KU LEUVEN

ARENBERG DOCTORAL SCHOOL
FACULTY OF ENGINEERING TECHNOLOGY
DEPARTMENT OF ELECTRICAL ENGINEERING

Efficient parametric modeling, identification and equalization of room acoustics

Giacomo Vairetti

Supervisor:
Prof. dr. ir. Toon van Waterschoot
Co-supervisors:
Prof. dr. ir. Marc Moonen
Prof. dr. ir. Søren Holdt Jensen
(Aalborg University, Denmark)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Technology (PhD)

July 2018

Efficient parametric modeling, identification and equalization of room acoustics

Giacomo VAIRETTI

Examination committee:
Prof. dr. ir. Hans Rediers, chair
Prof. dr. ir. Toon van Waterschoot, supervisor
Prof. dr. ir. Marc Moonen, co-supervisor
Prof. dr. ir. Søren Holdt Jensen, co-supervisor
(Aalborg University, Denmark)
Prof. dr. ir. Johan Suykens
Prof. dr. ir. Jan Swevers
Prof. dr. ir. Edwin Reynders
Dr. ir. Michael Catrysse
(CoEnCo, Belgium)
Prof. dr. ir. Richard Heusdens
(TU Delft, The Netherlands)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Technology

July 2018

Preface

Here I am, finally writing what feels like the final chapter of a very long story, and I am not only referring to the length of this manuscript, but to the journey that had led to it. It all started in a small town in the italian Alps, where a young boy with a passion for music started to take his first steps into the real world. Having realized that the career of the rock star was too arduous, he started to get interested in technologies related to sound and acoustics. "If you are not good enough at creating good music, at least you can help at making good music sound better", he told himself.

Fourteen years passed since then, and I couldn't have reached this point without the people that had a fundamental role in this story. Erasmus of Rotterdam, who pushed me to crawl out of my shell, to really leave my beloved mountains for the first time and fly to flatland Finland. Marie Skłodowska Curie, who supported me (financially) for years, made me experience new things, and introduced me to many interesting people.

My main supervisor Prof. Toon van Waterschoot (feels good to come right after Erasmus and Marie Curie, uhu?), who has been the best supervisor I could have hoped for. Apart from the excellent mentorship and his constant support, his enthusiasm and knowledge pushed me over the unavoidable difficulties and setbacks. I can't help but thinking of him not only as a supervisor, but also as a friend, and I will cherish many moments from these years, especially the unforgettable concert of our one-time band "the Reverberators" (featuring Nico and Mina).

My co-supervisor Prof. Marc Moonen for the excellent supervision he has provided throughout the years, for improving my scientific writing skills, for his ideas and suggestions that always turned out to be extremely valuable, and for his ability of showing me things from a different angle. I feel I could have taken more advantage of his expertise also in the second part of the Ph.D., but the cue for Friday meetings was often quite long.

My external co-supervisor, Prof. Søren Holdt Jensen, especially for hosting me in Aalborg and for his cheerful attitude. I wrote my first journal paper during my Danish winter, so I would call it a very fruitful visit.

A separate mention goes to Dr. Enzo De Sena. Working with him has been a pleasure and his help has been always very valuable. I wish him all the best for his academic career. Also to Prof. Vesa Välimäki from Aalto University (Finland), to whom I will always be grateful for giving me the opportunity to work on my Master's thesis in his department. I am quite sure this story would have taken a very different direction without his help.

Besides my supervisors, I would like to thank the chair and members of the examination committee for their time and their helpful feedback on my thesis. The preliminary defense was tough, but I enjoyed discussing with them about my research. Additionally, I would like to separately thank my assessors, Prof. Johan Suykens and Prof. Jan Swevers, for taking the time of following my progress throughout the Ph.D.

Next come my colleagues at ESAT. Five years is a long time, and long is the list of the people with whom I shared most of it. I hope I don't forget anyone. First the 'senators' of the DSP group: thanks Joe, Rodrigo, Bruno, Pepe, Paschalis, Gert C., Alex, and Johnny, for making me feel part of the group from the real beginning. Thanks to Marijn, Wouter B., Martijn, David, Jorge, Hassan, Hanne, and Rodolfo, who are now keeping up the name of the DSP group in Belgium and around Europe, and finally my current colleagues: Amin, Giuliano, Niccolò, Randy, Mina, Thomas, Jeroen, Filippos, Fernando, Wouter L., Robbe, Neetha, Duowei, Gert D., Maja and Mohit. Among these names are some of who I consider my best friends during my years in Leuven.

The DREAMS team, including both supervisors and researchers. If all our seasonal schools and other events have been so enjoyable and formative is mostly because of them. It has been a great experience from which I learned a lot, not only from a scientific perspective, but on many different levels. I would like to explicitly thank the fellows that survived the long Danish winter with me (Clément, Adam, Neo, and Adel), those who visited us in Leuven (Pablo and Ante), and Aldona for being so good at her job and such a positive presence at work. My thoughts also goes to Nejem, who will not be forgotten.

I had a great time in Leuven and made many friends. Leo, Oreste, Carlo, Daniele, Sophie, Alice, Ana, Dan, Daryna, Francesco, Marta, Baharak, Alessandra, Silvia, Maria, Ivana, Ewa, Gabriele, Federico, Enrico, Nina, Juan, Serkan, Matthew, Felipe, Iman, Lisa, Attilio, Marcello, and all the others I forgot to mention. Thank you all for the good times we spent together, all the fun, the drinking (maybe too much sometimes), the eating, the talking, the

chilling. I will never forget it, and I hope we will manage to stay in contact somehow. A special thank goes to Giuliano and Deniz, who I consider more than friends (cfr. Giuliano's Ph.D. thesis), and with whom I shared many things and moments. Thanks for your help during my 'writing days'. I am happy we will live only 130 km apart, even though in different countries.

And no, I didn't forget my friends in Italy, in particular those you really realize are special only when you are far away. I am thinking of you, Teo, Lomba, Lela, Bona and Silvia. But also Robi, Chicca, Aldo, Valerio, Claudio, my classmates from high-school, and all the other friends from Valsassina. And sorry Arpi, Cesare and Diego for breaking up the band, but thanks for all the fun we had playing together.

And then comes the family. Nothing would have been possible without my parents, Umberto and Margherita, who supported me on this journey. I am very thankful for that and I am sure you are proud of me, which makes me proud as well. I am especially grateful to my mom. She was not happy to see me leave at first, but she knew it was my wish and that it was for my best. Her constant presence on a distance has been a certainty all these years. Not less important are my siblings, for their encouragement and friendship. My brother Alessio, for passing me the passion for music, and my sister Cecilia, for being so kind and for raising two amazing young men, Davide and Stefano. And the rest of the family, of course, Paolo, Vichi, Giovanna and Ste, Vito and Corinne, Marina and Orazio, among the others.

And at last, because she is at the same time family and friend, comes Ece. I cannot thank her enough for her love and support in the difficult moments, especially in this last year, for all her patience and the time spent on the IC Brussels train to come and stay with me for the weekend. I hope I will be as supportive when her turn to finalize her Ph.D. comes. Thanks *sevgilim* for the amazing moments we already spent together, and the many more that are yet to come. After 4 years, with definitely too much commuting, our time has finally come to live and build something beautiful together.

I began by saying that this felt like the end of a long story, but I now realize it is just the beginning of a new chapter, full of new challenges, new experiences, new things, that will change my life for the better. I will bring with me what I learned and the great people I met in Belgium, and I hope the Netherlands will be as kind and generous to me as the countries that hosted me until now.

Giacomo Vairetti
July 2018

Abstract

Room acoustic signal enhancement (RASE) applications, such as digital equalization, acoustic echo and feedback cancellation, which are commonly found in communication devices and audio equipment, aim at processing the acoustic signals with the final goal of improving the perceived sound quality in rooms. In order to do so, signal processing algorithms require the acoustic response of the room to be represented by means of parametric models and to be identified from the input and output signals of the room acoustic system. In particular, a good model should be both accurate, thus capturing those features of room acoustics that are physically and perceptually most relevant, and efficient, so that it can be implemented as a digital filter and used in practical signal processing tasks.

This thesis addresses the fundamental question in room acoustic signal processing concerning the appropriateness of different parametric models for room acoustics. Most room acoustic signal processing algorithms rely on the simplicity and versatility of all-zero (AZ) models, which however may require a large number of parameters to approximate a room impulse response (RIR) with high accuracy. The main goal of this thesis is then to develop parametric models with the same modeling accuracy as AZ models, but with lower model complexity. Pole-zero (PZ) models and especially models based on orthonormal basis functions (OBFs) are investigated. The properties of OBF models, such as orthogonality and scalability, are exploited in the development of iterative scalable algorithms, which provide numerically well-conditioned estimates of the model parameters. The nonlinear problem of estimating the pole parameters from measured RIRs is approached with a grid-search method, which not only provides stable and accurate estimates, but also enables an arbitrary allocation of the spectral resolution. A reduction in the number of parameters of 50% compared to AZ models is achieved in full-band, and up to 75% in the low and mid frequencies. A further reduction is obtained by estimating a set of poles common to multiple RIRs, based on the physically-motivated assumption of the poles being independent of the loudspeaker and microphone positions.

In many algorithms for RASE applications, the RIR has to be identified from speech or audio input-output signals, typically using adaptive digital filters. Fixed-poles infinite impulse response (IIR) adaptive filters based on OBF models, or simply OBF filters, present interesting properties in terms of error performance and convergence of the filter coefficients, which are dependent on the number and position of the fixed poles. A grid-search approach has been adopted for the pole estimation also in the multi-channel identification case, thus avoiding the use of recursive nonlinear algorithms. The resulting iterative algorithm adapts the linear coefficients of the multi-channel OBF filter using a modified version of the normalized least mean squares (NLMS) algorithm, meant to deal with issues at very low model orders, whereas the standard NLMS is used to track correlation parameters, based on which a new pair of complex-conjugate poles is fixed in the filter. A significant improvement in terms of identification accuracy and convergence compared to finite impulse response (FIR) filters, as well as robustness with respect to changes in the microphone positions, is observed at low frequencies, especially in small or damped rooms. The reduction in the filter order and the use of a common set of poles also helps in addressing some of the issues encountered in RASE applications, such as echo path undermodeling in acoustic echo cancellation, or frequency allocation in inverse filtering for digital equalization.

Particular attention is addressed to the low-frequency region of modal resonances, where the acoustics of small rooms is typically more problematic. In this regard, a series of acoustic measurements have been performed in a rectangular room using a subwoofer as sound source. The issues of measuring RIRs at low frequencies, mostly related to high ambient noise and to the nonlinear distortions produced by the subwoofer, are addressed and partially solved by means of the exponential sine-sweep method, a careful calibration of the measuring equipment and postprocessing operations. Moreover, a novel procedure for estimating the frequency-dependent reverberation time is suggested.

Finally, two applications in the context of digital equalization are presented. The first introduces a design procedure for a low-order equalizer using parametric IIR filters with improved mathematical tractability of the equalization problem and other desirable properties, which is used for minimum-phase equalization of loudspeaker and room responses. The second application describes the implementation of an existing solution for nonminimum-phase multi-channel equalization of car cabin acoustics, which involves the modeling of different aspects of the acoustic transfer functions. The common-poles version of a modeling algorithm for PZ models is derived, and adapted for estimating excess-phase zeros, which are then used to compensate for nonminimum-phase distortions.

Korte Inhoud

Ruimteakoestische signaalverbetering (RASE), met toepassingen als digitale egalisatie, akoestische-echo- en feedback-onderdrukking die we terugvinden in heel wat communicatietoestellen en audio-apparatuur, beoogt de verwerking van akoestische signalen met als einddoel de verbetering van de geluidskwaliteit waargenomen in een ruimte. Daartoe maken signaalverwerkingsalgoritmes gebruik van parametrische modellen die de akoestische respons van de ruimte voorstellen en die geïdentificeerd worden op basis van input- en outputsignalen van het ruimteakoestische systeem. Een goed model moet enerzijds nauwkeurig zijn, om de fysisch en perceptueel meest relevante kenmerken van de ruimteakoestiek te kunnen weergeven, en anderzijds efficiënt zijn, zodat het kan worden geïmplementeerd als een digitaal filter voor gebruik in praktische signaalverwerkingstaken.

Dit proefschrift behandelt de fundamentele vraag in welke mate verschillende parametrische modellen voor ruimteakoestiek geschikt zijn voor ruimteakoestische signaalverwerking. De meeste ruimteakoestische signaalverwerkingsalgoritmes steunen op de eenvoud en veelzijdigheid van *all-zero* (AZ) modellen, hoewel die doorgaans een groot aantal parameters vereisen om een ruimte-impulsrespons (RIR) met hoge nauwkeurigheid voor te stellen. Het voornaamste doel van dit proefschrift bestaat bijgevolg in de ontwikkeling van parametrische modellen met dezelfde modelleringsnauwkeurigheid als AZ modellen maar met een lagere modelcomplexiteit. *Pole-zero* (PZ) modellen en in het bijzonder modellen gebaseerd op orthonormale basisfuncties (OBFs) worden hier onderzocht. De eigenschappen van OBF modellen, zoals orthogonaliteit en schaalbaarheid, worden aangewend in de ontwikkeling van iteratieve, schaalbare algoritmes die numeriek goed geconditioneerde schattingen van de modelparameters afleveren. Het niet-lineaire probleem om de poolparameters uit opgemeten RIRs te schatten, wordt benaderd via een *grid-search* methode die niet enkel stabiele en nauwkeurige schattingen oplevert maar ook een willekeurige allocatie van de spectrale resolutie toelaat. Ten opzichte van AZ modellen wordt een reductie van 50% in het aantal parameters behaald over de volledige bandbreedte, en tot

75% in de lage en middenfrequenties. Een verdere reductie wordt bekomen door een set van gemeenschappelijke polen voor meerdere RIRs te schatten, gebaseerd op de fysisch gemotiveerde veronderstelling dat de polen onafhankelijk zijn van de luidspreker- en microfoonposities.

In heel wat algoritmes voor RASE toepassingen dient de RIR geïdentificeerd te worden op basis van spraak- of audio-input-outputsignalen, typisch met behulp van adaptieve digitale filters. Adaptieve filters met een oneindige impulsrespons (IIR) en vaste polen gebaseerd op OBF modellen, kortweg OBF filters, bezitten interessante eigenschappen in termen van foutperformantie en convergentie van de filtercoëfficiënten, afhankelijk van het aantal en de positie van de vaste polen. Een *grid-search* methode wordt ook toegepast voor de schatting van de polen in het meerkanaals identificatiescenario, waardoor het gebruik van recursieve niet-lineaire algoritmes wordt vermeden. Het resulterende iteratieve algoritme past de lineaire coëfficiënten van het meerkanaals OBF filter aan door middel van een aangepaste versie van het *normalized least mean squares* (NLMS) algoritme, waardoor problemen bij erg lage modelordes kunnen worden vermeden, terwijl het standaard NLMS algoritme gebruikt wordt om correlatieparameters te volgen, op basis waarvan een nieuw paar complex toegevoegde polen wordt vastgezet in het filter. Een significante verbetering in termen van identificatienauwkeurigheid en convergentie vergeleken met eindige-impulsrespons (FIR) filters, alsook robuustheid ten opzichte van veranderingen in de microfoonposities, worden vastgesteld bij lage frequenties, in het bijzonder in kleine of gedempte ruimtes. De reductie in de filterorde en het gebruik van een gemeenschappelijke set polen draagt ook bij tot de oplossing van een aantal problemen die zich voordoen in RASE toepassingen, zoals ondermodellering van echopaden in akoestische-echo-onderdrukking of frequentieallocatie in inverse filtering voor digitale egalisatie.

Bijzondere aandacht wordt besteed aan het laagfrequente gebied van modale resonanties, een gebied waarin de akoestiek van kleine ruimtes vaak meer problematisch is. In deze context werd een reeks akoestische metingen uitgevoerd in een rechthoekige ruimte met een *subwoofer* als geluidsbron. De problemen inherent aan het meten van RIRs bij lage frequenties, die hoofdzakelijk samenhangen met sterke achtergrondruis en niet-lineaire vervormingen geproduceerd door de *subwoofer*, worden aangepakt en deels opgelost door middel van de exponentiële *sine sweep* methode, de zorgvuldige calibratie van de meetapparatuur en de nabewerking van de metingen. Daarnaast wordt een nieuwe procedure voorgesteld voor het schatten van de frequentieafhankelijke nagalmtijd.

Tot slot worden twee toepassingen in de context van digitale egalisatie voorgesteld. De eerste toepassing leidt een ontwerpprocedure in voor lage-orde egalisatie op basis van parametrische IIR filters. Deze procedure wordt gebruikt

voor minimumfase-egalisatie van luidspreker- en ruimteresponsen en vertoont wenselijke eigenschappen waaronder een elegante wiskundige beschrijving van het egalisatieprobleem. De tweede toepassing beschrijft de implementatie van een bestaande oplossing voor niet-minimumfase meerkanaalsegalisatie van de akoestiek in een wagen, waarin verschillende aspecten van de akoestische transferfuncties gemodelleerd worden. Een modelleringsalgoritme voor PZ modellen met gemeenschappelijke polen wordt afgeleid en aangepast om *excess-phase zeros* te schatten, die vervolgens gebruikt worden om niet-minimumfase vervormingen te compenseren.

Glossary

Acronyms

AD	analog-to-digital
AEC	acoustic echo cancellation
AFC	acoustic feedback cancellation
AIR	acoustic impulse response
ANC	active noise control
AP	all-pass
APA	affine projection algorithm
AR	auto-regressive
ARMA	auto-regressive moving-average
ATF	acoustic transfer function
AVR	acoustic virtual reality
AZ	all-zero
BB	block-based
BFGS	Broyden-Fletcher-Goldfarb-Shanno
BU	Brandenstein-Unbehauen
CAPR	common-acoustical-poles and their residues

CAPZ common-acoustical-poles and zeros

CR convergence rate

DA digital-to-analog

DCT discrete cosine transform

DFT discrete Fourier transform

DRC digital room correction

DTFT discrete-time Fourier transform

EDC energy decay curve

ERB equivalent rectangular bandwidth

ERLE echo return loss enhancement

ESS exponential sine-sweep

FIR finite impulse response

FPAF fixed-poles adaptive filter

FT Fourier transform

GF Green's function

GMP group matching pursuit

GN Gauss-Newton

HF high frequency

HP high-pass

IDFT inverse DFT

IF instantaneous frequency

IIR infinite impulse response

LASSO least absolute shrinkage and selection operator

LF low frequency

LIG linear-in-the-gain

LMS least mean squares

LP low-pass

LS least squares

LSDM log-spectral difference measure

LTI linear and time-invariant

MA moving-average

MFD matrix fraction description

MIMO multiple-input/multiple-output

MLS maximum-length sequence

MP matching pursuit

MSE mean square error

NF noise floor

NLIG nonlinear-in-the-gain

NLMS normalized least mean squares

NM normalized misalignment

NMSE normalized mean square error

NSSE normalized SSE

OBF orthonormal basis function

OMP orthogonal matching pursuit

PF parallel filter

PFE partial fraction expansion

PLDM perceptual linear distortion measure

PSD power spectral density

PZ pole-zero

RASE room acoustic signal enhancement

RHS right-hand side

RIM randomized image-source method

RIR room impulse response

RLS recursive least squares

RMS root mean square

RPE recursive prediction error

RRE room response equalization

RT reverberation time

RTF room transfer function

SAEC stereophonic acoustic echo cancellation

SB stage-based

SD steepest descent

SDM spectral distance measure

SFM spectral flatness measure

SIMO single-input/multiple-output

SISO single-input/single-output

SNR signal-to-noise ratio

SPL sound pressure level

SSE sum of squared errors

STMCB Steiglitz-McBride

TD transform-domain

TF transfer function

VAD voice activity detection

VSS variable step size

wBU warped BU

WFIR warped FIR

WIIR warped IIR

WN white noise

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Sound is all around us, and it surely is an important aspect of our lives. Through sound we communicate, with our voice and with music we express emotions, and from sound we extrapolate information about the environment we are in, becoming aware even of events out of our sight. Since, in modern life, we spend most of our time within four walls (actually six, considering floor and ceiling), what we most often hear is the combination of the sound emitted by a source and the result of the same sound interacting with the surfaces of the room. Indeed, the acoustic properties of the space, mostly its dimensions and the characteristics of the walls and objects within, determine how sound is modified before reaching our ears. Whenever a sound wave hits a surface, its energy is partially reflected in one or multiple directions and, after a number of reflections, it finally arrives at the listener. The buildup of all these reflected sounds, delayed in time and attenuated due to wall absorption, is called *reverberation*, and its importance to us is related to the features that it adds to the sound we hear.

Reverberation is sometimes a desirable property of rooms. We are so used to hearing sound in enclosed spaces that a complete lack of reverberation, as it can be experienced in an anechoic room, makes us uncomfortable. A moderate amount of reverberation is then favorable for the human voice to be perceived as more natural and pleasant. Moreover, reverberation is essential for music. Concert halls are specifically designed for reverberation to support the sound coming from the stage and give the listener the impression of being immersed in it [2]. On the other hand, 'poor' acoustics or excessive reverberation can be deleterious. Speech communication is adversely affected by strong reverberation, resulting in a reduction of intelligibility [3], whereas the perceived sound quality of music or speech is degraded if strong reflections are present and the acoustics

of the room is somewhat unbalanced. It is then often necessary to correct for the room acoustics and to compensate for its undesirable effects. And if optimization of the room dimensions and shape [4] or acoustic treatment using passive devices, such as absorbers and diffusers [5], are options for newly designed spaces and dedicated rooms, other solutions are needed in general.

In many practical *room acoustic signal enhancement* (RASE) applications, sound is reproduced in the acoustic environment through loudspeakers and captured by microphones. Digital signal processing tasks dealing with the enhancement of sound signals in rooms thus aim at processing the loudspeaker and microphone signals in the attempt to correct for the detrimental effects of reverberation, so as to improve the quality of the perceived sound signal.

Digital equalization of room acoustics [6, 7, 8, 9] aims at processing either the loudspeaker or the microphone signals in order to achieve a desired response at certain positions in the room. Two main scenarios are usually found in this context. In the *pre-equalization* scenario, commonly referred to as *digital room correction* (DRC), the signals are processed before being sent to the loudspeakers so as to pre-compensate for the undesired effects added to the sound by the room acoustics. Alternatively, *post-equalization*, also known as *dereverberation*, can be performed by processing the microphone signals in order to reduce the amount of reverberation and to partially restore the original source signal.

The purpose of *artificial reverberation* [10, 11] is not to reduce reverberation but rather to enhance it. For instance, the acoustic features of a room can be augmented by capturing sound and playing it back in the room after some manipulation [12, 13], or reverberation can be applied to a sound signal giving the impression that such signal was generated in a different acoustic environment, a very common practice in music and film post-production. This idea has been extended in recent years toward the development of *acoustic virtual reality* (AVR) systems [14, 15], where a simulated acoustic environment is reproduced through headphones or multiple loudspeakers to recreate a realistic listening experience, with applications in entertainment [16], acoustic design [17], and other fields [18, 19, 20, 21].

Another common room acoustic signal processing task is the elimination of artifacts such as echo and feedback from the microphone signals, which are familiar problems in hands-free communication and public address systems. Even though the purpose of *acoustic echo cancellation* (AEC) [22, 23] and *acoustic feedback cancellation* (AFC) [24] is not directly the correction of the room acoustics or the control of reverberation, the sound signal is modified by the acoustics of the room, so that the acoustic path between the loudspeaker and the microphone has to be simulated and applied to the loudspeaker signal to be able to remove the echo or feedback effect from the microphone signal.

In order to correct, enhance or synthesize the room acoustics, most of these applications require the acoustic response to be either measured or identified from the microphone and loudspeaker signals. The *room impulse response* (RIR) for given positions of the source and the receiver can be measured in various ways [25], from the recording of the room response to an impulsive source, produced e.g. by a handclap, a starting pistol or a popping balloon, to more sophisticated procedures involving the use of particular kinds of stimuli.

Once a RIR has been measured, it has to be represented in a form that can be used to process the sound signals in practice. *Parametric modeling* of room acoustics [26] aims at representing the input-output behavior of the system, i. e. its *room transfer function* (RTF), using a rational expression in the z-transform domain, which can be implemented as a digital filter. Models using *finite impulse response* (FIR) filters, whose parameters are the coefficients of the sampled and truncated RIR, are simple to use, but may require a large number of parameters in order to capture the true dynamics of the room response. Alternatively, a more compact representation can be achieved by employing models implemented as *infinite impulse response* (IIR) filters, which however present difficulties in the estimation of the filter parameters.

In recent years, parametric models based on *orthonormal basis functions* (OBFs) [27] received growing attention in the field of acoustic signal processing, since they allow to efficiently represent the resonant behavior of a room response by means of an IIR filter, without some of the difficulties of conventional models. Examples of their use are found in loudspeaker response equalization [28] and modeling of room, loudspeaker, and musical instrument responses [29, 30, 31, 32], speech synthesis [33], AEC [34] and AFC [35], and also in active noise control [36]. One partially unsolved issue, however, is the estimation of the nonlinear parameters of the model, which would normally require nonlinear optimization techniques.

Regardless of the model used, it is often impossible to first obtain a measurement of the room response. It is then necessary, in this case, to identify the parametric room model directly from the loudspeaker and microphone signals. Adaptive filters are normally used for this purpose, where the filter parameters are updated in steps at every new sample of the input signal based on some performance criterion, which may vary based on the specific task. Adaptive filters are also useful to track variations of the RTF in time, due to changes in the acoustics of the room or in the position of the source or the receiver. The added difficulty in the identification context is the fact that the input signal is often non-stationary and presents a non-white spectrum, as is the case for speech signals, so that the parameters of the adaptive filter may converge slowly toward an optimal solution.

The properties of OBF models provide some advantages compared to other IIR filters also in the identification context. Moreover, the adaptation of the linear filter coefficients follows the same rules under the same conditions as for FIR adaptive filters, such that most algorithms developed for FIR filters can be easily employed for the OBF counterpart with minor modifications. It follows that OBF models and the corresponding adaptive filters are good candidates to address some of the issues frequently encountered in RASE applications.

The rest of the chapter is organized as follows. Section 1.1 provides some basics and definitions of room acoustics useful for the understanding of the subsequent topics. A brief overview of methods for measuring RIRs is given in Section 1.2, whereas Sections 1.3 and 1.4 give some insight into modeling and identification of room acoustics systems with conventional and OBF parametric models. In Section 1.5, some of the RASE applications cited above are described. Finally, an overview and a chapter-by-chapter outline of the thesis is given in Section 1.6, in which the objectives and the contributions of this thesis are highlighted.

1.1 Room acoustics

The acoustics of an enclosed space is the result of the complex interaction of a number of different factors. The characteristics of the room, such as its dimensions, shape, and the acoustic properties of the walls and objects, determine in which ways a sound wave propagating in space is absorbed, reflected, or scattered in multiple directions.

In acoustic signal processing, the room is usually considered with good approximation as a linear system, where the output (the microphone signal) is the result of a linear transformation of the input (the loudspeaker signal), known as convolution. A room acoustic system can be then described by means of its RIR, i.e. the output response for an impulsive input signal, or its RTF, i. e. the mathematical function relating each value of the output of the room acoustics system to each value of the input signal.

Room modes

The RTF is defined in terms of the resonant vibrating *modes* of the room. Resonances are a characteristic of every enclosed space and are caused by standing waves resulting from the interaction of a sound wave traveling forward and backward between opposing walls. At points between the walls where the two waves interfere constructively (antinodes), the sound pressure of the standing wave is at its maximum level. Conversely, the destructive interference

of the two waves produces points in space with low sound pressure (nodes). Standing waves in rooms normally involve more than one pair of walls, with the result that points of high and low sound pressure are distributed unevenly in space, with patterns getting particularly complicated in rooms having an irregular shape.

The sound produced by the source is reinforced by a room mode only around the antinodes of its corresponding standing wave and only if the source driving frequency f_d is close to its resonant frequency f_i. Indeed, the frequencies at which standing waves are generated depend on the dimensions and shape of the room, so that the room modes are also unevenly distributed in frequency. Apart from the so-called *cavity mode* [37], i.e. the modal resonance centered at 0 Hz due to the air volume vibrations[1], the lowest resonant frequency of a mode in a given room is proportional to its largest dimension [1]. It follows that larger rooms have their first mode at a lower frequency compared to smaller rooms. Furthermore, room modes increase in number with increasing frequency, where the number of modes below frequency f in a room with volume V is approximated as [1]

$$N_f \approx \frac{4\pi}{3} V \frac{f^3}{c^3} \tag{1.1}$$

with c the sound velocity ($c \approx 343\,\mathrm{m/s}$ at $20\,^\circ\mathrm{C}$). As a consequence, smaller rooms have a lower number of modes at low frequencies compared to larger rooms.

As already mentioned, every time the sound wave with driving frequency f_d strikes a wall, part of its energy is reflected in the specular direction, giving rise to a standing wave if f_d is close to a modal frequency f_i. Another part is scattered or diffracted in different directions, if the size of the rough details and edges of the surface and the wavelength $\lambda = c/f_d$ are comparable, whereas some of the energy is transmitted outside the room directly through the walls or indirectly through other structures of the building. The remaining sound energy is dissipated into heat in the absorptive material covering the wall surface. The amount of energy that is absorbed depends on the properties of the material, such as its density, on the frequency and on the incident angle of the sound wave. Regarding its effect on room modes, sound absorption at the walls causes the energy of the standing wave to be reduced each time the backward and forward waves hit the surface. As a result, absorption defines the maximum amplitude that can be reached by the sound pressure of the standing wave, but also how its sound pressure decays in time.

[1] the effect of the cavity mode is a boost at low frequencies, which is heard in small rooms when a subwoofer is used.

Room impulse response

Following the description above, the RIR at a given position of the source $\boldsymbol{r}_s$ and of the receiver $\boldsymbol{r}$ is defined as the linear combination of the time decays of the energy of the room modes. Indeed, the sound pressure associated with each mode with resonance frequency f_i decreases exponentially with time according to the damping constant ζ_i, which is defined as the ratio between the sound energy absorbed by the wall and the total energy in the mode [37]. Under the reasonable assumption that $\zeta_i \ll f_i$, the RIR for a particular source-receiver position pair $\ddot{\boldsymbol{r}} = (\boldsymbol{r}_s, \boldsymbol{r})$ is then given by an infinite sum of exponentially decaying sinusoids as

$$h(\ddot{\boldsymbol{r}}, t) = \sum_{i=1}^{\infty} c_i(\ddot{\boldsymbol{r}}) e^{-\zeta_i t} \cos\big(2\pi f_i t + \phi_i(\ddot{\boldsymbol{r}})\big), \tag{1.2}$$

with the mode amplitude c_i and the phase ϕ_i depending on $\ddot{\boldsymbol{r}}$. Thus, changes in the positions of the source and the receiver result in variations of the amplitude and phase of the different modes, and consequently of the RIR [38]. As mentioned, the resonance frequency and the damping constant are determined by the properties of the room, and thus do not depend on the source-receiver positions. They do depend, however, on other factors, such as temperature, humidity, or changes in the position of objects (including people) in the room.

Reverberation time

The *reverberation time* (RT) of a room is defined as the time required for the sound pressure level to drop below 60 dB, and can be computed based on an average damping constant $\bar{\zeta}$ as

$$T_{60} = \frac{3\ln(10)}{\bar{\zeta}}. \tag{1.3}$$

However, absorption being frequency-dependent and non-uniformly distributed on the surfaces, different modes usually have different damping constants. Thus, some modes decay faster than others, possibly resulting in a non-uniform decay of the overall sound pressure. Typically, modes at low frequencies decay more slowly than modes at higher frequencies, where absorption of the walls (and in smaller part of the air) is more effective. It follows that the RT as defined above does not provide detailed information on how sound decays in a room at different frequencies. Larger rooms generally have longer RT, one reason being that the room dimensions are larger, so that a sound wave travels through space for a longer time before it strikes a wall. Another reason is that a larger

volume implies more surface area, usually having a limited amount of absorbing material, leading to a slower sound decay.

Modal overlap

The damping constant ζ_i also determines the *bandwidth* of a modal resonance, where the bandwidth is defined as the width of the resonance curve at -3 dB below the peak value ($B_i = \zeta_i/\pi$), with stronger absorption leading to larger bandwidths. The bandwidth of a resonance is commonly quantified by the dimensionless *Q-factor*, defined as $Q_i = {}^{f_i}/_{B_i}$, with larger resonances having smaller Q. A mode being characterized by a resonance curve with a certain bandwidth, it can be partially excited even when the driving frequency f_d does not coincide exactly with its modal frequency f_i. Furthermore, the resonant curves of multiple modes usually overlap, such that more than one mode can be partially excited at the same time. The amount of overlap of the modal resonances not only depends on the number of modes present at a certain frequency, according to the expression in (1.1), but on their bandwidth as well.

It follows that the overlap is relatively small at low frequencies, where absorption is usually lower, whereas modes tend to overlap more at higher frequency, where the mode bandwidth becomes larger. An indication about the separation between the two regions of weak and strong modal overlap is given by the so-called Schroeder frequency [39]

$$f_{\mathrm{Sch}} \approx 2000\sqrt{\frac{T_{60}}{V}} \tag{1.4}$$

where the information about the bandwidth of the modes is implicitly included in the RT. This expression is only indicative, and the distinction of the two regions not as clear, so that a transition region of moderately overlapping modes is typically considered in-between. In general, it suggests that a consistent modal overlap is found in large rooms already at very low frequencies, whereas small rooms tend to have a higher Schroeder frequency, below which room resonances are rather sharp and well separated. For this reason, small spaces are generally more problematic in the low frequency region, where the uneven frequency and space distribution of the room resonances produces large fluctuations of sound pressure, which introduce unpleasant *spectral coloration* in the perceived sound [40, 41, 42, 43, 44].

Room Transfer Function

As already mentioned, a room can be considered as a linear, causal, stable system with infinite degrees of freedom, whose input-output behavior can be characterized by its RTF, of which a graphical representation is given in Figure 1.1. Since our interest is in the processing of audio signals by means of digital filters, discrete-time signals are considered [45]. A digital audio signal is obtained either as a result of a discrete-time process or from a continuous signal by analog-to-digital conversion, i. e. by quantizing its amplitude and by sampling the signal at discrete time indexes $n = {}^{t}/_{T_s}$, where the sampling period T_s is the reciprocal of the sampling frequency f_s.

In a linear and time-invariant (LTI) system, the relation between the discrete-time input signal $u(n)$ and output signal $y(n)$ can be defined in terms of a *difference equation*, in which the value of the output signal at index n depends on a linear combination of its previous values and of the current and previous values of the input signal

$$\begin{aligned} y(n) = b_0\, u(n) + b_1\, u(n-1) + b_2\, u(n-2) + \dots \\ - a_1\, y(n-1) - a_2\, y(n-2) - \dots \end{aligned} \tag{1.5}$$

By introducing the backward shift operator q^{-1}, for which $u(n-1) = q^{-1}u(n)$, the above expression can be written as

$$(1 + a_1 q^{-1} + a_2 q^{-2} + \dots)\, y(n) = (b_0 + b_1 q^{-1} + b_2 q^{-2} + \dots)\, u(n)\,. \tag{1.6}$$

A similar relation is obtained by computing the z-transform of both sides of the difference equation in (1.6)

$$A(z)\, Y(z) = B(z)\, U(z) \tag{1.7}$$

where $Y(z)$ and $U(z)$ are the z-transforms of $y(n)$ and $u(n)$, respectively, and $A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \dots$ and $B(z) = b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots$ are polynomials in the complex variable z. The RTF is then the rational complex function relating the output to the input of the system, defined as

$$H(z) \triangleq \frac{Y(z)}{U(z)} = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots}{1 + a_1 z^{-1} + a_2 z^{-2} + \dots} = \frac{\sum\limits_{n=0}^{\infty} b_n\, z^{-n}}{1 + \sum\limits_{n=1}^{\infty} a_n\, z^{-n}}, \tag{1.8}$$

which corresponds to the z-transform of the sampled RIR sequence $h(n)$

$$H(z) = \sum_{n=0}^{\infty} h(n)\, z^{-n}\,. \tag{1.9}$$

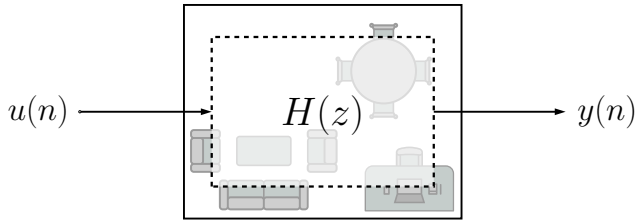

Figure 1.1: A representation of a LTI room acoustic system.

Notice that, if the RTF is evaluated on the unit circle in the complex z-plane, i. e. at $z = e^{j\omega T_s}$ with $\omega = 2\pi f$ the angular frequency, its *frequency response*, or *spectrum*, is obtained

$$H(e^{j\omega T_s}) = \sum_{n=0}^{\infty} h(n)\, e^{-j\omega T_s n}\,, \tag{1.10}$$

which is a continuous function of ω with period 2π, and corresponds to the *discrete-time Fourier transform* (DTFT) of $h(n)$. The spectrum $H(e^{j\omega T_s})$ is usually complex, such that it can be decomposed into its magnitude and phase spectra, as[2]

$$H(e^{j\omega}) = |H(e^{j\omega})| e^{j\angle H(e^{j\omega})}. \tag{1.11}$$

Poles and Zeros

The RTF can also be defined in terms of its zeros and poles. The numerator and denominator polynomials are factorized into first-order polynomials, giving

$$H(z) = b_0 \frac{(1 - q_1 z^{-1})(1 - q_2 z^{-1}) \dots}{(1 - p_1 z^{-1})(1 - p_2 z^{-1}) \dots} = b_0 \frac{\prod_{i=1}^{\infty} (1 - q_i z^{-1})}{\prod_{i=1}^{\infty} (1 - p_i z^{-1})}, \tag{1.12}$$

The numbers q_i and p_i are the roots of the numerator and denominator polynomials, respectively, and are called the *zeros* and the *poles* of the RTF.

[2] after normalizing, as commonly done, with respect to the sampling period T_s.

Another useful factorization of the RTF is obtained by performing a partial fraction expansion, thus obtaining an infinite summation of first-order terms as

$$H(z) = \sum_{i=1}^{\infty} \frac{R_i}{1 - p_i z^{-1}}, \tag{1.13}$$

where R_i is a (possibly complex) value called *residue* of the pole p_i. Since the coefficients of the polynomials $A(z)$ and $B(z)$ need to be real to have real-valued RIR, both the poles and their residues (as well as the zeros in (1.12)) must appear in the RTF either as real values or as pairs of complex-conjugate values. Poles and zeros are normally represented on the complex z-plane either in terms of their real and imaginary components or in polar form. A complex pole has the polar form $p_i = \rho_i e^{j\sigma_i}$, with $\rho_i = |p_i|$ its radius and σ_i its angle, and with $p_i^* = \rho_i e^{-j\sigma_i}$ its complex-conjugate (cfr. Figure 1.2).

First-order terms of the summation in (1.13) with pairs of complex-conjugate poles can be summed together to obtain second-order terms with real-valued coefficients

$$\begin{aligned} &\left\{ \frac{R_i}{1 - p_i z^{-1}} + \frac{R_i^*}{1 - p_i^* z^{-1}} \right\} = \frac{d_{i,0} + d_{i,1} z^{-1}}{1 - 2\rho_i \cos(\sigma_i) z^{-1} + \rho_i^2 z^{-2}} \\ &d_{i,0} = 2\,\mathrm{Re}\{R_i\} = 2|R_i| \cos(\angle R_i), \\ &d_{i,1} = 2\,\mathrm{Re}\{R_i p_i^*\} = 2|R_i|\, \rho_i \cos(\sigma_i - \angle R_i) \end{aligned} \tag{1.14}$$

By taking the inverse z-transform of the expression in (1.13), combining second-order terms together, the following expression is obtained,

$$h(n) = \sum_{i=1}^{\infty} 2|R_i| \rho_i^n \cos(\sigma_i n + \angle R_i), \tag{1.15}$$

which is an infinite summation of sampled exponentially decaying sinusoids and it is recognized as the discrete-time version of the RIR defined in (1.2). By comparing the two expressions, it is noticed that the amplitude c_i and phase ϕ_i of the mode responses in (1.2) are represented by the radius and angle of the complex-valued residues R_i, whereas, by substituting $t = nT_s$, the radius and angle of the poles represent the damping constants and modal frequencies, respectively, as

$$\rho_i = e^{-\zeta_i T_\mathrm{s}} \qquad \text{and} \qquad \sigma_i = 2\pi f_i / f_\mathrm{s}\,. \tag{1.16}$$

It follows that the position of the poles determines the frequency of oscillation and the time decay of the mode responses, whereas the residues (and thus the zeros) their amplitude and phase, which are a function of the source and the

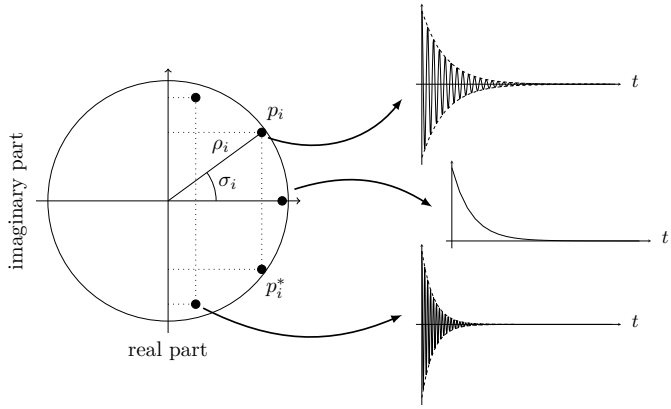

Figure 1.2: Modal responses in time associated to a real pole and two pairs of complex-conjugate poles.

receiver positions.

The fact that the mode responses decay with time implies that poles are distributed inside the unit circle ($\rho_i < 1$). Poles closer to the unit circle of the z-plane correspond to small damping constants ζ_i, and thus to room resonances with a narrower bandwidth B_i. Conversely, moving a pole towards the origin of the unit disc produces a wider resonance. A graphical representation of two pairs of complex-conjugate poles with different angle and radius, and their corresponding mode responses in time, are given in Figure 1.2. Finally, real poles correspond to non-oscillating exponentially-decaying functions with time decay determined by the pole radius and a low-pass characteristic in the frequency domain, so that a real pole corresponds to the 'cavity' mode at 0 Hz discussed in Section 1.1.

Regarding the zeros of the RTF, they generally appear both inside and outside the unit circle, such that the RTF can be factorized as

$$H(z) = \frac{B_{\text{in}}(z)B_{\text{out}}(z)}{A(z)} \tag{1.17}$$

with $B_{\text{in}}(z)$ and $B_{\text{out}}(z)$ the polynomials built from zeros inside and outside the unit circle, respectively. If $B_{\text{out}}(z) = 1$, the system is said to be *minimum-phase*. Minimum-phase systems, however, are rarely found in room acoustics. It was reported in [6] that the *nonminimum-phase* (or excess-phase) component of a RIR is mostly contained in the reverberation tail, and that rooms with very short RT can be approximately minimum-phase [46], especially if the direct

component has larger amplitude than the early reflections. Moreover, it has been noticed in [47] that the low-frequency response of a RIR is close to be minimum-phase. In all other cases, the RIR shows a significant nonminimum-phase component [48, 49], a fact that carries important implications in the equalization of the room response, as it will be discussed in Section 1.5.

