. Soc.*, vol. 61, no. 6, pp. 412–424, 2013.

[27] J. Choi and Y. 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, 2002.

[28] J.-W. Choi and Y.-H. Kim, “Manipulation of sound intensity within a selected region using multiple sources,” *J. Acoust. Soc. Am.*, vol. 116, no. 2, pp. 843–852, August 2004.

[29] J.-W. Choi, Y. Kim, S. Ko, and J.-H. Kim, “Super-directive loudspeaker array for the generation of personal sound zone,” in *Audio Engineering Society Convention 125*, 2008.

[30] P. Coleman, P. J. B. Jackson, M. Olik, M. 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, April 2014.

[31] P. Coleman, P. J. B. Jackson, M. Olik, and J. A. Pedersen, "Personal audio with planar bright zone," *J. Acoust. Soc. Am.*, vol. 136, no. 4, pp. 1725–1735, October 2014.

[32] ——, "Stereophonic personal audio reproduction using planarity control optimization," in *The 21st International Congress on Sound and Vibration*, July 2014.

[33] J. Donley and C. Ritz, "Multizone reproduction of speech soundfields: a perceptually weighted approach," in *2015 IEEE Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2015, pp. 342–345.

[34] J. Donley, C. Ritz, and W. B. Kleijn, "Reproducing personal sound zones using a hybrid synthesis of dynamic and parametric loudspeakers," in *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, December 2016.

[35] ——, "Multizone soundfield reproduction with privacy- and quality-based speech masking filters," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 6, pp. 1041–1055, June 2018.

[36] W. F. Druyvesteyn and J. Garas, "Personal sound," *J. Audio. Eng. Soc.*, vol. 45, no. 9, pp. 685–701, September 1997.

[37] S. J. Elliott, J. Cheer, J. Choi, and Y. Kim, "Robustness and regularization of personal audio systems," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 7, pp. 2123–2133, Sep. 2012.

[38] S. J. Elliott and M. Jones, "An active headrest for personal audio," *J. Acoust. Soc. Am.*, vol. 119, no. 5, pp. 2702–2709, 2006.

[39] S. J. Elliott, J. Cheer, H. Murfet, and K. R. Holland, "Minimally radiating sources for personal audio," *J. Acoust. Soc. Am.*, vol. 128, no. 4, pp. 1721–1728, October 2010.

[40] V. Emiya, E. Vincent, N. Harlander, and V. Hohmann, "Subjective and objective quality assessment of audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2046–2057, 2011.

[41] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*, Feb 2000.

[42] ——, "Advancements in impulse response measurements by sine sweeps," in *Audio Engineering Society Convention 122*, May 2007.

[43] J. Francombe, R. Mason, M. Dewhirst, and S. Bech, "Determining the threshold of acceptability for an interfering audio programme," in *Audio Engineering Society Convention 132*, April 2012.

[44] J. Francombe, "Perceptual evaluation of audio-on-audio interference in a personal sound zone system," Ph.D. dissertation, University of Surrey, 2014.

[45] J. Francombe, P. Coleman, M. Olik, K. Baykaner, P. J. B. Jackson, R. Mason, M. Dewhirst, S. Bech, and J. A. Pedersen, "Perceptually optimised loudspeaker selection for the creation of personal sound zones," in *Audio Engineering Society Conference: 52nd International Conference: Sound Field Control - Engineering and Perception*, September 2013.

[46] J. Francombe, R. Mason, M. Dewhirst, and S. Bech, "Elicitation of attributes for the evaluation of audio-on-audio interference," *J. Acoust. Soc. Am.*, vol. 136, no. 5, pp. 2630–2641, November 2014.

[47] ——, "A model of distraction in an audio-on-audio interference situation with music program material," *J. Audio Eng. Soc.*, vol. 63, no. 1/2, pp. 63–77, 2015.

[48] M. F. S. Gálvez and S. J. Elliott, "The design of a personal audio superdirective array in a room," in *Audio Engineering Society Conference: 52nd International Conference: Sound Field Control - Engineering and Perception*, September 2013.

[49] M. F. S. Gálvez, S. J. Elliott, and J. Cheer, "Loudspeaker arrays for family tv," in *Proceedings of the 19th International Congress on Sound and Vibration (ICSV 19)*, July 2012.

[50] ——, "A superdirective array of phase shift sources," *J. Acoust. Soc. Am.*, vol. 132, pp. 746–756, 2012.

[51] M. F. S. Gálvez and F. M. Fazi, "Room compensation for binaural reproduction," in *EuroRegio 2016*, June 2016.

[52] W.-S. Gan, J. Yang, and T. Kamakura, "A review of parametric acoustic array in air," *Applied Acoustics*, vol. 73, pp. 1211–1219, 2012.

[53] J. Garcia-Bonito, S. J. Elliott, and C. C. Boucher, "Generation of zones of quiet using a virtual microphone arrangement," *J. Acoust. Soc. Am.*, vol. 101, no. 6, pp. 3498–3516, June 1997.

[54] P. C. Hansen, *Discrete inverse problems: insights and algorithms.* Society for Industrial and Applied Mathematics, 2010.

[55] F. Jacobsen and P. M. Juul, *Fundamentals of General Linear Acoustics.* Wiley, 2013.

[56] F. Jacobsen, M. Olsen, M. Møller, and F. Agerkvist, "A comparison of two strategies for generating sound zones in a room," in *Proceedings of 18th International Congress on Sound and Vibration*, 2011.

[57] W. Jin and W. B. Kleijn, "Theory and design of multizone soundfield reproduction using sparse methods," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2343–2355, 2015.

[58] M. Jones and S. J. Elliott, "Personal audio with multiple dark zones," *J. Acoust. Soc. Am.*, vol. 124, no. 6, pp. 3497–3506, December 2008.

[59] P. Joseph, S. J. Elliott, and P. A. Nelson, "Near field zones of quiet," *Journal of Sound and Vibration*, vol. 172, no. 5, pp. 605–627, 1994.

[60] W. Jung, S. J. Elliott, and J. Cheer, "Estimation of the pressure at a listener's ears in an active headrest system using the remote microphone technique," *J. Acoust. Soc. Am*, vol. 143, no. 5, pp. 2858–2869, 2018.

[61] O. Kirkeby and P. A. Nelson, "Reproduction of plane wave sound fields," *J. Acoust. Soc. Am.*, vol. 94, no. 5, pp. 2992–3000, 1993.

[62] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 189–194, March 1998.

[63] O. Kirkeby, P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, “Local sound field reproduction using digital signal processing,” *J. Acoust. Soc. Am.*, vol. 100, no. 3, pp. 1584–1593, September 1996.

[64] W. Klippel, “Nonlinear large-signal behavior of electrodynamic loudspeakers at low frequencies,” *J. Audio Eng. Soc.*, vol. 40, no. 6, pp. 483–496, 1992.

[65] H. Kuttruff, *Room acoustics, 5th edt.* Spon Press, 2009.

[66] T. Lee, J. K. Nielsen, and M. G. Christensen, “Towards perceptually optimized sound zones: a proof-of-concept study,” in *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 136–140.

[67] 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 *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 491–495.

[68] X. Ma, P. J. Hegarty, and J. J. Larsen, “Mitigation of nonlinear distortion in sound zone control by constraining individual loudspeaker driver amplitudes,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 456–460.

[69] X. Ma, P. J. Hegarty, J. A. Pedersen, L. G. Johansen, and J. J. Larsen, “Personal sound zones: the significance of loudspeaker driver nonlinear distortion,” in *Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control*, Jul 2016.

[70] ——, “Assessing the influence of loudspeaker driver nonlinear distortion on personal sound zones,” in *Audio Engineering Society Convention 142*, May 2017.

[71] 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, 2018.

[72] M. B. Møller, J. K. Nielsen, E. Fernandez-Grande, and S. K. Olesen, “On the influence of transfer function noise on low frequency pressure matching for sound zones,” in *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, July 2018, pp. 331–335.

[73] ——, “On the influence of transfer function noise on sound zone control in a room,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 9, pp. 1405–1418, 2019.

[74] M. B. Møller and M. Olsen, “Sound zones: On performance prediction of contrast control methods,” in *Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control*, Jul 2016.

[75] ——, “Sound zones: on envelope shaping of fir filters,” in *24th International Congress on Sound and Vibration (ICSV 24)*, July 2017.

[76] ——, “On in situ beamforming in an automotive cabin using a planar loudspeaker array,” in *Proceedings of the 23rd International Congress on Acoustics*, September 2019, pp. 1109–1116.

[77] M. B. Møller and J. Østergaard, “A moving horizon framework for sound zones,” *Submitted for review at IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2019.

[78] P. M. Morse and K. U. Ingard, *Theoretical Acoustics*. McGraw-Hill, 1968.

[79] J. Mourjopoulos and M. A. Paraskevas, "Pole andd zero modeling of room transfer functions," *Journal of Sound and Vibration*, vol. 146, no. 2, pp. 281–302, 1991.

[80] J. L. Mueller and S. Siltanen, *Machine learning a bayesian and optimization perspective*. Academic Press, 2015.

[81] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Am.*, vol. 66, pp. 165–169, 1979.

[82] P. Nelson and S. Elliott, *Active control of sound*. Academic Press, 1993.

[83] J. K. Nielsen, T. Lee, J. R. Jensen, and M. G. Christensen, "Sound zones as an optimal filtering problem," in *2018 52nd Asilomar Conference on Signals, Systems and Computers*, 2018, pp. 1075–1079.

[84] T. Okamoto, "Generation of multiple sound zones by spatial filtering in wavenumber domain using a linear array of loudspeakers," in *2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, 2014, pp. 4733–4737.

[85] ——, "Analytical methods of generating multiple sound zones for open and baffled circular loudspeaker arrays," in *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2015, pp. 1–5.

[86] T. Okamoto and A. Sakaguchi, "Experimental validation of spatial fourier transform-based multiple sound zone generation with a linear loudspeaker array," *J. Acoust. Soc. Am.*, vol. 141, no. 3, pp. 1769–1780, 2017.