Reflectogram

In the opening section of this chapter, reverberation has been described as the sound coming directly from the source and the collection of sound reflected from the surfaces arriving at the receiver with a certain delay after the direct sound. Even though the RIR is defined as in (1.2), its interpretation in terms of reflections can be useful in the analysis of reverberation. From the graphical representation of a RIR, sometimes called *reflectogram* [1], of which an example taken from the SMARD database [50] is given in Figure 1.3, it is possible to identify strong reflections and to evaluate the overall time decay. Apart from the direct component, i. e. the sound arriving directly to the receiver after a certain traveling time (propagation delay), reflections in a reflectogram are divided into early and late reflections. *Early reflections* are normally stronger, as they reach the receiver after the sound was reflected only a limited number of times. Strong early reflections arriving at the receiver within 50-100 ms after the direct sound are desirable to a certain extent, as they support the energy of the direct sound and improve intelligibility in noisy environments [51, 52]. However, a strong reflection with a delay exceeding 100 ms is likely to be perceived as echo and thus have a negative effect on speech intelligibility and sound quality.

Reflections, which are initially quite sparse, increase in number as a cubic function of time (analogously to the increase with frequency of the number of modes), with the average temporal density of reflections arriving at time t approximately given as [1]

$$\frac{dN_t}{dt} \approx 4\pi \frac{c^3 t^2}{V}. \tag{1.18}$$

It follows that after a certain time instant, called the mixing time [53, 54], *late reflections* start to be so dense and coming randomly from all different directions, that the sound field can be considered to be diffuse, set aside the low-frequency region where the effects of modal behavior prevail. Due to absorption, the energy of the late reflections decays exponentially with time, until it eventually fades out. The diffuse reverberation tail is the main source of the effect for which sound is perceived as prolonged in time and 'distant' in space [44], such that sometimes the term 'reverberation' is used to refer to the late reflections only.

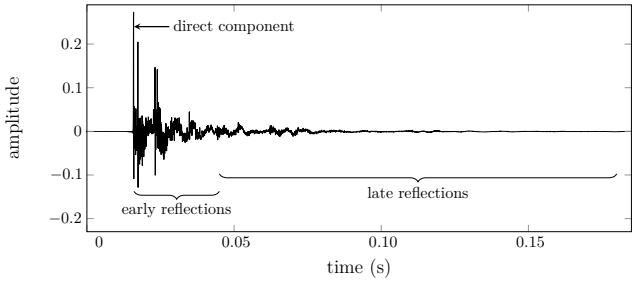

Figure 1.3: The reflectogram of a RIR.

1.2 Measuring room acoustics

Aside for rectangular rooms with very idealized conditions of their acoustic properties, for which analytical expression for the RIR and RTF can be obtained, information about the true poles and zeros of the RTF is generally not available. As a consequence, the RTF of a real space has to be obtained by other means. As the name suggests, a RIR can be obtained by recording the response of the room excited by an impulsive signal. However, exciting a room through a loudspeaker using an impulse with enough sound energy, such that the recorded response has a good signal-to-noise ratio (SNR) with respect to the ambient disturbances, is impossible in practice. Other ways of reproducing an impulsive signal, such as an exploding balloon or a starting pistol, present problems as well, such as lack of reproducibility and a non-flat response at all frequencies.

Nowadays, the availability of high-quality dedicated microphones and loudspeakers, together with advanced signal processing techniques, enabled the development of sophisticated methods[3], which can provide precise and reliable RIR measurements. A method for measuring RIR usually involves the generation of a particular kind of digital signal, which is amplified, converted to an analog signal and then reproduced inside the room through an omnidirectional loudspeaker with response as flat and linear as possible. The signal is then picked up by an omnidirectional microphone, also with response as flat and linear as possible, and then amplified and converted to the digital domain. Finally, postprocessing operations, depending on the method used, have to be performed in order to retrieve the RIR. Measuring RIRs in this way implies some assumptions to be made. The most important is that the acoustic system, comprising of the loudspeaker, the room, the microphone and other components, is considered to be a LTI system. However, as already mentioned, the RTF may vary in time and some components in the system may

[3] a review and some historical background can be found, for instance, in [25]

have a nonlinear behavior. This is the case, for instance, for the loudspeaker, which produces harmonic components and possibly other distortions [55] in the response, especially when it is driven at low frequency and at high levels.

When some of these assumptions are violated during the recording process, errors may appear in the retrieved RIR [56]. It follows that a reliable RIR measurement method should have the following properties:

1. controllability and reproducibility of the excitation signal,
2. robustness to RTF variations in time,
3. immunity to background and impulsive noise,
4. rejection of nonlinearities.

Many different methods have been suggested in the literature, each of which presents different characteristics with respect to the aspects just listed. Comparisons between the most commonly used methods have appeared in the literature [25, 57, 58]. Here two of these methods are mentioned.

The maximum-length sequence (MLS) method [59] uses an excitation signal which is a special binary sequence with flat magnitude spectrum and pseudorandom phase. If the sequence is long enough, its autocorrelation function approximates well an impulsive function. A good estimate of a RIR can be then obtained by circular convolution of the measured output with the time-reversed MLS sequence. Regarding the properties listed above, the MLS method is not able to reject nonlinear distortions, is not very robust to variations of the RTF, and it is not immune to the effect of noise. Indeed, as a consequence of its pseudorandom phase, deconvolving the MLS response evenly distributes the energy of any additional uncorrelated noise (stationary or impulsive) along the duration of the retrieved RIR. This results in a reduction of the SNR, which however can be increased by using longer sequences, although errors due to RTF variations are more likely to appear. The MLS method has been very popular in the past, as it was able to provide good RIR estimates in an inexpensive way.

Methods that gained wide-spread popularity in recent years involve the use of sweep signals [60]. Their main characteristic is that the instantaneous frequency of the signal increases with time. The RIR is retrieved by linear convolution of the measured output with the analytical inverse filter built from the time-reversed sweep signal, with some additional correction. Given that the ambient noise is normally more prominent at low frequencies, having a spectrum closer to 'pink' (-3 dB/octave) than to 'white', a better SNR is achieved using a sweep signal following an exponential time-frequency relation, hence its name

exponential sine-sweep (ESS) [61], which provides the system with more energy in the low part of the spectrum.

The time-frequency correspondence of the ESS signal provides a series of desirable properties. It is able to push a large part of the background noise into the non-causal part of the retrieved RIR, which is then discarded, as well as part of the energy of impulsive noise events, unless they occur during the final part of the sweep. Morever, also part of the nonlinear distortions produced by the loudspeaker are found in the non-causal part of the retrieved RIR. Finally, the ESS method proved to be less vulnerable to RTF variations in time than a linear sweep or the MLS method [58], such that extending the length of the sweep or averaging the responses obtained from multiple measurements can further increase the SNR without introducing significant errors.

Nevertheless, some of the nonlinear artifacts of the loudspeaker end up in the causal part of the retrieved RIR [58, 62]. In **Chapter 2**, the ESS method has been used to carry out RIR measurements in a rectangular room using a subwoofer as a source. This is a challenging scenario, as the subwoofer presents a strongly nonlinear behavior in the range of very low frequencies, where also the background noise is typically significant. It follows that, in order to obtain RIR measurements of good quality, the favorable properties of the ESS method have to be combined with a careful calibration of the measurement setup.

Analyzing room responses

Once a RIR is acquired, some analysis is necessary to first assess the quality of the measurement itself, and then to gain some insight into the acoustics of the room that has been measured. The RIR, or better its reflectogram, is analyzed to detect the presence of strong reflections, especially those that may be perceived as echo, to assess the density of reflections in different portions of the RIR [63], and to have a first impression of the reverberant characteristics of the room. If the ESS method was used, the non-causal part of the response is also of interest to assess the level of harmonic nonlinearities of the loudspeaker.

The analysis in the frequency domain provides other kinds of information. The spectrum of the retrieved RIR, defined in (1.10), is computed in practice for N discrete frequency values ω_k, uniformly distributed between 0 and 2π, giving the *discrete Fourier transform* (DFT),

$$H(e^{j\omega_k T_s}) = \sum_{n=0}^{N-1} h(n)\, e^{-j\omega_k T_s n} = \sum_{n=0}^{N-1} h(n)\, e^{-j\frac{2\pi k n}{N}} \,. \tag{1.19}$$

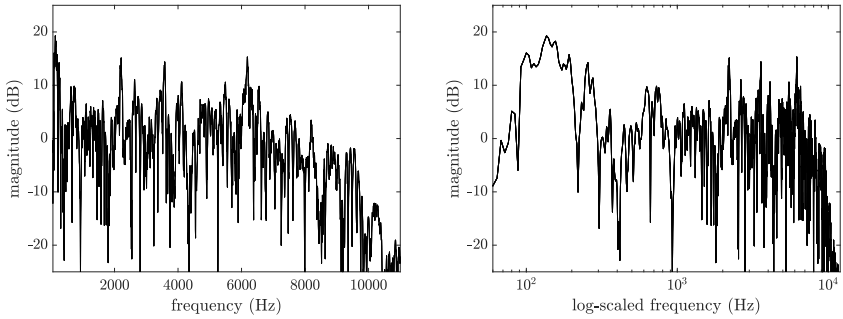

Figure 1.4: The magnitude frequency response of the RIR in Figure 1.3 on a linear (left) and logarithmic (right) frequency scale.

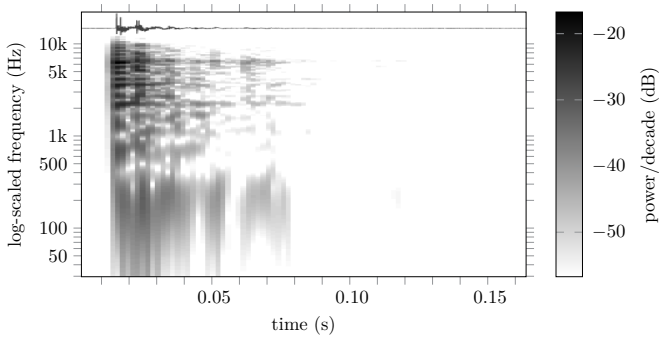

Figure 1.5: The spectrogram of the RIR in Figure 1.3.

The magnitude frequency response, obtained as the complex modulus of the DFT, provides insight into the main low-frequency resonances of the room, their bandwidth and the degree of modal overlap, and can give an indication of the amount of absorption present in the room at different frequency regions. The time and frequency representations are combined in the so-called *spectrogram*, i. e. the squared modulus of the short-time Fourier transform (STFT), obtained by computing the DFT of overlapping portions of the RIR sequence, with their duration determining a trade-off between resolution in time and in frequency. The analysis of the spectrogram is useful to asses the energy decay of the response at different frequencies and to identify particularly energetic and slowly decaying modes, as well as possible errors, such as impulsive noise or other artifacts, in the measured RIR. The analysis of Figures 1.4 and 1.5, for instance, reveals the presence of strong and slowly decaying resonances between 100 Hz and 200 Hz, but also few isolated components at higher frequencies.

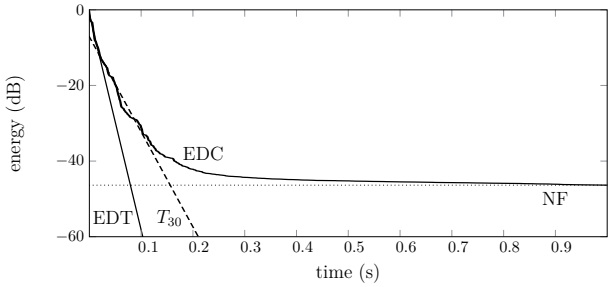

Figure 1.6: The EDC of the RIR in Figure 1.3, and the regression lines for the estimation of the EDT, the T_{30} and the NF.

Measuring reverberation time

Another important tool for RIR analysis is the *energy decay curve* (EDC), which can be obtained by backward integration of the RIR, converted to a logarithmic scale, as

$$\mathrm{EDC}(t) = 10\log_{10}\left(\frac{\int_t^\infty h^2(t)d\tau}{\int_0^\infty h^2(t)d\tau}\right)\ [\mathrm{dB}], \tag{1.20}$$

also known as Schroeder integral, which represents the normalized amount of energy remaining in the RIR at a given time t [64]. The RT, or T_{60}, is then defined as the time instant at which the EDC reaches -60 dB, or in other words the instant at which the energy of the modal responses have decayed, on average, below that level. However, the presence of noise in the measurement produces a bias in the EDC, which is useful to estimate the noise level, or *noise floor* (NF), but prevents the direct use of the definition to find the RT.

As discussed in the previous section, the energy of a RIR ideally decays exponentially with time, such that the EDC on a logarithmic scale presents a linear trend, at least until the NF is reached. It is then possible to obtain an estimate of the RT as the time instant at which a regression line fitting the first portion of the EDC reaches -60 dB. Frequency-dependent values of the RT are generally estimated by filtering the RIR with a bank of full-octave or one-third-octave band-pass filters. Moreover, by fitting the regression line on different portions of the EDC, more information can be inferred. The first 10 dB drop of the EDC determines the *early decay time* (EDT), which contains information about the level of early reflections, whereas fitting a regression line on different portions of the EDC, such as between -5 dB and -15, -25, or -35 dB, giving respectively the T_{10}, the T_{20}, and the T_{30}, provides a means of estimating

the RT in noisy conditions and of detecting complex modal decays, such as beating modes or double decays that often occur at low frequencies.

The estimation of the RT is indeed more involved at low frequency, where not only the EDC often shows complex decays, but also high levels of the NF. Moreover, fractional-octave filterbanks cannot be used, as the band-pass filter response is usually longer than the modal decays, leading to a significant over-estimation of the RT. Different procedures have been proposed for the estimation of the RT and the modal decay at low frequencies [65], by using modeling techniques or NF reduction methods, whereas the over-estimation problem has been addressed using different types of filterbanks [66]. In **Chapter 2**, a procedure for RT estimation at low frequency is described, which combines a fixed-bandwidth filterbank and a scalable modeling algorithm, described in Part II, allowing to obtain a noiseless approximation of the RIRs as OBF models and thus a more reliable estimate of the frequency-dependent RT.

1.3 Modeling room impulse responses

In order to perform room acoustic signal processing tasks, it is usually necessary to have a model of the room acoustics. Of interest in this work is the family of room acoustic models called *parametric models*, which are commonly used in practical applications. Parametric modeling of room acoustics aims at finding a meaningful approximation of the RTF, in one of its definitions given in Section 1.1, normally starting from measured RIRs. A RTF is the result of the combination of an infinite number of resonant responses. By limiting the frequency range of interest to the audible spectrum, i. e. by sampling the measured RIR at, for instance, $f_s = 48\,\text{kHz}$, the number of resonances becomes finite, but still extremely large, given the relation in (1.1) for $f = {}^{f_s}/_2$.

The idea of parametric modeling is then to represent a room response with a number of parameters that is large enough to model its essential features, but at the same time small enough to be able to perform real-time signal processing tasks. Indeed, an advantage of parametric models is that the approximated RTF can be directly implemented using a digital filter, which is then used, with or without manipulation, to process the audio signals in order to synthesize or control the acoustic response of the room.

1.3.1 Conventional parametric models

Different models within the family of parametric models have been developed, which can be distinguished by the form of the RTF they try to approximate.

All-zero models

The *all-zero* (AZ) model [26] (also known as moving average (MA) model) corresponds to the approximation of the RTF defined as in (1.8) with $A(z) = 1$, where the polynomial $B(z)$ is truncated to an order M. The transfer function of the AZ model then approximates the z-transform of the RIR $h(n)$ in (1.9), which is implemented as an FIR filter with M coefficients. The filter parameters are the coefficient values of the sampled RIR, truncated to the sample index M. Filtering a digital audio signal $u(n)$ using a causal FIR filter with M coefficients b_m (with $m = 0, \ldots, M-1$) corresponds to performing a discrete-time convolution, indicated with the symbol $*$, as [67]

$$y(n) = \sum_{m=0}^{M-1} b_m u(n-m) = \sum_{m=-\infty}^{\infty} h(m)u(n-m) \triangleq (h * u)(n) \tag{1.21}$$

which corresponds to the first M terms of the top row of the difference equation in (1.5). Notice that the output signal at previous time indexes does not contribute to its current value, or, in other terms, no feedback loop is present in the filter structure. As a consequence, the filter response has a finite duration of M samples. This is the main drawback of AZ models, which may require a large number of parameters to describe the essential properties of a RIR with a slow decay. Indeed, the RIR has an infinite impulse response, due to its resonant components, such that models implemented as IIR filters may be more appropriate.

All-pole and pole-zero models

The direct way to obtain an IIR filter is to model also the denominator polynomial of the RTF, so to obtain a recursive filter by including a feedback loop in the filter structure. Modeling the polynomial $A(z)$ then implies the estimation of the pole parameters. The *all-pole* model [26] (also known as autoregressive average (AR) model) corresponds to the approximation of the RTF defined as in (1.8) with $B(z) = b_0$, where the polynomial $A(z)$ has a finite order P. The all-pole model is able to approximate the resonant characteristics of the magnitude frequency response of the RTF, but generally not its phase response. Not having zeros in the transfer function implies that the all-pole

model can only approximate the minimum-phase characteristics of the system, making it not suitable to model nonminimum-phase RIRs.

If both the numerator and the denominator polynomials in (1.8) are used, with order M and P respectively, the *pole-zero* (PZ) model [68] (or ARMA model) is obtained, which is capable of modeling both the magnitude and phase response of the RTF. The approximation of a RTF using a PZ model can lead to a more compact representation, compared to the one achieved with an AZ model, but it also presents some difficulties. One aspect to be considered is the selection of the polynomial orders M and P, i.e. the number of model parameters. The choice can be made by either selecting orders which provide a good approximation of the target measured RIR, which involves performing the parameter estimation multiple times, or by relying on some a priori knowledge of the system, such as the number of room resonances to be modeled. Also, given that M and P are not required to be equal, estimating their optimal values with respect to the approximation of a given RTF is not a trivial task.

Another difficulty is that, differently from AZ and all-pole models for which a closed-form solution for the parameter estimation problem is available, the computation of the parameters of a PZ model is more involved. If the estimation algorithm aims to solve the *output-error* problem, i. e. to minimize the difference between the target RTF $H(z)$ and its approximation $\hat{H}(z)$,

$$E_{\mathrm{oe}}(z) \triangleq H(z) - \hat{H}(z) = H(z) - \frac{\hat{B}(z)}{\hat{A}(z)}, \tag{1.22}$$

nonlinear optimization algorithms have to be employed, with ensuing problems related to convergence to local minima and issues due to finite numerical precision [69].

Instead, linear regression methods, for which a closed-form solution is given by the least squares (LS) estimator, can be used if the so-called *equation-error* is minimized

$$E_{\mathrm{ee}}(z) \triangleq \hat{A}(z)H(z) - \hat{B}(z), \tag{1.23}$$

which is obtained from (1.22) by multiplying both the RTF and its approximation with the transfer function $\hat{A}(z)$. This operation introduces a frequency weighting in the estimation procedure, thus producing a biased estimate. In practice, less weight is given to parts of the spectrum with larger energy, such that the peaks of the RTF may not be modeled accurately.

The Steiglitz-McBride (STMCB) method [70] starts from an all-pole model, e.g. estimated using linear prediction methods [71] or the Prony's method [72], and then iteratively solves the equation-error problem. The bias issue is alleviated by compensating for the frequency weighting based on the denominator polynomial

estimated at the previous iteration. The biggest problem with this algorithm, especially for high model orders, is the potential instability resulting from poles estimated outside the unit circle.

A method similar to the STMCB method, but directly solving the output-error problem and less prone to instability issues, is the so-called Brandenstein-Unbehauen (BU) method [73]. This method consists of an FIR-to-IIR conversion algorithm, which iteratively estimates the denominator polynomial by minimizing the energy of the output of an all-pass filter with the same denominator, fed with the time-reversed RIR $h(-n)$, based on the concept of 'complementary signal' [74]. The numerator coefficients are then estimated by interpolation in closed-form according to a theorem by Walsh [75].

In all these cases, the RTF is approximated as a rational function with a form as in (1.8) which can be implemented in IIR filters in direct form. In other words, the estimated model parameters are the coefficients of the polynomials, which correspond to the coefficients of a direct-form filter. If different implementation forms are desired, such as cascaded or parallel forms [45], the estimated RTF has to be factorized into first- and second-order IIR filters sections. The factorization implies the computation of the roots of the polynomials $B(z)$ and $A(z)$, which becomes problematic when the order of the polynomials is high, because of numerical limitations. Algorithms for factorizing high-degree polynomials exist, which however involve nonlinear optimization techniques [76, 77]. For this reason, the estimation methods described above may not represent an optimal choice for approximating a RTF in pole-zero form.

Pole-zero models in parallel form

Particular attention is given here to the parallel form of PZ models [45], here referred to as *parallel filter* (PF) model, which corresponds to the finite-order approximation of the RTF as a finite summation of first-order terms as in (1.13) or of second-order terms as in (1.14). The advantage of PZ model in parallel form is the possibility of fixing the poles in the model structure, thus determining the frequency and damping of the exponentially-decaying sinusoidal components in (1.15), whereas their amplitude and phase parameters, which appear linearly in the model, can be estimated in closed-form by linear regression. The RIR is then approximated as a linear combination of P (complex- or real-valued) basis functions ψ_i, i. e. the exponentially-decaying sinusoids, built from a set of poles $\boldsymbol{p}$, each one weighted by a linear parameter d_i as

$$\hat{h}(n) = \sum_{i=1}^{P} \psi_i(n, \boldsymbol{p}) d_i \,. \tag{1.24}$$

It is possible in this way to allocate arbitrary frequency resolution by choosing appropriate radius and angle of the poles, for instance by matching the resolution of the human hearing using a logarithmic or a Bark frequency scale [78]. Another advantage over other PZ model forms is that PFs enable the use of parallel computing, providing a useful implementation tool for real-time audio processing applications [79].

These ideas have been exploited quite recently in many signal processing applications, such as equalization [80, 81, 82], artificial reverberation [83], active noise cancellation in headphones [84], head-related transfer function approximation [85] as well as modeling and synthesis of room, loudspeaker, and musical instrument responses [86, 87, 88]. However, some issues still remain, mostly related to the estimation of the model parameters from a target response. Even when the nonlinear problem inherent to pole-zero model estimation is avoided by predetermining the pole distribution [80, 89], the linear regression problem for estimating the numerator coefficients (i. e. the residues) may be very ill-conditioned, especially for high model orders, thus resulting in inaccurate estimates[4]. As for the estimation of the pole parameters from measured RIRs, the nonlinear methods cited above can be used, but the factorization of the denominator polynomial into its pole form is required, with the ensuing numerical issues. Finally, the PF filter is not suitable to model RTFs where the same pole appears more than once, in which case the filter structure should be modified accordingly [67].

Warped models

A mapping of the frequency resolution that resembles that of the auditory system can be also achieved by *frequency warping* [91]. Warping a RIR having N coefficients consists in filtering the RIR sequence with a series of N first-order *all-pass* (AP) filters with transfer function $D(z)$, corresponding to the mapping

$$z^{-1} \leftarrow D(z) = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}}, \tag{1.25}$$

with λ a pole parameter which can be tuned, for instance, to obtain a frequency mapping closely approximating the Bark scale [91, 78]. The so-called *warped FIR* (WFIR) filters and *warped IIR* (WIIR) filters are then obtained by substituting each unit delay z^{-1} in the transfer function of the standard FIR and IIR filters with a first-order AP filter as defined above[5].

[4] the regression matrix has large condition number, which results in numerical inaccuracies in the LS estimation [90].

[5] given the presence of AP filters in its transfer function, a WFIR filter, despite its name, has actually an IIR.

Also warped filters have been used for room response modeling [86], and other audio applications [91, 92, 93, 94, 95, 96, 97, 98, 99, 100]. In room response modeling, the RTF obtained from the truncated RIR as in (1.9) can be transformed to the warped domain using the mapping in (1.25), and modeled with a WFIR or a WIIR filter. Alternatively, and more commonly, the estimation of the filter parameters is not done with respect to the warped model structure. Instead, the estimation procedure consists in pre-warping the measured RIR, thus increasing resolution at low frequencies, estimating the model parameters in the warped domain using one of the methods designed for standard models, and finally remapping the estimated parameters to the original 'unwarped' domain. Also in this case, however, problems with high model orders may arise [91], and practical issues have to be considered for the implementation of WIIR filters [101, 102].

Pole-zero models with common poles

When modeling room acoustic systems with multiple inputs and multiple outputs, the RTF has to be estimated for each source-receiver pair. In order to further reduce the number of parameters required to approximate a set of RTFs, models based on common acoustical poles have been proposed, relying on the fact that the poles, i. e. the resonant frequency and bandwidth, do not depend on the source and receiver positions. The *common-acoustical-poles and zeros* (CAPZ) model [103] approximates multiple RTFs defined in the rational form in (1.8) by estimating a common set of denominator parameters and multiple sets of position-dependent numerator parameters. The parameters are computed by minimizing the average equation-error computed on the different RTFs, or by averaging multiple sets of denominator parameters obtained by modeling each RTF individually. The CAPZ model has been applied in different contexts, such as multi-channel equalization [104, 105], HRTF modeling [106, 107] and AEC [100, 108].

Similarly, the *common-acoustical-poles and their residues* (CAPR) model [109] approximates a set of RTFs in the pole-residue form in (1.13), where the poles are common to every RTF, while the residue for each individual RTF can be obtained via linear regression or linear prediction methods. By fixing the common poles in the second-order denominator polynomials, the residues values can also be used to define the position-dependent variations of the RTF, thus allowing to devise strategies for spatial interpolation and extrapolation of RTFs, as suggested in [109]. In the same work, the poles are estimated by assuming negligible absorption and by performing a search over a set of possible resonance frequencies distributed in the range of interest at intervals of 1 Hz. The common

poles are then obtained as a by-product of the minimization of the interpolation error computed at uniformly spaced positions.

Also output-error algorithms can be extended to the common denominator case. For instance, the BU method [73] is easily modified to the multi-channel case, as was proposed in the past for digital communication [110] and automation [111] applications, and recently employed in the context of room acoustic modeling [32] and of the characterization of the feedback path for AFC in hearing aids [35].

1.3.2 Models based on orthonormal basis functions

An alternative to conventional parametric models presented is provided by models based on OBFs [27], hereafter called *OBF models*. OBF models can be regarded as a generalization of some of these conventional models. For instance, they can be derived from an orthogonalization, followed by normalization, of the PF model discussed above. Orthogonalization is obtained by zero-pole cancellation using second-order AP filters, one for each second-order section, whereas every section response is orthonormalized with respect to each other[6]. It follows that, as for PF filters, the approximated RIR is the result of a linear combination of P basis functions φ_i built from poles in $\boldsymbol{p}$, which are the impulse responses of the second-order filter sections, orthonormal in this case, each weighted by a linear coefficient θ_i,

$$\hat{h}(n) = \sum_{i=1}^{P} \varphi_i(n, \boldsymbol{p})\theta_i \,. \tag{1.26}$$

As for the PF model and other fixed-denominator models in general, OBF models offer the possibility of incorporating prior knowledge about the underlying dynamics of the room acoustics system in the form of a set $\boldsymbol{p}$ of stable poles. Moreover, fixed-denominator models and OBF models with the same set of poles, span the same approximation space, such that the same approximated response can be obtained if the optimal values for the numerator coefficients are available. It is then reasonable to question whether orthonormality is such a desirable property to justify the choice of a model with a higher filter complexity. The answer is that, since the parameter estimation for non-orthogonal fixed-poles models is normally very ill-conditioned, using an orthogonal model structure is said to be the only practical way of fixing the poles in an IIR filter [112] and be able to obtain numerically accurate estimates for the numerator coefficients.

[6] A more detailed discussion about the properties of OBF models is provided in [27, 112, 113] as well as in Chapter 3.

Indeed, the poles being fixed and the numerator coefficients appearing linearly in the model, linear regression methods can be used. Also, because of orthogonality of the regressors, i. e. of the basis functions, a LS estimator for the linear coefficients is obtained from the inner product between the truncated basis functions and the N-samples response $h(n)$ to be modeled, without requiring any matrix inversion in the LS solution,

$$\hat{\theta}_i = \langle \boldsymbol{\varphi}_i, \boldsymbol{h} \rangle = \sum_{n=0}^{N-1} \varphi_i(n) h(n) \,. \tag{1.27}$$

This expression corresponds to the zero-lag one-sided cross-correlation between $\varphi_i(n)$ and $h(n)$ and thus it is a measure of their similarity.

The use of rational orthonormal bases appeared in the context of approximation theory back in the 1920s [114, 115], and later further developed by Walsh [75]. Following the work of Wiener [116], most of the early applications of OBF models are found in the field of continuous [117] and discrete-time [118] network synthesis (i. e. filter design). Later on, the theory of OBF models started to be developed in the context of system identification [119, 120], and their use began to appear in control applications [121, 122]. From there, the popularity of OBF models started to increase exponentially, with applications found in different fields, such as system identification [123, 124, 125, 126, 127, 128, 129, 130], signal processing [131, 132, 133, 134, 135], model approximation [136, 128], or adaptive control [137], and seems to be an active topic even nowadays [138, 139, 140, 141, 142, 143, 144, 145, 146, 147].

Different models belong to the family of OBF models, normally referred to with the names of their inventors or the name of the orthogonal polynomials they are derived from. If the pole set $\boldsymbol{p}$ contains the same repeated real pole, the *Laguerre* model, which is implemented as a prefiltered version of the WFIR filter, or the *Legendre* model is obtained. If different real poles are included, the model is called *Takenaka-Malmqvist*, whereas if $\boldsymbol{p}$ contains the same repeated pair of complex-conjugate poles, the so-called *2-parameters Kautz* model is obtained.

Modeling room acoustics with OBF models

Of particular interest in room acoustic modeling is the model built from a set of different pairs of complex-conjugate poles, sometimes referred to as *Kautz* model. Indeed, the idea is to model the resonant response of the RTF with a combination of resonances, thus estimating poles close to the true poles of the system, but without the numerical problems of non-orthogonal models. OBF models have been applied to acoustic and audio signal processing applications

only recently. In particular, general multi-pole OBF models were used for speech synthesis [33], loudspeaker response equalization [28], and modeling of room and musical instrument responses [29, 30, 31, 32, 148, 149, 150, 33]. In the rest of this thesis, when not specified otherwise, the term OBF model is used to refer to the general model containing repeated and non-repeateted complex-conjugate poles (such as the Kautz model or other possible realizations [112]), and possibly real poles[7].

The main problem concerning room acoustic modeling with OBF models is the estimation of the nonlinear pole parameters. An option would be to rely on prior information about the true poles of the system [27], which however is often not available for room acoustic systems. Moreover, selection strategies, based on optimality conditions [134, 135] or nonlinear recursive estimation algorithms [151], are limited to OBF models with repeated single poles or to low model orders [125, 152, 153].

The nonlinear pole estimation problem for the general OBF model is more involved, especially for high model orders, and not many solutions have been proposed in the literature. The use of the BU method was suggested in [148, 29] for the approximation of a RIR $h(n)$, motivated by some analogy found between OBF models and the estimation algorithm. More specifically, without getting into the details discussed in [148, 29, 31] and reviewed in Chapter 7, it was noticed that minimizing the energy of the 'complementary signal', as done by the BU method, corresponds to a minimization of the error produced by approximating the RIR with an OBF model. The BU method, however, approximates the RTF using a linear frequency resolution. In order to increase the accuracy in the approximation at low frequency, where it is normally more important to obtain a good model of the RTF, the *warped BU* (wBU) method was introduced in [150, 29], in which the BU method is applied to the warped RIR. The estimated parameters are mapped back to the original frequency scale, analogously to the estimation procedure using warped models discussed previously.

A scalable *matching pursuit* (MP) algorithm, named OBF-MP and described in **Chapter 3**, which consists of a greedy grid-search approach, was proposed in [154, 113][8]. The algorithm avoids the nonlinear problem by defining a set of candidate poles and exploiting orthogonality to select at each iteration the pole (or pair of poles) for which the approximation error is minimized. Advantages with respect to the BU method consist in the possibility of arbitrarily allocating

[7] real poles are actually not very useful, given that a measured response always presents a band-pass characteristic, due to the cut-off of the loudspeaker response at low frequencies and to the anti-aliasing filter at high frequencies introduced in the AD conversion.

[8] A similar approach was adopted in [144] in a system identification framework, although limited to low-order models applied to very simple systems.

frequency resolution and to determine the order of the approximation during the estimation. Moreover, the algorithm delivers unconditionally stable pole estimates even at high model orders, for which the BU method may exhibit some instability, and the parameters are estimated already in the complex pole form, such that the factorization of high-order polynomials is not required.

Also OBF models can implement the idea of modeling multiple RTFs with a common denominator. In this context, recent works proposed the use of common-poles in OBF models for subband modeling of RIRs [32] and feedback path characterization in hearing aids [35], in which the modification of the BU method for the estimation of a common denominator mentioned above and its version including frequency weighting [155] are used. Also the OBF-MP algorithm can be easily extended to modeling with a common set of poles, as introduced in [156] and described in **Chapter 4** of this thesis.

1.4 Identification of room acoustic systems

In many practical applications, it is not possible to first obtain a measurement of the RIRs at a number of different source-receiver positions inside the room and to model the RTF directly. It follows that the system RTF has to be identified from input-output signals. Moreover, in tasks such as AEC or AFC, it is necessary to track the variations in the RTF, due to changes in the source or receiver positions. For this reason, *adaptive filters* [157, 158] are normally used to identify the RTF from the most recent values of the input and output signals.

Adaptation algorithms

In the identification scenario, depicted in Figure 1.7, the coefficients α_i of the adaptive filter $\hat{F}(z, \boldsymbol{\alpha}(n))$ with M coefficients $\boldsymbol{\alpha}(n) = [\alpha_1(n), \ldots, \alpha_M(n)]$, are updated at each incoming sample of the input signal $u(n)$, in the attempt of recursively minimizing some performance criterion, normally defined as a function of the error signal

$$e(n) = y(n) - \hat{y}(n) = y(n) - \hat{F}(q, \boldsymbol{\alpha}(n))\, u(n). \tag{1.28}$$

If, after a number of recursions, the identification is successful, the transfer function of the filter $\hat{F}(z, \boldsymbol{\alpha}(n))$ should represent a good approximation of the RTF $H(z, n)$. The generic form of the update rule is given as

$$\boldsymbol{\alpha}(n+1) = \boldsymbol{\alpha}(n) + \Delta\boldsymbol{\alpha}(n, e(n), \boldsymbol{u}(n)) \tag{1.29}$$

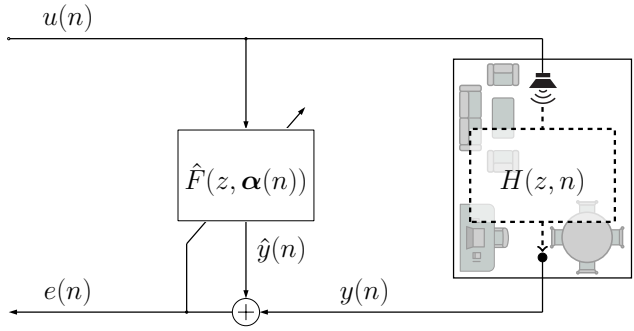

Figure 1.7: Room acoustic system identification scenario.

where $\Delta\boldsymbol{\alpha}$ is the change in the coefficient values that occurs at sample n based on the cost function of the adaptation algorithm adopted, which is a function of the error signal $e(n)$ and on the current and past values of the input signal, included in the vector $\boldsymbol{u}(n)$.

The most commonly used algorithms try to minimize the *mean square error* (MSE) cost function $J = \mathbb{E}\{e(n)^2\}$, which simplifies the adaptation task. Indeed, such cost function has a quadratic error surface, which is differentiable and guarantees the existence of a global minimum. These adaptation algorithms, of which a detailed discussion is out of the scope of this section [157, 158], can be divided into two main categories. *Gradient-based* algorithms, such as the least mean squares (LMS) algorithm or its normalized version, descend step-by-step in the direction of the gradient of the instantaneous squared error $e^2(n)$ computed with respect to the filter coefficients. *LS-based* algorithms, such as the recursive least squares (RLS) algorithm, minimize a cost function defined using a (weighted) sum of a given number of previous samples of the squared error signal, which allows to approach the optimal solution in a smaller number of recursions, at the expense of a higher computational complexity.

The choice of a particular adaptation algorithm should be made based on the application requirements. The performance of an adaptation algorithm can be defined with respect to a number of different factors [157], i. e.

1. the *accuracy* of the identification at different recursions, quantified by the misalignment (or misadjustment) between the actual RTF $H(z, n)$ and the filter transfer function $\hat{F}(z, \boldsymbol{\alpha}(n))$,

2. the *variability* of the filter coefficients after convergence, quantified by the misalignment at steady-state (for $n \to \infty$), which describes the behavior

of the algorithm in response to variations of the characteristics of the input signal or to changes of the RTF,

3. the *convergence behavior* of the algorithm, defining how fast the value of the filter coefficients approaches the steady-state solution, quantified by the rate or time of convergence,

4. the *computational complexity*, defined as the number of operations involved in a single recursion, and the memory requirements,

5. the *robustness* and *stability* of the algorithm with respect to internal or external disturbances, and the ability to avoid local optimum solutions.

Adaptive filters

An important choice to be made even before selecting a particular adaptation algorithm, pertains to the RTF model to be used. Among the different RTF models discussed in the previous section and their relative filter implementation, FIR filters are normally preferred over IIR filters for a number of reasons. First, FIR filters being linear in their parameters, the implementation of adaptive algorithms is simpler. For instance, the computation of the gradient vector is not even necessary, since it corresponds to the input signal vector. Second, global convergence and stability of the filter estimate in gradient-based algorithms is normally guaranteed, unless a too large step is taken in the gradient direction. Finally, a large selection of algorithms and theoretical results about their behavior are available mostly for FIR adaptive filters.

However, some of the factors listed above are not independent from each other, such that some of the requirements may not be met at the same time. For instance, higher accuracy would require an FIR filter with a large number of coefficients, which however normally implies slower convergence, higher variability and increased complexity. On the other hand, if a low-order filter is required, the tail of the RIR that is not modeled contributes in reducing the accuracy of the identification and in increasing variability [159, 160, 161, 162]. It follows that in some cases, the use of adaptive IIR filters would be desirable to obtain good accuracy using less filter parameters. However, IIR adaptive filters in direct form are more difficult to handle than FIR filters, especially in the output-error form configuration [163], which is the reason why it is quite rare to find them used in practical applications. They require to check for stability and to compute the gradient vector analytically, which are both computationally demanding operations, and they are not guaranteed to converge to a global minimum.

Alternative forms of IIR filters can be used, such as the lattice form or the parallel form [164], which make the stability monitoring easier, but still require involved computations of the gradient vectors with respect to the denominator coefficients and suffer from convergence to local minima. For this reason, *fixed-poles adaptive filters* (FPAFs) [165] represent an interesting option. Indeed, by fixing the poles in the filter structure, the result is linear in the parameters, such that the same adaptation algorithms developed for FIR filters can be adopted with the same implementation complexity. Moreover, the gradient vector with respect to the linear parameters is readily available, stability is guaranteed by fixing the poles inside the unit circle, and the algorithm converges to the global minimum under the same conditions as for FIR filters. Thus, the performance of a FPAF mostly depends on the location of the poles and on the characteristics of the input signal. As a matter of fact, the identification accuracy and the convergence properties depend largely on how far the filter poles are from the real poles of the system and how the spectral characteristics of non-stationary input signals may impact the adaptation of the linear coefficients of each section of the filter [130]. Unfortunately, these dependencies are difficult to analyze and thus the behavior of FPAFs hard to predict.

OBF adaptive filters

As for modeling, the orthogonality property of IIR filters based on OBFs, hereafter named *OBF filters*, provide some advantages compared to other IIR filters also in the adaptive context [166, 167]. For instance, the convergence behavior of FPAFs depends on the correlation matrix of the responses of the filter sections to the input signal, and more specifically on its numerical properties, which in turn is related to the number and position of the poles, and to the energy of the input signal at different frequencies. It turns out that, due to orthogonality, OBF filters produce correlation matrices with better numerical properties, and thus better convergence, not only for white input signals, but for a large range of input spectra. Furthermore, orthogonality is also the key aspect to enable the use of analysis tools, similar to the ones developed for FIR adaptive filters [168]. It follows that, for given fixed poles, the performances of an OBF adaptive filter with respect to the input signal characteristics, are completely determined by the position of the poles. In addition, the analysis results obtained for OBF filters, can be extended to FPAPs built from the same set of poles [167, 169].

The problem, once again, is then to determine a set of poles which can provide good performances with as few poles as possible, such that good accuracy and fast convergence can be obtained simultaneously. Not many examples of methods for the identification of the poles from input-output data were found,

especially in the context of room acoustics. Recursive algorithms [170] should be adopted in this case, such as the method in [151], which adapts the coefficients of a single-pole OBF filter using a combination of RLS and a nonlinear recursive algorithm. For the general OBF model, however, the analytical expressions of the gradient vectors with respect to the poles are very involved [171], and a strategy for the adaptation of the poles not very practical.

In order to avoid the nonlinear problem, the idea of using a grid search, as in the estimation algorithms described in Part II, was also adopted to identify a set of poles, possibly common to more source-receiver positions, from either white noise or speech signals. The resulting scalable algorithm, described in **Chapter 5**, uses the normalized LMS algorithm and a modified version of it to search among a set of candidate poles the one that reduces the instantaneous squared error the most.

Other system identification and data-driven modeling approaches

System identification consists of representing dynamical systems using mathematical models obtained from measured input-output data. The most well-established approach in system identification is based on *parametric prediction error* and *maximum likelihood methods* [172, 173], both in time [174] and in frequency domain [175]. These methods first rely on the selection of a model structure of a given order, which is adequate to describe the system at hand. If the model and its order are well chosen, the system can be identified with good accuracy. Methods in this framework are the most widely used in room acoustic system identification and other related fields, such as modal analysis in structural engineering, where the aim is to estimate the modal parameters (eigenfrequency, damping constant, mode shapes) from vibration data of a structure [176]. Methods specifically developed for modal analysis are also available, but tend to fail in cases of low SNR and for systems with high dynamic range and modal density [177]. The system identification literature on OBF models under this classical framework is well-established, even though mostly limited to Laguerre and 2-parameter Kautz models, or to Generalized OBF models, consisting of repeated sets of a finite number of poles (see references in Section 1.3.2).

Another common approach worth mentioning, even though its application in room acoustic modeling is quite limited [178], is subspace identification [179, 180]. *Subspace identification methods* aim at estimating a state-space model by means of projections onto certain subspaces generated from input-output data, normally using a singular value decomposition (SVD) algorithm. Advantages of these methods are that no specific model structure has to be selected, that

compact models can be easily achieved by model reduction, and that iterative (nonlinear) optimization techniques are not required. On the contrary, it is more difficult to include prior knowledge of the system, the estimation may be less accurate than with prediction error methods [181], and the recursive update of the SVD algorithm is not well-suited for online identification [182].

The already vast 'classic' system identification literature recently became even larger with the introduction of concepts borrowed from the fields of statistics and machine learning, such as kernel-based, Bayesian, and regularized estimation methods. Rather than relying on finite-dimensional models, *kernel methods* [183] formulates the identification problem in an infinite-dimensional space, consisting of all possible impulse responses, which are modeled as zero-mean Gaussian processes. Prior knowledge is included by means of a covariance function, also known as kernel, whereas ill-conditioning problems are avoided using *regularization methods* [184, 185]. Kernel methods are potentially very useful in RASE applications when also the nonlinearities of the loudspeaker needs to be identified [186, 187]. The identification can be performed in the infinite-dimensional space in an adaptive way [188, 189, 190, 191] without problems of convergence to local minima, while obtaining a filter which is nonlinear in the input space. In this context, the kernel could be constructed based on OBFs [147], which imposes stability and may be particularly suited for loudspeaker/room impulse response estimation. The RIR estimation problem was also tackled using *Bayesian estimation methods* [192, 193], relying on the assumed sparseness of the RIR coefficients.

A recent trend in room acoustic modeling is indeed to formulate the estimation problem in the compressive sensing framework [194], with the aim of obtaining a sparse representation of RIRs. The concept of sparsity in room acoustic modeling can refer to the early part of the time-domain RIR, in which sound reflections are discrete and scarce [195, 196], to the low portion of the spectrum of the RTF, where the modal density is low [197], or to the spatial distribution of room modes, which can be approximated by a linear combination of a finite number of basis functions, such as plane waves [197, 198], spherical harmonics [199] or spherical waves [200]. In this context, a RIR estimation algorithm aiming at obtaining a sparse solution by selecting OBFs out of a large dictionary using sparsity-promoting regularization and convex optimization was proposed in [201], but not included in this thesis. In the thesis, sparsity in the RIR representation is achieved using a different concept [202, 203, 204] by means of analytical dictionaries built from OBFs.

1.5 Room acoustic signal enhancement

Modeling and identification of room acoustics are at the basis of all digital signal processing tasks intended to correct, modify or synthesize the response of the system, with the purpose of enhancing the desired qualities of sound signals. In this section, three common RASE applications are described, namely digital equalization, artificial reverberation and AEC.

1.5.1 Digital equalization

Equalization in room acoustics [6, 7, 8, 9] aims at improving the objective and subjective quality of sound reproduced in rooms using digital signal processing techniques[9]. In order to do so, the detrimental effects introduced by the loudspeakers and the room acoustics should be corrected for. Ideally, a digital equalizer should be designed in order to invert the loudspeaker-room response, such that the source signal can be faithfully reproduced at the receiver position. In practice, mainly due to the nonminimum-phase characteristics of the RIR, perfect equalization at one or multiple positions inside a room is difficult to achieve.

Minimum-phase equalization

A stable and causal inverse RTF is only achievable for minimum-phase systems, since a RTF with zeros outside the unit circle, i. e. with $B_{\text{out}}(z) \neq 1$ in the RTF in (1.17), would result in an unstable inverse RTF. A minimum-phase inverse, however, can be achieved by first decomposing the RTF into a minimum-phase component $H_{\text{mp}}(z)$ and an AP component $D(z)$ as

$$H(z) = H_{\text{mp}}(z)D(z) = \frac{B_{\text{in}}(z)B_{\text{out}}(z^{-1})}{A(z)} \frac{B_{\text{out}}(z)}{B_{\text{out}}(z^{-1})} \tag{1.30}$$

which is obtained by multiplying and dividing by the polynomial built by reflecting excess-phase zeros inside the unit circle ($z \to z^{-1}$), and then inverting the minimum-phase part $H_{\text{mp}}(z)$ of the RTF[10]. Given that an AP transfer function has a flat magnitude response, the magnitude response of the minimum-phase component corresponds to the one of the original RTF. It follows that

[9] an exhaustive overview and classification of methods developed in the last 40 years can be found in [9]

[10] in practice, a minimum-phase/all-pass decomposition can be obtained, for instance, using the homomorphic method described in [45, 48].

an equalization filter with transfer function equal to $H_{\mathrm{mp}}^{-1}(z)$ is able to correct for the spectral characteristics of the loudspeaker/room response.

The exact minimum-phase equalizer, however, presents some issues. The inverse response, indeed, is characterized by a very long response, due to the fact that deep and narrow notches in the room magnitude response correspond to prominent and long-ringing resonances in the inverse response. For this reason, a high-order equalizer is normally required for exact inversion. Moreover, the inverse equalizer built from a given RTF measured at specific source-receiver positions is effective only at those positions and only if the acoustic conditions are almost unchanged [38]. Indeed, variations in the RTF can change the position of the peaks and notches in the room magnitude response, especially at higher frequencies, so that a sharp resonance in the equalizer aiming to correct a notch in the measured RTF may actually create a detrimental boost in the response.

To deal with these problems, the room magnitude response is normally smoothed to a certain degree in order to level out deep notches and sharp resonances, which extends the area of effective equalization and also results in a reduction of the required length for the equalization filter. Another option is to design the equalizer based on an approximated model of the RTF, in which peaks and notches are coarsely modeled. For this purpose, warped models [93, 99] and fixed-poles models, such as the PF [80, 87, 205] and OBF [28] models, have been suggested, which also allow to obtain a desired frequency resolution, such as the Bark frequency scale. By properly selecting poles [89, 205] it is possible to unevenly allocate resolution and to control the sharpness of the resonances in the equalizer response in different frequency regions. For instance, at low frequencies, where it is important to correct for the modal behavior of the room, high resolution and sharp resonances can be assigned to the filter, whereas at higher frequencies, where the variability of the RTF is higher and notches are less perceivable, a lower resolution and a gentler equalization is sufficient. An alternative is to recur to multirate approaches [206, 207] in which different subbands can be treated differently.

The robustness with respect to measurement errors and variations in the RTF can be improved also by designing an equalizer based on the response measured at multiple positions. A multi-point equalization filter is computed, for instance, by minimizing the average equalization error at multiple microphones using LS or adaptive filters [208], or by first obtaining a prototype response by averaging [209] or clustering [210] techniques, which is then used in the design. An alternative approach relies on the modeling of multiple RTFs using the CAPZ model [104, 105, 211], where the common denominator $A(z)$ is inverted to be used as an FIR pre-equalizer, such that common room resonances are partially compensated.

Special kinds of filters are often found in loudspeaker and room response equalization, such as those used in graphic [82, 212] and parametric equalizers [213, 214], normally having low filter orders. These equalizers are quite popular, being implemented in many commercial products, such that equalizer design methodologies that would allow to automatically correct for deviations from a desired response without relying on the manual tuning of trained experts are of interest [215, 216]. In this context, a novel procedure for equalization using IIR biquadratic filters, which designs an efficient low-order equalizer while focusing on the equalization of magnitude peaks, is described in **Chapter 6**.

Nonminimum-phase equalization

Even though the unbalanced magnitude response is the main source of coloration and reduced perceived sound quality, it is often important to correct also for the phase response [49, 217], especially if some reduction of the reverberation tail is desired. In this case, a mixed-phase equalizer, correcting both the minimum-phase and the nonminimum-phase components of the RTF, is required, and the equalization task is sometimes referred to as *dereverberation*[11]. Such an equalizer is unstable or noncausal by nature, but it can be made stable and causal by introducing a modeling delay. Such delay should be chosen as the duration of the noncausal inverse filter response, which is normally as long as the original RIR [6]. Using a shorter delay, as required in practice, not only provides only partial equalization, but also introduces errors in the equalized response, known as *pre-ringing* or pre-echo, which appear in the equalized response before the direct sound. Moreover, the variability of the RTF and errors in the measured RIRs tend to amplify the problem [218].

Some of the methods cited above, such as the multi-point adaptive equalization [208], and multi-channel methods based on the exact inversion of RTF either in the time domain or in the frequency domain, e.g. [219, 220], have been applied to the design of mixed-phase equalizers, but they normally exhibit the aforementioned problems. More robust methods [221, 222] that mitigate these issues, but only obtain partial dereverberation, have appeared in recent years. Other approaches attempted to reduce the effective length of the mixed-phase inverse filter, by using complex smoothing techniques [223] or IIR filters with uneven frequency resolution. Also in the mixed-phase case, the use of fixed-poles filters may be potentially useful to reduce the equalizer order and partially control the behavior of the equalization, provided that poles are distributed in a meaninful way. The use of OBF filters has been previously investigated [28]

[11] in case the source signals are not available, for instance when the signal to enhance is coming from a person talking inside the room, the equalization task is told to be 'blind' [3].

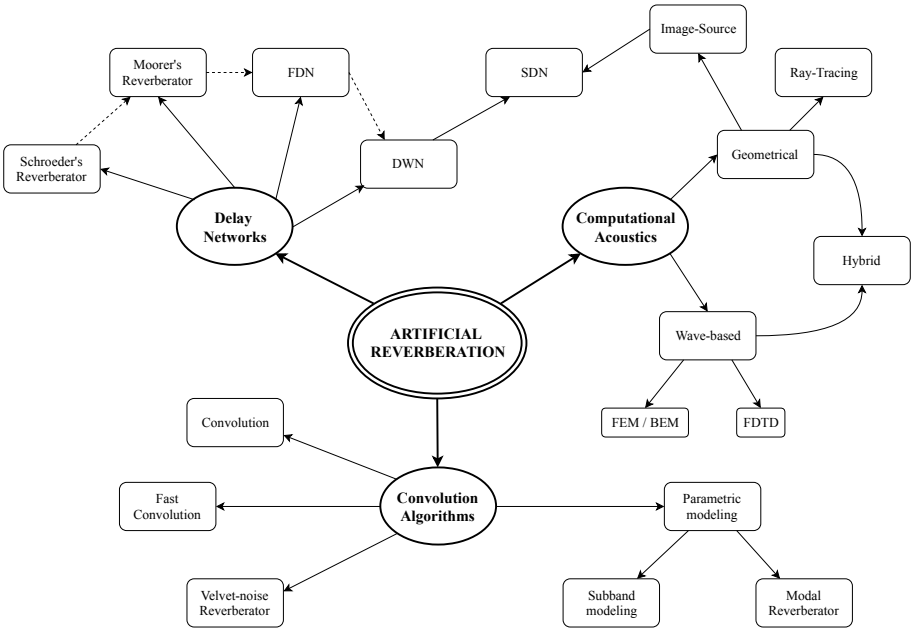

Figure 1.8: Overview of artificial reverberation methods.

for fixed configurations of the poles. In **Chapter 5**, a method is suggested for identifying the poles of the equalizer implemented as an OBF filter.

A different multi-channel approach [224, 225, 226] has been proposed, which uses a polynomial-based control systems framework providing analytical expressions for the optimal filter. The aim is to partially invert the RTF of all the acoustic channels, and then use sound field superposition to equalize a *primary* loudspeaker with the aid of a number of *support* loudspeakers to reach the desired target response at a number of control points and in their vicinity. Also, the level of the pre-ringing introduced is controlled by means of an AP filter designed from nearly common excess-phase zeros, whereas robustness to RTF variability and RIR errors is addressed by modeling the RTFs as a sum of a deterministic and a stochastic part. The method in [225] has been applied for the design of a multiple-input/multiple-output (MIMO) equalizer for improving the acoustics inside a car cabin [227], as described in **Chapter 7**, where the design procedure is outlined and methods for modeling the RTFs and their excess-phase components are suggested.

1.5.2 Artificial reverberation

The term *artificial reverberation* refer to a large variety of different approaches [10, 11] that try to simulate the acoustic response of a room in order to enhance or manipulate its features. Artificial reverberation finds application in different fields, such as acoustic design and analysis of the interior acoustics of buildings, music (both live and recorded), film post-production, and, more recently, virtual reality. These methods, of which a schematic overview is given in Figure 1.8, can be divided into three main categories (see [10, 11] and references therein):

1. *delay networks* methods, in which tapped delay lines and digital filters (such as comb and AP filters) are used to simulate early reflections and late reverberation. They are based on a perceptual approach, i. e. not relying on measured RIR nor trying to simulate the acoustics of an existing room, where the reverberation characteristics, such as RT, echo density and diffuseness, are controlled by the parameters of the filters;

2. *computational acoustics* methods, which rely on physically-based room models. Starting from a room with certain dimensions and properties of the surfaces, the sound field at low frequencies is obtained by wave-based methods [228], which predict how acoustic waves propagate in the room using numerical methods. At higher frequencies, where wave-based methods are too demanding, geometrical acoustics methods [229, 230, 231] are used instead, which assume a ray-like behavior of sound;

3. *convolution algorithms*, which apply the measured, estimated or modeled room response to an audio signal by convolution or filtering.

The third category is closely related to the scope of this thesis. Indeed, the general approach, at least conceptually, is to convolve a 'dry' audio signal with a RIR, so to apply to the signal the acoustic features of the space in which the RIR was recorded. In practice, linear convolution, defined in (1.21), is a very demanding operation, so that fast and efficient methods for convolving long RIRs in real time have been developed, based on fast convolution in the frequency domain using blocks [45] and partitioning of the RIR [232].

An alternative to the complexity problem of FIR-based direct convolution is to model the measured RIR using techniques described in Section 1.3 and perform the convolution using IIR filters with a reduced number of coefficients, thus requiring less operations [233]. A possibility, useful to avoid problems related to modeling with high model orders, is to recur to subband modeling using multirate techniques [206, 207]. This way, not only the subband responses

can be modeled with less parameters compared to the fullband case, but also the number of parameters for each subband may vary, for instance because of differences in modal superposition or time decays at different frequency regions, thus leading to computational savings.

In recent years, the use of the PF, discussed in Section 1.3, was proposed as a means of producing reverberation, where each filter section synthesizes the response of a modal resonance. What is interesting in this *modal reverberator* [83, 11] is that it provides a means of controlling the parameters of each mode, i. e. frequency, damping constant and amplitude, in an interactive way (in real-time and with no latency). For instance, movements of the source or the listener inside the room could be simulated by modifying the mode amplitude parameters, which are position dependent. Or the acoustic features of a RIR, measured and modeled as a PF, could be altered by modifying the pole parameters of the filter, and possibly other kind of effects [234]. The parameters of the PF filter can be selected based on the desired decays of modes in different frequency regions, or by modeling a measured RIR. In the original work [83], the modal frequencies are estimated as the frequencies of the most prominent spectral peaks, while the damping constants are determined based on the decay times estimated in subbands. The amplitude parameters are then found by a weighted LS estimation, where a weighting function is used to obtain a good fit in the early part of the response. Alternatively, modeling methods, such as those described in Section 1.3 and in Chapters 3 and 4, could be used instead.

1.5.3 Acoustic echo cancellation

One common task of room acoustic signal processing is the suppression of the echo that often arises in hands-free applications, which has a very detrimental effect on the quality of communication. The situation is depicted in Figure 1.9, where the speech signal $u(n)$ of the speaker in the transmission room on the left is reproduced, with a certain transmission delay Δ_T, in the receiving room on the right, where it is modified by the room acoustics before being picked up by the microphone. The role of AEC [22, 23] is to model and identify the time-varying acoustic echo path $F(z, n)$, i. e. the RTF, between the loudspeaker and the microphone in the receiving room using an adaptive filter $\hat{F}(z, n)$. In this way, the echo signal $y(n)$, which is normally corrupted by some additive uncorrelated noise $v(t)$, can be canceled and only the echo-free speech signal of the speaker in the receiving room (in case he is talking) is sent to the loudspeaker in the transmission room. The adaptive filter thus needs to identify the echo path with good accuracy as quickly as possible, be able to track its variations in time in an effective way, and be robust against the influence of noise.

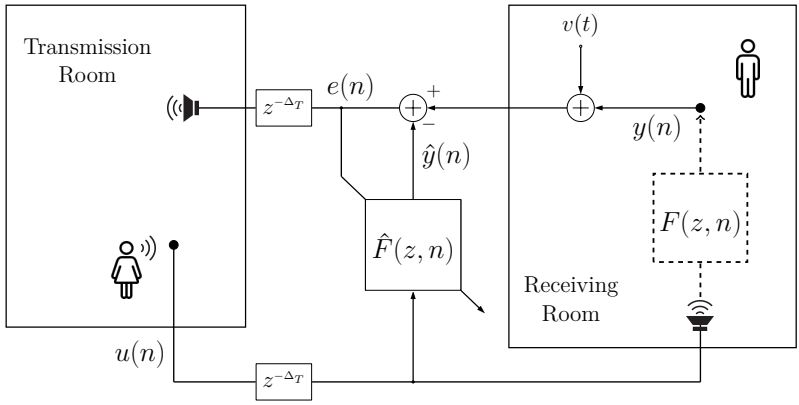

Figure 1.9: Acoustic echo cancellation scenario.

Problems in the context of AEC arise when both speakers are talking at the same time. This *double-talk* situation results in slow convergence, or even divergence, of the adaptive filter. A common solution is the use of a double-talk detector [23], which freezes the filter adaptation when both speakers are concurrently active, although other robust approaches have been suggested [235]. Another issue is related to the *poor excitation* characteristics of speech signals. The non-whiteness and non-stationarity of speech results in large variations of the signal power, which translate to an ill-conditioned autocorrelation matrix of the input signals. If a large model order is required for the adaptive filter, as is the case for reverberant environments, the autocorrelation matrix gets larger, resulting in an aggravation of the ill-conditioning problem and thus a slower filter convergence. A common solution in this case is to recur to regularization methods, either fixed or dependent on prior knowledge of the room acoustics [185], in order to reduce the condition number of the autocorrelation matrix and speed up convergence. Often a limited number of filter coefficients is available. In case the echo canceler is implemented as an FIR filter with less coefficients than the number of samples of the RIR, the unmodeled part of the echo path can be interpreted as additional noise [159, 162]. As a consequence, the identified echo path may present a bias and a high variance [160], which have a negative impact on the performance of the echo canceler.

The *undermodeling* and the ill-conditioning problems are even more critical in case two (or more) loudspeakers are present in the receiving room. In *stereophonic acoustic echo cancellation* (SAEC) [161, 236], the two loudspeaker signals from the transmission room are strongly correlated with each other, which results in a highly ill-conditioned stereo autocorrelation matrix. This

is translated not only to very slow convergence of the coefficients of the two parallel echo cancelers, but also to a problem of identifiability of the echo paths between the two loudspeakers and each microphone present in the receiving room. For this reason, an additional stage is often necessary in order to decorrelate the input signals, at least partially, before they are sent to the loudspeakers. Decorrelation is normally achieved by processing the stereo signals with a nonlinear operation [161], by adding spectrally-masked noise [236], or by means of time-varying AP filters [237], where the degree of decorrelation is limited by the level of degradation of the speech quality introduced by this operation.

Some of the problems inherent to AEC and SAEC could be alleviated by implementing the echo canceler as an IIR filter [103, 238]. Indeed, if the acoustic echo path can be estimated without incurring in undermodeling problems using a reduced number of filter coefficients, that would result in a better conditioned autocorrelation matrix. However, the use of IIR filters in AEC have been discouraged in the past by their difficulties in the filter adaptation and possibly by the wide-spread belief that only a limited improvement over FIR filters can be achieved [239, 240]. The first argument is no longer valid for FPAFs, even though slower convergence is still a possibility.

Also in this application, the use of OBF adaptive filters can bring some benefit. More precisely, their orthogonality property can reduce the ill-conditioning problems due to the non-whiteness of the input signal, at least for the single-channel case. However, practical advantages over FIR filters are only achieved if the poles of the filter are selected to be close enough to the true poles of the system, such that a small number of coefficients is required to estimate the echo path. It follows that, once again, the main issue to be tackled is the estimation of the pole parameters. A few examples of the use of single-pole OBF filters in echo cancellation are found in the literature [34, 131, 241], whereas the use of general OBF filters has been suggested in [171] but not verified on realistic scenarios. In **Chapter 5**, the identification algorithms developed are applied to a simple AEC scenario, where the poles are estimated from speech signals.

1.6 Overview of the thesis

The main aim of this thesis is the development of efficient parametric models and identification methods for room acoustics signal enhancement (RASE) applications, with a focus on equalization. Pole-zero (PZ) models with fixed-poles are investigated, as their infinite impulse response (IIR) nature can bring advantages over all-zero (AZ) models in terms of efficiency, especially in the low

frequencies, without some of the typical problems with conventional PZ models. A consistent part of this thesis is devoted to orthonormal basis function (OBF) models and adaptive filters, whose properties make them particularly suited to model room acoustics and to tackle some of the issues commonly encounter in RASE applications.

1.6.1 Research objectives

The main research objectives of this thesis can be stated as follows:

1. Characterization of the room transfer function (RTF) in the modal frequency region, where the acoustics of a room is more problematic, especially in small spaces. The focus is on the measurement of room impulse responses (RIRs) and the analysis of related issues due to the ambient noise and the nonlinear behavior of the loudspeaker when driven with high levels at low frequencies.

2. Development of efficient room acoustic parametric models with the same modeling accuracy but a lower model complexity compared to the all-zero model. Models based on OBFs are investigated in order to assess their potential in providing a compact yet accurate representation of the acoustic system. The effectiveness of the proposed models and the related parameter estimation algorithms with respect to other methods are evaluated by comparing their performance in approximating measured target RIRs.

3. Application of the developed models and algorithms in a system identification framework using adaptive filters, towards their implementation in real RASE tasks. For this purpose, the problem of identifying a model of the room response from both white noise and speech signals is addressed.

4. Design and implementation of equalization methods addressing the problem of compensating the unbalanced response of a loudspeaker and of a multiple position room acoustic system. The focus is on equalization using IIR filters with reduced filter orders, so as to deliver effective solutions with low complexity.

1.6.2 General overview

This thesis is divided in four parts, each addressing one of the four main research objectives stated above, and each one related to one of the four main topics

treated, namely room acoustic measurements, modeling, identification, and equalization.

Part I (chapter 2) focuses on the characterization of room acoustics at very low frequencies. The discussion is centered on the difficulties encountered when performing acoustic measurements in the modal frequency range, such as prominent ambient noise and nonlinear distortions of the loudspeaker, and on the countermeasures to be used to obtain reliable room impulse response (RIR) measurements. The problem of estimating the reverberation time in the low-frequency range is also addressed.

Part II (chapters 3 and 4) deals with the topic of modeling room acoustics using parametric models. The focus is on models based on orthonormal basis functions (OBFs), which present interesting properties compared to other conventional parametric models. Some of these properties have been exploited for the design of a scalable matching pursuit (MP) algorithm, named OBF-MP, for the estimation of the pole parameters of an OBF model from measured RIRs. An extended version of the algorithm, named OBF-GMP, is also presented, dealing with the problem of estimating poles common to a set of room transfer functions (RTFs) measured at different locations inside the room, based on the concept of common-acoustical-poles.

Part III (chapter 5) investigates the use of IIR adaptive filters based on OBF models in a system identification framework, motivated by their good numerical properties. The theory of OBF adaptive filters is reviewed, with emphasis on their performance in relation to critical aspects, such as the filter order and the characteristics of the input spectrum. A scalable identification algorithm, inspired by the modeling algorithms described in Part II and named stage-based (SB) OBF-GMP, is introduced, which is able to iteratively estimate the poles of the OBF filter from white noise and speech multi-channel input signals. The potential of OBF adaptive filters and of the proposed algorithm is assessed by means of simple scenarios in the context of room response equalization (RRE) and acoustic echo cancellation (AEC).

Part IV (chapters 6 and 7) presents two specific room acoustic signal enhancement (RASE) applications in the context of digital equalization. The first is an iterative procedure for the design of a low-order parametric equalizer using constrained IIR filters, such as peaking and shelving filters. The procedure is applied to the minimum-phase equalization (magnitude-only) of loudspeaker and room responses. The second application is the implementation of a nonminimum-phase response equalization method, which uses a polynomial-based multi-channel framework for acoustic modeling and control design. The focus is on the RTF modeling and on the design of mixed-phase filters enabling nonminimum-phase partial inversion.

1.6.3 Thesis outline

Chapter 2 introduces a new RIR database measured in a rectangular room with subwoofers as sound sources. The measurements are performed with the exponential sine-sweep (ESS) method, whose characteristics are suitable to deal with the issues encountered at very low frequencies. It is shown that, indeed, the recorded responses present high levels of ambient noise and nonlinear distortions, both harmonic and impulsive, whose effects are partially mitigated by a careful calibration of the measurement equipment and by postprocessing operations. A procedure for estimating the reverberation time is also proposed, dealing with the over-estimation and noise floor problems encountered by standard procedures at low frequencies. A bank of narrow band-pass filters is used in combination with an approximation of the RIRs using OBF models.

Chapter 2 has been published, as an engineering report, as:

- **G. Vairetti**, N. Kaplanis, E. De Sena, S. H. Jensen, S. Bech, M. Moonen, and T. van Waterschoot, "The Subwoofer Room Impulse Response (SUBRIR) database," *J. Audio Eng. Soc.*, vol. 65, no. 5, pp. 389–401, May 2017.

Chapter 3 addresses the problem of modeling room responses using OBF models, whose orthogonality property can bring additional advantages over conventional models, such as model efficiency, stability and scalabilty. The latter is found to be related to the analogy between OBF models and the definition of the RIR as an infinite summation of exponentially decaying sinuosoids. These properties are exploited in a novel MP estimation algorithm, named OBF-MP, where the nonlinear problem of estimating the pole parameters is avoided by means of a grid-search. The algorithm, whose performance is compared to state-of-the-art modeling methods, not only delivers efficient, scalable and stable model estimates, but also provides an added layer of flexibility in the allocation of frequency resolution.

Chapter 3 has been published as:

- **G. Vairetti**, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. van Waterschoot, "A scalable algorithm for physically motivated and sparse approximation of room impulse responses with orthonormal basis functions," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 7, pp. 1547—1561, Jul. 2017.

Chapter 4 presents a simple extension of the OBF-MP algorithm, intended to estimate a set of poles common to multiple RTFs measured in the same

acoustic space. Even though the poles estimated with this extended algorithm, named OBF-GMP, cannot be claimed to be the true poles of the system, a more compact representation of a set of RTFs is achieved, as shown by simulation results performed on the low-frequency measurements presented in Chapter 2.

Chapter 4 is based on a conference paper published as:

- **G. Vairetti**, E. De Sena, T. van Waterschoot, M. Moonen, M. Catrysse, N. Kaplanis, and S. H. Jensen, "A physically-motivated parametric model for compact representation of room impulse responses based on orthonormal basis functions", in *Proc. 10th Eur. Congr. Expo. Noise Control Eng. (EuroNoise 2015)*, Maastricht, The Netherlands, pp. 149–154, Jun. 2015.

Chapter 5 deals with the topic of room acoustic system identification using adaptive filters based on OBF models. Contrary to standard IIR adaptive filters, the orthogonality property of OBF filters with fixed poles enables an analysis of their adaptation performance, showing good numerical properties and similarities with finite impulse response (FIR) adaptive filters. The properties of OBF adaptive filters are reviewed and an identification algorithm is introduced, named SB-OBF-GMP, able to identify a common set of poles from low-frequency multi-channel input-output data, both for white noise and speech signals. Simulation results show that better performances compared to FIR filters with the same number of adaptive coefficients can be achieved, with the extent of this improvement dependent on the characteristics of the room acoustics system, such as the room volume and the reverberation time. Possible applications of OBF adaptive filters and the proposed algorithm are suggested in the context of AEC, showing fast convergence and robust results with respect to the undermodeling problem, and of single-channel RRE, where the frequency resolution of the equalizer is directly determined by the identification of the poles of the inverse filter.

Chapter 5 has been submitted for publication, as a tutorial paper, as:

- **G. Vairetti**, E. De Sena, S. H. Jensen, M. Moonen, and T. van Waterschoot, "Orthonormal basis functions adaptive filters for room acoustic signal enhancement," Submitted for publication to *Signal Process.*, Elsevier, Apr. 2018.

Chapter 6 presents an automatic design procedure for a low-order parametric equalizer. Differently from state-of-the-art methods, which aim at minimizing the distance in magnitude between the system and the target responses, the proposed design procedure of the minimum-phase equalizer is based on the

minimization of the sum of squared errors, leading to an improved mathematical tractability of the equalization problem and a stronger emphasis on the equalization of the more perceptually relevant spectral peaks. Examples of loudspeaker and room responses equalization show that an effective design of a low-order equalizer is also achieved.

Chapter 6 has been submitted for publication as:

- **G. Vairetti**, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. van Waterschoot, "An automatic design procedure for low-order IIR parametric equalizers," Submitted for publication to *J. Audio Eng. Soc.*, Apr. 2018.

Chapter 7 deals with the problem of equalization of an acoustic system, specifically an audio reproduction system inside a car cabin. An existing solution has been adopted, which designs a robust nonminimum-phase MIMO equalizer able to correct the response of a primary speaker at different receivers within a given listening region, with the help of a number of support loudspeakers. The approach, which is based on a polynomial-based control system framework, strongly relies on modeling techniques. A possible implementation of the suggested 'probabilistic modeling' of the RTFs, intended to provide robustness to RTF variations, is described. The common-denominator (CD) BU method for modeling RTFs with shared common poles has been derived, independently from its recent appearence in [32], with the inclusion of a regularization parameter to mitigate ill-conditioning problems. The BU method has been also adapted for the modeling of the all-pass (AP) component of the RTFs, whose estimates are required for the design of an all-pass filter meant to remove phase distortions common to all positions in the listening area.

Chapter 7 is based on the final report for the IWT (Agency for Innovation by Science and Technology) project *RAVENNA: Proof-of-concept of a Rationed Architecture for Vehicle Entertainment and NVH Next-generation Acoustics*, in collaboration with Premium Sound Solutions N.V.:

- **G. Vairetti**, T. Dietzen, D. Pelegrin Garcia, M. Moonen, and T. van Waterschoot, "Automatic Calibration of Car Cabin Acoustics in a Multi-Channel Equalization Framework," KU Leuven, Tech. Report, July 2017.

Chapter 8 concludes the thesis by restating the research objectives, summarizing the contributions of this work, and suggesting possible directions for future research.

Part I

Measurements

Chapter 2

Measuring room impulse responses at low frequency

The Subwoofer Room Impulse Response (SUBRIR) database

Giacomo Vairetti, Neofytos Kaplanis, Enzo De Sena, Søren Holdt Jensen, Søren Bech, Marc Moonen, and Toon van Waterschoot

Published in *J. Audio Eng. Soc.*, vol. 65, no. 5, pp. 389–401, May 2017, DOI: 10.17743/jaes.2017.0007.

G. Vairetti, N. Kaplanis, E. De Sena, S. H. Jensen, S. Bech, M. Moonen, and T. van Waterschoot, "The Subwoofer Room Impulse Response (SUBRIR) database," *J. Audio Eng. Soc.*, vol. 65, no. 5, pp. 389–401, May 2017.
Changes include layout, representation, and minor editing aspects.

The candidate's contributions as first author include: co-implementation of the analysis software tools for the measured data, co-analysis of the measured data, co-formulation of the conclusions, text redaction and editing.

Abstract

This chapter introduces a new database of room impulse responses (RIRs) measured in an empty rectangular room using subwoofers as sound sources. The purpose of this database, publicly available for download, is to provide acoustic measurements within the frequency region of modal resonances. Performing acoustic measurements at low frequencies presents many difficulties, mainly related to ambient noise and to unavoidable nonlinearities of the subwoofer. In this chapter, it is shown that these issues can be addressed and partially solved by means of the exponential sine-sweep method and a careful calibration of the measurement equipment. A procedure for estimating the reverberation time at very low frequencies is proposed, which uses a cosine-modulated filterbank and an approximation of the RIRs using parametric models in order to reduce problems related to low signal-to-noise ratio and to the length of typical band-pass filter responses.

2.1 Introduction

Room impulse response (RIR) measurements are essential to assess the performance of acoustic signal enhancement algorithms, e.g. for applications such as dereverberation [242], source separation [243], source localization [50], blind acoustic parameter estimation [244], convolutive reverb [245], and many others. Several available RIR databases [246, 247, 242, 243, 50, 244, 245] are intended for different audio signal processing tasks, each requiring a different choice of measurement method and of the measuring equipment. For instance, the databases in [246] and [247] contain binaural and head-related RIRs, and are useful in hearing-aids applications. Other databases present specific configurations of the microphones, usually arranged into arrays. What is common to all these databases is that they use full-range loudspeakers, whose frequency response typically has a lower bound of 50-100 Hz. While these databases cover a frequency range sufficient for the development and evaluation of speech enhancement algorithms, information about a significant portion of the modal response of the room is missing.

Nowadays, home audio systems generally include a subwoofer, which is intended for the reproduction of low-frequency content typically in the region between 20 Hz and 150 Hz. In this frequency range, small-sized typical rooms operate within the modal frequency region [1]. In small-sized rooms, most of the acoustical problems are actually due to poor acoustics at very low frequencies (LFs). The modal resonances are usually well separated, energetic, and detectable by the human ear [248], thus degrading the perceived sound quality.

A subwoofer with small enough lower cut-off frequency can even partially excite the so-called cavity mode (i.e. the modal resonance centered at 0 Hz). Therefore, algorithms for home audio system applications, such as room compensation algorithms, should be validated also on RIRs measured within the frequency region of modal resonances. Moreover, such RIRs may provide new insights and be useful to validate physical models of room acoustics, although detailed information about the boundaries conditions are not available. To the authors' best knowledge, a RIR database measured at very LFs is not yet available.

The Subwoofer Room Impulse Response (SUBRIR) database introduced in this chapter is a collection of RIRs measured in a standard domestic listening room using a subwoofer as the sound source. Two subwoofers with different characteristics and two types of omnidirectional microphones were used to measure the RIR at different locations, for a total of 96 measurements[1]. Performing acoustic measurements at very LFs presents some difficulties, mainly related to LF ambient noise and to unavoidable nonlinear distortions of the subwoofer [55].

Nonlinear distortions can be divided into two categories: *regular* nonlinear distortions refer to systematic and reproducible distortions, such as harmonic spectral components, whose impact to the overall performance of the loudspeaker can be controlled in the design process [249]. *Irregular* nonlinear distortions are instead due to loudspeaker defects and are less easily reproducible and controllable [250]. The main irregular distortion artifact noticed in the measurements presented in this chapter was recognized as the so-called *rub & buzz* distortion [251, 250, 252]. This is a signal-dependent distortion caused by defects due to manufacturing errors, aging or overload. Possible causes of this type of distortion are buzzing parts (e.g. a loose glue joint), the voice coil rubbing or bottoming (i.e. hitting the backplate due to over-displacement), loose particles, air leakages, etc.

The family of methods for measuring RIRs known to have a high immunity against distortion artifacts is the one where a sweep is used as the excitation signal [253, 60, 254]. This chapter shows that the exponential sine-sweep (ESS) method [61] is particularly suitable for measuring good quality LF-RIR measurements regardless of all the difficulties mentioned above. The ESS is known to provide a better signal-to-noise ratio (SNR) and a better rejection of distortion artifacts than other RIR measurement methods [25, 57, 58, 255].

This chapter also outlines a procedure to estimate reverberation time (RT) at very LFs. Indeed, the standard specifications [256] are not applicable in this frequency region due to the low SNR [65] and to the influence of the response

[1] A subset of this database for one subwoofer and one microphone was already presented shortly in [156].

of the band-pass filters of the filterbank [257]. The proposed approach uses a cosine-modulated filterbank, which reduces the bias introduced by typical filterbanks at LFs, and a representation of the RIRs using orthonormal basis function (OBF) models [27], which allows to remove the effect of the noise floor.

The chapter is structured as follows. In Section 2.2, a brief summary of the ESS method is given, together with comments on advantages and disadvantages of the method. Section 2.3 describes the room in which the measurements were performed, together with details of the measurement equipment. In Section 2.4, an analysis of the measurements performed is given; the recorded signals and the retrieved RIRs are analyzed and guidelines on how to obtain good quality measurements are provided. In Section 2.5, values for the frequency-dependent RT at LFs are estimated with the proposed approach. Section 2.6 concludes the chapter and summarizes the recommendations for performing LF-RIR measurements.

2.2 The exponential sine-sweep (ESS) measurement method: a summary

This section reviews the key points of the ESS method and discusses its applicability in measuring LF-RIRs. A detailed treatment of the ESS method can be found in [61, 25].

The excitation signal used by the ESS method is a sweep signal with instantaneous frequency (IF) increasing exponentially with time. The IF at time t of the sweep signal of duration T is given by

$$f(t) = e^{(1-(t/T))\ln(f_a)+(t/T)\ln(f_b)} = f_a \left(\frac{f_b}{f_a}\right)^{(t/T)}, \tag{2.1}$$

where f_a and f_b are the starting frequency and stopping frequency, respectively. The instantaneous phase is obtained by integrating (2.1) between 0 and t, and used as the argument of a sinusoidal function, leading to the excitation signal,

$$s(t) = \sin\left(\frac{2\pi T}{\ln\left(\frac{f_b}{f_a}\right)}\left(f(t) - f_a\right)\right). \tag{2.2}$$

The excitation signal, $s(t)$, is fed to the loudspeaker and the response $y(t)$ is recorded with a microphone. The RIR $\hat{h}(t)$ is retrieved by linear convolution of the recorded signal $y(t)$ with the so-called inverse signal $v(t)$ ($\hat{h}(t) = y(t) \otimes v(t)$, with $\otimes$ indicating convolution). The inverse signal is built such that the linear

convolution of the sweep signal with the inverse signal produces a shifted delta function $s(t) \otimes v(t) = \delta(t - T)$. The inverse signal can be obtained by time-reversing the sweep signal, plus an amplitude scaling to compensate for the different energy content at various frequencies, as

$$v(t) = C \cdot \left(\frac{f_b}{f_a}\right)^{-(t/T)} s(T - t). \tag{2.3}$$

Here, C is a normalization constant, modified from [255] to include start and stop frequencies different from 0 and the Nyquist frequency, respectively, as

$$C = \frac{2\, f_b \, \ln(f_b/f_a)}{(f_b - f_a)\, T}. \tag{2.4}$$

The excitation signal used in the measurements presented in this chapter is the sweep signal defined in (2.2), with start frequency $f_a = 0.1\,\text{Hz}$ and stop frequency $f_b = f_s/2$, where $f_s = 48\,\text{kHz}$ is the sampling frequency. The duration of the sweep signal was set to $T = 5\,\text{s}$, followed by one second of silence, to ensure that the reverberant tail in the recorded signal has faded out.

The beginning and the end of the excitation signal is usually smoothed out using a tapering window in order to force the sweep to start and stop with zero phase, thus avoiding switching noise. In this way, ringing and ripples effects are reduced, at the expense of a slight deviation from the desired magnitude spectrum [25]. The tapering window used consisted of two ramp functions of length 1000 samples. The one at the beginning of the sweep signal was defined as a quarter of a cycle of a sinusoidal function (as suggested in [25]), while the one at the end of the sweep signal was a linear ramp function.

The spectrogram of the sweep signal is given in Figure 2.1 (using the `spgrambw` function included in the `voicebox` toolbox [258]), while the magnitude responses of the sweep signal, of the inverse signal and of the result of the convolution of the two is shown in Figure 2.2. From the latter, a slight deviation from the ideal uniformly flat magnitude response can be noticed. This effect is due to the tapering window and is only noticeable below 5 Hz, i.e. outside the frequency range of the subwoofers. The code for generating the sweep signal and its inverse was adapted from the code provided in [50].

The main sources of error in measuring RIRs are the presence of ambient noise, the nonlinear distortions caused by the loudspeaker, and the time-variance of the acoustic system due to changes in the room temperature or in the position of people. The ESS method is known to be robust in tackling these issues [58, 62]. According to (2.1), the IF grows faster as time advances, with the result that the excitation signal has a magnitude spectrum with a pink characteristic (-3

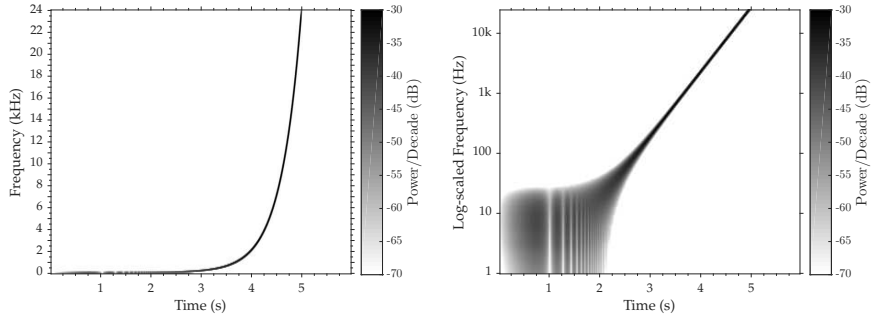

Figure 2.1: The spectrogram of the sweep signal in a linear frequency scale (left) and in a logarithmic frequency scale (right). In both plots, the power resolution is linear.

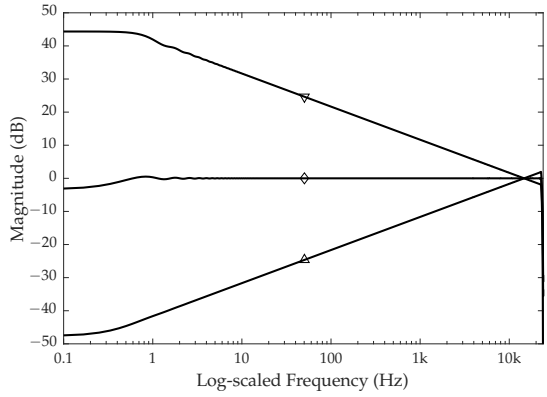

Figure 2.2: The magnitude responses of the sweep signal ($\triangledown$), of the inverse signal ($\triangle$) and of the linear convolution between the two ($\diamond$).

dB/octave). High SNR can be achieved because also the ambient noise normally has a spectrum with a pink characteristic, rather than white.

A characteristic of the ESS method is that the time-frequency correspondence of the sweep signal guarantees that at time t all the spectral components with frequency above the IF of the sweep are shifted before the causal RIR after convolution [61, 25, 58, 62]. As a consequence, the ambient noise having frequency above the IF of the sweep is pushed in the acausal part of the retrieved response, thus contributing to the increase of the SNR in the causal part. Moreover, this characteristic of the ESS method makes it quite robust against

impulsive noise, provided that the impulsive event does not occur towards the end or just after the sweep signal, in which case the energy of the impulsive noise would overlap with the causal room impulse response [58].

The same principle explains the ability of the ESS method to partially reject regular nonlinear harmonic distortions caused by the loudspeaker when driven beyond its linear operating range [251]; each order of distortion creates a sweep with IF proportional to its order, e.g. the second-order distortion has IF increasing twice as fast as the IF of the sweep signal. It follows that the linear convolution with the inverse signal pulls back these distortions into the non-causal part of the RIR. However, this is not true for all harmonic distortion artifacts; each order of distortion also creates sweeps with IF proportional to submultiples of its order, which means that odd-order distortions produce artifacts with the same IF as the sweep signal, that overlap with the causal part of the retrieved RIR. The same arguments are valid for irregular distortions caused by defects [250], such as rub & buzz; the ESS method is able to reject all the distortions with IF above the IF of the sweep.

A final consideration pertains to the sensitivity of the measurement method to the time-variance of the acoustic system. This is important because a better measurement SNR can be achieved by synchronous averaging of multiple measurements recorded for the same source-receiver position pair [61, 25, 57, 58]. It was shown in [58] that the ESS method is more robust to time variations, compared to other methods, so that an improvement of the SNR of 3 dB can be obtained by doubling the number of measurements (or alternatively the duration of the sweep signal) without introducing significant errors. In addition, the time variance is more prominent at high frequencies, so that synchronous averaging of multiple measurements can be safely applied to increase the SNR of the retrieved RIR at LFs.

2.3 Measurement Setup

2.3.1 Room description

The measurements were conducted in an empty small-sized room, aiming to model a typical domestic listening environment. The room dimensions were 4.09 m L × 6.35 m W × 2.40 m H, which satisfy the IEC 60268-13 specifications [259] and ensure a reasonably uniform distribution of low-frequency room modes. The theoretical values of the central frequencies of the first 20 room modes are given in Table 2.1 [1]. The structure is based on a brick construction comprising of lightly plastered painted walls, a wooden acoustic floating floor, and a

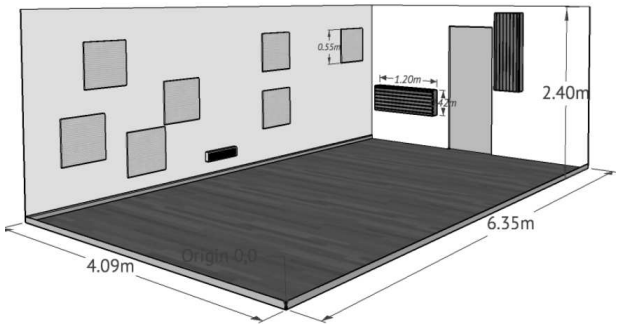

Figure 2.3: A sketch of the room at B&O headquarters, Struer, Denmark.