[87] 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 *Audio Engineering Society Conference: 52nd International Conference: Sound Field Control - Engineering and Perception*, September 2013.

[88] M. Olik, P. J. B. Jackson, P. Coleman, and J. A. Pedersen, "Optimal source placement for sound zone reproduction with first order reflections," *J. Acoust. Soc. Am.*, vol. 136, no. 6, pp. 3085–3096, 2014.

[89] F. Olivieri, F. M. Fazi, P. A. Nelson, M. Shin, S. Fontana, and L. Yue, "Theoretical and experimental comparative analysis of beamforming methods for loudspeaker arrays under given performance constraints," *Journal of Sound and Vibration*, vol. 373, pp. 302–324, July 2016.

[90] F. Olivieri, F. M. Fazi, M. Shin, and P. A. Nelson, "Pressure-matching beamforming method for loudspeaker arrays with frequency dependent selection of control points," in *Audio Engineering Society Convention 138*, May 2015.

[91] M. Olsen and M. B. Møller, "Sound zones: on the effect of ambient temperature variations in feed-forward systems," in *Audio Engineering Society Convention 142*, May 2017.

[92] M. Olsen and M. Møller, "Sound zones," Master's thesis, The Technical University of Denmark, 2011.

[93] M. Olsen and M. B. Møller, "Sound zones: scattering study with head and torso simulator," in *Audio Engineering Society Conference: 52nd International*

Conference: Sound Field Control - Engineering and Perception, September 2013.

[94] J. Park, J. Choi, and Y. 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, June 2013.

[95] 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 *Proceedings of the 20th International Congress on Acoustics, ICA 2010*, 2010.

[96] J. A. Pedersen, “Adaptive bass control - the abc room adaptation system,” in *Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction*, May 2003.

[97] M. Poletti, “An investigation of 2-d multizone surround sound systems,” in *Audio Engineering Society Convention 125*, Oct 2008.

[98] J. Rämö, S. Bech, and S. H. Jensen, “Real-time perceptual model for distraction in interfering audio-on-audio scenarios,” *IEEE Signal Processing Letters*, vol. 24, no. 10, pp. 1448–1452, 2017.

[99] J. Rämö, L. Christensen, S. Bech, and S. H. Jensen, “Validating a perceptual distraction model using a personal two-zone sound system,” *Proceedings of Meetings on Acoustics*, vol. 30, no. 1, 2017.

[100] 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 *Audio Engineering Society Convention 141*, September 2016.

[101] J. Rämö, S. Bech, and S. H. Jensen, “Validating a real-time perceptual model predicting distraction caused by audio-on-audio interference,” *J. Acoust. Soc. Am.*, vol. 144, no. 1, pp. 153–163, 2018.

[102] F. Rumsey, *Spatial Sound*. Focal Press, 2001.

[103] D. H. M. Schellekens, M. B. Møller, and M. Olsen, “Time domain acoustic contrast control implementation of sound zones for low-frequency input signals,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 365–369.

[104] M. Schneider and W. Kellerman, “Multichannel acoustic echo cancellation in the wave domain with increased robustness to nonuniqueness,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 3, pp. 518–529, 2016.

[105] 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, 2014.

[106] M. F. Simón Gálvez, S. J. Elliott, and J. Cheer, “Time domain optimizatiion of filters used in a loudspeaker array for personal audio,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 11, pp. 1869–1878, 2015.

[107] S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, and F. Zotter, “Spatial sound with loudspeakers and its perception: a review of the current state,” *Proceedings of the IEEE*, vol. 101, no. 9, pp. 1920–1938, 2013.

[108] S. Spors, H. Buchner, R. Rabenstein, and W. Herbordt, "Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering," *J. Acoust. Soc. Am*, vol. 122, no. 1, pp. 354–369, July 2007.

[109] Z. Tu, J. Lu, and X. Qiu, "Robustness of a compact endfire personal audio system against scattering effects (l)," *J. Acoust. Soc. Am.*, vol. 140, no. 4, pp. 2720–2724, 2016.

[110] J. W. Marshall Leach, *Introduction to Electroacoustics and Audio Amplifier Design, 3rd edt.* Kendall/Hunt, 2003.

[111] D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 6, pp. 697–707, 2001.

[112] S. Widmark, "Causal mse-optimal filters for personal audio subject to constrained contrast," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 5, pp. 972–987, 2019.

[113] Y. J. Wu and T. D. Abhayapala, "Spatial multizone soundfield reproduction: theory and design," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1711–1720, 2011.

[114] ——, "Multizone soundfield reproduction using multiple loudspeaker arrays," in *Proceedings of 20th International Congress on Acoustics, ICA 2010*, August 2010.

[115] 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, 2017.

[116] ——, "Robust reproduction of sound zones with local sound orientation," *J. Acoust. Soc. Am.*, vol. 142, no. 1, pp. EL118–EL122, 2017.

[117] ——, "Robust personal audio reproduction based on acoustic transfer function modelling," in *Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control*, Jun 2016.

ISSN (online): 2446-1628
ISBN (online): 978-87-7210-512-3
AALBORG UNIVERSITY PRESS