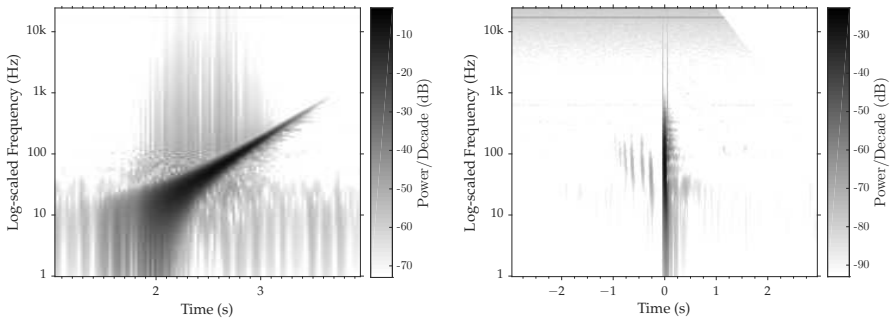

Figure 2.4: The spectrogram of the near-field recording $S_4^A M_{\text{nf}}^C R_1$ (left) and of the retrieved RIR (right). Notice the rub & buzz distortions above the sweep signal in the left plot, and the harmonic nonlinear distortions in the anti-causal part of the RIR in the right plot.

wooden suspended false ceiling filled with absorptive material. The IEC 60268-13 standard requires the room to be filled with ordinary room furnishings, semi-covered floor and reflective roof to achieve a certain degree of diffusion and absorption and meet a 'typical' RT (e.g. $\text{RT}_{200\text{Hz}-4\text{kHz}} = 0.3 - 0.6\,\text{s}$). During the measurements described here, the room was empty but included a total of 16 high-frequency acoustic panels (8 panels on each side wall), measuring $0.5 \times 0.5 \times 0.025\,\text{m}$ each, and 2 Helmholtz absorbers ($1.20 \times 0.42 \times 0.13\,\text{m}$) with resonance frequency 200 Hz and 300 Hz, attached on the rear wall. A sketch of the room is given in Figure 2.3. The air conditioning was kept off to limit possible low-frequency noise, but the room temperature was kept monitored at 21 °C (±1°C).

f_n (Hz)	n_x	n_y	n_z	f_n (Hz)	n_x	n_y	n_z
0	0	0	0	83.91	2	0	0
27.02	0	1	0	87.19	1	1	1
41.95	1	0	0	88.16	2	2	1
49.91	1	1	0	89.63	0	2	1
54.05	0	2	0	91.28	1	3	0
68.42	1	2	0	98.96	1	2	1
71.50	0	0	1	99.81	2	2	0
76.44	0	1	1	108.09	0	4	0
81.07	0	3	0	108.10	0	3	1
82.90	1	0	1	110.24	2	0	1

Table 2.1: The theoretical value of the eigenfrequencies, with the corresponding mode index numbers [1].

q	x	y	z	p	x	y	z
1	1.12	1.56	1.50	1	3.84	3.84	0.53
2	0.77	4.04	1.80	2	2.90	0.80	0.53
3	2.04	2.47	0.90	3	3.63	5.83	0.53
4	1.62	5.32	0.60	4	2.35	4.55	1.13
5	3.05	3.06	1.50				
6	3.09	5.07	1.00				

Table 2.2: Source-receiver positions (in meters). The source position corresponds to the center of the subwoofer cone.

2.3.2 Measurement equipment

Two types of subwoofers were used as sound sources. The first, denoted here as Subwoofer A, was a purpose-made loudspeaker based on a closed-box design (Genelec 1094), comprising of an 18" driver in a rigid wooden cabinet (V≈168 ℓ) and capable of reproducing frequencies well below 20 Hz (-6 dB_{SPL} at 14 Hz, based on near-field measurements described below). The second, denoted here as Subwoofer B, was a Genelec 7050B comprising of an 8" driver in a spiral bass reflex design and a metallic cylindrical cabinet, having a high-pass filter with cut-off frequency of 25 Hz and a low-pass filter with cut-off frequency set

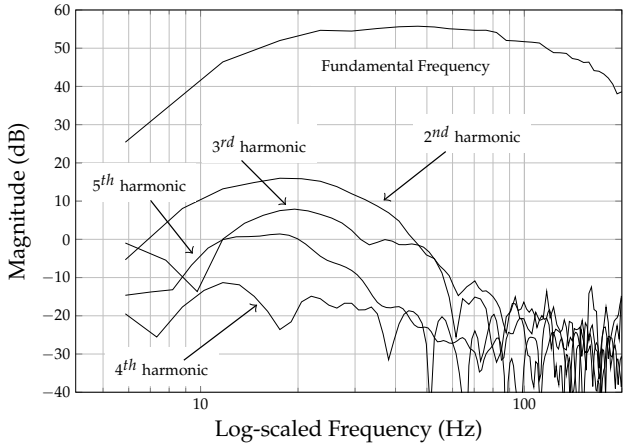

Figure 2.5: The harmonic distortion magnitude response for subwoofer A up to the fifth order.

at 120 Hz [260].

The responses were recorded by two microphones connected to a B&K 2669 preamplifier and a B&K NEXUS 2690-A conditioner. The first microphone, denoted here as Microphone C, was a B&K 4939 (1/4"), with a 0° incidence frequency range from 4 Hz to 100 kHz (±2 dB), thermal noise level of 28 dBA and sensitivity of 4 mV/Pa. The second microphone, denoted here as Microphone D, was a B&K 4133 (1/2"), with a 0° incidence frequency range from 4 Hz to 40 kHz (±2 dB), thermal noise level of 20 dBA and sensitivity of 12.5 mV/Pa. Microphones and subwoofers were connected to an RME UCX audio interface. No signal processing was enabled within the signal chain.

A total of 96 RIRs were measured in the room using the two subwoofers and the two omnidirectional microphones. Each subwoofer was placed at four positions in the room and measured at six microphone positions, completing a set of 24 source-receiver combinations, in conformity with ISO 3382-2 [256] for precision measurements. The source-receiver positions are summarized in Table 2.2. The notation $S_p^s M_q^m R_r$ will be used to refer to a particular recorded signal, with $s = \{A, B\}$ indicating the two subwoofers and $m = \{C, D\}$ indicating the two microphones, $p = \{1, \ldots, 4\}$ and $q = \{1, \ldots, 6\}$ indicating the source and receiver positions, respectively (see Table 2.2), and $r = \{1, \ldots, 10\}$ indicating the number of a particular recording. The subwoofer at position $p = 2$ is placed facing the door, whereas at the other positions it is placed facing the wall opposite the door.

2.3.3 Near-field and calibration measurements

In general, measuring the free-field response of a LF source requires rooms with very large dimensions. Keele [261] suggested that such measurements could be realized within a non-anechoic environment, by placing the receiver at a point of maximum pressure i.e. at the apex of the driver. The near-field measurements presented here were performed for subwoofer A placed at position $p = 4$ (see Table 2.2) with the microphone capsule placed at a distance of 5 mm on axis from the driver's cone at maximal outward displacement, as recommended in [261]. For subwoofer B, information is provided by the manufacturer.

Figure 2.4 shows the spectrogram of the near-field recording and of the retrieved RIR. In the spectrogram on the recorder signal, impulsive noise can be seen above the sweep. This artifact, which is not visible in the retrieved RIR, is often referred as rub & buzz distortion and is likely generated by the voice coil periodically beating some internal parts of the speaker, such as connection wires, loose particles or other defects [251, 250]. These distortions have a low level compared to the recorded sweep signal, approximately -50 dB below the peak of the signal , and will be either shifted in the non-causal part of the RIR or made not visible in the spectrogram of the retrieved RIR by the presence of the room resonances. It should be noticed that, being these types of distortion deterministic, averaging over multiple measurements will not decrease their level [250, 251].

Harmonic regular nonlinear distortions cannot be easily noticed in the spectrogram of the recorded signal, but become visible in the spectrogram of the retrieved RIR in the right plot of Figure 2.4; distortions at least up to the fifth order appear in the anti-causal part of the RIR. The level of the harmonic distortions is reported in Figure 2.5, where the magnitude response of the linear component and of the first four higher harmonics are depicted on a logarithmic frequency scale. Notice that the harmonic distortions are more prominent between 10 and 50 Hz, and tend to decay at higher frequencies. What is recorded above 90 Hz is practically ambient noise (the measured SNR was around 70 dB). A similar plot for Subwoofer B is provided in [260].

The microphones were calibrated with a B&K 4231. The output level of each subwoofer was then adjusted so that the sound level at 0.50 m was equal for the two subwoofers (56 dBC_{RMS} / peak 70 $\mathrm{dB\,SPL}_{\mathrm{at\,53\,Hz}}$)[2] when placed at the center of the room. Some of these calibration measurements are included in the database for reference.

[2]C-weighted root mean square (RMS) value obtained by reproducing pink noise at equal output level as the sine-sweep. Peak sound pressure level (SPL) obtained by reproducing sine-sweeps.

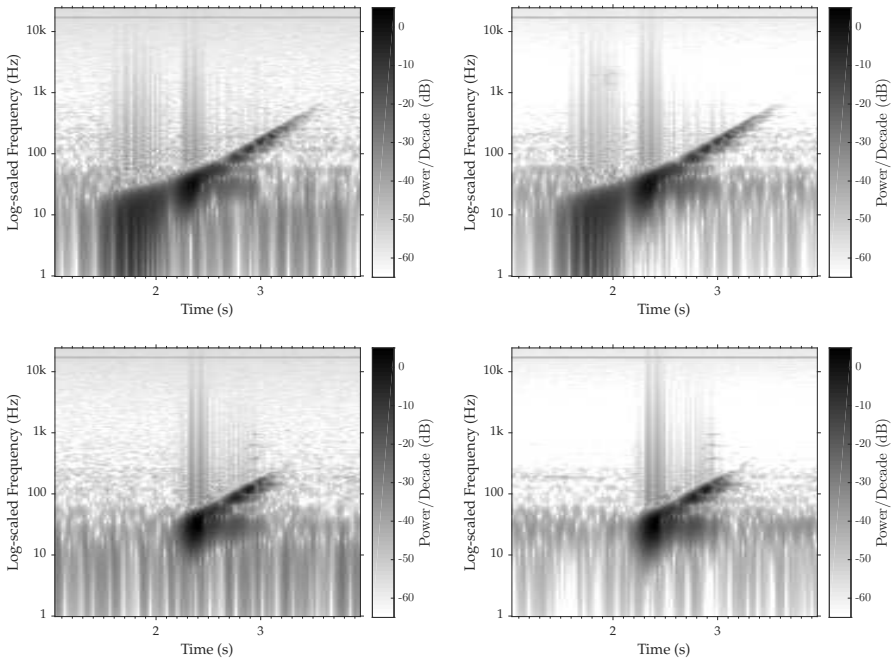

Figure 2.6: The spectrogram of the recorded signals $S_4^A M_6^C R_1$ (top left), $S_4^A M_6^D R_1$ (top right), $S_4^B M_6^C R_1$ (bottom left), and $S_4^B M_6^D R_1$ (bottom right). Notice the differences in the frequency response of the two subwoofers (top vs. bottom) and in the level of the ambient noise (left vs. right), and the steady component at 16 kHz. Also notice the wide power range.

2.4 Measurement analysis and postprocessing

2.4.1 Recorded signals

For each source-receiver position pair, 10 recordings were performed sequentially. The analysis of the recorded signals is important to detect possible issues and assess the quality of the measurements.

Figure 2.6 shows the spectrograms of the first recordings for the position pair $(p, q) = (4, 6)$ and for the four combinations of subwoofers and microphones. The following considerations apply in general for the other recordings and for the other source-receiver position pairs. The sweep signal is only partially

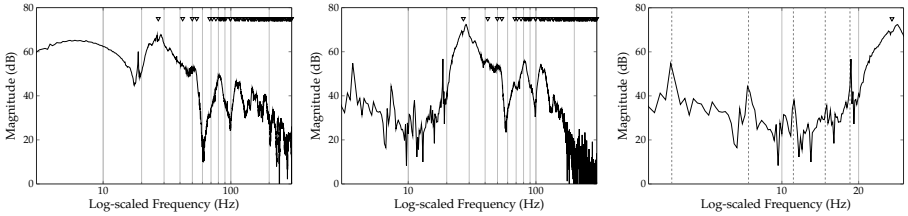

Figure 2.7: The magnitude response of the recorded signals $S_4^A M_6^D R_1$ (left) and $S_4^B M_6^D R_1$ (center). The frequency range between 3 Hz and 30 Hz (right) of the latter, showing the harmonic noise component (dashed lines). In all plots, the theoretical values of the eigenfrequencies ($\triangledown$) are shown (see Table 2.1).

reproduced, according to the frequency range of the subwoofer response (see Section 2.3). In comparison with the synthesized sweep signal in the right plot of Figure 2.1 or with the near-field measurement in Figure 2.4, it can be noticed how the recorded sweep is smeared out in time due to reverberation; in particular, from these plots we can expect a strong resonance between 20 and 30 Hz, corresponding to the first axial room mode (see Table 2.1). In these plots, all the difficulties inherent to LF-RIR measurements discussed earlier are visible. First, the LF ambient noise and the pink characteristic of its spectrum are evident. Second, irregular nonlinear distortion artifacts (or rub & buzz) for both subwoofers can be observed above the recorded sweep signal, as discussed for the near-field measurements (cfr. Section 2.3.3 and Figure 2.4). Finally, a steady component appearing in all measurements at 16 kHz can be observed in Figure 2.6. This disturbance, which is well above the frequency region of interest, was generated by a power adapter of one of the devices used for the measurements. From the comparison between different combinations of subwoofer and microphone, it can be seen how the $1/2$" microphone (M^D) (plots on the right in Figure 2.6) provides a lower noise level (≈ 5 dB difference), which is in agreement with specifications (see Section 2.3).

Figure 2.7 shows the magnitude response of recordings for the source-receiver position pair $(p, q) = (4, 6)$ with microphone D. It is clear that subwoofer A has a larger operational frequency range than subwoofer B. In particular, subwoofer A is able to partially excite the cavity mode (left plot); subwoofer B, on the other hand, has a frequency range between 25 Hz and 120 Hz (center plot). The same plot shows the presence of LF noise, which is not visible due to the cavity modal resonance in the left plot. Strong noise components are present at very LFs and have a harmonic structure, with fundamental frequency at 3.7 Hz (see right plot); as these components occur below the operating range of the subwoofer, they are unlikely related to the nonlinearities of the subwoofer, and are probably due to some external disturbance. Regarding the rub & buzz

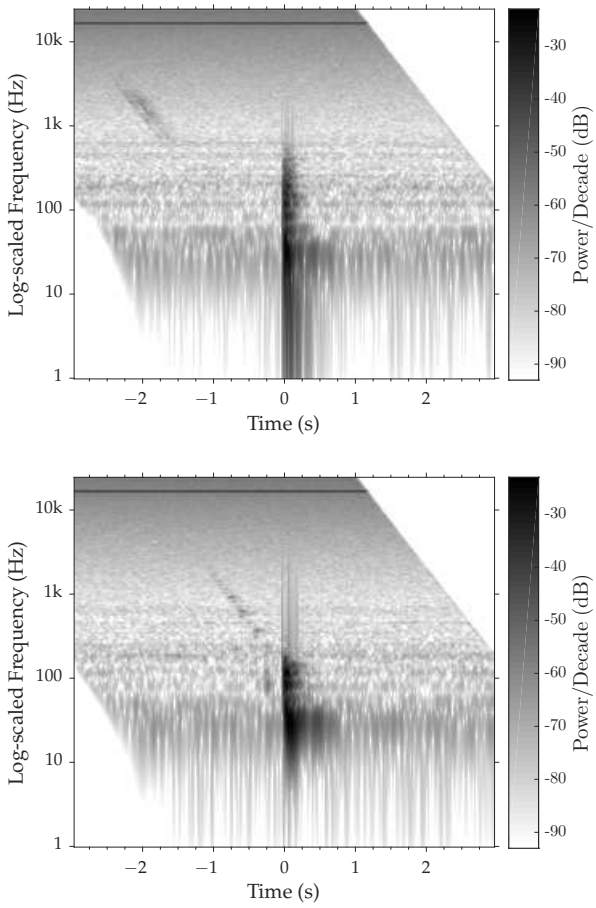

Figure 2.8: The RIRs retrieved from the recorded signals $S_4^A M_6^D R_1$ (top), $S_4^B M_6^D R_1$ (bottom).

distortion artifacts noticed in Figure 2.6, their characteristic impulsive nature does not allow them to be seen in the magnitude response, since they mix up with the ambient noise. According to Klippel [251, 250], these types of distortion would produce a harmonic spectrum if driven with a constant tone, which is not the case for a sweep with time-varying IF like the ESS sweep signal.

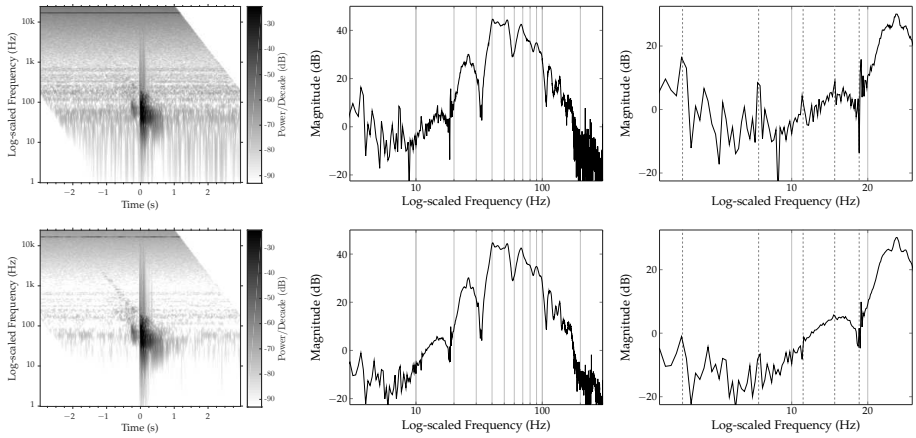

Figure 2.9: Synchronous averaging. The spectrogram and the magnitude response of the RIR retrieved from a single recording $S_3^B M_5^D R_1$ (top row) and the corresponding responses after averaging over 10 recordings. The frequency range between 3 Hz and 30 Hz (right) showing the harmonic noise component (dashed lines).

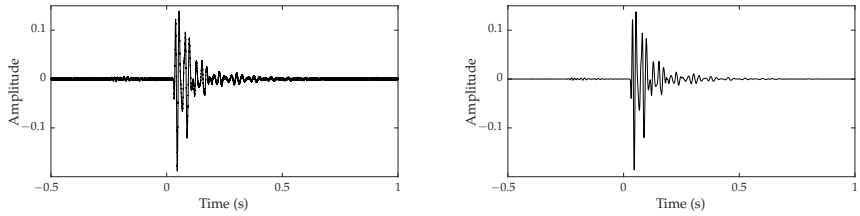

Figure 2.10: The retrived RIR $S_3^B M_5^D$ before (left) and after postprocessing (i.e. synchronous averaging over 10 recordings and low-pass filtering) (right).

2.4.2 Retrieved room impulse responses

The linear convolution necessary to retrieve the RIR is performed in the frequency domain by multiplying the discrete Fourier transform (DFT) of the recorded signal and of the inverse signal, computed with a DFT size equal to twice the number of samples of the signals ($2(T+1)f_s$), and then performing an inverse DFT. Figure 2.8 shows the spectrograms of the RIRs retrieved from the signals recorded at source-receiver position pair $(p, q) = (4, 6)$ using microphone D only (see right column of Figure 2.6). Compared to the spectrograms of the recorded signals, the LF noise in the retrieved RIRs is significantly reduced, as

a consequence of the higher SNR achieved with the ESS method at LFs. On the other hand, the ambient noise at high frequencies is amplified in the retrieved RIRs, as well as the 16 kHz steady component; this is probably due to the fact that the ambient noise spectrum is not exactly pink.

Another effect is visible in these spectrograms; an impulsive event appears in both cases as a downward slanted line starting in the anti-causal part of the response, likely to be attributed to a strong occurrence of the rub & buzz distortion. It is not clear if the impulsive event affects the linear causal part as well, its level being close to the ambient noise level. The same can be said for regular harmonic nonlinear distortions, which are not clearly distinguishable from the background noise (except for a 2^{nd} harmonic appearing in the bottom plot). Finally, well-separated room resonances with long decay are particularly noticeable as a smearing in time of the response in the causal part.

Postprocessing

In order to limit the presence of nonlinear distortions, a relatively low sound level of the subwoofer has been set (see Section 2.3.3). As a consequence, the SNR of the RIRs retrieved from a single recording is not very high. In order to increase the SNR, the following postprocessing operations are suggested. First, it is strongly recommended to perform a synchronous averaging over the RIRs retrieved from different recordings for a given source-receiver position pair and for a given subwoofer-microphone combination; as discussed already in Section 2.2, the robustness to time variations of the ESS method, especially at LFs, allows to perform such an averaging over the different recordings, thus obtaining an SNR improvement of 3 dB per doubling of the number of realizations [25, 58]. Notice that synchronous averaging could also be performed on the recorded signals before retrieving the RIRs by linear convolution, and that an alternative would be to double the length of the sweep signal.

The ESS method, however, has a poor noise rejection at high frequencies; a simple low-pass filtering can be applied to get rid of the high frequency noise (as well as the 16 kHz component). Finally, the non-causal part of the RIR can be discarded, if the interest is limited to the causal part only. A ready-to-use set of postprocessed RIRs, measured with subwoofer B and microphone D, for which a low-pass filter with cut-off frequency at 1 kHz and 100 Hz roll-off has been used, is available for download[3].

An example of the result of averaging is given in Figure 2.9, comparing the spectrogram and magnitude response of the RIR retrieved from a single

[3] https://lirias.kuleuven.be/bitstream/123456789/572970/3/SUBRIR_SpB_MicD_RIRs.zip (password: subrir2016)

recording (top) and after synchronous averaging over 10 recordings (bottom), with source-receiver position pair $(p, q) = (3, 5)$, and with subwoofer B and microphone D. From the magnitude responses, computed over the causal part of the RIR, it can be seen how averaging is able to reduce the noise level by at least 10 dB, including the very LF disturbance already noticed in Figure 2.7. From the spectrograms, it can be observed how the reduction in the noise level makes the nonlinear distortions more visible; the fact that the impulsive occurrences of the rub & buzz effect are not reduced in level after averaging, is a confirmation of the deterministic nature of these events. As a consequence, great care has to be taken in the setup of the subwoofer sound level during calibration, so that nonlinear distortions are kept to a minimum. The effect of synchronous averaging can be also seen in Figure 2.10, showing the RIR measured at position pair $(p, q) = (3, 5)$ for a single recording and after averaging over 10 recordings.

2.5 Reverberation time

The RT (or T_{60}) is defined as the time instant when the RIR energy decays by 60 dB from its peak value. This is usually calculated on the basis of the energy decay curve (EDC), i.e. the total amount of energy remaining in the impulse response at a given time [64]. The RT is taken as the time instant when the EDC drops below -60 dB. In most measurements, however, the noise floor level is above -60 dB and therefore this definition cannot be used in practice. In these cases, the RT is calculated using linear regression analysis and the least-squares fit procedure [256]. The decay curve is approximated by a line interpolating the EDC instead of using the EDC itself: the T_{10} is defined by interpolating the EDC between -5 and -15 dB, the T_{20} between -5 and -25 dB, and the T_{30} between -5 and -35 dB. The slope of the line interpolating the EDC within a given integration interval provides the decay rate d (in dB/s), from which an estimate of the RT is given as $-60/d$ [256]. The ISO 3382-2 standard [256] also requires the noise floor level to be at least 10 dB below the lower limit of integration, so that the the T_{30} can be reliably estimated only for an SNR of at least 45 dB.

Frequency-dependent values of the RT are generally estimated using a bank of full-octave or one-third-octave band-pass filters [256]. Estimating the RT in subbands at very LFs is problematic. The main issues are related to low SNR, to complex modal decays (such as beating modes or double decays) [65], and to the influence of the bandpass filters of the filterbank [257]. Let us first focus on the latter. At very LFs, typical one-third-octave filterbanks have band-pass filters with a very narrow bandwidth, resulting in a long decay which may exceed the RT of the RIR, especially if the attenuation requirement described

by the standard [262] are met. The central frequency $f_c(b)$ of the b^{th} band-pass filter and its bandwidth $B(b)$ are defined with respect to the 0^{th} band centered at 1000 Hz as [262]

$$f_c(b) = 2^{b/3}\, 1000 \text{ [Hz]}, \tag{2.5}$$

$$B(b) = f_c(b)\frac{2^{1/3}-1}{2^{1/6}} \text{ [Hz]}. \tag{2.6}$$

As the central frequency decreases, the bandwidth decreases exponentially, so that, e.g., the band $b = -16$ centered at $f_c(-16) = 25$ Hz has a bandwidth of just $B(-16) = 5.8$ Hz. A very narrow band-pass filter has poles very close to the unit circle of the z-transform domain, determining a slow decay of its impulse response. As a result, one-third-octave filterbanks yield a strong overestimation of the RT at LFs (at least for bands below $b = -12$, $f_c(-12) = 62.5$ Hz).

In order to reduce the influence of the filters, a cosine-modulated filterbank with all filters having the same bandwidth can be used. The cosine-modulated filterbank used has 10 channels evenly distributed over the range 0 Hz to 200 Hz, and was generated with a finite impulse response (FIR) prototype filter designed using the approach in [263], with a stop-band attenuation of 60 dB. The so-obtained band-pass filters have a fixed bandwidth of 20 Hz and a decay rate of 135 ms, which is expected to be lower than the RT of the room.

Another issue is associated to the low SNR of the RIR measurements, which results in a dynamic range not sufficient for the estimation of the T_{30}. Figure 11 shows that the T_{30}^{RIR} estimate is strongly biased due to the presence of noise, while the T_{10} estimate remains largely unaffected. A conservative choice would then involve using T_{10} for all frequency bands. However, as explained later in this section, the T_{10} estimates sometimes fail to capture phenomena such as double decays and beating modes. An alternative is to visually inspect the EDCs in each frequency bands (or estimate their noise floor level) and choose the most appropriate definition of the RT in each case.

In order to overcome this issue, an approach similar to [65] has been used. Here, instead of calculating the RT of the noisy RIR directly, it is calculated based on a best-fitting noiseless parametric room model. More specifically, the RIRs of the database are first approximated by an OBF model [27], which provides a representation of a RIR as a linear combination of resonant responses. The model parameter values are estimated using the OBF-GMP (group matching pursuit) algorithm described in [156], which is a scalable greedy algorithm with no limitations in the model order. The number of resonances used in the approximation was set to 70, which provided an accurate approximation (average normalized mean square error of -37 dB) without overfitting. This resulted in a nearly noiseless representation of the RIRs,

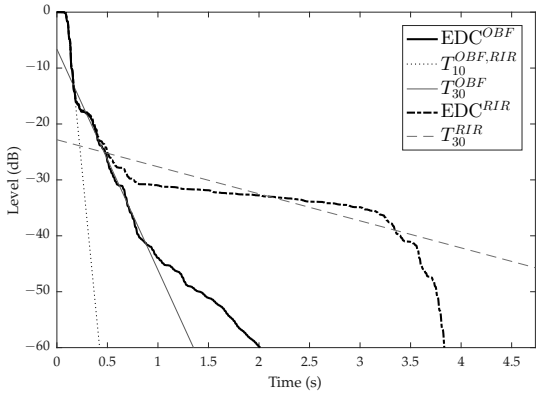

Figure 2.11: The EDCs calculated from a RIR ($S_2^A M_3^D$) after postprocessing and from its OBF approximation for the subband centered at 30 Hz. The interpolation lines for estimating the T_{30} and the T_{10} are also shown.

as shown in Figure 2.11. The figure shows the EDCs of a postprocessed RIR and of its OBF approximation for the subband centered at 30 Hz. Here, it is clear that the T_{30} value (which is calculated by interpolating the EDC between -5 and -35 dB) greatly overestimate the RT. On the other hand, the value obtained from the EDC of the OBF approximations is largely unaffected by noise. Notice also that the T_{10} is correctly estimated in both cases, as shown in Figure 2.11, with the two interpolating lines for the T_{10} overlapping.

Figure 2.12 shows the average RT values in each subband estimated from the OBF approximation of the RIRs retrieved from the signals recorded with microphone D (for microphone C, similar curves are obtained). Only the subbands centered within the limits of the frequency response of the subwoofers are considered. It can be seen that, while the T_{30} is around 400 ms above 75 Hz, it has much higher values at very LFs. This is probably due to the fact that the first axial mode, the one with theoretical frequency at 27 Hz, is very prominent. The influence of this mode can be clearly seen in both plots of Figure 2.12 in the T_{30} curve, where the highest values for the RT correspond to the band centered at 30 Hz. The T_{10} is also of interest in the modal region, where the low modal density gives rise to double decays and fluctuations due to beating modes [65]. A particularly large difference between the two decay rates is observed in Figure 2.11 for the frequency band around 30 Hz, and this is the reason why the T_{10} fails to capture the room resonant behavior in that region, as indicated in Figure 2.12.

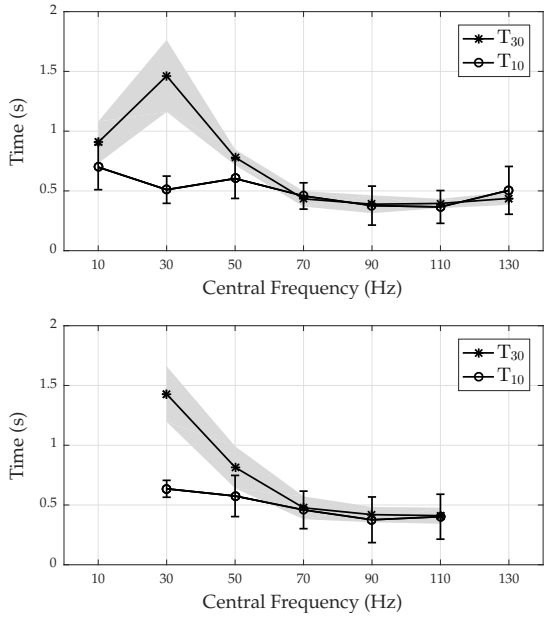

Figure 2.12: The average T_{30} ($*$) and the average T_{10} ($\circ$) for subwoofer A (top) and B (bottom) estimated from the OBF approximations of the RIRs retrieved from the signals recorded using microphone D. The shaded area and the vertical lines show the standard deviation for the T_{30} and the T_{10}, respectively.

2.6 Conclusion

A new RIR database measured with subwoofers as sound sources has been introduced, filling the gap of available acoustic measurements at LFs. Common difficulties in performing acoustical measurements at LFs have been addressed. The main issues proved to be a prominent LF ambient noise and the presence of impulsive irregular nonlinear distortions due to defects of the subwoofer (rub & buzz).

The ESS method has been chosen to estimate the RIRs, due to its robustness to nonlinear distortions and its capability of providing a higher SNR at LFs. However, not all distortions can be isolated using the ESS method, with impulsive distortions and odd-order harmonic distortions partially overlapping with the causal RIR. For this reason, near-field and calibration measurements become important to verify the nonlinear behavior of the subwoofer and to set the subwoofer level accordingly, so as to avoid distortion artifacts or at least to

reduce them to an acceptable level.

Synchronous averaging of the recordings for the same source-receiver position pair is also recommended, since it allows to achieve an SNR increase of 3 dB for each doubling of the number of recordings. The same increase can be achieved by doubling the length of the sweep signal, but with an increased risk of impulsive events occurring during the sweep.

Common difficulties in estimating the frequency-dependent RT at very LFs have been also addressed. The influence of the band-pass filters has been reduced by using a fixed-bandwidth cosine-modulated filterbank, while the problem of low SNR has been tackled by estimating the RT from a noiseless approximation of the RIRs obtained with OBF models.

The SUBRIR database is available for download[4] and it is expected to find application in the testing of acoustic signal enhancement algorithms intended for music reproduction and in the validation of physical models for room acoustics. The database has already been used in the validation of algorithms for multi-channel room acoustic system identification with fixed-pole adaptive digital filters [156, 264, 154, 113].

Acknowledgments

The authors would like to thank Bang & Olufsen A/S for the use of their premises and equipment.

[4] https://lirias.kuleuven.be/handle/123456789/572970 (password: subrir2016)

Part II

Modeling

Chapter 3

Modeling room impulse responses using orthonormal basis functions

A scalable algorithm for physically motivated and sparse approximation of room impulse responses with orthonormal basis functions

Giacomo Vairetti, Enzo De Sena, Michael Catrysse, Søren Holdt Jensen, Marc Moonen, and Toon van Waterschoot

Published in *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 7, pp. 1547—1561, Jul. 2017, DOI: 10.1109/TASLP.2017.2700940

G. Vairetti, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. van Waterschoot, "A scalable algorithm for physically motivated and sparse approximation of room impulse responses with orthonormal basis functions," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 7, pp. 1547—1561, Jul. 2017.
Changes include layout, representation, and minor editing aspects.

The candidate's contributions as first author include: literature study, co-development of the presented algorithms, software implementation and computer simulations, co-design of the evaluation experiments, co-formulation of the conclusions, text redaction and editing.

Abstract

Parametric modeling of room acoustics aims at representing room transfer functions (RTFs) by means of digital filters and finds application in many acoustic signal enhancement algorithms. In previous work by other authors, the use of orthonormal basis functions (OBFs) for modeling room acoustics has been proposed. Some advantages of OBF models over all-zero and pole-zero models have been illustrated, mainly focusing on the fact that OBF models typically require less model parameters to provide the same model accuracy. In this chapter, it is shown that the orthogonality of the OBF model brings several additional advantages, which can be exploited if a suitable algorithm for identifying the OBF model parameters is applied. Specifically, the orthogonality of OBF models does not only lead to improved model efficiency (as pointed out in previous work), but also leads to improved model scalability and model stability. Its appealing scalability property derives from a previously unexplored interpretation of the OBF model as an approximation to a solution of the inhomogeneous acoustic wave equation. Following this interpretation, a novel identification algorithm is proposed that takes advantage of the OBF model orthogonality to deliver efficient, scalable and stable OBF model estimates, which is not necessarily the case for nonlinear estimation techniques that are normally applied.

3.1 Introduction

Parametric modeling of room acoustics aims at representing room transfer functions (RTFs) by means of rational expressions in the z-transform domain, implemented through digital filters, and finds application in a variety of acoustic signal enhancement tasks, e.g. echo cancellation, feedback cancellation, and dereverberation, as well as in auralization systems. The most common parametric models are all-zero (AZ) models [26], which define a finite impulse response (FIR) filter as a truncation of a sampled room impulse response (RIR). AZ models enable achieving an arbitrary degree of accuracy, but a good approximation of a RIR usually requires a large number of model parameters. Pole-zero (PZ) models [68], which produce an infinite impulse response (IIR), are used sometimes in order to reduce the number of parameters. PZ models have a more meaningful motivation from a physical point of view, in the sense that the resonant behavior of room acoustic responses can be represented by means of complex-conjugate poles in the transfer function. This is particularly true when a PZ model is implemented using the parallel form of fixed-pole IIR filters [45, p.359]. This parallel filter (PF), proposed in recent years for

RTF modeling and audio equalization [80, 81, 86, 265, 82], consists of second-order all-pole filters, each of which is defined by a pair of complex-conjugate poles. Its transfer function is given by a linear combination of resonances, in analogy with the physical definition of a RTF as an infinite summation of room modes [1, 37, 109]. However, since RTFs are characterized by a complicated time-frequency evolution and a large number of room resonances, the improvement in modeling efficiency obtained with PZ models compared to AZ models is in some cases only marginal [239]. Moreover, PZ models often suffer from instability and ill-conditioning issues in the estimation of the model parameters, especially for high model orders, which is why AZ models are usually preferred.

In order for models producing an IIR to become a valid alternative to AZ models, models with improved model efficiency and with stable and numerically well-conditioned identification algorithms (and possibly other interesting properties) are sought. Fixed-pole models based on orthonormal basis functions (OBFs) [27, 126, 112, 266] can be derived directly from an orthogonalization of PF models. OBF models span the same approximation space of PF models for the same set of poles, with the difference that the outputs of each second-order all-pole filter are made orthogonal to each other by a sequence of all-pass filters (i.e. by zero-pole cancellation). The use of single-pole OBF models for acoustic echo cancellation [131, 34], and of multiple-pole OBF models for loudspeaker response equalization and modeling of room and musical instrument responses [148, 149, 150, 29, 28] have been previously motivated by the possibility of positioning the poles anywhere inside the unit circle, thus providing stability of the filter and giving freedom in the allocation of the frequency resolution. It has been shown that these properties, together with orthogonality, provide a more accurate representation of the RTF for a given number of model parameters, compared to conventional models. Differently from PF models, orthogonality makes the estimation of the parameters that appear linearly in OBF models straightforward and numerically well-conditioned. The poles, on the other hand, appear nonlinearly in the model, which makes their estimation a difficult problem, requiring in principle nonlinear estimation techniques. In [34], the pole parameters of single-pole OBF models were optimized using the Gauss-Newton method. In [148, 149, 150, 29, 28], multiple poles were estimated with a nonlinear iterative algorithm for FIR-to-IIR filter conversion, called the Brandenstein-Unbehauen (BU) method [73], resembling the Steiglitz-McBride (STMCB) method for PZ modeling [70]. The BU method exploits the orthogonality of OBF filters by minimizing the energy of a target RIR with a sequence of all-pass filters. Although this method is capable of producing accurate pole estimates, it is not exempt from numerical problems for high model orders, in which case the algorithm can converge to a local minimum and even produce unstable poles. Modifications of the BU method have been proposed to overcome this problem, such as through prewarping of

the target RIR, here called warped BU (wBU) [150, 91], in order to approximate a desired frequency resolution, or partitioning of the target RIR in frequency subbands or in time [30]. Furthermore, the BU method and its variants require the model order to be determined before estimation, resulting in a non-scalable algorithm that has to be run every time the number of poles to be estimated changes.

The nonlinear problem of estimating the poles was bypassed in [201] by applying convex optimization to a discrete grid of candidate stable poles. A sparse solution was obtained by selecting basis functions out of a large non-orthogonal dictionary. In [154], a matching-pursuit-based algorithm called OBF-matching pursuit (MP) was introduced. A similar algorithm was also suggested in [197] for the estimation of the poles of a RIR model described as the linear combination of sampled exponentially decaying sinusoids, but not considering any particular filter implementation of the model (if not the implicit use of FIR filters); however, the choice of this model implies a non-orthogonal dictionary and, consequently, ill-conditioning problems in the estimation of the parameters, which would require the use of computationally more complex versions of the algorithm, such as Orthogonal MP as in [267], or suboptimal iterative procedures [197]. The OBF-MP algorithm [154], instead, exploits the appealing properties of OBF models, i.e. orthogonality, stability and numerical well-conditioning, in order to deliver efficient, scalable and stable OBF model estimates for room acoustic modeling, which can be directly implemented through a stable IIR filter. It is shown in the present work that the scalability property of the algorithm stems from a previously unexplored interpretation of OBF models as an approximation to a solution of the inhomogeneous acoustic wave equation. Indeed, OBF models are physically motivated in the modal region, where the RTF is a linear combination of room resonances, sparse in frequency. The OBF-MP algorithm thus provides a sparse approximation of the most dominant modes in the low-frequency region of the RTF, while approximating the spectral envelope at higher frequencies. In this chapter, the OBF-MP algorithm is further investigated and its performance in terms of efficiency and computational complexity is studied for a large set of measured RIRs.

The chapter is organized as follows. In Section 3.2, fundamentals of the theory of room acoustics are briefly reviewed, together with an overview of conventional parametric models. In Section 3.3, the OBF models are reviewed in detail, as well as their use in the approximation of a target RIR. The OBF-MP algorithm is described in Section 3.4 and its computational complexity is analyzed. In Section 3.5, the concept of model and filter complexity of different parametric models is introduced. Simulation results are shown in Section 3.6, comparing the performance in the approximation of a large set of measured RIRs of OBF models estimated using the OBF-MP algorithm with respect to conventional

models and OBF models estimated using the BU method. A discussion of the results and future work can be found in Section 3.7, which also concludes the chapter.

3.2 Parametric modeling of room acoustics

This section reviews elements of room acoustics and provides an overview of conventional parametric models.

3.2.1 Fundamentals of room acoustics

The RTF between an omnidirectional point source $s(\boldsymbol{r}, t) = s(t)\delta(\boldsymbol{r} - \boldsymbol{r}_s)$ at position $\boldsymbol{r}_s = (x_s, y_s, z_s)$ (with $s(t)$ a given source function and $\delta(\cdot)$ the Kronecker delta function) and a receiver at position $\boldsymbol{r} = (x, y, z)$, can be seen as a linear superposition of room modes, mutually orthogonal in the space dimension, with the mode amplitudes depending on $\boldsymbol{r}$ and $\boldsymbol{r}_s$. This is described by the Green's function (GF) of the inhomogeneous acoustic wave equation, which, neglecting higher-order terms such as the variability of the temperature and of the density of the medium [1, 37], is given by

$$P(\boldsymbol{r}, \boldsymbol{r}_s, \omega) = G(\omega) \sum_{i=1}^{\infty} \frac{\psi_i(\boldsymbol{r})\psi_i(\boldsymbol{r}_s)\ j\omega}{\omega^2 - \omega_i^2 - 2j\zeta_i\omega_i + \zeta_i^2}, \tag{3.1}$$

with $P(\boldsymbol{r}, \boldsymbol{r}_s, \omega)$ the sound pressure in a room at the driving frequency ω for given receiver and source positions $\boldsymbol{r}$ and $\boldsymbol{r}_s$, and $G(\omega)$ a frequency-dependent gain constant. The eigenfrequencies ω_i, also called resonance frequencies [1], are the values of ω for which the acoustic wave equation has non-zero solutions satisfying the boundary conditions. The eigenfunction ψ_i corresponding to eigenfrequency ω_i defines a three-dimensional standing wave, called a room mode. A given room mode is dominant when the driving frequency ω is close to its resonance frequency ω_i, while it has little contribution to the sound field when the source or the receiver is placed on one of its nodal surfaces, i.e. where either $\psi_i(\boldsymbol{r}_s)$ or $\psi_i(\boldsymbol{r}_0)$ is close to zero. The damping constant ζ_i accounts for frequency-dependent energy losses at the walls and determines the -3 dB half-bandwidth of the room resonance, which is $B = \zeta_i/\pi$ (in Hz) [1, 37].

The inverse Fourier transform (FT) of (3.1) gives the RIR, which, for $t \geq 0$, is a sum of exponentially decaying sinusoids,

$$h(\boldsymbol{r}, \boldsymbol{r}_s, t) = \sum_{i=1}^{\infty} c_i e^{-\zeta_i t} \cos(\omega_i t + \phi_i), \tag{3.2}$$

where the i^{th} sinusoid has amplitude c_i and phase ϕ_i, resonance frequency ω_i and a decay determined by the damping constant ζ_i. The GF describes the sound field for any possible position of the source and the receiver inside any kind of room. However, a closed-form analytical expression for ψ_i exists only for simple room shapes and for idealized boundary conditions.

The problem of modeling a RIR presents many challenges, mainly because of its complicated time-frequency structure. The RIR measured in a reverberant room has typically a very long duration and presents a complicated pattern of the arrival of reflections. An example [268] of a typical RIR is shown in Fig. 3.1. Furthermore, the modal density increases with the square of the frequency ω, i.e. the approximate number of eigenfrequencies per Hz is given by

$$n_{\omega_i}(\omega) \approx \frac{V}{c^3\pi}\omega^2\,, \tag{3.3}$$

with V the volume of the room and c the sound velocity. The expression (3.3) is derived for rectangular rooms, but is asymptotically valid for rooms of any shape [1, 37]. It follows that the modes are well separated only at low frequencies, while they tend to overlap at higher frequencies. The so-called 'Schroeder frequency' [39] gives an indication as to where the transition between these two regions occurs:

$$f_{\text{Sch}} \approx 2000\sqrt{\frac{T}{V}}\,, \tag{3.4}$$

where T is the reverberation time, defined as the time it takes for the RIR to decay to 60 dB below its starting level, which depends on the damping characteristics of the walls [1]. This expression shows that the overlap is strong already at low frequencies especially for large halls and for rooms with highly absorptive surfaces, for which resonances have larger bandwidth. A consequence of the overlap is that in diffuse field conditions, i.e. above the Schroeder frequency f_{Sch}, the number of magnitude peaks in the RTF in a given range is much lower than the theoretical number of modes [1]. The idea of modeling a RTF using OBF models is then to use a finite number of resonant responses, as opposed to the infinite summation in equations (3.1) and (3.2), to model accurately low-frequency well-separated dominant room modes and to approximate the spectral envelope of overlapping modes at higher frequencies.

3.2.2 Conventional parametric models for room acoustics

Parametric modeling of room acoustics aims at approximating the GF in (3.1) by a rational function in the z-domain,

$$H(\ddot{\boldsymbol{r}}, z) = \frac{B(\ddot{\boldsymbol{r}}, z)}{A(\ddot{\boldsymbol{r}}, z)} = \frac{\sum_{i=0}^{Q} b_i(\ddot{\boldsymbol{r}})z^{-i}}{1 + \sum_{i=1}^{P} a_i(\ddot{\boldsymbol{r}})z^{-i}}, \tag{3.5}$$

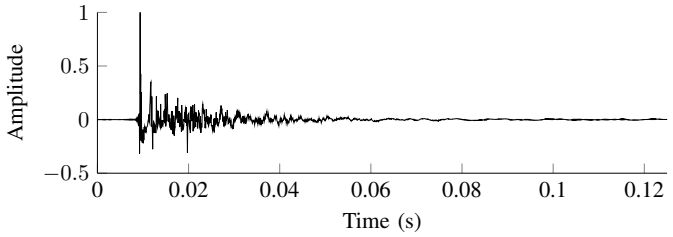

Figure 3.1: RIR measured in the Speech Lab at KU Leuven.

where $\ddot{\boldsymbol{r}} = (\boldsymbol{r}_s, \boldsymbol{r})$ denotes a particular source-receiver position pair. Common assumptions to be made are stability, causality, linearity, and time-invariance of the acoustic system. The expression in (3.5) can be rewritten in a pole-zero form by factorizing the numerator and denominator polynomials, yielding

$$H(\ddot{\boldsymbol{r}}, z) = \frac{B(\ddot{\boldsymbol{r}}, z)}{A(\ddot{\boldsymbol{r}}, z)} = b_0(\ddot{\boldsymbol{r}}) \frac{\prod_{i=1}^{Q} \{1 - q_i(\ddot{\boldsymbol{r}}) z^{-1}\}}{\prod_{i=1}^{P} \{1 - p_i(\ddot{\boldsymbol{r}}) z^{-1}\}} \,. \tag{3.6}$$

The zeros q_i represent anti-resonances and time delays in the RIR, while poles p_i are associated with room resonances.

AZ models [26], for which $A(\ddot{\boldsymbol{r}}, z) = 1$, can achieve an arbitrary degree of accuracy by using a high-order FIR filter. The main problem is that the number of parameters of the filter necessary to model the resonant behavior of the system often has to be quite large, depending on the sampling frequency f_s and the reverberation time T. Furthermore, the RIR strongly depends on the source and receiver position, so that the parameter values obtained for approximating a RIR at a given source-receiver position $\ddot{\boldsymbol{r}}_1 = (\boldsymbol{r}_{s_1}, \boldsymbol{r}_1)$ are in general significantly different from those for a RIR at another position $\ddot{\boldsymbol{r}}_2 = (\boldsymbol{r}_{s_2}, \boldsymbol{r}_2)$.

Models producing an IIR are used in an attempt to reduce the number of parameters needed to approximate a target RIR [269]. PZ models [68] uses both zeros and poles, so that both room resonances and time delays can be modeled, as well as the non-minimum-phase components of the RTF. However, since both $A(\ddot{\boldsymbol{r}}, z)$ and $B(\ddot{\boldsymbol{r}}, z)$ in (3.6) are non-constant polynomials in z^{-1}, no closed-form solution exists to the model parameter estimation problem and nonlinear optimization methods are required. These methods usually start from the estimation of an all-pole model and then iteratively compute optimal parameter values in the Least Squares (LS) sense. The most popular one is the so-called STMCB method [70], which, however, is not guaranteed to converge and may become unstable, especially for high model orders. Another difficulty lies in determining the optimal values for Q and P in (3.5) or (3.6), i.e. the order of the numerator and denominator polynomial, respectively.

PF models, which use the parallel form of fixed-pole IIR filters [45, p.359] consisting of a parallel of second-order all-pole filters, result from a partial fraction expansion (PFE) of the transfer function in (3.5), which, for $Q < P$, can be written as

$$H(\ddot{\boldsymbol{r}}, z) = \sum_{i=1}^{P} \frac{R_i(\ddot{\boldsymbol{r}})}{(1 - p_i(\ddot{\boldsymbol{r}})z^{-1})}, \tag{3.7}$$

where R_i are the residues of the poles p_i. If $Q \geq P$, an FIR filter of order $Q - P + 1$ should be added to the right-hand side of the equation [45, pp.112-114],[67, 265]. When the coefficients of $A(z)$ and $B(z)$ in (3.5) are real, complex poles will occur in conjugate pairs, so that for each one-pole filter defined by (R_i, p_i) there will be a one-pole filter defined by (R_i^*, p_i^*). These two terms can be added together to form a real second-order section, so that (3.7) becomes

$$H(\ddot{\boldsymbol{r}}, z) = \sum_{i=1}^{P/2} \left\{ \frac{R_i(\ddot{\boldsymbol{r}})}{1 - p_i(\ddot{\boldsymbol{r}})z^{-1}} + \frac{R_i^*(\ddot{\boldsymbol{r}})}{1 - p_i^*(\ddot{\boldsymbol{r}})z^{-1}} \right\}, \tag{3.8}$$

whose impulse response, with $n = tf_s$ the discrete time variable, is given by

$$h(\ddot{\boldsymbol{r}}, n) = \sum_{i=1}^{P/2} \left\{ R_i(\ddot{\boldsymbol{r}}) \left[p_i(\ddot{\boldsymbol{r}})\right]^n + R_i^*(\ddot{\boldsymbol{r}}) \left[p_i^*(\ddot{\boldsymbol{r}})\right]^n \right\}, \tag{3.9}$$

which is a finite sum of pairs of geometric series, each for a pair of complex-conjugate poles. After some elaborations, this can be shown to be equivalent to

$$h(\ddot{\boldsymbol{r}}, n) = \sum_{i=1}^{P/2} 2|R_i(\ddot{\boldsymbol{r}})|\rho_i^n \cos(\sigma_i n + \angle R_i(\ddot{\boldsymbol{r}})), \tag{3.10}$$

with ρ_i and σ_i respectively the radius and the angle of the pole $p_i = \rho_i e^{j\sigma_i}$, which is a finite linear combination of exponentially decaying sinusoids sampled in time, with amplitude and phase determined by the residues R_i. It is evident by comparing the expressions in (3.10) and (3.2) that a RIR can be approximated by a PF model using poles with radius and angle defined by the damping constants ζ_i and the resonance frequencies ω_i as

$$\begin{aligned} \rho_i &= e^{-\zeta_i / f_\mathrm{s}}, \\ \sigma_i &= \omega_i / f_\mathrm{s}\,. \end{aligned} \tag{3.11}$$

Notice that, when ρ_i is small and σ_i is close to either 0 or π, the resonance generated by p_i is influenced by the resonance generated by p_i^*, so that their magnitude peaks have frequency slightly different from $\pm\omega_i$ [67, 213]. The PF model as an approximation of the GF was first discussed in [109] in relation

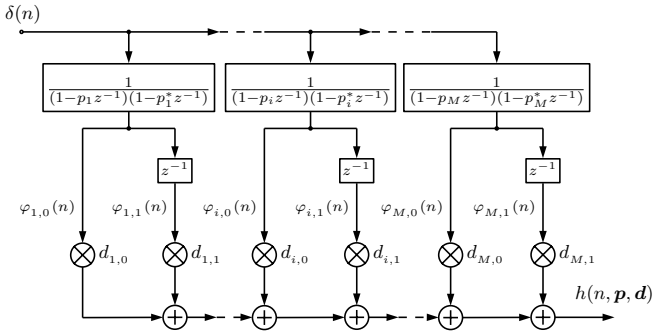

Figure 3.2: The PF model structure (with $M = P/2$). The impulse responses to the second-order IIR filters, denoted by $\varphi_{i,0}$ and $\varphi_{i,1}$ for $i = 1, \ldots, M$, are used as basis functions in a linear-in-the-parameters model structure.

to the modeling of a RTF by using common-acoustical-poles and their residues (CAPR). It has been shown that the GF in (3.1) is a normalized mean square error (NMSE) for the resonance frequencies, which can be approximated by a PF model, assuming $\zeta_i \ll \omega_i$. It is also shown that the residues $R_i(\ddot{\boldsymbol{r}})$ are related to the eigenfunctions ψ_i of the GF, thus expressing the variation of the RTF at different source and receiver positions.

The transfer function of the PF in (3.8) can be rearranged as

$$H(\ddot{\boldsymbol{r}}, z) = \sum_{i=1}^{P/2} \left[\frac{d_{i,0}(\ddot{\boldsymbol{r}}) + d_{i,1}(\ddot{\boldsymbol{r}}) z^{-1}}{(1 - p_i(\ddot{\boldsymbol{r}}) z^{-1})(1 - p_i^*(\ddot{\boldsymbol{r}}) z^{-1})} \right],$$

$$d_{i,0}(\ddot{\boldsymbol{r}}) = \mathrm{Re}\{R_i(\ddot{\boldsymbol{r}})\} = |R_i(\ddot{\boldsymbol{r}})| \cos(\angle R_i(\ddot{\boldsymbol{r}})), \tag{3.12}$$

$$d_{i,1}(\ddot{\boldsymbol{r}}) = \mathrm{Re}\{R_i(\ddot{\boldsymbol{r}}) p_i^*(\ddot{\boldsymbol{r}})\} = |R_i(\ddot{\boldsymbol{r}})\, p_i| \cos(\sigma_i - \angle R_i(\ddot{\boldsymbol{r}})),$$

and implemented as a parallel of second-order filters, shown in Fig. 3.2, which is linear in the parameters $\{d_{i,0}, d_{i,1}\}$, but nonlinear in the poles $\{p_i, p_i^*\}$. Each second-order section models a room resonance, with resonance frequency and bandwidth determined by the position of $\{p_i, p_i^*\}$, within the unit circle in order to ensure stability. Particular attention should be given to repeated poles, which produce polynomial amplitude envelopes on the decaying exponentials [67], the order of which is determined by the multiplicity of the repeated pole. It should be noticed that, in the presence of repeated poles, the model structure in Fig. 3.2 has to be modified accordingly.

3.3 Orthonormal basis function models

Parametric models based on OBFs can be derived from an orthogonalization of PF models. The orthogonality of the basis functions, together with the linearity in the parameters, introduces some desirable properties which bring a number of advantages in terms of efficiency and numerical stability in the modeling of RIRs. In this section, OBF models are also described as a generalization of other parametric models. Furthermore, their properties are described along with their application in the approximation of a target RIR.

3.3.1 Construction of OBF models

OBF models are derived with a Gram-Schmidt orthonormalization procedure applied to one- and two-pole filters [27, 126, 112]. Starting from a normalized first-order IIR filter with pole p_1 and transfer function

$$\Psi_1(z, p_1) = \frac{A_1}{1 - p_1 z^{-1}}, \tag{3.13}$$

where $A_1 = \sqrt{1 - |p_1|^2}$ is a normalization factor, a second-order filter with poles $[p_1, p_2]$ and transfer function orthogonal to (3.13) can be obtained as

$$\Psi_2(z, [p_1, p_2]^T) = \frac{A_2(z^{-1} - p_1^*)}{(1 - p_1 z^{-1})(1 - p_2 z^{-1})}, \tag{3.14}$$

with $A_2 = \sqrt{1 - |p_2|^2}$ and with * indicating complex conjugation. The orthogonality of Ψ_1 and Ψ_2 is provided by the zero in $z = 1/p_1^*$ and can be investigated using Cauchy's residual theorem via the inner product on the Hardy space on the unit circle $\mathcal{H}_2(\mathbb{T})$ (with $\mathbb{T} \triangleq \{z : |z| = 1\}$) as (see [112])

$$\langle \Psi_1, \Psi_2 \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Psi_1(e^{j\omega}) \Psi_2^*(e^{j\omega}) d\omega = \frac{1}{2\pi j} \oint_{\mathbb{T}} \Psi_1(z) \Psi_2^*(1/z^*) \frac{dz}{z} = 0. \tag{3.15}$$

The transfer function in (3.14) can be seen as the product of a normalized first-order IIR filter defined by p_2 and a first-order all-pass filter defined by p_1. By repeating the procedure for a set of poles $\boldsymbol{p}_i = \{p_1, \ldots, p_i\}$, the i^{th} transfer function will consist of a normalized first-order IIR filter defined by p_i and a sequence of first-order all-pass filters defined by the pole set $\boldsymbol{p}_{i-1} = \{p_1, \ldots, p_{i-1}\}$,

$$\Psi_i(z, \boldsymbol{p}_i) = \left(\frac{\sqrt{1 - |p_i|^2}}{1 - p_i z^{-1}} \right) \prod_{l=1}^{i-1} \left(\frac{z^{-1} - p_l^*}{1 - p_l z^{-1}} \right), \tag{3.16}$$

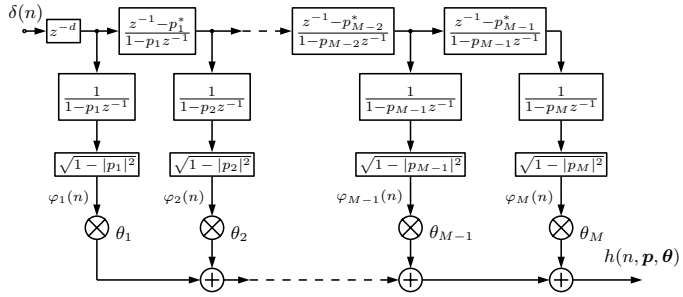

Figure 3.3: The Takenaka-Malmquist OBF model structure for M real poles.

which is also known as the Takenaka-Malmquist function [27]. The corresponding model structure is shown in Fig. 3.3, where the model output $h(n, \boldsymbol{p}, \boldsymbol{\theta})$ is a linear combination of the responses of the basis functions, weighted by the linear parameters θ_i.

An OBF model based on the functions in (3.16) can be seen as a generalization of other well-known models. If all the poles are identical and real, the Laguerre model [123] is obtained, which is in turn a normalized version of a so-called warped FIR filter model [91], with the value of the warping parameter the repeated real pole. If the pole is placed in the origin, then the Laguerre filter simplifies to an AZ model.

When the pole set $\boldsymbol{p}_i$ contains complex poles, the basis functions in (3.16) are generally complex-valued and are thus not useful for the identification of real systems. As for PF models, two real-valued basis functions can be obtained by combining pairs of complex-conjugate poles, and by orthogonalizing each pair of basis functions with respect to each other (plus a normalization factor). Different realizations of an OBF model can be obtained for particular choices of these normalization factors, as explained in [112]. A combination of a Takanaka-Malmquist model and the so-called Kautz model can be used, as suggested in [149], modeling real and complex poles, respectively. This model structure, henceforth called mixed-Kautz model, is shown in Fig. 3.4 for $\dot{m}$ real poles and $\ddot{m}$ pairs of complex-conjugate poles. The basis functions of a mixed-Kautz model are defined for a real pole p_i as

$$\Psi_i(z, \boldsymbol{p}_i) = \left(\frac{A_i}{1 - p_i z^{-1}} \right) \prod_{l=1}^{i-1} \left(\frac{z^{-1} - p_l^*}{1 - p_l z^{-1}} \right), \tag{3.17}$$

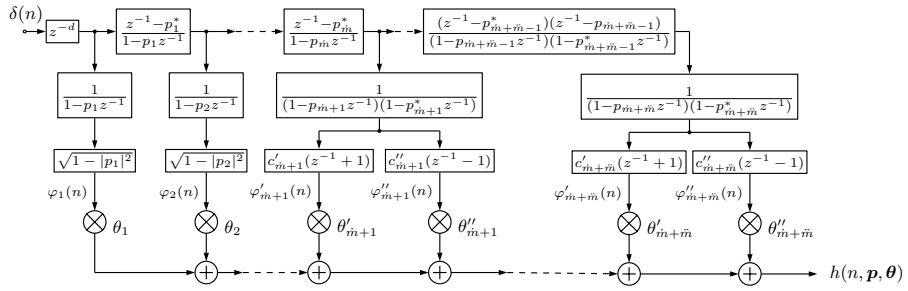

Figure 3.4: The mixed-Kautz model structure for $\dot{m}$ real poles and $\ddot{m}$ pairs of complex-conjugate poles. For convenience, the basis functions corresponding to real poles defined in (3.17) are followed by the basis functions corresponding to complex-conjugate pole pairs defined in (3.18).

or for a complex-conjugate pole pair $\{p_{i-1}, p_i\} = \{p_i, p_i^*\}$ as

$$\begin{aligned}
\Psi'_i(z, \boldsymbol{p}_i) &= \frac{c'_i(z^{-1}+1)}{(1-p_i z^{-1})(1-p_i^* z^{-1})} \prod_{l=1}^{i-2} \frac{(z^{-1}-p_l^*)}{(1-p_l z^{-1})}, \\
\Psi''_i(z, \boldsymbol{p}_i) &= \frac{c''_i(z^{-1}-1)}{(1-p_i z^{-1})(1-p_i^* z^{-1})} \prod_{l=1}^{i-2} \frac{(z^{-1}-p_l^*)}{(1-p_l z^{-1})}.
\end{aligned} \tag{3.18}$$

with $A_i = \sqrt{1-|p_i|^2}$, and normalization factors $c'_i = |1-p_i|A_i/\sqrt{2}$ and $c''_i = |1+p_i|A_i/\sqrt{2}$. Notice that the pair of basis functions in (3.18) are built as a product of a sequence of $i-2$ first-order all-pass filters given by the poles in $\boldsymbol{p}_{i-2}$, a second-order all-pole filter defined by $\{p_i, p_i^*\}$ and a normalization term, so that the model structure for complex-conjugate poles is given by a parallel of orthonormalized second-order IIR filters. However, real poles may not be of much interest in the approximation of measured RTFs; even though positive real poles would be useful for modeling the cavity mode of a room response, a measured RTF has a band-pass characteristic, with a cut-off at low frequencies determined by the response of the high-pass filter of the loudspeaker, and a cut-off at high frequencies given by the low-pass behavior of the loudspeaker or the anti-aliasing filter. For this reason, only complex-conjugate poles can be considered, thus resulting in the use of a Kautz model.

3.3.2 Properties of OBF models

The orthogonality of the basis functions provides some desirable properties. First, the OBFs form a complete set in $\mathcal{H}_2(\mathbb{T})$, under the assumption that $\sum_{i=0}^{\infty}(1-|p_i|)=\infty$ [112]. Thus, by decomposing a target RIR in terms of an orthogonal expansion, the approximation error can be made arbitrarily small by choosing a large enough number of poles.

Second, orthogonality provides flexibility, which results from the fact that poles can be arbitrarily positioned inside the unit circle (for the sake of stability), and that frequency resolution can be allocated unevenly in different regions of the spectrum without numerical conditioning problems, regardless of the model order. This is not the case, for example, for PZ models, where problems of ill-conditioning and instability can arise for high model orders.

Third, OBF models are linear in the parameters θ_i, which means that linear regression can be applied in order to estimate their optimal values. Moreover, due to the orthogonality of the basis functions, it is not necessary to carry out a matrix inversion, which is often a source of numerical problems. Another consequence of orthogonality is the fact that the parameters θ_i for each IIR filter are independent from the ones for others filters in the structure, so that a model of lower order can be obtained from a model of higher order only by truncation, and similarly additional poles can be included without recomputing the values of the θ_i's corresponding to the poles already used. An additional advantage of OBF models over PF models is that the same pole can be included more than once (e.g. to model modes with a double decay) without the need to modify the structure. These properties are exploited in the scalable algorithm described in Section 3.4.

3.3.3 Approximation of a RIR with an OBF model

The approximation of a target RIR $h(n)$ using an OBF model consists in estimating the parameters in the pole set $\boldsymbol{p}=\{p_i\}$ and in the set of parameters $\boldsymbol{\theta}=\{\theta_i\}$, with $i=1,\ldots,M$ (cfr. Fig. 3.4 where $M=\dot{m}+2\ddot{m}$), that minimize the distance between a target RIR $h(n)$ and the model response $h(n,\boldsymbol{p},\boldsymbol{\theta})$ for $n=1,\ldots,N$. For a fixed set of poles $\boldsymbol{p}$, the problem of estimating $\boldsymbol{\theta}$ is linear and can be solved in closed form. The response $h(n,\boldsymbol{p},\boldsymbol{\theta})$ of an OBF model for an impulse input signal $\delta(n)$ is the linear combination of the responses $\varphi_i(n,\boldsymbol{p}_i)$

of the M basis functions $\Psi_i(z, \boldsymbol{p}_i)$ (see e.g. Fig. 3.3 or Fig. 3.4),

$$\begin{aligned} h(n, \boldsymbol{p}, \boldsymbol{\theta}) &= \sum_{i=1}^{M} \theta_i \Psi_i(z, \boldsymbol{p}_i) \delta(n) \\ &= \sum_{i=1}^{M} \theta_i \varphi_i(n, \boldsymbol{p}_i) = \boldsymbol{\varphi}(n, \boldsymbol{p})^T \boldsymbol{\theta}, \end{aligned} \tag{3.19}$$

where $\boldsymbol{\varphi}(n, \boldsymbol{p})$ is a vector containing the responses $\varphi_i(n, \boldsymbol{p}_i)$ at time n. By stacking all the vectors $\boldsymbol{\varphi}(n, \boldsymbol{p})$ for $n = 1, \ldots, N$ in a matrix $\boldsymbol{\Phi}(\boldsymbol{p})$ of size $N \times M$, the optimal values for $\boldsymbol{\theta}$ for a given input-output set $\{\boldsymbol{\delta}, \boldsymbol{h}\} = \{\delta(n), h(n)\}_{n=1}^{N}$ can be estimated in LS sense as

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\Phi}(\boldsymbol{p})^T \boldsymbol{h}. \tag{3.20}$$

Note that the LS estimation does not require any matrix inversion, given that the orthonormality of the basis functions implies $\boldsymbol{\Phi}(\boldsymbol{p})^T \boldsymbol{\Phi}(\boldsymbol{p}) = \boldsymbol{I}_M$. It can be seen from (3.20) that the optimal estimate for $\boldsymbol{\theta}$ corresponds to the correlation of the basis functions in $\boldsymbol{\Phi}(\boldsymbol{p})$ with the target RIR vector $\boldsymbol{h}$, so that $\hat{\boldsymbol{\theta}}$ gives the degree of similarity between each basis function and the target RIR.

The problem of estimating the optimal pole set $\hat{\boldsymbol{p}}$ can be then regarded as finding the poles that generate basis functions that are maximally correlated with the target RIR, so that the approximation error between the model response and the target RIR is minimized. However, no closed-form solution to the pole estimation problem is available. The state-of-the-art approach for the multiple-poles case is the BU method [148, 149, 150, 29, 28], an iterative nonlinear method based on FIR-to-IIR filter conversion [73]. Frequency prewarping of the target response has been proposed for audio applications in order to match a particular frequency-scale mapping, such as the Bark scale [78]. The BU method exploits the orthogonality of OBF models and provides accurate estimates for the pole parameters. However, the model order has to be predetermined, and stability problems can arise from numerical issues at high model orders.

3.4 The OBF-MP algorithm

The problem of sparse linear approximation of a signal consists in finding a compact representation by a combination of functions taken from an overcomplete basis. These functions are usually called predictors or atoms, which altogether form a basis, sometimes called dictionary. The most popular methods for sparse approximation can be divided in two main categories

[203]. In the first one, convex optimization techniques are used to minimize a functional, such as the ℓ_1 norm in the least absolute shrinkage and selection operator (LASSO) [270]. The second category includes iterative greedy algorithms, such as orthogonal matching pursuit (OMP) [271, 272, 204].

Our approach aims to find a sparse approximation of a target RIR as a linear combination of a finite number of OBFs. A RIR cannot be considered a sparse time-frequency signal itself, with a certain degree of sparsity only in the modal region. By first modeling dominant low-frequency modes and the spectral envelope at higher frequencies, the proposed algorithm is able to provide a sparse approximation of a RIR using a finite-order OBF model. In this section, an OMP-based greedy algorithm, which is termed here OBF-MP [154], is used to iteratively select poles from a large set of candidate poles distributed over the unit disc, thus bypassing the inherent nonlinear problem. At each iteration of the OMP algorithm, the predictor that has the highest correlation with the current residual response is selected. The problem in the OBF-MP algorithm, given that OBFs are defined by previous poles in the structure, consists in defining a dictionary of candidate predictors, where the dictionary has to be updated at each iteration using the predictor selected at the previous iteration. The advantage of OBF-MP over the conventional OMP algorithm is that the orthogonal projection of the current residual response onto the set of predictors selected at the previous iterations is not necessary. The predictors, in fact, are already orthogonal to each other by construction. This ensures that the algorithm does not contain any matrix inversion, thus avoiding ill-conditioning problems. Moreover, since the candidate predictors are orthogonal to the predictors selected at previous iterations, computing the correlation with the current residual response is equivalent to computing the correlation with the target RIR.

Another consequence of orthogonality is the scalability of the algorithm, from which it follows that the number of parameters of the final model structure does not have to be defined in advance. A pole and the related linear coefficient are estimated at each iteration, independently of poles selected at previous iterations. It follows that additional poles can be estimated just by running extra iterations of the algorithm, without any problem of instability or numerical ill-conditioning. This scalability property of the algorithm is also a consequence of the fact that, similarly to what was discussed at the end of Section 3.2 for PF models, also OBF models can be regarded as a way of approximating a RTF. It has been already mentioned that, for the same set of non-repeated poles, the basis functions of a PF model and the ones of an OBF model span the same approximation space, so that it is possible to convert the values of the linear parameters from one model form to the other by simply a linear transformation [81].

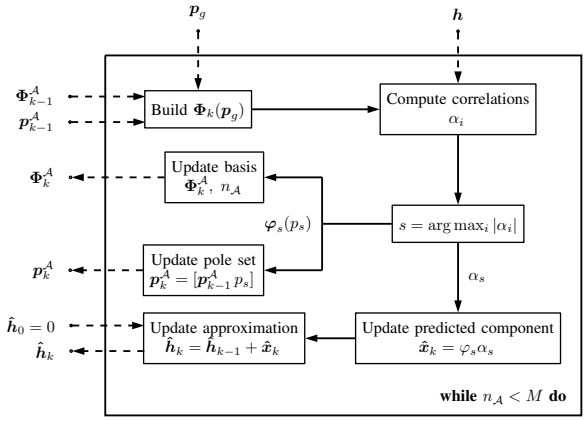

Figure 3.5: The OBF-MP algorithm block diagram. Inbound dashed lines represent initial conditions and inputs, while outbound dashed lines represent outputs.

Following the above interpretation, the idea of the OBF-MP algorithm is to iteratively compute a sparse approximation $\hat{\boldsymbol{h}}$ of a target RIR $\boldsymbol{h}$ of length N samples as a linear combination of length-N OBFs, analogously to the definition of a RIR as a summation of exponentially decaying functions, independent one from each other. The OBFs are selected from a dictionary $\boldsymbol{\Phi}_k$ of candidate predictors $\boldsymbol{\varphi}_i$ $(i = 1, \ldots, D)$ and included in the basis $\boldsymbol{\Phi}_k^{\mathcal{A}}$. At each iteration k, D candidate predictors $\boldsymbol{\varphi}_i$, orthogonal to the predictors in the current basis $\boldsymbol{\Phi}_{k-1}^{\mathcal{A}}$ constructed with the current set of active poles $\boldsymbol{p}_k^{\mathcal{A}}$, are built from G poles placed arbitrarily in a grid $\boldsymbol{p}_g$ inside the upper half of the unit disc. The matrix $\boldsymbol{\Phi}_k$ has dimensions $N \times D$, with $D = \dot{m} + 2\ddot{m}$ where $\dot{m}$ and $\ddot{m}$ denote respectively the number of real poles and complex poles in the grid $\boldsymbol{p}_g$, so that $G = \dot{m} + \ddot{m}$.

The OBF-MP algorithm is described in detail below, and a graphical representation is depicted in Fig. 3.5. First, a grid of G candidate poles $\boldsymbol{p}_g$ is defined, similarly to [201, 154, 197], with poles distributed according to a desired frequency resolution or prior knowledge about the system. In [201, 154, 197], the angle and the radius of the poles were distributed either uniformly or logarithmically on the unit disc, with the latter option intended to increase the resolution at low frequencies. Here, a different pole grid is used, depicted in Fig. 3.6, henceforth referred to as Bark-exp grid; the radius ρ_i of the poles decreases exponentially at the increase of the angle σ_i, as suggested in [28], according to $\rho_i = \varrho^{\frac{\sigma_i}{\pi}}$, with ϱ the value of the radius defined at the Nyquist frequency. Regarding the values for ϱ, it is suggested here to set the number

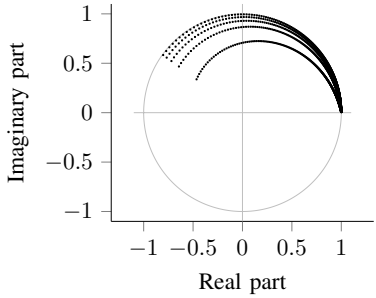

Figure 3.6: The Bark-exp pole grid for the OBF-MP algorithm (here with 5 radii and 400 angles and upper angle limited to 0.8π).

of radii for each angle and distribute them logarithmically in order to increase density toward the unit circle. Differently from [28], in which the angles follow a logarithmic scale, the Bark frequency scale [78] was chosen. The Bark scale further increases the resolution at low frequency, providing an effect similar to the prewarping of the RIR used in the wBU method. In this way, a higher density of poles close to the unit circle is achieved at low frequencies, allowing a more accurate approximation of energetic and narrow-bandwidth resonances, while at higher frequencies poles sparser in frequency and more distant from the unit circle provide a coarser approximation.

At the first iteration, the current basis $\boldsymbol{\Phi}_0^{\mathcal{A}}$ and the set of active poles $\boldsymbol{p}_0^{\mathcal{A}}$ are empty (with the number of predictors in the basis $n_{\mathcal{A}} = 0$). Also the target approximation vector $\hat{\boldsymbol{h}}_0$, is initially set to zero. At each iteration k, the matrix of candidate predictors $\boldsymbol{\Phi}_k(\boldsymbol{p}_g)$ is updated according to the mixed-Kautz structure in Fig. 3.4. The matrix $\boldsymbol{\Phi}_k$ has always dimension $N \times D$ (since an OBF model admits repeated poles, a pole that is selected by the algorithm is not removed from the pole grid $\boldsymbol{p}_g$) and its columns are the OBFs built from the poles in $\boldsymbol{p}_g$ with transfer functions as in (3.17) and in (3.18), thus orthonormal to the predictors in the current basis $\boldsymbol{\Phi}_{k-1}^{\mathcal{A}}$ built from the poles in the current active pole set $\boldsymbol{p}_{k-1}^{\mathcal{A}}$. The predictor(s) in $\boldsymbol{\Phi}_k$ that has the largest absolute correlation α_i with the target RIR vector $\boldsymbol{h}$ is selected and added to the basis $\boldsymbol{\Phi}_k^{\mathcal{A}}$, while the corresponding pole is included in the set of active poles $\boldsymbol{p}_k^{\mathcal{A}}$. The correlation for real and complex poles is computed in two different ways. For real poles, the correlation is the projection of $\boldsymbol{h}$ onto the predictor $\boldsymbol{\varphi}_i$ ($\alpha_i = \boldsymbol{\varphi}_i^T \boldsymbol{h}$). For a pair of complex-conjugate poles $\{p_i, p_i^*\}$ the correlation is the projection of $\boldsymbol{h}$ on the plane defined by predictors $\boldsymbol{\varphi}_i'$ and $\boldsymbol{\varphi}_i''$(see Fig. 3.7), which are mutually orthogonal, and is given by

$$\alpha_i = \sqrt{\alpha_i'^2 + \alpha_i''^2} = \sqrt{(\boldsymbol{\varphi}_i'^T \boldsymbol{h})^2 + (\boldsymbol{\varphi}_i''^T \boldsymbol{h})^2}. \tag{3.21}$$

Algorithm 1 OBF-MP algorithm

$\boldsymbol{p}_g = \{p_1, \ldots, p_G\}$ ▷ Define poles in the pole grid
$\boldsymbol{\Phi}_0^{\mathcal{A}} = \emptyset, n_{\mathcal{A}} = 0$ ▷ Initialize set of active predictors (basis)
$\boldsymbol{p}_0^{\mathcal{A}} = \emptyset$ ▷ Initialize the set of active poles
$\hat{\boldsymbol{h}}_0 = \mathbf{0}$ ▷ Initialize target approximation vector
$k = 1$ ▷ k: iteration counter
while $n_{\mathcal{A}} < M$ **do** ▷ M: desired number of OBFs
  Build $\boldsymbol{\Phi}_k(\boldsymbol{p}_g)$ ▷ Build matrix of candidate predictors $\boldsymbol{\varphi}_i$
  $s = \arg\max_i |\alpha_i|$ ▷ Find $\boldsymbol{\varphi}_i$ max. correlated with $\boldsymbol{h}$
  $\boldsymbol{p}_k^{\mathcal{A}} = [\boldsymbol{p}_{k-1}^{\mathcal{A}}\ p_s]$ ▷ Add selected pole to active pole set
  if p_s real **then** ▷ Update basis and predicted component
    $\boldsymbol{\Phi}_k^{\mathcal{A}} = [\boldsymbol{\Phi}_{k-1}^{\mathcal{A}}\ \ \boldsymbol{\varphi}_s],\ n_{\mathcal{A}} = n_{\mathcal{A}} + 1$
    $\hat{\boldsymbol{x}}_k = \boldsymbol{\varphi}_s \alpha_s$
  else if p_s complex **then**
    $\boldsymbol{\Phi}_k^{\mathcal{A}} = [\boldsymbol{\Phi}_{k-1}^{\mathcal{A}}\ \ \boldsymbol{\varphi}_s'\ \ \boldsymbol{\varphi}_s''],\ n_{\mathcal{A}} = n_{\mathcal{A}} + 2$
    $\hat{\boldsymbol{x}}_k = [\boldsymbol{\varphi}_s'\, \boldsymbol{\varphi}_s''][\alpha_s'\, \alpha_s'']^T$
  end if
  $\hat{\boldsymbol{h}}_k = \hat{\boldsymbol{h}}_{k-1} + \hat{\boldsymbol{x}}_k$ ▷ Update target approximation vector
  $k = k + 1$
end while

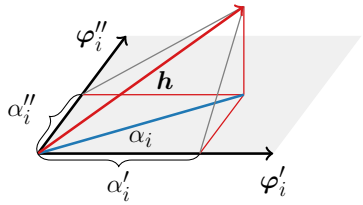

Figure 3.7: Graphical interpretation of the correlation between the target RIR vector $\boldsymbol{h}$ and the predictors of a pair of complex-conjugate poles $\{p_i, p_i^*\}$.

The k^{th} predicted component $\hat{\boldsymbol{x}}_k$ is obtained from the last added predictor(s) using the maximum correlation α_s as regression coefficient (with $s = \arg\max_i |\alpha_i|$), by $\hat{\boldsymbol{x}}_k = \boldsymbol{\varphi}_s \alpha_s$, if the selected pole is real, or by $\hat{\boldsymbol{x}}_k = [\boldsymbol{\varphi}_s'\, \boldsymbol{\varphi}_s''][\alpha_s'\, \alpha_s'']^T$, otherwise. The current target RIR approximation vector $\hat{\boldsymbol{h}}_k$ is obtained by adding the predicted component $\hat{\boldsymbol{x}}_k$ to the previous target RIR approximation vector $\hat{\boldsymbol{h}}_{k-1}$. As a consequence of its scalability property, the

algorithm can terminate when the desired number M of predictors in the basis is reached, or alternatively when the approximation error falls below a desired value.

3.4.1 Algorithmic complexity analysis

Here the asymptotic computational complexity of the OBF-MP algorithm is analyzed, assuming for simplicity that only complex poles are included in the pole grid. With reference to Algorithm 1, there are two operations that determine the asymptotic behavior of the algorithm. Building the matrix $\boldsymbol{\Phi}_k$ of candidate predictors (step 7) at each iteration is the most demanding operation, which involves the generation of D predictors of length N, which sums up to a complexity of $\mathcal{O}(3ND)$ multiplications (cfr. the expressions in (3.18) and Figure 3.4). The second operation to consider is the computation of the correlation coefficients (step 8), which is a multiplication of the matrix $\boldsymbol{\Phi}_k$ with the vector $\boldsymbol{h}$ of length N, which results in $\mathcal{O}(ND)$ multiplications. The computational complexity associated to vector updates and other operations is negligible. The overall complexity of the OBF-MP algorithm after $k = M/2$ iterations, is $\mathcal{O}(2MND)$ multiplications, i.e. linearly proportional to the three variables considered. In other words, the computational complexity increases linearly with the length of the impulse response, the number of candidate poles, and the number of iterations. This is comparable with the complexity of the BU method, whose most demanding operation is represented by the solution to a set of overdetermined linear equations, which implies a QR factorization of a large $N \times M$ rectangular matrix (complexity $\mathcal{O}(NM^2)$), followed by a back-substitution of a $M \times M$ triangular matrix (complexity $\mathcal{O}(M^2)$) [273]. This is performed for I iterations, with the overall computational complexity of the BU method summing up to $\mathcal{O}(INM^2)$, which is quadratic with respect to the number of estimated poles M.

3.5 Model and filter complexity

In this section, the complexity of the parametric models presented in the previous sections will be analyzed from two different perspectives. First, the *model complexity* (or *representation complexity*) C_m is considered, which is the number of parameters that is necessary to represent the system under study. Second, the *filter complexity* (or *simulation complexity*) C_f is considered as the number of operations that are required to obtain the filter output signal for a given input signal when the parameter values are available. While a measure often used in the literature is the *model order*, it is believed that the two

concepts just proposed are less prone to misinterpretation and thus preferable for the comparison of different parametric models in terms of complexity. For simplicity, OBF models and PF models having complex-conjugate pole pairs only are considered.

Model complexity

The calculation of the model complexity C_{m} is straightforward for AZ and PZ models. By referring to (3.5) and (3.6), the number of parameters for AZ models corresponds to the number of numerator coefficients ($C_{\mathrm{m}} = Q + 1$), while for PZ models it is the sum of denominator and numerator coefficients ($C_{\mathrm{m}} = P + Q + 1$). For PF models, if $P/2$ is the number of complex-conjugate poles pairs, the number of parameters required is $C_{\mathrm{m}} = 2P$, since each second-order section can be represented with one pole p_i (which is a complex number defined by two parameters, while p_i^* is given by complex conjugation) and two linear parameters (denoted by $d_{i,0}$ and $d_{i,1}$ in (3.12) and in Fig. 3.2). The same is obtained for OBF models, in which the all-pass filters and the normalization factors can be computed from the knowledge of the poles (see e.g. Fig. 3.4). The model complexity C_{m} is summarized in the left column of Table 3.1.

Filter complexity

The filter complexity C_{f} is calculated here as the number of multiplications required to compute the filter output for a given input signal. For AZ and PZ models, one multiplication is required for each coefficient, so that $C_{\mathrm{f}} = C_{\mathrm{m}}$. This is true also for PF models, in which four multiplications are required for each second-order section, two for the second-order IIR filter and two for the linear parameters (see Fig. 3.2). In case of repeated poles, the structure has to be modified, but the number of multiplications remains the same [67]. For OBF models, the normalization coefficients c_i' and c_i'' can be combined together with the related linear parameters θ_i' and θ_i'', so that the only difference between OBF models and PF models in terms of filter complexity is determined by the orthogonalization. By including the second-order all-pass filters in the structure, two more multiplications per section have to be included (assuming that the input of the all-pass filter is the output of the previous second-order IIR filter), summing up to six per section, so that the filter complexity for $P/2$ pairs $\{p_i, p_i^*\}$ is $C_{\mathrm{f}} = 3P$. The filter complexity C_{f} is summarized in the right column of Table 3.1. Notice that an OBF model is more complex than a PF model. However, these two models span the same approximation space for the same set of poles, thus leading exactly to the same filter response when the optimal linear coefficients are computed using the ℓ_2 norm in LS design. It

Table 3.1: Model and filter complexity

model	C_m	C_f
AZ	$Q+1$	$Q+1$
PZ	$P+Q+1$	$P+Q+1$
PF	$2P$	$2P$
OBF	$2P$	$3P$

would be then possible to convert an OBF model into a PF model with lower filter complexity, as was also suggested in [80].

3.6 Simulation results

The modeling performance of the OBF-MP algorithm described in Section 3.4 was tested on $R = 41$ RIRs measured for several source-receiver positions in three different rooms with different reverberation times. The RIRs were taken from three publicly available databases, namely MARDY [242], SMARD [50], and MIRD [243]. A fourth database of 24 low-frequency RIRs, called SUBRIR [274], was used separately to evaluate the modeling performance of the algorithm in the modal frequency region, as will be discussed later in this section. Their specifications, such as the room volume V, the surface area S, the reverberation time T, the Schroeder frequency f_Sch computed as in (3.4), and the mixing time t_m, are listed in Table 3.2. According to [53], the most accurate estimate of the mixing time t_m, i.e. the time instant at which the diffuse reverberation tail begins, is given by a formula related to the concept of mean free path length, given by $t_\mathrm{m} = 20V/S + 12$ (in ms). Notice that the MIRD database includes RIRs measured in a room where the reverberation time is controlled by means of movable acoustic panels, resulting in 3 different values of T in Table 3.2. All target RIRs are sampled at $f_\mathrm{s} = 48\,\mathrm{kHz}$ and truncated to $N = 6000$ samples. This corresponds to the shortest 'useful duration', defined as the time instant where the SNR of the measured RIR is 10 dB [65]. In order to compute the SNR value, the decay curve and the noise floor level were estimated with the method by Lundeby et al. [56]. Since the modeling of the delay of the RIR is not part of the scope of this chapter, the direct path component was considered as the starting point of the RIR. However, a simple delay could be easily included in the model structure of the OBF model by setting the parameter d in Fig. 3.4.

Table 3.2: Database specifications

database	V (m^3)	S (m^2)	t_{m} (ms)	T (s)	f_{Sch} (Hz)	RIRs
SMARD	170.4	207.3	28.4	0.15	59	8
MARDY	208.8	255.6	28.0	0.45	93	9
				0.16	86	8
MIRD	86.4	129.6	25.3	0.36	129	8
				0.61	168	8
SUBRIR	62.3	102.1	24.2	0.5-1.5	>180	24

In the simulations presented in this section, an approximated response $\hat{\boldsymbol{h}}^{(r)}$, with $r = 1, \ldots, R$, was computed for each target RIR $\boldsymbol{h}^{(r)}$ using the OBF-MP algorithm. OBF models obtained with OBF-MP were compared to AZ and PZ models and to OBF models obtained with the wBU method suggested in [29], henceforth called OBF-wBU models. The measure used to compare the performance of different models with the same model complexity C_{m} is the NMSE, averaged over all R RIRs, which in the time domain is given by

$$h_{\mathrm{NMSE}}(\mathrm{dB}) = 10\log_{10}\frac{1}{R}\sum_{r=1}^{R}\frac{\|\boldsymbol{h_r} - \hat{\boldsymbol{h}}_{\boldsymbol{r}}\|_2^2}{\|\boldsymbol{h_r}\|_2^2}, \tag{3.22}$$

while the average frequency response NMSE is defined as

$$H_{\mathrm{NMSE}}(\mathrm{dB}) = 10\log_{10}\frac{1}{R}\sum_{r=1}^{R}\frac{\|\boldsymbol{H_r} - \hat{\boldsymbol{H}}_{\boldsymbol{r}}\|_2^2}{\|\boldsymbol{H_r}\|_2^2}, \tag{3.23}$$

with $\boldsymbol{H}_r$ and $\hat{\boldsymbol{H}}_r$ the discrete Fourier transform (DFT) of $\boldsymbol{h}_r$ and $\hat{\boldsymbol{h}}_r$, respectively. The NMSE was computed on the complete time response ($h_{\mathrm{NMSE}}^{\mathrm{full}}$), and on the early ($h_{\mathrm{NMSE}}^{\mathrm{early}}$) and late ($h_{\mathrm{NMSE}}^{\mathrm{late}}$) responses separately. The time instant separating the two parts was set to 25 ms, corresponding to the shortest mixing time t_m for the three rooms considered (see Table 3.2). Also the NMSE in the frequency response was analyzed for the frequency range between 0 Hz and 20 kHz ($H_{\mathrm{NMSE}}^{\mathrm{full}}$), as well as at low/mid frequencies between 0 and 4 kHz ($H_{\mathrm{NMSE}}^{\mathrm{low/mid}}$), and at high frequencies between 4 kHz and 20 kHz ($H_{\mathrm{NMSE}}^{\mathrm{high}}$), in order to show the differences in performance of the models in different frequency ranges. Although the Schroeder frequency in (3.4) would have been a more natural choice for separating the frequency range, its value in the databases considered was found to be below or just above the lower cut-off frequency of the loudspeaker used for the measurements. The upper limit of 20 kHz was chosen to avoid considering the frequency range dominated by the influence of the anti-aliasing filter.

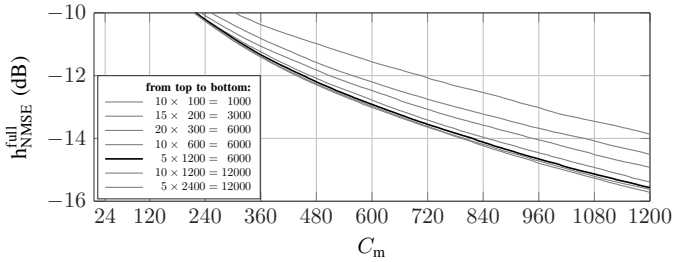

Figure 3.8: The average time-domain NMSE in (3.22) for different pole allocations and densities of a Bark-exp grid. The darker line indicates the grid chosen for the simulations.

The Bark-exp grid used in these simulations counts $G = 6000$ poles with 5 different radii distributed logarithmically with values ϱ at Nyquist from 0.5 to 0.99, and 1200 different angles placed from 48 Hz to 19.2 kHz according to the Bark frequency scale [78] with Bark-warping factor $w = 0.766$. The limits on the angle were chosen to avoid approximating the response below the cut-off frequency of the loudspeakers and above the cut-off frequency of the anti-aliasing filter. As a result, the grid contains only complex poles. The reason of such an uneven allocation of the number of radii and angles is due to the frequency resolution that is required to approximate low-frequency resonances and to the observation that increasing the resolution in the angle is more important than in the radius. Using 1200 angles provides a constant resolution of 2.5 Hz below 500 Hz; this seems to be already a sufficient resolution, as confirmed by the results depicted in Fig. 3.8, showing the average full-response time-domain NMSE over a selection of 10 RIRs, computed for different allocations of radii and angles of poles in the Bark-exp grid. It should be noted in the figure that doubling the number of poles in the grid from 6000 to 12000 does not provide a significant increase in the accuracy.

The wBU method, also using Bark-warping factor $w = 0.766$, was slightly modified in order to avoid numerical instabilities. Although it has been proved in [73] that the BU method provides a stable IIR filter by conversion of an FIR filter, we observed cases where the solution with the minimum conversion error contains poles outside the unit circle. In the same paper, instabilities were noticed in those cases where the FIR filter was maximum-phase. However, in those cases, the conversion error was supposed to be high, so that it was always possible to find a stable solution with low conversion error. In [29], numerical limitations were observed in the wBU method and in the computation of the roots of the poles for a number of parameters C_m above 300. Since higher values of C_m were considered in our simulations, the wBU method was modified by

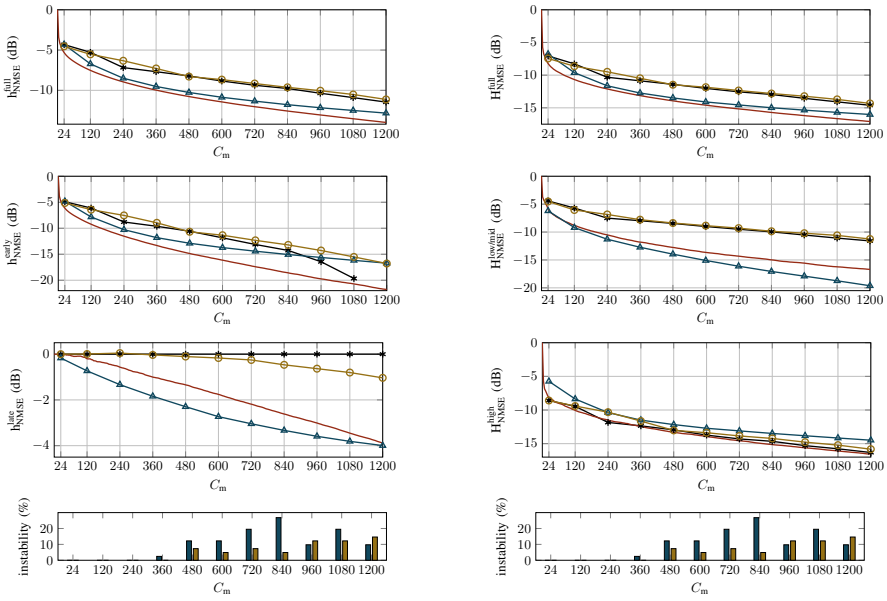

Figure 3.9: The average NMSE vs. the model complexity C_m. (left) The average time-domain NMSE in (3.22) for the entire response (top) and for the early (middle) and late response (bottom). (right) The average frequency-domain NMSE in (3.23) for the entire frequency range (top) and for the frequency regions $[0, 4]$ kHz (middle) and $[4, 20]$ kHz (bottom). AZ models (∗), PZ models (○), OBF models obtained with OBF-wBU (△) and with OBF-MP (—). At the bottom, occurrences of unstable solutions for OBF models obtained with OBF-wBU (left bars) and PZ models (right bars) are reported (same plot on both columns).

choosing the first stable solution with minimum conversion error. The number of iterations was set to 100. For PZ models, the STMCB method [70] with $P = Q + 1$ parameters has been used, with the initial estimate obtained with Prony's method [72]. In order to reduce the number of unstable solutions, only three iterations were executed.

Fig. 3.9 presents simulation results comparing the performance of the different models for varying model complexity C_m. The MATLAB code for generating the results presented in this section is available online[1]. The NMSE produced by OBF models obtained with OBF-MP was computed at each iteration, while for other models the NMSE was computed only for given values of C_m. In the bottom row of the figure, the occurrences of unstable solutions given by the

[1] https://lirias.kuleuven.be/handle/123456789/581178

wBU method and the STMCB method are reported. For the wBU method, the first stable solution was used, as described above, while for PZ models, unstable solutions were removed from the calculation of the average NMSE. It is clear that both methods suffer from instability due to ill-conditioning above certain values of C_{m}.

In the left column of Fig. 3.9, results for the NMSE in the time domain defined in (3.22) are given for the complete response, for the early reflections and for the late reverberation. The plot on top shows that OBF models provide in general a better approximation of the target RIR over N samples, with OBF-MP outperforming AZ models even in the approximation of the early response (middle plot), except when AZ models achieve perfect modeling (at $C_{\mathrm{m}} = 1200$, $h^{\mathrm{early}}_{\mathrm{NMSE}}(\mathrm{dB}) = -\infty$ for AZ models). Focusing on OBF models, OBF-MP shows an overall improvement over OBF-wBU, with the former having a better performance in the early part of the response and the latter performing better in the late part (bottom plot).

The plots in the right column of Fig. 3.9 show results in the frequency domain. Results in the frequency range between 0 and 4 kHz show a clear improvement in the approximation of the low/mid frequencies given by OBF models, with OBF-MP and OBF-wBU having a similar performance for small C_{m}, but with an increased accuracy for increasing C_{m} provided by the wBU method. This does not imply a degraded performance in the higher part of the spectrum, where OBF models give an error comparable to the error of AZ and PZ models, with OBF-MP providing an improvement over OBF-wBU and the other models, as can be seen in the plot at the bottom (this improvement is less visible than above, given the larger frequency range considered).

In general, differences in the performance of OBF-MP and OBF-wBU are a result of the inherent discretization of the OBF-MP algorithm; its limited resolution prevents the OBF-MP algorithm from perfectly matching the frequency and bandwidth of some magnitude peaks. As a consequence, these peaks are approximated using poles with a slightly shorter radius, which corresponds to a larger bandwidth and a shorter decay of the time response; which is the reason why OBF-wBU shows better performances at low frequencies and in the late response. On the other hand, OBF-MP has a higher resolution at higher frequencies and, as a result, a better performance in that frequency region and in the approximation of the early response.

These results can be visualized on the approximated frequency magnitude responses of the example in Fig. 3.10, showing the more accurate approximation of low-frequency resonances provided by OBF models with a Bark-scale resolution compared to AZ and PZ models, with the OBF model obtained with OBF-wBU performing better than the OBF-MP, for the reason explained

above. However, a large error is introduced by OBF-wBU at high frequencies, while OBF-MP is able to better approximate the envelope of the magnitude response. Looking at the selected pole sets for the different models, some differences can be observed: for PZ models, the poles are evenly distributed in the entire Nyquist interval, while for OBF models, the poles are mostly concentrated in the low frequencies (and closer to the unit circle), with a larger concentration in the very low frequencies for OBF-wBU models. It should be noted that, while the wBU method allows to control the frequency resolution only by means of the warping parameter, the pole grid of the OBF-MP algorithm offers more flexibility in the selection of the candidate poles and the possibility of incorporating prior knowledge about the characteristics of the room.

As discussed in Section 3.5, another important aspect to consider when comparing different parametric models is their filter complexity C_f. While the filter complexity C_f for AZ and PZ models equals the model complexity C_m, OBF models require extra computations (cf. Table 3.1). As an illustrative example, Fig. 3.11 shows the error of the different models as a function of the filter complexity C_f, and should be compared to the top-left plot of Fig. 3.9. The corresponding model complexity C_m for OBF models is reported on the axis underneath. It can be seen that using OBF models gives a smaller average NMSE compared to AZ and PZ models also in terms of filter complexity. Given the equivalence in terms of filter response between OBF models and PF models, the number of multiplications for OBF models can be reduced by using a PF implementation (in which case C_f equals C_m).

In order to perform the same kind of analysis in the modal region, similar simulations were run on the SUBRIR database [156, 274]. The SUBRIR database is a collection of RIRs measured at low frequency using a subwoofer as a source. Here, a subset of $R = 24$ RIRs measured with a Genelec 7050B subwoofer (with frequency range 25-120 Hz) and a B&K 4133 (1/2") microphone was used. The RIRs were downsampled to $f_\mathrm{s} = 800$ Hz and truncated to 1.5 s (corresponding to the maximum reverberation time, as reported in Table 3.2 and in [274]).

The OBF-MP grid used is this case has 600 angles uniformly placed from 0 to π (the Bark scale below 500 Hz has uniform resolution) and 10 radii logarithmically distributed from 0.75 to 0.999, while the BU method is applied without prewarping. The top plot of Fig. 3.12 shows the error of the different models as a function of the filter complexity C_f, similarly to Fig. 3.11. As in the previous examples, although OBF-MP models and OBF-BU models perform similarly, the BU method leads to numerical conditioning problems and to unstable solutions for values of the model complexity as low as $C_\mathrm{m} = 240$. PZ models were not considered here, as the STMCB algorithm provided unstable solutions almost in every situation. In this case, the improvement obtained with

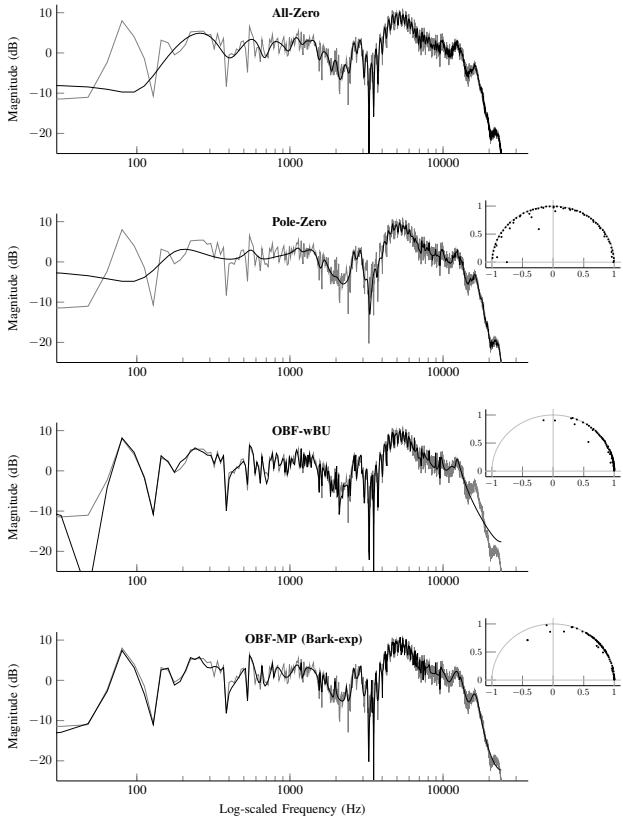

Figure 3.10: Approximated magnitude responses for (from top to bottom) an AZ model, a PZ model, OBF models with the wBU method and with the proposed method using a 5×1200 Bark-exp pole grid, together with the corresponding selected pole set ($C_{\mathrm{m}} = 300$). The target response (from MARDY) is shown in gray.

OBF models with respect to AZ models is more accentuated than in the previous examples. The reason for this is that in the modal region the number of room resonances is low and the models based on OBFs provide a more meaningful approximation of a RTF than AZ models, as discussed in Section 3.3.

The comparison of the performance of OBF-MP and OBF-BU in the frequency region of the loudspeaker, instead, presents significant differences (bottom plot). Our interpretation is that the ill-conditioning problems of the BU method are worsened by the fact that the spectrum above 130 Hz contains only noise. The

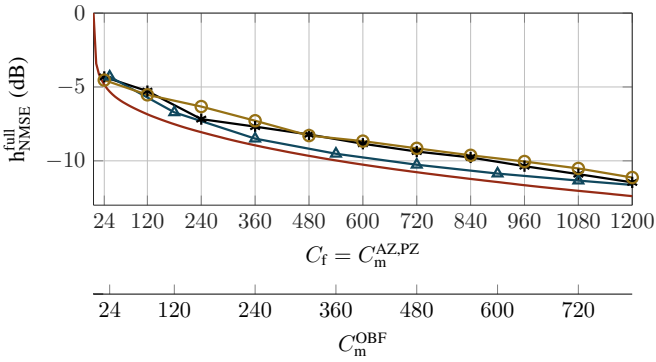

Figure 3.11: The average time-domain NMSE in (3.22) for the entire response for different values of filter complexity C_{f}. AZ (∗), PZ (○), OBF-wBU (△) and OBF-MP (—) models. The filter complexity corresponds to C_{m} for AZ and PZ models, while the corresponding values of C_{m} for OBF models are shown in the additional axis.

result is that also poles above that frequency are estimated. The poles selected by the well-conditioned OBF-MP algorithm, even though the poles in the grid are placed from 0 to π, are instead well concentrated within the range of the loudspeaker response, as shown in Fig. 3.13.

3.7 Conclusion and future work

The use of OBF models for obtaining a compact and accurate approximation of a target RIR has been motivated by the desirable properties derived from orthogonality, such as an improved numerical conditioning in the estimation of the numerator parameters of the transfer function for a fixed denominator. However, also OBF models are nonlinear in the parameters, so that the estimation of the poles is still a nonlinear problem. The state-of-the-art technique, the BU method, based on an FIR to IIR conversion which exploits the orthogonality property of OBF models, has some restrictions.

In this chapter, the novel algorithm, termed OBF-MP, has been studied and compared to the BU method in terms of modeling performance. Simulation results for RIRs measured in different rooms showed that OBF models are able to achieve a reduction in the approximation error compared to conventional parametric models for the same model and filter complexity, provided that the estimation of the pole parameters is accurate. Although the two algorithms

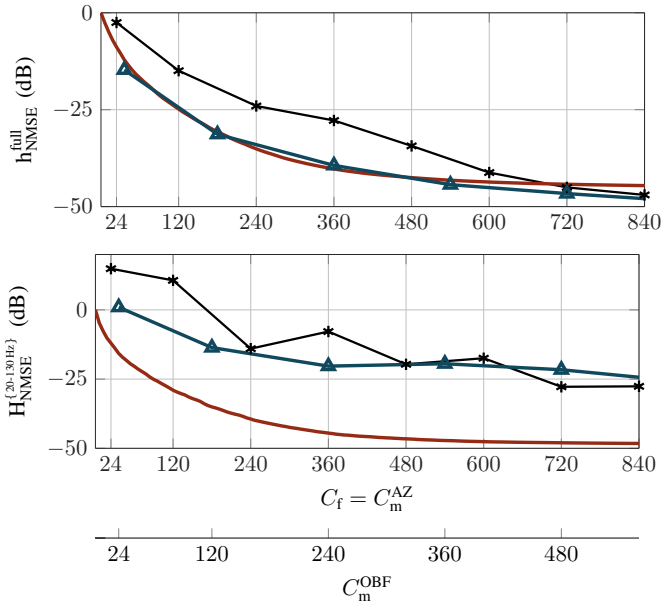

Figure 3.12: SUBRIR database. (top) The average time-domain NMSE in (3.22) for the entire response w.r.t. the filter complexity C_f. (bottom) The average frequency-domain NMSE in (3.23) between 20 Hz and 130 Hz w.r.t. the filter complexity C_f. AZ (∗), OBF-BU (Δ) and OBF-MP (—) models.

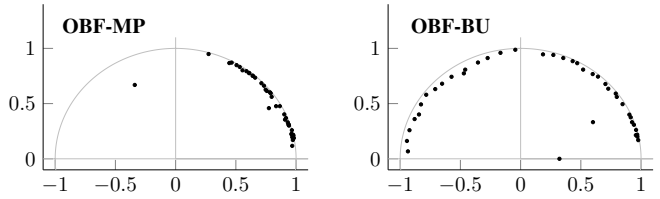

Figure 3.13: SUBRIR database. The set of 40 complex-conjugate pole-pairs ($C_m = 160$) obtained with OBF-MP (left) and the BU method (right) in the approximation of one RIR ($f_s = 800$ Hz).

considered for the estimation of the poles in an OBF model seem to have similar modeling capabilities, they present many differences. While the BU method suffers from numerical conditioning problems and instability, the OBF-MP algorithm always delivers stable and well-conditioned OBF model estimates. Indeed, the OBF-MP algorithm bypasses the nonlinear problem of estimating

the poles of an OBF model by defining a set of candidate stable poles and by selecting a complex-conjugate pole pair at each iteration based on the correlation between the target RIR and the basis functions built from the candidate poles. Orthogonality of the basis functions assures that this operation is numerically well-conditioned.

Moreover, while the BU-method requires the number of poles to be determined before estimation, the OBF-MP algorithm is scalable, in the sense that a new pair of complex-conjugate poles can be estimated independently of the poles estimated at previous iterations. Scalability turned out to be related to the analogy between OBF models and the definition of the RIR as an infinite summation of exponentially decaying sinuosoids independent from each other. The OBF-MP algorithm follows this interpretation by creating an approximation of the target RIR by adding a pair of OBFs and reducing the approximation error at each iteration. Differences in the performance between OBF-MP and OBF-BU in different time and frequency regions are a consequence of the approach to the pole estimation problem. While the BU method does not make any assumption on the position of the poles and controls the frequency resolution only by prewarping of the target RIR, the grid of candidate poles of the OBF-MP algorithm is an important design choice that adds a layer of flexibility. Any desired frequency resolution could be obtained, motivated by prior knowledge about the acoustics of the room or by application requirements. In this chapter, the Bark-exp grid was introduced to provide an accurate approximation at low frequencies, following the physical interpretation described above. However, the Bark-exp grid provides low resolution at high frequency, so that for larger model complexities, i.e. after the dominant modes and the spectral envelope have been approximated, pole grids with higher resolution at high frequencies can become more efficient. A possibility to overcome this issue could be to refine the estimation of the poles at each iteration using numerical optimization methods. This possibility and the inclusion of prior knowledge about the system in the estimation problem is left for future work.

The computational complexity of the OBF-MP algorithm is determined by the length of the target RIR sequence, the number of poles in the grid and the number of model parameters, and it is comparable with the algorithmic complexity of the BU method. Different approaches have been presented for reducing the complexity of the BU method and overcome its limitations, such as subband modeling [150], polyphase design and successive segmentation in the time domain [30]. It is believed that such extensions could be applied to the OBF-MP algorithm as well. Another interesting aspect is the possibility of exploiting the concept of common acoustical poles, as considered e.g. in [109] and [103]. The OBF-MP algorithm was modified in [156] in order to estimate a common set of poles from measurements taken for different source-receiver

positions inside a room. It has been shown that a significant reduction in the number of parameters necessary to model the RTF for different source-receiver positions can be achieved. A block-based version of the OBF-MP algorithm has been proposed in [264] and applied in [154] to the estimation of the poles of an adaptive IIR filter based on OBFs from input-output data of a SIMO room acoustic system. Results show that poles can be accurately estimated from white input-output data as well, offering a reduced approximation error compared to FIR filters, with the same convergence rate and complexity of the adaptation algorithm for the linear coefficients, but an improved robustness to the variability of the RTF for different source-receiver positions. Further research will focus on understanding the relation between the estimated common poles and the acoustic characteristics of the room, and on estimating the poles from nonstationary input-output data.

Chapter 4

Common-poles modeling of room impulse responses

A physically-motivated parametric model for compact representation of room impulse responses based on orthonormal basis functions

Giacomo Vairetti, Enzo De Sena, Toon van Waterschoot, Marc Moonen, Michael Catrysse, Neofytos Kaplanis, and Søren Holdt Jensen

Published in *Proc. 10th Eur. Congr. Expo. Noise Control Eng. (EuroNoise 2015)*, Maastricht, The Netherlands, pp. 149–154, Jun. 2015.

G. Vairetti, E. De Sena, T. van Waterschoot, M. Moonen, M. Catrysse, N. Kaplanis, and S. H. Jensen, "A physically-motivated parametric model for compact representation of room impulse responses based on orthonormal basis functions", in *Proc. 10th Eur. Congr. Expo. Noise Control Eng. (EuroNoise 2015)*, Maastricht, The Netherlands, pp. 149–154, Jun. 2015.
Changes include layout, representation, and other editing aspects. Specifically, some portions of the paper have been omitted to avoid repetition of the content presented in previous chapters. The mathematical notation has been uniformed to that of the previous chapter.

The candidate's contributions as first author include: literature study, co-development of the presented algorithms, software implementation and computer simulations, co-design of the evaluation experiments, co-formulation of the conclusions, text redaction and editing.

Abstract

A room impulse response (RIR) shows a complex time-frequency structure, due to the presence of sound reflections and room resonances at low frequencies. Many acoustic signal enhancement applications, such as acoustic feedback cancellation, dereverberation and room equalization, require simple yet accurate models to represent a RIR. Parametric modeling of room acoustics attempts at approximating the room transfer function (RTF), for given positions of source and receiver inside a room, by means of rational functions in the z-domain that can be implemented through digital filters. However, conventional parametric models, such as all-zero and pole-zero models, have some limitations. In this chapter, fixed-pole infinite impulse response (IIR) filters based on orthonormal basis functions (OBFs) are used as an alternative, motivated by their analogy to the physical definition of the RIR as a Green's function of the acoustic wave equation. An accurate estimation of the model parameters allows arbitrary allocation of the spectral resolution, so that the room resonances can be described well and a compact representation of a target RIR can be achieved. The model parameters can be estimated by a scalable matching pursuit (MP) algorithm called OBF-MP, which selects the most prominent resonance at each iteration. A modified version of the algorithm, called OBF-GMP (group matching pursuit), is introduced for the estimation of a common set of poles from multiple RIRs measured at different positions inside a room. Simulation results using a database of RIRs measured using a subwoofer show that, in comparison to OBF-MP, the OBF-GMP significantly reduces the number of parameters necessary to represent the RIRs.

4.1 Introduction

A room impulse response (RIR) shows a complex time-frequency structure, due to the presence of room resonances at low frequencies and the intricate temporal structure of sound reflections. Parametric models are used in all those acoustic signal enhancement applications that require the RIR to be represented in a simple yet accurate way. Examples of these applications are acoustic feedback cancellation, dereverberation, and room equalization. In parametric modeling, a room transfer function (RTF), corresponding to a Green's function of the acoustic wave equation for specific positions of the loudspeaker and the microphone inside a room, is represented by means of a rational function in the z-domain and implemented through digital filters. This rational function can be written in terms of zeros and poles by computing the complex-valued roots of the numerator and denominator polynomials,

respectively. However, conventional parametric models, such as all-zero and pole-zero models, present some limitations. The all-zero (AZ) model [26] uses a finite impulse response (FIR) filter to approximate the sampled RIR, with the number of parameters corresponding to the sample index at which the RIR is truncated. A zero approximation error is obtained up to the truncation index, but a large number of parameters is generally required in order to capture the resonant characteristics of the room, especially when the reverberation time is high. Moreover, the parameter values are strongly dependent on the source and receiver positions. All-pole and pole-zero (PZ) models are used in an attempt to overcome these limitations. These models use pairs of complex-conjugate poles to represent resonances in the RTF. The infinite impulse response (IIR) nature of these models enables to reduce the number of parameters and potentially to obtain parameter values less sensitive to changes in the source and receiver positions. The common-acoustical-poles and zeros (CAPZ) model [103] exploits the fact that room resonances are independent of the position of the source and receiver, but are rather a characteristic of the room itself. As the name suggests, the RTFs measured at different positions in the room are parametrized by a common set of poles, while differences between these responses are described by different sets of zeros. In this way, a more compact representation of a group of RIRs is obtained.

An alternative to conventional parametric models is provided by a particular family of models based on orthonormal basis functions. Orthonormal basis function (OBF) models [27, 126, 112] define a fixed-pole IIR filter, which is an orthonormalized parallel connection of second-order resonators, whose impulse responses represent damped sinusoids. Then, the RIR approximation is built as a linear superposition of a finite number of exponentially decaying sinusoids, whose frequency and decay rate is determined by the position of the poles inside the unit circle. The analogy with the definition of the RTF is clear. Each term of the Green's function corresponds to a resonator whose impulse response is a sinusoid, oscillating at a particular resonance frequency and damped with a particular damping constant [1]. OBF models possess many other desirable properties, such as orthogonality and stability. These models are also very flexible, in the sense that poles can be distributed arbitrarily inside the unit circle of the z-plane, thus giving freedom in the allocation of the spectral resolution. However, since OBF models are nonlinear in the pole parameters, estimating the poles that provide a good approximation of a given RIR is a nonlinear problem. Nonlinear optimization techniques have been proposed for the optimization of the poles in different applications, such as acoustic echo cancellation [34], and loudspeaker and room modeling [29]. In [275, 113], the nonlinear problem was avoided by iteratively selecting poles from a discrete grid using a scalable matching pursuit algorithm, called OBF-MP.

This chapter introduces a modified version of the OBF-MP that aims at estimating a set of poles common to multiple RIRs measured at different positions in the same room. It is shown that the modified algorithm, termed here OBF-GMP, approximates the set of RIRs more accurately, for the same total number of parameters, compared to the case when the poles are estimated individually for each RIR or when the all-zero model is used. Simulations have been performed on the database of low-frequency RIRs, introduced in Chapter 2.

The chapter is structured as follows. In Section 4.2, the OBF-GMP algorithm is introduced. Simulation results are presented in Section 4.3, whereas Section 4.4 concludes the chapter and indicates possible directions for future work.

4.2 The OBF-GMP algorithm

The OBF-MP algorithm [154, 113], described in Section 3.4, is a matching pursuit algorithm which, at each iteration, selects the predictors, i.e. the pair of basis functions, that are mostly correlated with the target RIR. The candidate predictors are generated based on a grid $\boldsymbol{p}_g = \{p_1, \ldots, p_G\}$ of poles distributed inside the unit circle, each pole representing also its complex-conjugate. The distribution of the poles in the grid can be dictated by the desire of a higher spectral resolution in the frequency range of interest or by other considerations based on prior knowledge about the acoustics of the room. Two examples are given in Figure 4.1. The left example shows a pole grid with angles distributed logarithmically, which yields a higher resolution at low frequencies. The right example shows a pole grid with radii distributed logarithmically, which yields a higher resolution close to the unit circle. A third example was given in Figure 3.6.

At each iteration, the matrix $\boldsymbol{\Phi}_k(\boldsymbol{p}_g)$ is built with the basis functions computed for each pole in the grid $\boldsymbol{p}_g$ and orthogonalized to the basis functions added in previous iterations. The OBF-MP algorithm is scalable. In fact, since the resulting filter structure is orthogonal by construction, the linear coefficients do not have to be recomputed at each iteration. As a consequence, the model order does not have to be determined beforehand and more poles can be added just by running extra iterations. At each iteration, the approximation error is reduced and the algorithm can be stopped when the desired degree of accuracy is obtained. Orthogonality also implies that the linear coefficients correspond to the correlation of the basis functions with the RIR. It follows that no matrix inversion operation is involved in the algorithm, avoiding any problem of numerical ill-conditioning.

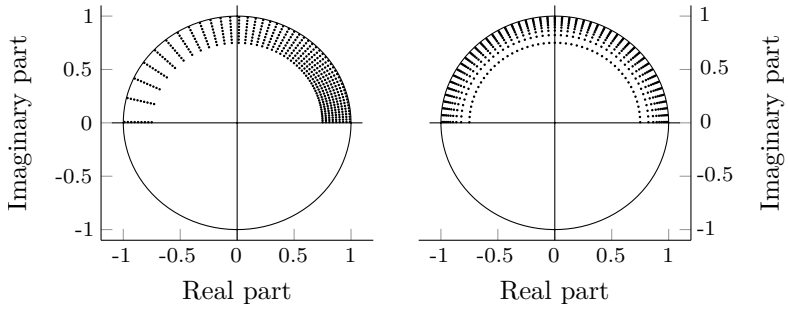

Figure 4.1: Pole grids using 500 poles, with 50 values for the angle $[1, f_s/2-1]$ Hz and 10 values for the radius [0.75,0.99]. (left) Logarithmic angles. (right) Logarithmic radii.

Here, the OBF-MP algorithm is modified in order to estimate a set of poles which is common to a set of R RIRs measured in the same room. The modified algorithm, called OBF-GMP (Group Matching Pursuit), is intended to reduce the number of parameters necessary to represent the RIRs by identifying the resonant characteristics of the room, common to all RIRs. The OBF-GMP algorithm, listed below, works as follows. First, a grid $\boldsymbol{p}_g$ of G candidate poles has to be defined. Then, the R target RIRs $\boldsymbol{h}^r = \{h_r(n)\}_{n=1}^N$ are stacked in a matrix $\boldsymbol{\Upsilon} = [\boldsymbol{h}^1, \ldots, \boldsymbol{h}^R]$. At each iteration k, the algorithm selects the pair of predictors in $\boldsymbol{\Phi}_k$ having maximum correlation, on average, with the target RIRs in $\boldsymbol{\Upsilon}$. As for the OBF-MP algorithm, the correlation α_i^r with each RIR vector $\boldsymbol{h}^r$ is defined as the projection of $\boldsymbol{h}^r$ on the plane defined by the predictors $\boldsymbol{\varphi}_i'$ and $\boldsymbol{\varphi}_i''$ of a pair of complex-conjugate poles $\{p_i, p_i^*\}$ (see Figure 4.2), and is given by

$$\alpha_i^r = \sqrt{(\alpha_i^{r\prime})^2 + (\alpha_i^{r\prime\prime})^2} = \sqrt{\left(\boldsymbol{\varphi}_i'^T \boldsymbol{h}^r\right)^2 + \left(\boldsymbol{\varphi}_i''^T \boldsymbol{h}^r\right)^2}, \tag{4.1}$$

where the correlation coefficients $\alpha_i^{r\prime}$ and $\alpha_i^{r\prime\prime}$ are obtained from the matrix product $\boldsymbol{\Lambda}_k = \boldsymbol{\Phi}_k^T \boldsymbol{\Upsilon}$, with $\alpha_i^{r\prime}$ corresponding to the element of $\boldsymbol{\Lambda}_k$ at column r and row $2i-1$, and $\alpha_i^{r\prime\prime}$ to the element at column r and row $2i$.

For each pair of complex-conjugate poles, the correlations with all the R target RIRs are then summed together, and the pole pair $\{p_s, p_s^*\}$ selected is the one with index

$$s = \arg\max_i \sum_{r=1}^{R} |\alpha_i^r|. \tag{4.2}$$

The pole $p_s \in \boldsymbol{p}_g$ is included in the set of active poles $\boldsymbol{p}_k^{\mathcal{A}}$ and the corresponding pair of predictors $\{\boldsymbol{\varphi}_s' \, \boldsymbol{\varphi}_s''\}$ added to the basis $\boldsymbol{\Phi}_k$ built from $\boldsymbol{p}_k^{\mathcal{A}}$. Each k^{th} predicted component $\hat{\boldsymbol{x}}_k^r$ is obtained from the last added pair of predictors

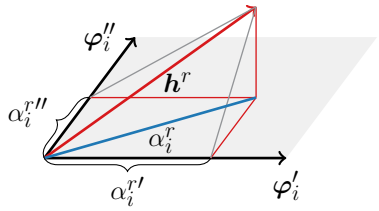

Figure 4.2: Graphical interpretation of the correlation between the r^{th} target RIR $\boldsymbol{h}^r$ and the predictors of a pair of complex-conjugate poles.

Algorithm 2 OBF-GMP algorithm

$\boldsymbol{\Upsilon} = [\boldsymbol{h}^1, \ldots, \boldsymbol{h}^R]$, $\hat{\boldsymbol{\Upsilon}}_0 = \boldsymbol{0}$ ▷ Define matrices of target and approx. RIRs

$\boldsymbol{p}_0^{\mathcal{A}} = \emptyset$, $\boldsymbol{p}_g = \{p_1, \ldots, p_G\}$ ▷ Initialize the set of active poles and the pole grid

$\boldsymbol{\Phi}_0^{\mathcal{A}} = \emptyset$, $k = 1$ ▷ Initialize basis, k: iteration counter

while $2k \leq M$ **do** ▷ M: desired model order

  Build $\boldsymbol{\Phi}_k(\boldsymbol{p}_g)$ ▷ $\boldsymbol{\Phi}_k$: matrix of candidate predictors built from $\boldsymbol{p}_g$

  $\boldsymbol{\Lambda}_k = \boldsymbol{\Phi}_k^T \boldsymbol{\Upsilon}$ ▷ Compute matrix of correlation coefficients

  $\alpha_i^r = \sqrt{(\alpha_i^{r\prime})^2 + (\alpha_i^{r\prime\prime})^2}$ ▷ Compute correlations for each pair of predictors

  $s = \arg\max_i \sum_{r=1}^{R} |\alpha_i^r|$ ▷ Find average max. correlation

  $\boldsymbol{p}_k^{\mathcal{A}} = [\boldsymbol{p}_{k-1}^{\mathcal{A}}\ p_s]$ ▷ Add selected pole to active pole set

  $\boldsymbol{\Phi}_k^{\mathcal{A}} = [\boldsymbol{\Phi}_{k-1}^{\mathcal{A}}\ \ \boldsymbol{\varphi}_s'\ \ \boldsymbol{\varphi}_s'']$ ▷ Update current basis

  $\hat{\boldsymbol{X}}_k = [\hat{\boldsymbol{x}}_k^1, \ldots, \hat{\boldsymbol{x}}_k^R]$, $\hat{\boldsymbol{x}}_k^r = [\boldsymbol{\varphi}_s'\, \boldsymbol{\varphi}_s''][\alpha_s^{r\prime}\, \alpha_s^{r\prime\prime}]^T$ ▷ Update predicted components

  $\hat{\boldsymbol{\Upsilon}}_k = \hat{\boldsymbol{\Upsilon}}_{k-1} + \hat{\boldsymbol{X}}_k$ ▷ Update matrix of RIR approx. vectors

  $\boldsymbol{E}_k = \boldsymbol{\Upsilon} - \hat{\boldsymbol{\Upsilon}}_k$ ▷ Compute current approx. error matrix

  $k = k + 1$

end while

using the corresponding correlation coefficients as regression coefficient by $\hat{\boldsymbol{x}}_k^r = [\boldsymbol{\varphi}_s'\, \boldsymbol{\varphi}_s''][\alpha_s^{r\prime}\, \alpha_s^{r\prime\prime}]^T$, and stored in the matrix $\hat{\boldsymbol{X}}_k = [\hat{\boldsymbol{x}}_k^1, \ldots, \hat{\boldsymbol{x}}_k^R]$. Each current RIR approximation vector $\hat{\boldsymbol{h}}_k^r$ is then obtained by adding the corresponding predicted component $\hat{\boldsymbol{x}}_k^r$ to the previous RIR approximation vector $\hat{\boldsymbol{h}}_{k-1}^r$, or in matrix form $\hat{\boldsymbol{\Upsilon}}_k = \hat{\boldsymbol{\Upsilon}}_{k-1} + \hat{\boldsymbol{X}}_k$, with $\hat{\boldsymbol{\Upsilon}}_k = [\hat{\boldsymbol{h}}_k^1, \ldots, \hat{\boldsymbol{h}}_k^R]$. The algorithm terminates when the desired number M of functions in the basis is reached or when the approximation error falls below a desired value.

4.3 Simulation results

The simulation results presented here aim at comparing the OBF-GMP algorithm with the OBF-MP algorithm and the all-zero modeling. The obtained models are compared in terms of their ability to approximate a set of R RIRs for a given number of parameters. For the all-zero modeling, the number of parameters is R times the number of taps used in the FIR filter for each RIR. For the OBF models with only complex-conjugate poles, the number of parameters is the number of complex-conjugate poles P plus the number of linear coefficients P, which sum up to $2P$ coefficients (see Section 3.5). When estimating $P/2$ pole pairs individually for each RIR with the OBF-MP algorithm, the total number of parameters is then $C = 2PR$. In case $P/2$ pole pairs are estimated jointly for all RIRs with the OBF-GMP algorithms, only one common set of P poles is necessary, and the total number of parameters becomes $C = P+PR = P(R+1)$.

The different models were tested on $R = 23$ RIRs taken from the SUBRIR database [274]. Each RIR was downsampled to $f_s = 800\,\text{Hz}$ and truncated to $N = 1600$ samples from the first strong peak, selected as its starting point. The OBF-GMP pole grid used $G = 1000$ poles with 20 different radii distributed logarithmically from 0.75 to 0.995 and with 50 different angles placed linearly in the range $[1, f_s/2 - 1]\,\text{Hz}$ (right plot of Figure 4.1). The error measure used to compare the performance of different models with the same number of parameters is the normalized mean square error (NMSE) in the time domain, averaged over all RIRs, given by

$$h_{\text{NMSE}}(dB) = 10\log_{10}\frac{1}{R}\sum_{r=1}^{R}\frac{\|\boldsymbol{h}^r - \hat{\boldsymbol{h}}^r\|_2^2}{\|\boldsymbol{h}^r\|_2^2}. \tag{4.3}$$

Figure 4.3 shows the average NMSE produced by the OBF models using the two algorithms and by the all-zero modeling, for the same total number C of model parameters divided by the number of RIRs. It can be seen that the OBF models provide a better approximation compared to the all-zero model. Moreover, there is a significant reduction in the approximation error when the OBF-GMP algorithm is used instead of the OBF-MP algorithm, mainly resulting from the use of a larger number of poles (almost double). The fact that this improvement is less noticeable when the number of parameters is small can be explained by observing that the OBF-MP algorithm tends to select poles closer to the unit circle, which approximate better the main strong resonances of the target magnitude response with a small number of poles.

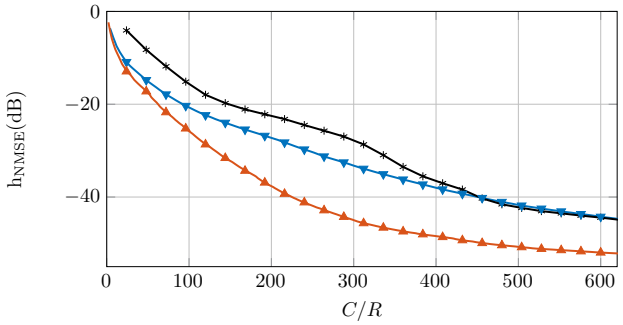

Figure 4.3: The average NMSE w.r.t. the total number of model parameters divided by the number of RIRs. AZ models (∗) and OBF models with poles estimated using the OBF-GMP algorithm (▲), and the OBF-MP algorithm (▼).

4.4 Conclusion and future work

In this chapter, the OBF-MP algorithm for the estimation of the poles was modified in order to approximate multiple RIRs jointly. The idea is also exploited in the CAPZ model, with the main difference that the CAPZ model requires the model order to be determined in advance. Simulation results on a set of low-frequency RIRs measured in a rectangular room show that the OBF-GMP algorithm allows to reduce the number of parameters, obtaining a more compact representation of multiple RIRs.

Future research will further investigate the topic in the pursuit of a better understanding of the behavior of the OBF-GMP algorithm w.r.t. different configurations of the pole grid, which could be informed by prior knowledge about the characteristics of the room, different numbers of RIRs and different pole selection criteria, also including a comparison with the CAPZ modeling.

Part III

Identification

Chapter 5

Room acoustic system identification using OBF adaptive filters

Orthonormal basis functions adaptive filters for room acoustic signal enhancement

Giacomo Vairetti, Enzo De Sena, Søren Holdt Jensen, Marc Moonen, and Toon van Waterschoot

Submitted for publication to *Signal Processing*, Elsevier, Apr. 2018.

The candidate's contributions as first author include: literature study, co-development of the presented algorithms, software implementation and computer simulations, co-design of the evaluation experiments, co-formulation of the conclusions, text redaction and editing.

Abstract

Room acoustic signal enhancement applications strongly rely on algorithms for identifying the acoustic response of the room (or its inverse) at one or multiple locations of the sound sources and microphones. As shown in recent studies, stable infinite impulse response (IIR) filters based on orthonormal basis functions (OBFs) (hereafter called *OBF filters*) are well-suited for approximating room responses, especially at low frequencies where the approximation is well motivated also from a physical point of view. Moreover, when used in an adaptive algorithm, the orthogonality property of OBF filters enables an analysis of their tracking performance, with theoretical results showing faster convergence compared to other fixed-poles IIR filters for a wide range of input spectra. In this chapter, the theory of OBF adaptive filters is reviewed in relation to the identification of room acoustic systems. An iterative algorithm is proposed for the identification of room acoustic systems at low frequencies, able to accurately estimate the characteristic poles of the room transfer functions from white noise and speech signals. It is shown that the actual advantage compared to the use of finite impulse response filters mostly depends on the characteristics of the room itself, e.g. its dimensions and reverberation time. Finally, the applicability of OBF adaptive filters to acoustic signal enhancement algorithms is discussed by means of examples in the context of acoustic echo cancellation and room equalization.

5.1 Introduction

Applications of room acoustic signal enhancement (RASE), such as acoustic echo cancellation (AEC) [22], acoustic feedback cancellation (AFC) [24], or room response equalization (RRE) [9], normally rely on the modeling and identification of the room impulse response (RIR) or its inverse at one or multiple locations of the sound sources and microphones. The most commonly used RIR model is the all-zero model, in which the model parameter values are identified as the coefficients of a finite impulse response (FIR) adaptive filter [26]. The reasons for the popularity of FIR adaptive filters are their simplicity, a well-consolidated theory, and a large variety of algorithms available for each specific application. Far less popular is the use of infinite impulse response (IIR) filters, which implies modeling a room transfer function (RTF) as a pole-zero model [68]. In theory, IIR filters would allow to reduce the number of coefficients required to adequately model the RTFs, thus providing more efficient algorithms. However, their adoption in RASE applications has been discouraged by their higher complexity, the added difficulty in the adaptive identification

of the pole parameters, potential instability issues and convergence to local minima [163], and possibly by the wide-spread belief that no relevant advantage can be obtained compared to FIR filters [239, 240].

Recent years have witnessed an increasing interest into pole-zero models based on orthonormal basis functions (OBFs) [27, 126, 112] for modeling room acoustics [113, 156, 29, 32]. The idea consists in modeling a RTF as a weighted combination of second-order resonators, whose frequencies and bandwidths are determined by the position of the pole parameters, while their amplitude is controlled by a set of linear coefficients (or weights). As for other fixed-poles IIR filters [165], advantages of using IIR filters based on OBFs (hereafter called *OBF filters*) are related to the increase in accuracy obtained by having the poles of the filter transfer function (TF) closer to the true poles of the system [27, 130, 167], while easily ensuring filter stability. In addition, the orthogonality property of OBF filters provides numerical well-conditioning and fast convergence of the filter adaptation [130, 169, 276], and is the key aspect in enabling an analysis of the error performance and of the dynamic behavior of adaptive algorithms, analogously to the FIR case.

Even though the theory is well established within the field of system identification, OBF filters have been adopted for room acoustic modeling only recently. Effective modeling algorithms have been suggested for the estimation of the poles, showing that a set of stable and accurate poles can be obtained, with the possibility of allocating frequency resolution unevenly in different regions of the spectrum [113, 29]. These algorithms have been modified for the estimation of a common set of poles from multiple RTFs [156, 32] and for modeling in subbands [32] and in time-domain partitions [30]. Although improvements in the approximation accuracy can be obtained over all-zero models on the entire audible spectrum [113], OBF models are particularly suited at low frequencies, where the RTFs are characterized by moderately overlapping and slowly decaying modes [113, 156].

Nonetheless, the use of OBF filters in RASE applications is still very limited. Some examples are found in the context of AEC [276], AFC [35] and RRE [28, 277], all relying on the availability of measured or pre-identified RIRs. The few cases in which the filter pole parameter values are directly estimated from input-output data have been found applied to the identification of acoustic echo systems. Methods have been proposed for the on-line estimation of both the poles and the linear coefficients [151, 34, 241], but limited to the case of an OBF filter with a single repeated pole (known as *Laguerre* and *2-parameter-Kautz* filters). The method in [241] is said to be applicable to OBF filters with non-repeated pairs of complex-conjugate poles (also known as *Kautz filters*) using approximated expressions for the gradients [171], but this possibility has not yet been verified on realistic AEC scenarios.

The purpose of this chapter is to discuss the applicability of OBF adaptive filters to RASE algorithms. In Section 5.2, a review of OBF models and OBF adaptive filters is provided. The focus is on the frequency domain analysis of the error performance and dynamic behavior of adaptive algorithms with respect to the characteristics of the input and noise signals. It is shown that adaptive algorithms developed for FIR filters are easily extended to the OBF filter case when poles are fixed, whereas the adaptation of the poles requires nonlinear recursive identification algorithms. An alternative version of the normalized least mean squares (NLMS) algorithm is introduced to deal with the adaptation of the linear coefficients when the filter order is small. The proposed modification, called here OBF-NLMS, normalizes the response of each OBF second-order section individually, in an analogy with transform-domain (TD) adaptive filters [278, 279, 280, 281, 282].

In Section 5.3, an identification algorithm is introduced, inspired by the scalable group matching pursuit (GMP) modeling algorithm, named OBF-GMP and described in [113, 156], which ameliorates the previously proposed block-based (BB)-OBF-GMP identification algorithm in [154]. The algorithm, named here stage-based (SB)-OBF-GMP, performs an iterative grid search that avoids the nonlinear problem and gives the possibility of estimating a common set of poles from multiple microphone signals to be then fixed and used in RASE applications. The newly proposed identification algorithm uses the NLMS and OBF-NLMS adaptation algorithms in order to overcome the problems of the BB algorithm in dealing with non-stationary and non-white input signals.

Section 5.4 provides identification results at low frequencies performed on measured and simulated RIRs. The aim of this section is to analyze the relation between the characteristics of the room, such as its dimensions and reverberation time, and the advantages over FIR filters that can be expected by using OBF adaptive filters with a common set of poles. In Section 5.5 examples are given in order to illustrate possible uses of OBF adaptive filters and the proposed identification algorithm in the context of AEC and RRE. Finally, Section 5.6 discusses the applicability of OBF adaptive filters to RASE applications and Section 5.7 concludes the chapter.

5.2 OBF adaptive filters

A RTF can be expressed as an infinite summation of room resonances, each with a certain central frequency, bandwidth and amplitude, whose density increases with frequency [1]. It can be approximated by means of an OBF filter as a finite summation of second-order all-pole filters (or resonators), having TF as

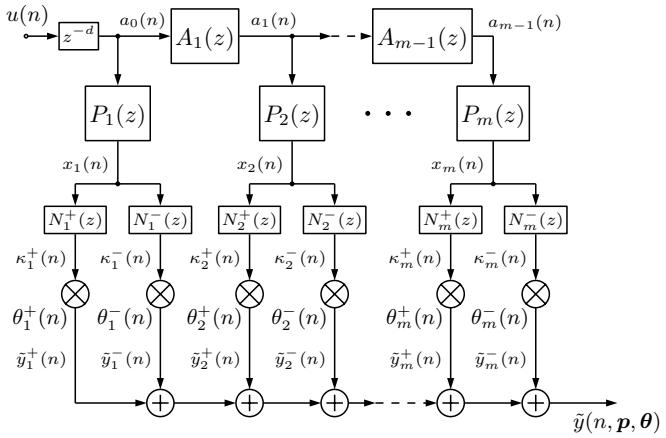

Figure 5.1: The OBF adaptive filter for m pole pairs.

in (5.1), each weighted by a pair of linear amplitude coefficients.

$$P_i(z) = \frac{1}{D_i(z)} = \frac{1}{(1 - p_i z^{-1})(1 - p_i^* z^{-1})} \ , \tag{5.1}$$

$$A_i(z) = \frac{\bar{D}_i(z)}{D_i(z)} = \frac{(z^{-1} - p_i)(z^{-1} - p_i^*)}{(1 - p_i z^{-1})(1 - p_i^* z^{-1})} \ , \tag{5.2}$$

$$N_i^{\pm}(z) = |1 \pm p_i| \sqrt{\frac{1 - |p_i|^2}{2}} (z^{-1} \mp 1) \ . \tag{5.3}$$

Each resonator is defined by a pair of complex-conjugate poles $\boldsymbol{p}_i = [p_i, p_i^*] = \rho_i e^{\pm j\vartheta_i}$, with radius $\rho_i = e^{-\zeta_i / f_s}$ ($\rho_i < 1$ for stability) related to the bandwidth ζ_i, and angle $\vartheta_i = \omega_i / f_s$ related to the resonance frequency ω_i (f_s being the sampling frequency and * indicating complex conjugation). The responses of the resonators are orthogonalized with respect to each other by means of a series of all-pass filters built from the same pole pairs $\boldsymbol{p}_i$ with TF as in (5.2) (the influence of the poles p_i and p_i^* is canceled by the nonminimum-phase zeros in $1/p_i$ and $1/p_i^*$ of the all-pass TF [112]), whereas a pair of first-order all-zero filters with TF as in (5.3) produces mutually orthonormal responses. Given that a RTF presents multiple resonances with different frequencies and a band-pass characteristic due to the band-pass response of the loudspeaker (and the anti-aliasing filter), only OBF filters with multiple complex poles are considered here. For more details about the construction of OBF models and their relation to room acoustics and other parametric models, the reader is referred to [113].

The OBF adaptive filter structure is depicted in Figure 5.1. Each i^{th} resonator section has a pair of orthonormal TFs, or basis functions, given by $\Psi_i^{\pm}(z, \tilde{\boldsymbol{p}}_i)$, which are weighted by a pair of time-varying amplitude coefficients $\boldsymbol{\theta}_i(n) = [\theta_i^+(n), \theta_i^-(n)]$ (with $i = 1, \ldots, m$ and m the number of complex-conjugate pole pairs), giving the overall filter TF at discrete time-instant $n = {}^{t}/_{f_s}$

$$G(z, \boldsymbol{p}, \boldsymbol{\theta}(n)) = z^{-d} \sum_{i=1}^{m} \left[\theta_i^+(n)\Psi_i^+(z, \tilde{\boldsymbol{p}}_i) + \theta_i^-(n)\Psi_i^-(z, \tilde{\boldsymbol{p}}_i)\right]$$

$$\text{with } \Psi_i^{\pm}(z, \tilde{\boldsymbol{p}}_i) = N_i^{\pm}(z) P_i(z) \prod_{\iota=1}^{i-1} A_\iota(z), \tag{5.4}$$

$\boldsymbol{p} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_m]^T$ a set of m pairs of complex-conjugate poles, $\boldsymbol{\theta}(n) = [\boldsymbol{\theta}_1(n), \ldots, \boldsymbol{\theta}_m(n)]^T$, both with dimensions $M \times 1$ ($M = 2m$), and $\tilde{\boldsymbol{p}}_i = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_i]^T$. Figure 5.2 shows the power spectrum of the basis functions $\Psi_i^{\pm}(z, \tilde{\boldsymbol{p}}_i)$ generated from a set of $m = 5$ pole pairs with different radii and angles.

The intermediate signals $\boldsymbol{\kappa}_i(n, \tilde{\boldsymbol{p}}_i) = [\kappa_i^+(n, \tilde{\boldsymbol{p}}_i), \kappa_i^-(n, \tilde{\boldsymbol{p}}_i)]$ are filtered versions of the input signal $u(n)$ (i.e. $\kappa_i^{\pm}(n, \tilde{\boldsymbol{p}}_i) = \Psi_i^{\pm}(q, \tilde{\boldsymbol{p}}_i)u(n)$, with q^{-1} the backward time-shift operator for which $q^{-1}u(n) = u(n-1)$), and the output signal of the filter is a weighted summation of the intermediate signals given as

$$\begin{aligned}
\tilde{y}(n, \boldsymbol{p}, \boldsymbol{\theta}) &= \sum_{i=1}^{m} \left[\tilde{y}_i(n, \tilde{\boldsymbol{p}}_i, \boldsymbol{\theta}_i)\right] \\
&= \sum_{i=1}^{m} \left[\tilde{y}_i^+(n, \tilde{\boldsymbol{p}}_i, \theta_i^+) + \tilde{y}_i^-(n, \tilde{\boldsymbol{p}}_i, \theta_i^-)\right] \\
&= \sum_{i=1}^{m} \left[\kappa_i^+(n, \tilde{\boldsymbol{p}}_i)\theta_i^+(n) + \kappa_i^-(n, \tilde{\boldsymbol{p}}_i)\theta_i^-(n)\right] \ ,
\end{aligned} \tag{5.5}$$

or in vector form as

$$\begin{aligned}
\tilde{y}(n, \boldsymbol{p}, \boldsymbol{\theta}) &= \boldsymbol{\kappa}^T(n, \boldsymbol{p})\boldsymbol{\theta}(n) \\
\text{with} \quad \boldsymbol{\kappa}(n, \boldsymbol{p}) &= [\boldsymbol{\kappa}_1(n, \tilde{\boldsymbol{p}}_1), \ldots, \boldsymbol{\kappa}_m(n, \tilde{\boldsymbol{p}}_m)]^T.
\end{aligned} \tag{5.6}$$

Finally, a d-samples delay can be included to take the acoustic delay of the RIR into account (cfr. leftmost block in Figure 5.1).

The main reason for using an orthonormal model structure in place of other non-orthogonal fixed-pole ones is related to numerical considerations: even though

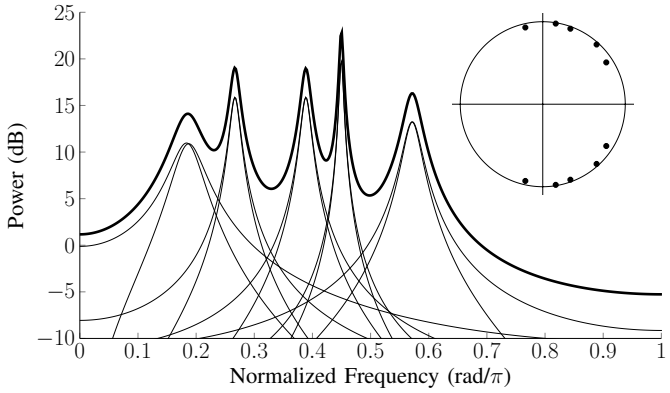

Figure 5.2: The power responses of 5 pairs of basis functions generated from the pole pairs depicted in the top-right corner. The resulting $\gamma_M(\omega, \boldsymbol{p})$ in (5.12) is also shown (thick line).

orthogonal and non-orthogonal models with the same fixed poles span the same approximation space, it is found that the estimation of a RTF with respect to the non-orthogonal structure can be very ill-conditioned. Orthogonal model structures, instead, provide a well-conditioned estimation problem for a wide range of input power spectral density (PSD) (not only white), so that their use is said to be the only practical way of fixing the poles in a filter structure [130]. For the same reason, OBF adaptive filters also show faster convergence than other fixed-poles adaptive filters.

To keep the discussion as simple as possible, a RTF $H(z)$ is assumed to be linear and time-invariant, so that a microphone signal can be defined as

$$y(n) = H(q)u(n) + v(n), \tag{5.7}$$

with $v(n)$ a zero-mean additive white noise (WN) signal with variance $S_v(\omega) = \mathbb{E}\{v^2(n)\} = \sigma_v^2$, $\forall\omega$, and $u(n)$ the loudspeaker input signal having spectral density

$$S_u(\omega) = \sum_{\tau=-\infty}^{\infty} R_u(\tau)e^{-j\omega\tau}, \tag{5.8}$$

with covariance function $R_u(\tau) = \mathbb{E}\{u(n)u(n-\tau)\}$ ($\mathbb{E}\{\cdot\}$ denotes the expected value). The input PSD plays an important role in the behavior of OBF adaptive filters, as it will be explained later on. Indeed, the power σ_ι^2 of each intermediate signal $\kappa_\iota(n)$ (corresponding to $\kappa_i^+(n)$ for ι odd and to $\kappa_i^-(n)$ for ι even, with $i = {}^{(\iota+\iota \bmod 2)}/_2$ and $\iota = 1, \ldots, M$) is determined by the product between the input PSD and the power of the corresponding OBF frequency response

$\Psi_\iota(e^{j\omega}, \tilde{\boldsymbol{p}}_i)$,

$$\mathbb{E}\{|\kappa_\iota(n)|^2\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_u(\omega)\left|\Psi_\iota(e^{j\omega}, \tilde{\boldsymbol{p}}_i)\right|^2 d\omega. \tag{5.9}$$

This means that an intermediate signal will have small power at a given frequency ω whenever either $S_u(\omega)$ or the power of the OBF frequency response is small. Also notice that, because of normalization,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \left|\Psi_\iota(e^{j\omega}, \tilde{\boldsymbol{p}}_i)\right|^2 d\omega = 1\,. \tag{5.10}$$

Finally, it is noticed here that the total power of the intermediate signals is given by the sum of the power of each intermediate signal as

$$\begin{aligned}\mathbb{E}\{\|\boldsymbol{\kappa}(n,\boldsymbol{p})\|^2\} &= \sum_{\iota=1}^{M}\frac{1}{2\pi}\int_{-\pi}^{\pi} S_u(\omega)\left|\Psi_\iota(e^{j\omega}, \tilde{\boldsymbol{p}}_i)\right|^2 d\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} S_u(\omega)\sum_{\iota=1}^{M}\left|\Psi_\iota(e^{j\omega}, \tilde{\boldsymbol{p}}_i)\right|^2 d\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} S_u(\omega)\gamma_M(\omega,\boldsymbol{p})d\omega, \end{aligned} \tag{5.11}$$

$$\text{with}\;\; \gamma_M(\omega,\boldsymbol{p}) = \sum_{\iota=1}^{M}\left|\Psi_\iota(e^{j\omega}, \tilde{\boldsymbol{p}}_i)\right|^2 \tag{5.12}$$

a frequency-dependent term that will be essential in the analysis of the main properties of OBF adaptive filters, summarized in the remainder of the present section. For more extensive explanations, including the case for time-varying systems, the reader is referred to [130, 167, 169, 27].

5.2.1 Estimation accuracy: bias and variance errors

An estimate of $H(z)$ obtained using an OBF filter as in (5.4) with poles fixed in $\boldsymbol{p}$ produces a prediction error given by

$$\varepsilon(n,\boldsymbol{p},\boldsymbol{\theta}) = y(n) - \tilde{y}(n,\boldsymbol{p},\boldsymbol{\theta}) = y(n) - \boldsymbol{\kappa}^T(n,\boldsymbol{p})\boldsymbol{\theta}(n). \tag{5.13}$$

If N samples of the input and output signals are available and the following quadratic cost function is adopted

$$V_N(\boldsymbol{\theta},\boldsymbol{p}) = \frac{1}{2N}\sum_{n=1}^{N}\varepsilon^2(n,\boldsymbol{\theta},\boldsymbol{p}), \tag{5.14}$$

a least squares (LS) estimate $\hat{\boldsymbol{\theta}}_N$ of the linear amplitude coefficients $\boldsymbol{\theta}$ can then be found by minimizing (5.14) with respect to $\boldsymbol{\theta}$. The estimation error produced is made of two terms, a *bias error* $E_\beta(\omega)$ and a *variance error* $E_\nu(\omega)$, and it can be written (in the frequency domain, where $\omega \equiv e^{j\omega}$) as

$$\begin{aligned} E(\omega, \boldsymbol{p}, \hat{\boldsymbol{\theta}}_N) &= H(\omega) - G(\omega, \boldsymbol{p}, \hat{\boldsymbol{\theta}}_N) \\ &= [H(\omega) - G(\omega, \boldsymbol{p}, \boldsymbol{\theta}_o)] + [G(\omega, \boldsymbol{p}, \boldsymbol{\theta}_o) - G(\omega, \boldsymbol{p}, \hat{\boldsymbol{\theta}}_N)] \\ &= E_\beta(\omega, \boldsymbol{p}, \boldsymbol{\theta}_o) + E_\nu(\omega, \boldsymbol{p}, \boldsymbol{\theta}_o, \hat{\boldsymbol{\theta}}_N), \end{aligned} \tag{5.15}$$

with $\boldsymbol{\theta}_o = \lim_{N\to\infty} \hat{\boldsymbol{\theta}}_N$ the Wiener solution ($\boldsymbol{\theta}_o = \boldsymbol{R}_{\boldsymbol{\kappa}}^{-1}\boldsymbol{r}$, where $\boldsymbol{R}_{\boldsymbol{\kappa}} = \mathbb{E}\{\boldsymbol{\kappa}(n, \boldsymbol{p})\boldsymbol{\kappa}^T(n, \boldsymbol{p})\}$ is the autocorrelation matrix of the intermediate signals and $\boldsymbol{r} = \mathbb{E}\{\boldsymbol{\kappa}(n, \boldsymbol{p})y(n)\}$ is the cross-correlation vector between the intermediate signals and the microphone signal, assuming $u(n)$ and $y(n)$ being jointly wide-sense stationary stochastic processes, both with zero mean). This means that the bias error depends on the model structure chosen and on its order (i.e. the dimension of $\boldsymbol{p}$ and $\boldsymbol{\theta}$), whereas the variance error depends on the actual estimation of the parameters, and is thus influenced by the characteristics of the input and noise signals. It can be seen from (5.15) that both terms also depend on the pole set $\boldsymbol{p}$, which means that not only the number of poles, but also their position, affects the estimation error.

It is quite intuitive to expect a decrease in the bias error when the poles are moved away from the origin of the z-plane closer to the true poles of the system. A result valid for a constant $S_u(\omega) = \sigma_u^2$, but extendable to other cases [130], formalizes this idea showing that the bias error tends to zero for $M \to \infty$ (more specifically, it decreases geometrically in the model order M) and that it is proportional to the distance between the K true poles $\xi_\varkappa$ of the system ($\varkappa = 1, \ldots, K$) and the M poles p_i of the model structure ($i = 1, \ldots, M$), according to

$$E_\beta(\omega, \boldsymbol{p}, \boldsymbol{\theta}_o) \leq \sum_{\varkappa=1}^{K} \left| \frac{\aleph_\varkappa}{e^{j\omega} - \xi_\varkappa} \right| \prod_{i=1}^{M} \left| \frac{\xi_\varkappa - p_i}{1 - p_i^* \xi_\varkappa} \right|, \tag{5.16}$$

with $\aleph_\varkappa$ the residues of the partial fraction expansion of $H(z)$ (see [130]).

The dependency of the asymptotic variance error on the poles is instead less intuitive. It turns out [167] that the influence of the pole location is quantified by the frequency-dependent term $\gamma_M(\omega, \boldsymbol{p})$ in (5.12) in such a way that the variance can be approximated as

$$\mathbb{E}\{|E_\nu(\omega, \boldsymbol{p}, \boldsymbol{\theta}_o, \hat{\boldsymbol{\theta}}_N)|^2\} \approx \frac{1}{N} \frac{\sigma_v^2}{S_u(\omega)} \gamma_M(\omega, \boldsymbol{p}), \tag{5.17}$$

which is a generalization of the FIR case, for which $\gamma_M(\omega, \boldsymbol{p}) = M$, $\forall\omega$ [168]. The implications are that increasing the number of poles in order to reduce the bias error in a certain frequency region would also increase the sensitivity to noise in that same region. As a consequence, a good estimate of $H(z)$ would have as few poles, as close as possible to the actual poles of the RTF.

5.2.2 Adaptation of the linear coefficients

In most RASE applications, an estimate of a RTF has to be obtained adaptively, as new samples of the source and microphone signals are available. The adaptation rule for the recursive estimation of the linear filter parameter vector $\boldsymbol{\theta}(n)$ is given by

$$\hat{\boldsymbol{\theta}}(n+1) = \hat{\boldsymbol{\theta}}(n) + \boldsymbol{g}(n, \boldsymbol{p})\varepsilon(n, \boldsymbol{p}, \hat{\boldsymbol{\theta}}), \tag{5.18}$$

with $\boldsymbol{g}(n, \boldsymbol{p})$ a gain vector. When considering the linear $\boldsymbol{\theta}$ parameters, OBF models are linear regression models, as can be seen in (5.5) (the regression vectors are independent from the previous estimates of the parameters). Thus, it is possible to apply standard adaptive algorithms developed for FIR filters [157], with the only difference that the regression vector for an OBF adaptive filter is represented by the vector of intermediate signals $\boldsymbol{\kappa}(n, \boldsymbol{p})$, instead of $\boldsymbol{u}(n) = [u(n), u(n-1), \ldots, u(n-M+1)]$ (the last M samples of the input signal $u(n)$). The increase in complexity is only given by the filtering of the input signal to compute $\boldsymbol{\kappa}(n, \boldsymbol{p})$, and not in the adaptation scheme itself.

Based on how the gain vector $\boldsymbol{g}(n, \boldsymbol{p})$ is computed, different algorithms are obtained: the least mean squares (LMS) algorithm is obtained for

$$\boldsymbol{g}(n, \boldsymbol{p}) = \mu\boldsymbol{\kappa}(n, \boldsymbol{p}), \tag{5.19}$$

with μ the step size. Normalization is usually necessary to avoid that large values in $\boldsymbol{\kappa}(n, \boldsymbol{p})$ would lead to large variations in $\hat{\boldsymbol{\theta}}(n+1)$. The vector of intermediate signals $\boldsymbol{\kappa}(n, \boldsymbol{p})$ can hence be normalized, leading to the NLMS gain vector

$$\boldsymbol{g}(n, \boldsymbol{p}) = \frac{\tilde{\mu}}{\delta + \|\boldsymbol{\kappa}(n, \boldsymbol{p})\|^2}\boldsymbol{\kappa}(n, \boldsymbol{p}), \tag{5.20}$$

where δ is a small regularization term to avoid instability or convergence problems resulting from poor excitation [185, 283]. Another common, yet more complex, adaptation algorithm is the recursive least squares (RLS) algorithm [165, 130], for which the gain vector is given by

$$\boldsymbol{g}(n, \boldsymbol{p}) = \boldsymbol{\Upsilon}_{\boldsymbol{\kappa}}(n)\boldsymbol{\kappa}(n, \boldsymbol{p}), \text{ with}$$

$$\boldsymbol{\Upsilon}_{\boldsymbol{\kappa}}(n+1) = \frac{1}{\lambda}\left\{\boldsymbol{\Upsilon}_{\boldsymbol{\kappa}}(n) - \frac{\boldsymbol{\Upsilon}_{\boldsymbol{\kappa}}(n)\boldsymbol{\kappa}(n, \boldsymbol{p})\boldsymbol{\kappa}^T(n, \boldsymbol{p})\boldsymbol{\Upsilon}_{\boldsymbol{\kappa}}(n)}{\lambda + \boldsymbol{\kappa}^T(n, \boldsymbol{p})\boldsymbol{\Upsilon}_{\boldsymbol{\kappa}}(n)\boldsymbol{\kappa}(n, \boldsymbol{p})}\right\}, \tag{5.21}$$

$\lambda = 1 - \mu$ a forgetting factor and $\boldsymbol{\Upsilon}_{\boldsymbol{\kappa}}(n)$ an estimate of the inverse of the autocorrelation matrix $\boldsymbol{R}_{\boldsymbol{\kappa}} = \mathbb{E}\{\boldsymbol{\kappa}(n,\boldsymbol{p})\boldsymbol{\kappa}^T(n,\boldsymbol{p})\}$ of the intermediate signals. A trade-off between convergence and complexity is obtained with the affine projection algorithm (APA) [284], for which the update rule in (5.18) becomes

$$\hat{\boldsymbol{\theta}}(n+1) = \hat{\boldsymbol{\theta}}(n) + \mu \boldsymbol{K}_Q(n)\big(\delta \boldsymbol{I}_Q + \boldsymbol{K}_Q^T(n)\boldsymbol{K}_Q(n)\big)^{-1}\boldsymbol{\varepsilon}_Q(n) \tag{5.22}$$

where $\boldsymbol{I}_Q$ is the $Q \times Q$ identity matrix used for regularization with δ a small constant, $\boldsymbol{K}_Q(n) = [\boldsymbol{\kappa}(n,\boldsymbol{p}), \boldsymbol{\kappa}(n-1,\boldsymbol{p}), \ldots, \boldsymbol{\kappa}(n-Q+1,\boldsymbol{p})]$ of size $M \times Q$, and $\boldsymbol{\varepsilon}_Q(n) = [\varepsilon(n,\boldsymbol{p},\boldsymbol{\theta}), \varepsilon(n-1,\boldsymbol{p},\boldsymbol{\theta}), \ldots, \varepsilon(n-Q+1,\boldsymbol{p},\boldsymbol{\theta})]$, with Q the so-called projection order. Notice that for $Q = 1$, the APA corresponds to the NLMS algorithm. Most algorithms developed for FIR adaptive filters can be derived easily for OBF filters as well. For instance, variable step size (VSS) algorithms [285, 286, 287] or regularized algorithms [185] can be used for highly non-stationary signals, or the Kalman filter [130] for time-varying systems.

Dynamic behavior: transient and steady-state errors

As mentioned earlier, orthogonality offers the possibility of studying the error behavior and transient performance of OBF filters in adaptive algorithms and the relation to step size, noise power and input PSD. The following results have been derived in [130] for $M \to \infty$, but proved (empirically) to be valid also for small model orders.

The steady-state error (after convergence, i.e. for $n \to \infty$) is similar to the expression of the variance in (5.17), with the introduction of the step size μ,

$$\mathbb{E}\{|E(\omega,\boldsymbol{p},\hat{\boldsymbol{\theta}}(\infty))|^2\} \approx \frac{\mu\sigma_v^2}{[S_u(\omega)]^r}\gamma_M(\omega,\boldsymbol{p}), \tag{5.23}$$

where $r = 0$ for LMS and $r = 1$ for RLS (with $\mu = 1 - \lambda$). This expression shows that the LMS algorithm depends on the choice of the step size and on the noise variance, but it is invariant to the input PSD $S_u(\omega)$, whereas for RLS the steady-state estimation error is inversely proportional to $S_u(\omega)$; for NLMS, expression (5.23) with $r = 0$ is still valid, if the normalization in (5.20) is considered to be included in the step size μ, which depends on $S_u(\omega)$ according to (5.11). As for the variance, also the steady-state error in (5.23) depends on $\gamma_M(\omega,\boldsymbol{p})$, i.e. it increases for a larger number of poles and for larger poles densities, and it is equal to the FIR case for $\gamma_M(\omega,\boldsymbol{p}) = M$ [168].

The transient error for the LMS algorithm, i.e. the estimation error at iteration $n+1$ with respect to the error at iteration n, can be approximated as

$$\mathbb{E}\{|E(\omega, \boldsymbol{p}, \hat{\boldsymbol{\theta}}(n+1))|^2\} \approx [1-\mu S_u(\omega)]^2 \mathbb{E}\{|E(\omega, \boldsymbol{p}, \hat{\boldsymbol{\theta}}(n))|^2\} + \mu^2 \sigma_v^2 S_u(\omega) \gamma_M(\omega, \boldsymbol{p}), \tag{5.24}$$

which shows the error dependency on two terms: when $\mu S_u(\omega)$ is large, the error is reduced in the first term, but it increases based on the second term. The presence of the frequency-dependent term $\gamma_M(\omega, \boldsymbol{p})$, i.e. the number and location of the poles, affects the second term of (5.24), which also depends on the noise variance. Once again, the expression in (5.24) is a generalization of the FIR case [168].

Convergence rate: step size and numerical conditioning

The orthogonality property of OBF filters ensures better-behaved and faster convergence of the adaptation algorithm, compared to non-orthogonal fixed-pole adaptive filters [130]. For the LMS algorithm, the convergence speed is determined by the choice of the step size μ and by the condition number C of the intermediate signals correlation matrix $\boldsymbol{R}_{\boldsymbol{\kappa}}$, defined as the spread of its eigenvalues as $C(\boldsymbol{R}_{\boldsymbol{\kappa}}) = {}^{\lambda_{\max}}/_{\lambda_{\min}}$, with $\lambda_{\max}$ and $\lambda_{\min}$ the maximum and minimum eigenvalues, respectively [273]. Convergence is controlled by the exponential factor $(1-\mu\lambda_{\min})^n$, with a larger value for $\lambda_{\min}$ yielding faster convergence [157], which decays to zero for $\mu < {}^{1}/_{\lambda_{\max}}$. It follows that the convergence rate in the mean for the LMS algorithm can be no faster than [130]

$$\left(1 - \frac{\lambda_{\min}}{\lambda_{\max}}\right)^n = \left(1 - \frac{1}{C(\boldsymbol{R}_{\boldsymbol{\kappa}})}\right)^n. \tag{5.25}$$

For OBF filters with a WN input signal, i.e. with constant PSD $S_u(\omega) = \sigma_u^2$, the convergence rate is optimal as $C(\boldsymbol{R}_{\boldsymbol{\kappa}}) \approx 1$ ($\boldsymbol{R}_{\boldsymbol{\kappa}} \approx \sigma_u^2 \boldsymbol{I}$, with $\boldsymbol{I}$ the identity matrix). For colored input signals, the optimal conditioning of the correlation matrix is lost. However, by virtue of orthogonality, a bound on $C(\boldsymbol{R}_{\boldsymbol{\kappa}})$ in relation to the input PSD is derived as

$$\min_{\omega\in[-\pi,\pi]} S_u(\omega) \leq \lambda(\boldsymbol{R}_{\boldsymbol{\kappa}}) \leq \max_{\omega\in[-\pi,\pi]} S_u(\omega), \tag{5.26}$$

with $\lambda(\boldsymbol{R}_{\boldsymbol{\kappa}})$ the set of eigenvalues of $\boldsymbol{R}_{\boldsymbol{\kappa}}$, so that an upper bound on the average convergence rate can be found as [130]

$$\left(1 - \frac{\min\limits_{\omega\in[-\pi,\pi]} S_u(\omega)}{\max\limits_{\omega\in[-\pi,\pi]} S_u(\omega)}\right)^n. \tag{5.27}$$

This means that OBF adaptive filters are particularly robust in terms of numerical well-conditioning for a wide range of input PSD [169], so that the condition number is expected to be smaller compared to the case in which a non-orthogonal structure is used, and thus the convergence rate faster.

5.2.3 The OBF-NLMS and its analogy to TD-NLMS

As mentioned above, the NLMS algorithm is meant to avoid that large values in the regression vector would result in large variations in the parameter update. Whereas in FIR adaptive filters, this is accomplished by normalizing with respect to the power of the previous M samples of the input signal (or, by considering $\tilde{\mu} = \mu M$, with respect to the mean value of $\|\boldsymbol{u}(n)\|^2$) [283], the normalization in the OBF filter case has a different interpretation. Indeed, the NLMS update in (5.20) normalizes the regression vector $\boldsymbol{\kappa}(n, \boldsymbol{p})$ with respect to the instantaneous power of all the intermediate signals at time n, or equivalently, by considering $\tilde{\mu} = \mu M$, by their mean power. This means that the update rule has no memory of previous samples of $\boldsymbol{\kappa}(n, \boldsymbol{p})$, so that large values of the input signal may result in large variations of the linear coefficients. It is possible, indeed, especially when the filter order M is very small, that the power σ_ι^2 at time n of different intermediate signals $\kappa_\iota(n)$ is similar, and so is their mean power.

For this reason an alternative version of the NLMS algorithm (named OBF-NLMS) is introduced here, which normalizes each intermediate signal $\kappa_\iota(n)$ individually based on an estimate of its power σ_ι^2. The gain vector $\boldsymbol{g}(n)$ for the OBF-NLMS is then

$$\boldsymbol{g}(n) = \mu[\delta \boldsymbol{I}_M + \hat{\boldsymbol{\Sigma}}_M(n, \boldsymbol{p})]^{-1} \boldsymbol{\kappa}(n, \boldsymbol{p}) \,, \tag{5.28}$$

with $\hat{\boldsymbol{\Sigma}}_M(n, \boldsymbol{p})$ an $M \times M$ diagonal matrix, whose ι^{th} diagonal element $\hat{\sigma}_\iota^2$ is an estimate of the the intermediate signal power σ_ι^2. The power estimates are computed using an exponential window update, implemented as a one-pole filter with pole $0 \ll \beta < 1$ (i.e. the forgetting factor) as

$$\hat{\sigma}_\iota^2(n) = \beta \hat{\sigma}_\iota^2(n-1) + (1-\beta)|\kappa_\iota(n)|^2 \tag{5.29}$$

Each linear coefficient is then updated individually as

$$\hat{\theta}_\iota(n+1) = \hat{\theta}_\iota(n) + \frac{\mu}{\delta + \hat{\sigma}_\iota^2(n)} \kappa_\iota(\tilde{\boldsymbol{p}}_i, n) \varepsilon(n, \boldsymbol{p}, \boldsymbol{\theta}) \,. \tag{5.30}$$

The only disadvantages compared to the standard NLMS are a small increase in complexity due to (5.29) and the requirement of a reasonable initial estimate $\hat{\sigma}_\iota^2(0)$ at the beginning of the adaptation, in order to avoid slow convergence

of the power estimates. This second issue is easily overcome by computing a short-term mean on the first few samples of $|\kappa_\iota(n)|^2$ before starting to adapt the filter coefficients.

The necessity of introducing the OBF-NLMS algorithm for low model orders M is illustrated in Figure 5.3, showing the identification of a low-frequency RIR simulated using the randomized image-source method (RIM) [230] (reverberation time (RT) $T_{60} = 0.25$ s, room M in Table 5.1) with male speech input signal downsampled to $f_s = 800$ Hz. The curves represent the normalized misalignment (NM) obtained using the standard NLMS and the OBF-NLMS algorithm, which is defined as

$$\mathrm{NM}(n) = 10\log_{10}\left(\frac{\|\boldsymbol{h} - \hat{\boldsymbol{h}}(n, \boldsymbol{p}, \hat{\boldsymbol{\theta}})\|_2^2}{\|\boldsymbol{h}\|_2^2}\right) \text{ dB}, \tag{5.31}$$

where $\boldsymbol{h}$ is the true RIR vector of length N samples, and $\hat{\boldsymbol{h}}(n, \boldsymbol{p}, \hat{\boldsymbol{\theta}}) = \boldsymbol{\Psi}_N(\boldsymbol{p})\hat{\boldsymbol{\theta}}(n)$ is the estimated RIR at time n, with the columns of the $N \times M$ matrix $\boldsymbol{\Psi}_N(\boldsymbol{p})$ being the N-samples OBF responses to an impulsive input signal (see [113]). It can be seen in the top plot that large values of the input signal results in large variations of the steady-state error for the NLMS algorithm when the model order is small (here $M = 6$). The OBF-NLMS algorithm, on the other hand, provides faster convergence and low parameter variability in the steady-state. For higher model orders, instead, the NLMS adaptation rule in (5.20) becomes effective in dealing with large values of the input signal, with comparable performance at a lower complexity, as can be seen in the bottom plot for $M = 20$.

An analogy between the OBF-NLMS algorithm and the time-domain implementation of TD adaptive algorithms [278, 279, 280, 281, 282] is noticed. TD algorithms apply an orthonormal transform, such as the discrete Fourier transform (DFT) or the discrete cosine transform (DCT) to name the most common, in order to partially decorrelate the samples of the input signal vector $\boldsymbol{u}(n)$ and thus accelerate the rate of convergence of the FIR filter parameters when using the NLMS algorithm. It has been shown [279] that the convergence actually improves when, instead of normalizing all the intermediate signals (i.e. the transformed input signals) with respect to the instantaneous power of $\boldsymbol{u}(n)$ as in the standard NLMS algorithm, the normalization is performed at each intermediate signal with respect to the inverse of a short-time average of its power, as in (5.30). It is actually this normalization and not the transformation itself that reduces the eigenvalue spread and speeds up convergence.

Another, more conceptual, analogy between OBF adaptive filters and TD algorithms is in their interpretation as filterbanks [280, 282]: the orthonormal transforms, such as the DFT or the DCT, can indeed be seen as a parallel

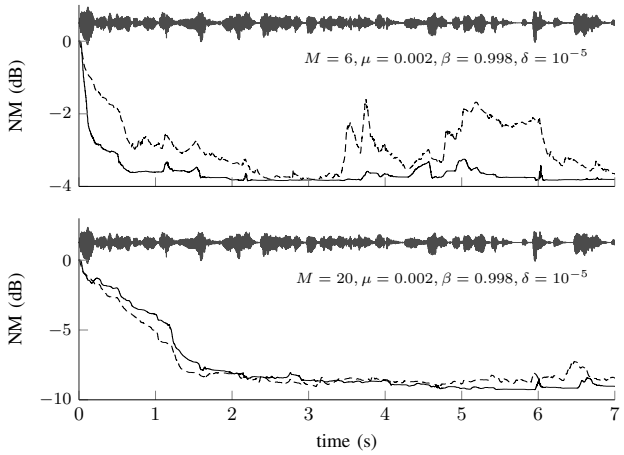

Figure 5.3: The comparison between the NM of NLMS (dashed) and OBF-NLMS (solid) for $M = 6$ (top) and $M = 20$ (bottom).

of band-pass filters with a uniform distribution of their central frequencies. The decorrelation performance of different transforms depends mainly on the characteristics of the input PSD [280]. For instance, the DCT is particularly suitable for input signals with a low-pass characteristic. Also OBF adaptive filters can be seen as a parallel of band-pass filters which partially decorrelate the intermediate signals, with the difference that the filter central frequencies and bandwidth (i.e. the poles) are chosen with respect to the system to be identified and not to the expected input PSD, even though both the pole location and the input PSD influence the tracking behavior of the algorithm, as explained in the previous section.

5.2.4 Adaptation of the poles

Gradient-based algorithms can also be used as well in the adaptation of the denominator coefficients of the TF of IIR filters [163]. However, difficulties arise from the fact that a pole-zero model is a nonlinear regression model, in the sense that the regressors depend on previous values of the filter coefficients (at least in the output-error approach) [163, 168, 170]. This issue is normally circumvented in direct-form pole-zero models by disregarding this dependency and treating the problem as in linear regression (in which case they are called *pseudolinear* regression models). This is not a possibility for OBF adaptive filters, the reason being that each pole pair $\boldsymbol{p}_i$ appears in the all-pass sequence

of the successive $m-i$ TFs (5.4) and in the normalization filters as well, so that they cannot be regarded as pseudolinear regression models.

When minimizing the sum of squared errors in (5.14), recursive prediction error algorithms [170], or Gauss-Newton-type recursive algorithms, should be then used to try to adjust the filter coefficients. For instance, an algorithm was proposed in [151], in which linear coefficients and poles are updated in an alternating fashion using RLS and a recursive Gauss-Newton algorithm with a backtracking strategy for determining the optimal step-size, respectively. The idea of recursive nonlinear identification algorithms is to update the pole parameters $\boldsymbol{p}(n)$ along the search direction $\boldsymbol{q}(n)$ according to

$$\boldsymbol{p}(n+1) = \boldsymbol{p}(n) + \mu \boldsymbol{q}(n), \tag{5.32}$$

where $\boldsymbol{q}(n)$ for the quadratic cost function in (5.14) has the form

$$\begin{aligned} \boldsymbol{q}(n) &= 2\boldsymbol{B}_{\boldsymbol{p}}^{-1}(n)\nabla\varepsilon_{\boldsymbol{p}}(n)\varepsilon(n,\boldsymbol{p},\boldsymbol{\theta}), \quad \text{with} \\ \nabla\varepsilon_{\boldsymbol{p}}(n) &= {}^{\partial\varepsilon(n,\boldsymbol{p},\boldsymbol{\theta})}\!/_{\partial\boldsymbol{p}(n)} = -{}^{\partial\tilde{y}(n,\boldsymbol{p},\boldsymbol{\theta})}\!/_{\partial\boldsymbol{p}(n)}. \end{aligned} \tag{5.33}$$

Different algorithms differ on how $\boldsymbol{B}_{\boldsymbol{p}}(n)$ is chosen. For instance, the steepest descent algorithm chooses $\boldsymbol{B}_{\boldsymbol{p}}(n) = \boldsymbol{I}$, while in the Gauss-Newton algorithm $\boldsymbol{B}_{\boldsymbol{p}}(n) = \nabla\varepsilon_{\boldsymbol{p}}(n)\nabla\varepsilon_{\boldsymbol{p}}^{H}(n)$ is an approximation of the Hessian matrix [288] ($\{\cdot\}^H$ indicating the Hermitian transpose). Common to all these algorithms is the computation of the gradient vector $\nabla\varepsilon_{\boldsymbol{p}}(n)$, i.e. the derivatives of the error with respect to the pole parameters. The expressions for the gradient with respect to the k^{th} pole pair $\boldsymbol{p}_k$ are obtained by computing (and leaving out the dependency of $\tilde{y}(n)$ and $\tilde{y}_i(n)$ on $\boldsymbol{p}$ and $\boldsymbol{\theta}$ for brevity)

$$\frac{\partial\tilde{y}(n)}{\partial\boldsymbol{p}_k(n)} = \sum_{i=1}^{m} \frac{\partial\tilde{y}_i(n)}{\partial\boldsymbol{p}_k(n)}, \tag{5.34}$$

which can be divided in three parts, as

$$\frac{\partial\tilde{y}(n)}{\partial\boldsymbol{p}_k(n)} = \sum_{i=1}^{k-1} \frac{\partial\tilde{y}_i(n)}{\partial\boldsymbol{p}_k(n)} + \frac{\partial\tilde{y}_k(n)}{\partial\boldsymbol{p}_k(n)} + \sum_{i=k+1}^{m} \frac{\partial\tilde{y}_i(n)}{\partial\boldsymbol{p}_k(n)}. \tag{5.35}$$

The first term is zero, since $\tilde{y}_i(n)$ for $i = 1, \ldots, k-1$ is independent from $\boldsymbol{p}_k$. Even though analytic expressions for the other two terms can be derived, the nonlinear dependency of the filter output on the pole parameters makes the expressions involved and expensive to compute. Also, the pole parameters being complex-valued, a parametrization of the poles is necessary in order to obtain real-valued expressions for the gradients. Although it would be natural to parametrize the poles in terms of their radius ρ_i and angle ϑ_i, the computation

of the gradients becomes slightly simpler by considering the parameters $\zeta_i = -(p_i + p_i^*) = -2\rho_i \cos\vartheta_i$ and $\eta_i = p_i p_i^* = \rho_i^2$, for which $D_i(z)$ in (5.1) becomes $D_i(z) = 1 + \zeta_i z^{-1} + \eta_i z^{-2}$ (in [151] the real and imaginary part of the pole parameters were used instead). The expressions for the gradients with respect to ζ_i and η_i were derived in [171] for a slightly different realization of OBF filters and for orthogonal filters.

The full and approximated expressions for the gradients in the realization used in this chapter are given in Appendix A.1. It can be seen that these expressions are quite complicated, especially for the third term in (5.35). A computationally reasonable, but suboptimal way of adapting the pole parameters is by approximating the gradients (with χ_k either ζ_k or η_k) as

$$\frac{\partial \tilde{y}(n)}{\partial \chi_k(n)} \approx \frac{\partial \tilde{y}_k(n)}{\partial \chi_k(n)}\,, \tag{5.36}$$

which assumes slow convergence of the parameters and poles close to the unit circle (see [171]). A further simplification is obtained by renouncing to the normality property of OBF filters by redefining $N_i^{\pm}(z) = (z^{-1} \mp 1)$, in which case the gradient becomes

$$\frac{\partial \tilde{y}(n)}{\partial \chi_k(n)} \approx \frac{\partial \tilde{y}_k(n)}{\partial \chi_k(n)} = -\frac{1}{D_k}\frac{\partial D_k}{\partial \chi_k}\tilde{y}_k(n)\,, \tag{5.37}$$

with ${}^{\partial D_k}/_{\partial \zeta_k} = z^{-1}$ and ${}^{\partial D_k}/_{\partial \eta_k} = z^{-2}$. This simplification of the OBF model to an orthogonal model would also allow to regard it as a pseudolinear model, with the regressors defined by the input signal $u(n)$ and the various outputs of resonators and all-pass filters (signals $a_i(n)$, $x_i(n)$ and their previous samples), where the dependency of the regressors on previous values of the parameters to be estimated is ignored. In the following, these ideas for the adaptation of the poles are not developed further. Instead, an iterative identification algorithm is proposed, which estimates the poles of one or multiple RTFs avoiding the nonlinear problem by employing a grid-based matching pursuit approach.

5.3 The SB-OBF-GMP identification algorithm

It was shown in the previous section that the identification of the linear coefficients of an OBF adaptive filter is governed by the same conditions and with the same implementation complexity as for the identification with an FIR filter (with differences in frequency resolution and noise sensitivity based on $\gamma_M(\omega, \boldsymbol{p})$ defined in (5.12)). The identification of the poles, on the other hand, is a nonlinear problem and, even though it is possible to devise recursive algorithms

as discussed above, problems such as slow convergence or convergence to local minima may occur. Slow convergence is also related to the number of poles and the choice of the initial values in recursive algorithms, which should be based on some prior (usually unavailable) knowledge about the system. Another issue is the computational complexity of the recursive algorithm, especially if a backtracking strategy for the selection of the step-size in (5.32) is used to speed up convergence [288].

Here, a different approach to the identification of the poles is taken. Instead of adapting the pole parameters, the inherent nonlinear problem is avoided by using a grid search and by selecting poles one by one in an order-recursive fashion. The iterative algorithm proposed here is similar to the BB-OBF-GMP algorithm in [154], in which a dictionary is built by collecting N_b samples of candidate intermediate signals, with poles defined on a grid spanning a portion of the unit disc. In each block b, one pair of complex-conjugate poles $\boldsymbol{p}_b$ is selected from the grid as the one that produces the pair of intermediate signals that is mostly correlated, on average, with the last N_b samples of the prediction error signals $\varepsilon_r(n)$ produced in each acoustic channel r considered ($r = 1, \ldots, R$) using LS estimation. The pole-pair $\boldsymbol{p}_b$ is then added to the previously selected common poles in the multi-channel OBF adaptive filter, whose number m of resonator sections increases by one ($m \leftarrow m+1$), whereas the linear coefficients for each acoustic channel are adapted with respect to each $\varepsilon_r(n)$ using the LMS update rule in (5.18) with gain vector as in (5.19). A description of the algorithm is detailed in Appendix A.2.

The BB-OBF-GMP algorithm proved to be capable of accurately identifying a common set of poles from WN input signals in a single-input/multiple-output (SIMO) room acoustic system. The LMS algorithm, however, works well as long as the input PSD is constant and the step size is correctly chosen according to it. As a consequence, when the input signal is non-stationary and non-white, the step size may be too small or too large, and the algorithm would either converge very slowly or become erratic. Also, when the linear coefficients adapt too slowly and the size of the block is not sufficient, the prediction error signals $\varepsilon_r(n)$ compared to which the correlation with the candidate intermediate signals in the dictionary is computed, may not have reached the steady-state, resulting in a poor identification performance.

The SB-OBF-GMP algorithm proposed here introduces a series of modifications meant to deal with these issues. First, instead of collecting N_b samples of both the candidate intermediate signals and the residual signals, the correlation between them is tracked in time for a given number of samples N_s (with $N_s < N_b$ required), before one new pole pair is selected. The NLMS algorithm is used, as it is equivalent to an adaptation of the correlation between the error signals and the candidate intermediate signals using an exponential window update, as

it will be explained later on. Another modification pertains to the introduction of the OBF-NLMS adaptation rule (5.28) in the multi-channel OBF adaptive filter. The reason is to deal with large values of the intermediate signals when the model order $M = 2m$ is small, i.e. when the multi-channel OBF adaptive filter has a small number m of resonator sections, as explained in Section 5.2.3. For higher model orders, the performance of OBF-NLMS and NLMS are similar, with the latter being less expensive computationally, so that when a sufficiently large number of poles m has been included in the multi-channel OBF adaptive filter, it is possible to employ the NLMS update with gain vector as in (5.20).

The idea of the proposed algorithm is indeed that, since the poles are considered to be a characteristic of the room itself and thus approximately time-invariant [103], the identification of a common set of poles can be performed once for a given setup or environment at the beginning of the session of a RASE task, and then kept fixed afterwards, with RTF variations tracked by the adaptive linear coefficients only. The algorithm is designed for SIMO room acoustic systems, but it can be extended to the multiple-input/multiple-output (MIMO) case, as suggested in Section 5.4. It could be also used to find good initial values and to determine the order M, as required by nonlinear recursive algorithms, which could be also used to adapt the poles to track slow variations in time of the room acoustics.

5.3.1 Algorithm description

The proposed algorithm aims to build a SIMO OBF adaptive filter including one common pole pair at each stage. A *stage* is defined as the period at the end of which a pole pair $\boldsymbol{p}_c$ is selected and included in the multi-channel OBF adaptive filter. A new pole pair $\boldsymbol{p}_c$ is selected based on the correlation coefficients, which values are tracked using the NLMS algorithm for the duration of a stage, between the prediction error signals $\varepsilon_r(n)$ and the candidate intermediate signals. The pole pair in the grid associated with the pair of candidate intermediate signals with the highest correlation, averaged over the R acoustic channels and evaluated at the end of the stage, is selected and added to the active pole set $\boldsymbol{p}_{m+1}^{\mathcal{A}} = [\boldsymbol{p}_m^{\mathcal{A}}, \boldsymbol{p}_c]$ of the multi-channel OBF adaptive filter, thus adding a new resonator section. At the beginning of each new stage ($m \leftarrow m+1$), the linear coefficients $\hat{\boldsymbol{\theta}}_m^r(n) = [\theta_m^{r+}(n), \theta_m^{r-}(n)]$ associated with the intermediate signals built from the pole pair $\boldsymbol{p}_c$ selected at the previous stage, start to be adapted for each channel r, together with $\hat{\boldsymbol{\theta}}_i^r(n) = [\theta_i^{r+}(n), \theta_i^{r-}(n)]$ with $i = 1, \ldots, m-1$. In other words, the size of each r^{th} set of linear coefficients increases by two at the end of each stage, and their value adapted using either the OBF-NLMS or the NLMS update. The algorithm stops whenever the number of selected poles m reaches a maximum value $m_{\max}$ or some stopping criterion based on the

average power of the prediction error signals is satisfied. Readers uninterested in the details of the algorithm may skip the following description and resume the reading from Section 5.3.2 or even from Section 5.4, without compromising their understanding of the rest of the chapter.

The aim of the SB-OBF-GMP algorithm is to minimize the sum of the instantaneous squared errors, over the R channels,

$$\begin{aligned}\underset{\boldsymbol{p}_m^{\mathcal{A}},\hat{\boldsymbol{\Theta}}_M(n)}{\text{minimize}}\ \|\boldsymbol{\epsilon}_m(n)\|^2 &= \left\|\boldsymbol{y}(n)-\hat{\boldsymbol{y}}_m(n,\boldsymbol{p}_m^{\mathcal{A}},\hat{\boldsymbol{\Theta}}_M)\right\|^2 \\ &= \left\|\boldsymbol{y}(n)-\boldsymbol{\kappa}(n,\boldsymbol{p}_m^{\mathcal{A}})^T\hat{\boldsymbol{\Theta}}_M(n)\right\|^2, \end{aligned} \tag{5.38}$$

where $\boldsymbol{\epsilon}_m(n) = [\varepsilon_m^1(n),\ldots,\varepsilon_m^R(n)]$ represents the vector of prediction error signals $\varepsilon_m^r(n)$ (with $r = 1,\ldots,R$) for the R acoustic channels at time n, $\boldsymbol{y}(n) = [y^1(n),\ldots,y^R(n)]$ is the vector of output signals and $\hat{\boldsymbol{y}}_m(n,\boldsymbol{p}_m^{\mathcal{A}},\hat{\boldsymbol{\Theta}}_M) = [\hat{y}_m^1(n,\boldsymbol{p}_m^{\mathcal{A}},\hat{\boldsymbol{\theta}}^1(n)),\ldots,\hat{y}_m^R(n,\boldsymbol{p}_m^{\mathcal{A}},\hat{\boldsymbol{\theta}}^R(n))]$ the vector of estimated outputs of the multi-channel OBF adaptive filter, with the linear filter coefficient vectors defined as $\hat{\boldsymbol{\theta}}^r(n) = [\hat{\boldsymbol{\theta}}_1^r(n),\ldots,\hat{\boldsymbol{\theta}}_m^r(n)]^T$ and $\hat{\boldsymbol{\theta}}_i^r(n) = [\hat{\theta}_i^{r+}(n),\hat{\theta}_i^{r-}(n)]$. The symbol $\{\hat{\cdot}\}$ is used instead of $\{\tilde{\cdot}\}$ to indicate the fact that $\hat{\boldsymbol{y}}_m(n,\boldsymbol{p}_m^{\mathcal{A}},\hat{\boldsymbol{\Theta}}_M)$ is an estimate. The vector of estimated output signals $\hat{\boldsymbol{y}}_m(n,\boldsymbol{p}_m^{\mathcal{A}},\hat{\boldsymbol{\Theta}}_M)$ is obtained by the linear combination of the $M = 2m$ intermediate signals $\boldsymbol{\kappa}(n,\boldsymbol{p}_m^{\mathcal{A}}) = [\boldsymbol{\kappa}_1(n),\ldots,\boldsymbol{\kappa}_m(n)]^T$ weighted by the linear filter coefficient vectors $\hat{\boldsymbol{\theta}}^r(n)$, which are stacked in the $M\times R$ matrix $\hat{\boldsymbol{\Theta}}_M(n) = [\hat{\boldsymbol{\theta}}^1(n),\ldots,\hat{\boldsymbol{\theta}}^R(n)]$. The intermediate signals in $\boldsymbol{\kappa}(n,\boldsymbol{p}_m^{\mathcal{A}})$ are the output signals of the orthonormalized resonator sections built from the active pole set $\boldsymbol{p}_m^{\mathcal{A}}$ and having TFs $\Psi(z,\boldsymbol{p}_m^{\mathcal{A}}) = \{\Psi_1^{\pm}(z,\tilde{\boldsymbol{p}}_1^{\mathcal{A}}),\ldots,\Psi_m^{\pm}(z,\tilde{\boldsymbol{p}}_m^{\mathcal{A}})\}$, where $\Psi_i^{\pm}(z,\tilde{\boldsymbol{p}}_i^{\mathcal{A}})$, defined as in (5.4), is built from the first i poles $\tilde{\boldsymbol{p}}_i^{\mathcal{A}} = [\boldsymbol{p}_1,\ldots,\boldsymbol{p}_i]^T \in \boldsymbol{p}_m^{\mathcal{A}}$.

A schematic of the SB-OBF-GMP algorithm is depicted in Figure 5.4 and listed in Algorithm 3 with a slightly simplified notation, both containing elements explained below.

Multi-channel OBF linear filter coefficient adaptation

As already mentioned, the proposed OBF-NLMS update rule in (5.28) is used, at least while M is small, for the adaptation of the linear coefficients of the multi-channel OBF filter, which in matrix form for the multi-channel case becomes

$$\hat{\boldsymbol{\Theta}}_M(n+1) = \hat{\boldsymbol{\Theta}}_M(n) + \mu[\delta\boldsymbol{I}_M + \hat{\boldsymbol{\Sigma}}_M(n,\boldsymbol{p}_m^{\mathcal{A}})]^{-1}\boldsymbol{\kappa}(n,\boldsymbol{p}_m^{\mathcal{A}})\boldsymbol{\epsilon}_m(n)\,. \tag{5.39}$$

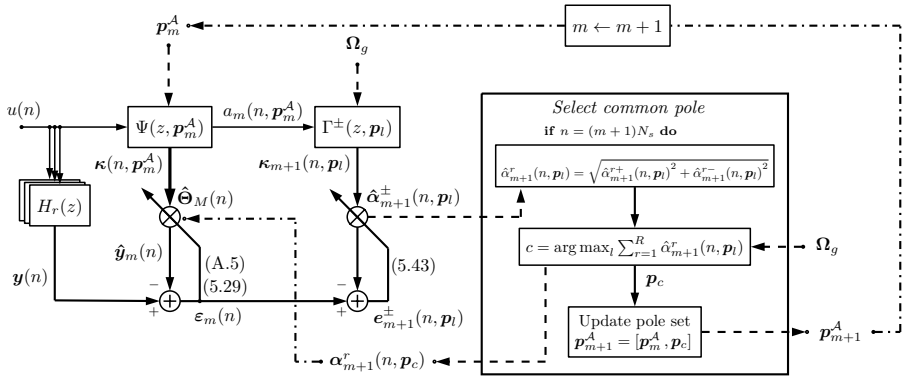

Figure 5.4: The simplified schematics of the SB-OBF-GMP algorithm.

Initially ($m = 0$), the active pole set $\boldsymbol{p}_m^{\mathcal{A}}$ is empty, so that the vector of estimated output equals to all zeros ($\hat{\boldsymbol{y}}_0(n) = \boldsymbol{0}$) and $\boldsymbol{\epsilon}_0(n) = \boldsymbol{y}(n)$. At the beginning of each new stage, a new pole pair $\boldsymbol{p}_c$ is added to the multi-channel OBF filter based on the selection strategy described below, and a pair of linear coefficients per each channel is included in each coefficient vector $\hat{\boldsymbol{\theta}}^r(n)$, thus augmenting the size of the square matrices $\boldsymbol{I}_M$ and $\hat{\boldsymbol{\Sigma}}_M(n, \boldsymbol{p}_m^{\mathcal{A}})$ and the number of rows of $\hat{\boldsymbol{\Theta}}_M(n)$ in (5.39) by two ($M \leftarrow M + 2$).

Common-poles selection strategy

The main purpose of the SB-OBF-GMP algorithm is to identify a set of poles common to all R acoustic channels to be fixed into a multi-channel OBF adaptive filter. The poles are estimated using a matching pursuit approach. First, a grid $\boldsymbol{\Omega}_g$ of L candidate pairs of complex-conjugate poles is defined on the unit disc based on some prior knowledge of the room acoustic system or some particular desired frequency resolution [113]. For each pole pair $\boldsymbol{p}_l \in \boldsymbol{\Omega}_g$ (with $l = 1, \ldots, L$), the pair of candidate intermediate signals $\boldsymbol{\kappa}_{m+1}(n, \boldsymbol{p}_l) = [\kappa^+_{m+1}(n, \boldsymbol{p}_l), \kappa^-_{m+1}(n, \boldsymbol{p}_l)]$ is obtained as the $(m+1)$-th intermediate signals of an OBF filter built from the pole set $[\boldsymbol{p}_m^{\mathcal{A}}, \boldsymbol{p}_l]$, i.e. by filtering the input $u(n)$ with the TFs $\Psi^{\pm}_{m+1}(z, \boldsymbol{p}_l) = N^{\pm}_{m+1}(z, \boldsymbol{p}_l) P_{m+1}(z, \boldsymbol{p}_l) \prod_{i=1}^{m} A_i(z, \boldsymbol{p}_i)$, where the product corresponds to the series of m second-order all-pass filters defined by the pole pairs $\boldsymbol{p}_i \in \boldsymbol{p}_m^{\mathcal{A}}$ (cfr. Figure 5.1). Equivalently, $\boldsymbol{\kappa}_{m+1}(n, \boldsymbol{p}_l)$ can be computed by filtering the output of the all-pass series $a_m(n, \boldsymbol{p}_m^{\mathcal{A}}) = \prod_{i=1}^{m} A_i(q, \boldsymbol{p}_i) u(n)$ with pairs of filters having TFs $\Gamma^{\pm}(z, \boldsymbol{p}_l) = N^{\pm}_{m+1}(z, \boldsymbol{p}_l) P_{m+1}(z, \boldsymbol{p}_l)$.

Algorithm 3 SB-OBF-GMP algorithm

$\mathbf{\Omega}_g = \{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_L\}$ ▷ Define pole grid

$\boldsymbol{p}_0^{\mathcal{A}} = \emptyset\,, m = 0$ ▷ Initialize the set of active poles

$\boldsymbol{\epsilon}_0(n) = \boldsymbol{y}(n),\ a_0(n) = u(n-d)$ ▷ Initialize signals

$\hat{\boldsymbol{\alpha}}_l^{\pm}(0) = \mathbf{0},\ \ l = 1, \ldots, L$ ▷ Set correlation coefficients

while $m < m_{\max}$ **do** ▷ $m_{\max}$: max number of pole pairs

Update multi-channel OBF adaptive filter

  if $m > 0$ **then**

    $\boldsymbol{\epsilon}_m(n) = \boldsymbol{y}(n) - \boldsymbol{\kappa}(n, \boldsymbol{p}_m^{\mathcal{A}})^T \hat{\mathbf{\Theta}}_M(n)$ ▷ (5.38)

    $\hat{\sigma}_\iota^2(n) = \beta \hat{\sigma}_\iota^2(n-1) + (1-\beta)|\kappa_\iota(n)|^2$ ▷ (5.29)

    $\hat{\mathbf{\Theta}}_M(n+1) = \hat{\mathbf{\Theta}}_M(n) + \mu[\delta \boldsymbol{I}_M + \hat{\mathbf{\Sigma}}_M(n, \boldsymbol{p}_m^{\mathcal{A}})]^{-1} \boldsymbol{\kappa}(n, \boldsymbol{p}_m^{\mathcal{A}}) \boldsymbol{\epsilon}_m(n)$ ▷ (5.39)

  end if

Update candidate intermediate signals and correlation coeffs.

  $\kappa_{m+1}^{\pm}(n, \boldsymbol{p}_l) = \Gamma^{\pm}(q, \boldsymbol{p}_l) a_m(n), \quad \forall \boldsymbol{p}_l \in \mathbf{\Omega}_g$

  $\hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n+1, \boldsymbol{p}_l) = \lambda \hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n, \boldsymbol{p}_l) + (1-\lambda) \frac{\kappa_{m+1}^{\pm}(n, \boldsymbol{p}_l) \boldsymbol{\epsilon}_m(n)}{\|\boldsymbol{\kappa}_{m+1}(n, \mathbf{\Omega}_g)\|^2 / 2L}$ ▷ (5.43)

Select common pole and set variables

  if $n = (m+1)N_s$ **then**

    $\hat{\alpha}_{m+1}^r(n, \boldsymbol{p}_l) = \sqrt{\hat{\alpha}_{m+1}^{r+}(n, \boldsymbol{p}_l)^2 + \hat{\alpha}_{m+1}^{r-}(n, \boldsymbol{p}_l)^2}$ ▷ (5.44)

    $c = \arg\max_l \sum_{r=1}^{R} \hat{\alpha}_{m+1}^r(n, \boldsymbol{p}_l)$ ▷ (5.45)

    $\boldsymbol{p}_{m+1}^{\mathcal{A}} = [\boldsymbol{p}_m^{\mathcal{A}}\,,\ \boldsymbol{p}_c]$ ▷ Add $\boldsymbol{p}_c$ to active pole set

    $\sigma_\iota^2 = \|\boldsymbol{\kappa}_{m+1}(n, \mathbf{\Omega}_g)\|^2 / 2L$ ▷ Set power estimate

    $\hat{\boldsymbol{\theta}}_{m+1}^r(n) \leftarrow \hat{\boldsymbol{\alpha}}_{m+1}^r(n, \boldsymbol{p}_c)\ \ (r = 1, \ldots, R)$ ▷ Set new $\boldsymbol{\theta}$

    $m \leftarrow m + 1$ ▷ Move to next stage

    $\hat{\boldsymbol{\alpha}}_l^{\pm}(n) = \emptyset,\ \forall l$ ▷ Reset correlation coefficients

  end if

end while

The result is a parallel of L pairs of candidate intermediate signals $\boldsymbol{\kappa}_{m+1}(n, \boldsymbol{p}_l)$ with a common input $a_m(n)$, which can be stacked in a vector $\boldsymbol{\kappa}_{m+1}(n, \mathbf{\Omega}_g) = [\boldsymbol{\kappa}_{m+1}(n, \boldsymbol{p}_1), \ldots, \boldsymbol{\kappa}_{m+1}(n, \boldsymbol{p}_L)]^T$ of size $2L \times 1$. The idea is that, while the linear coefficients $\hat{\mathbf{\Theta}}_M(n)$ of the OBF filter are being adapted using the OBF-NLMS rule to minimize the power of the error signals in $\boldsymbol{\epsilon}_m(n)$, the algorithm updates also the correlation coefficients $\hat{\alpha}_{m+1}^{r\pm}(n, \boldsymbol{p}_l)$ between each

error signal $\varepsilon_m^r(n)$ and each candidate intermediate signal $\kappa_{m+1}^{\pm}(n,\boldsymbol{p}_l)$. In order to minimize the instantaneous squared error signals in $\boldsymbol{e}_{m+1}^{\pm}(n,\boldsymbol{p}_l) = [\boldsymbol{e}_{m+1}^{1\pm}(n,\boldsymbol{p}_l),\ldots,\boldsymbol{e}_{m+1}^{R\pm}(n,\boldsymbol{p}_l)]$, averaged over the R channels, produced by each candidate intermediate signal,

$$\underset{\boldsymbol{p}_l,\hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n)}{\text{minimize}} \left\|\boldsymbol{e}_{m+1}^{\pm}(n,\boldsymbol{p}_l)\right\|^2 = \left\|\boldsymbol{\epsilon}_m(n) - \kappa_{m+1}^{\pm}(n,\boldsymbol{p}_l)\hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n,\boldsymbol{p}_l)\right\|^2, \quad (5.40)$$

each vector of correlation coefficients $\hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n,\boldsymbol{p}_l) = [\hat{\alpha}_{m+1}^{1\pm}(n,\boldsymbol{p}_l),\ldots,\hat{\alpha}_{m+1}^{R\pm}(n,\boldsymbol{p}_l)]$ is adapted in time. It follows from simple calculations that the LMS adaptation rule minimizing (5.40) can be written as

$$\hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n+1,\boldsymbol{p}_l) = \left(1-\mu|\kappa_{m+1}^{\pm}(n,\boldsymbol{p}_l)|^2\right)\hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n,\boldsymbol{p}_l) + \mu\kappa_{m+1}^{\pm}(n,\boldsymbol{p}_l)\boldsymbol{\epsilon}_m(n). \quad (5.41)$$

Normalizing the step size μ with respect to the instantaneous power of the regressor, gives (for $\lambda = 1-\mu$)

$$\hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n+1,\boldsymbol{p}_l) = \lambda\hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n,\boldsymbol{p}_l) + (1-\lambda)\frac{\kappa_{m+1}^{\pm}(n,\boldsymbol{p}_l)\boldsymbol{\epsilon}_m(n)}{|\kappa_{m+1}^{\pm}(n,\boldsymbol{p}_l)|^2}, \quad (5.42)$$

which is recognized as an exponential window update of the normalized instantaneous correlation between $\kappa_{m+1}^{\pm}(n,\boldsymbol{p}_l)$ and $\boldsymbol{\epsilon}_m(n)$ with forgetting factor λ.

The update rule in (5.42) is not immune to large values of the intermediate signals $\kappa_{m+1}^{\pm}(n,\boldsymbol{p}_l)$, which would result in large variations of the estimates $\hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n+1,\boldsymbol{p}_l)$. Since tracking the power of each $\kappa_{m+1}^{\pm}(n,\boldsymbol{p}_l)$ as proposed for the OBF-NLMS would be too expensive, an effective solution is proposed, which consists of normalizing the instantaneous correlations by the instantaneous average of the overall power of all candidate intermediate signals, giving

$$\hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n+1,\boldsymbol{p}_l) = \lambda\hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n,\boldsymbol{p}_l) + (1-\lambda)\frac{\kappa_{m+1}^{\pm}(n,\boldsymbol{p}_l)\boldsymbol{\epsilon}_m(n)}{\|\boldsymbol{\kappa}_{m+1}(n,\boldsymbol{\Omega}_g)\|^2/2L}. \quad (5.43)$$

The average of the instantaneous power is in this case a good estimator of the actual power of the candidate intermediate signals, given that the poles in the grid are numerous and distributed within a wide frequency range. This is the same argument to explain the comparable performance of the NLMS algorithm with respect to OBF-NLMS for higher model orders (cfr. Figure 5.3).

At the end of each stage, i.e. at sample $n = (m+1)N_s$, the selection of the pole pair is performed. Given that (5.43) is a correlation update, the choice is based on which pair of candidate intermediate signals is mostly correlated on average

with the error signals $\boldsymbol{\epsilon}_m(n)$. The correlation of each pair of intermediate signals for the r-th channel is computed as

$$\hat{\alpha}^r_{m+1}(n, \boldsymbol{p}_l) = \sqrt{\hat{\alpha}^{r+}_{m+1}(n, \boldsymbol{p}_l)^2 + \hat{\alpha}^{r-}_{m+1}(n, \boldsymbol{p}_l)^2}. \tag{5.44}$$

A pole pair $\boldsymbol{p}_{m+1} \leftarrow \boldsymbol{p}_c$ is then chosen from the grid as the one giving intermediate signals with the highest average correlation on the R acoustic channels as

$$c = \arg\max_l \sum_{r=1}^{R} \hat{\alpha}^r_{m+1}(n, \boldsymbol{p}_l) \tag{5.45}$$

and added to the active pole set $\boldsymbol{p}^{\mathcal{A}}_{m+1} = [\boldsymbol{p}^{\mathcal{A}}_m, \boldsymbol{p}_{m+1}]$ of the OBF adaptive filter. The linear filter coefficients $\hat{\boldsymbol{\theta}}^r_{m+1}(n) = [\hat{\theta}^{r+}_{m+1}(n), \hat{\theta}^{r-}_{m+1}(n)]$ are set equal to the correlation coefficients $\hat{\boldsymbol{\alpha}}^r_{m+1}(n, \boldsymbol{p}_c) = [\hat{\alpha}^{r+}_{m+1}(n, \boldsymbol{p}_c), \hat{\alpha}^{r-}_{m+1}(n, \boldsymbol{p}_c)]$, so that they are already close to their steady-state value. Moreover, if the OBF-NLMS is used, the power estimates for the newly added poles are set equal to $\sigma^2_\iota = \|\boldsymbol{\kappa}_{m+1}(n, \boldsymbol{\Omega}_g)\|^2 / 2L$ (with ι equal to $M+1$ and $M+2$). Finally, the algorithm moves to the next stage ($m \leftarrow m+1$), all the correlation coefficients are reset to zero, and another pole pair is estimated as described above, until a desired number of poles $m = m_{\max}$ has been selected or some other stopping criterion based on the error in (5.38) is satisfied.

5.3.2 Algorithm evaluation

The final goal of the SB-OBF-GMP algorithm is to determine a set of poles, common to multiple acoustic paths, which are close to the true poles of the system considered. Since this knowledge, as well as the knowledge of the optimal pole parameters of an OBF filter, is usually not available, the algorithm is evaluated with respect to the OBF-GMP algorithm [113, 156]. Also, it should be noticed that the poles identified do not have a one-to-one correspondence with the true poles. In many cases, due to the mode overlapping, the finite resolution of the grid, and the iterative nature of the algorithm, one pole pair with slightly smaller radius is selected in the vicinity of two or more true pole pairs, so that on average the distance between the estimated poles and the true poles is reduced, as well as the bias error in (5.16).

Most estimation methods, including the OBF-GMP algorithm, for estimating the pole parameters of an OBF filter rely on the availability of measured RIRs. Thus, the purpose of this evaluation is to verify whether the proposed identification algorithm is able to achieve the same approximation error that is obtained with the OBF-GMP algorithm using the same pole grid. Obtaining

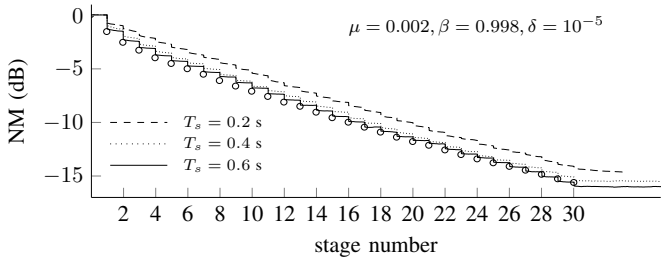

Figure 5.5: The averaged NM for the OBF-GMP algorithm (∘), which has access to the measured RIRs, and for the SB-OBF-GMP algorithm with different stage lengths T_s. Room M, $T_{60} = 0.5$ s, WN input signals.

comparable results is a confirmation of the performance of the SB-OBF-GMP algorithm, given that estimating the parameters from input-output signals is more involved than from measured RIRs.

Identification from white noise signals

The evaluation is performed with respect to the identification of a set of simulated RIRs generated using the RIM [230] at sampling frequency $f_s = 800$ Hz at one loudspeaker and 4 microphone positions ($R = 4$, room M, T_{60}= 0.5 s, see Section 5.4 for details). First, the RIRs are modeled using the OBF-GMP algorithm [113, 156]. The pole grid used has 400 angles uniformly placed from 1 Hz to 399 Hz and 5 radii logarithmically distributed from 0.7 to 0.9925, summing up to $L = 2000$ candidate pole pairs. The NM in (5.31), averaged over the 4 RIRs, is computed for each estimation of a new pole pair and depicted in Figure 5.5 using ∘. Then, the SB-OBF-GMP algorithm is tested for 10 realizations of an input WN sequence with unit variance, using the same pole grid described above. The NM is averaged over the 4 RIRs and over the different realizations of the input signal. Three different values for the stage length $T_s = {}^{N_s}/_{f_s}$ have been used, with a new pole pair added to the active pole set every 0.2, 0.4, and 0.6 seconds, corresponding to the three curves of the averaged NM in Figure 5.5. It is seen that results very similar to the OBF-GMP algorithm can be obtained already with stages of $T_s = 0.4$ s (a significant improvement with respect to the BB-OBF-GMP algorithm [154] requiring blocks of more than 2 s).

Identification from speech signals

The identification from non-stationary and non-white signals, such as speech, is more challenging. One difficulty is related to the slower convergence rate due to the non-constant input PSD, as described in Section 5.2.2, for which longer stages may be required to reduce the misalignment. Moreover, since OBF-NLMS and NLMS are used to counteract problems related to the non-stationarity of speech, the non-constant input PSD increases the steady-state error (see (5.23)), both for the linear coefficients in $\hat{\boldsymbol{\Theta}}_M$ and the correlation coefficients in $\hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n, \boldsymbol{p}_l)$. Convergence rate can be increased using a larger step size μ (see (5.24)), at the expense of a larger steady-state error, which also leads to higher misalignment and less accurate pole identification. On the other hand, greater accuracy is achieved by reducing the step size, which however requires longer stages for the coefficients to converge.

In any case, regardless of the stage length T_s chosen, the short-term frequency spectrum within one stage is usually far from flat, resulting in an uneven excitation of the frequency range of interest. In a given stage, some of the candidate OBF TFs are not sufficiently excited for the corresponding correlation coefficient to converge sufficiently fast. It follows that the pole selection is influenced by the frequency content of the input signal in the current stage, so that deviations from the behavior of the modeling algorithm are normally expected.

Another issue is the long-term frequency range excited by a speech signal at low frequencies. The voiced speech of an adult male typically has fundamental frequency between 85 and 180 Hz, whereas that of an adult female between 165 and 255 Hz [289], so that a speech signal rarely has sufficient power to excite the lower modes of the system. Relatively small rooms already have modal frequencies well below the cut-off frequency of a speech signal. Thus, the capabilities of the algorithm of identifying the system from speech signals at low frequencies depend on both the characteristics of the signal itself and of the modal characteristics of the room response, as discussed further in Section 5.4.

To illustrate these concepts, the SB-OBF-GMP algorithm is tested on 10 long sequences of male speech taken from an audiobook "A Tramp Abroad" by Mark Twain[1], downsampled to $f_s = 800$ Hz, where the silent portions of the signals were removed using a voice activity detection algorithm [290, 258]. The same pole grid as in the WN case is used. Three different stage lengths $T_s = \{0.5, 1, 2\}$ s have been chosen, corresponding to the three curves of the NM, averaged over the 4 RIRs and over the 10 different input signal sequences

[1] publicly available at http://librivox.org/a-tramp-abroad-by-mark-twain/, MP3-files at 128kbps were converted to WAV.

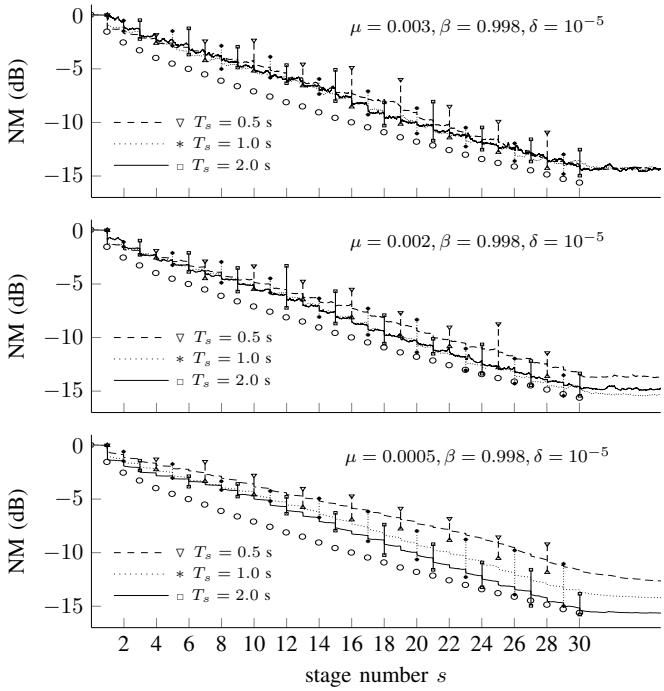

Figure 5.6: The averaged NM for the OBF-GMP algorithm ($\circ$) and for the SB-OBF-GMP algorithm with different step sizes μ and different stage lengths T_s. Room M, $T_{60} = 0.5$ s, male speech input signal (librivox).

(the vertical lines correspond to the range between the maximum and minimum values). The experiment is repeated for three different values of the step size μ, corresponding to the three plots in Figure 5.6. In the top plot ($\mu = 0.003$), roughly the same result, with similar deviations, is obtained for different stage lengths, even though the performance of the OBF-GMP algorithm is not attained. This bias is mostly due to the large steady-state error, and not due to the low convergence rate, given that a longer stage does not provide almost any improvement. As already mentioned, a lower misalignment is achieved by employing a smaller step size, thus reducing the parameter variability. The convergence rate, however, decreases, so that improvements are obtained only for stages long enough to allow the coefficients to converge. An overall improvement can be noticed in the middle plot, especially for $T_s = 1$ s and $T_s = 2$ s. By further reducing the step size, the parameter variability decreases, but also the convergence rate, so that a further reduction in the misalignment may be difficult even for long stages, as suggested from the bottom plot of Figure 5.6

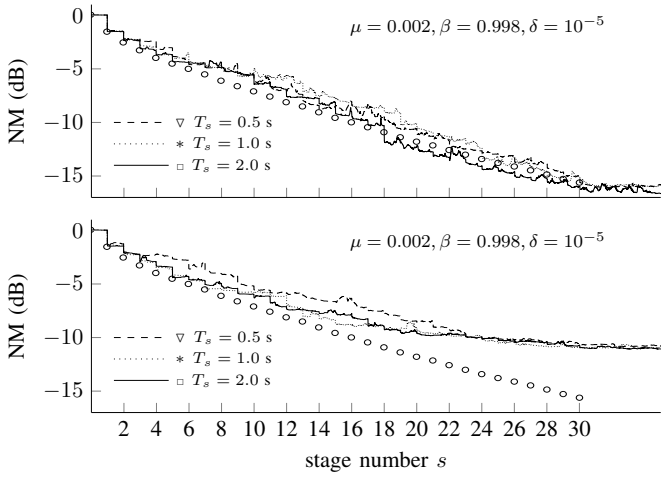

Figure 5.7: The averaged NM for the OBF-GMP algorithm (○) and for the SB-OBF-GMP with different stage lengths T_s. Room M, $T_{60} = 0.5$ s, male (top) and female (bottom) speech input signal (EBU-SQAM).

for $\mu = 0.0005$ and $T_s = 2$ s. Extensive simulations on different materials show that, in general, a good trade-off between convergence rate and steady-state error is provided by a step size between $\mu = 0.002$ and $\mu = 0.001$, which gives reasonably low misalignment for a stage length T_s below 1 s.

Another example is presented showing the same experiment performed on a male speech sequence taken from the EBU-SQAM database [291] (see top plot of Figure 5.7). In this case, the SB-OBF-GMP algorithm with stage length 1 s, is able to obtain similar results in terms of NM as the OBF-GMP algorithm[2]. What differentiates this case from the previous ones is the approximately flat long-term PSD above 80 Hz, which also corresponds to the cut-off frequency of the frequency response of the room. To conclude, the algorithm is tested for a female speech sequence from the same database. As mentioned above, the fundamental frequency of female speech is normally above 165 Hz. As a consequence, not enough signal power is available below that frequency, and the system cannot be fully identified. This is shown in the bottom plot of Figure 5.7, where, after a certain number of stages, adding a new pole pair does not significantly reduce the NM.

[2]the SB-OBF-GMP algorithm gives a small improvement over the OBF-GMP algorithm in some cases because the iterative pole selection strategy of both algorithms is optimal only in relation to the poles already selected.

5.4 Identification results at low frequencies

In all system identification tasks, the first step is to decide which model is the most adequate for the problem at hand. Trying out different models is normally cumbersome and in some cases not even possible. It is then important to have an indication about which model to use, based on some prior knowledge of the system. For applications in room acoustics, the prior information that may be available regards the room dimensions, the characteristics of the surfaces in the room and/or the reverberation time.

The question regarding the adequateness of IIR models in RASE applications (especially AEC) is a recurrent topic in the literature [238, 240, 239]. In these works, one common conclusion is that the advantage given by the superior modeling capabilities of IIR filters over FIR filters is not significant enough to justify their use, given the higher filter complexity and the added difficulty in the approximation/identification process. This result was explained by observing that a RTF is composed of a large number of resonances, and since an IIR filter models each resonance by a second-order TF, the number of filter parameters that is required to accurately identify the system may be not far from the required number of filter taps in an FIR filter.

In [113], it was shown by simulation results on a number of RIRs measured in different rooms that the use of OBF filters gives an advantage over FIR filters, even when the increased filter complexity is accounted for; this advantage is more significant at low frequencies, where a RTF normally presents sharper resonances and a lower degree of modal overlap [1]. The actual advantage of using OBF filters is then dependent on these two elements, which in turn are related to the room dimensions, the reverberation time, and the number of poles used.

In this section, simulation results are shown in order to analyze the use of OBF and FIR adaptive filters in relation to the characteristics of the room at low frequency. Moreover, the estimation of poles common to multiple acoustic channels, which brings computational savings [113, 156] (since each loudspeaker signal is filtered by the same OBF filter, having microphone-dependent linear coefficients), is evaluated at positions in the room (validation microphones) others than the ones at which the poles are estimated (training microphones). The analysis is first performed on simulated scenarios, and then verified on measured RIRs.

Table 5.1: RIM simulated room specifications

room	V (m^3)	$N_{f_s/2}$	T_{60} (s)	f_{Sch} (Hz)	$N_{f_{\text{Sch}}}$
S	17	113	0.15, 0.25, 0.50	188, 243, 343	12, 25, 71
M	50	332	0.25, 0.50, 0.75	141, 200, 245	15, 42, 76
L	300	1993	0.50, 0.75, 1.00	82, 100, 115	17, 31, 47

5.4.1 Simulated rooms

Three different rooms were simulated at $f_s = 800$ Hz using the RIM [230] with a random displacement of 1 cm, each room with three different RTs, resulting in 9 different cases. The characteristics of the rooms are given in Table 5.1, listing the room volume V, the theoretical number of modes [1] below half the sampling frequency,

$$N_{f_s/2} \approx \frac{4\pi}{3} V \left(\frac{f_s/2}{c} \right)^3 \tag{5.46}$$

(with $c = 343\,\text{m/s}$ the sound velocity), the 3 different RTs T_{60} per each room, their corresponding Schroeder frequency $f_{\text{Sch}} \approx 2000\sqrt{T_{60}/V}$, and the theoretical number of modes $N_{f_{\text{Sch}}}$ below the Schroeder frequency. The particular choice of the sampling frequency was made to include the highest Schroeder frequency among the 9 cases considered.

Multiple loudspeakers and microphones are distributed in a fixed configuration in the rooms as illustrated in Figure 5.8, also showing the room width W, length L and height H. Full dots represent the validation microphones, consisting of four arrays V_a ($a = 1, \ldots, 4$) of 4 microphones each, with inter-microphone spacing of 25 cm (note that V_4 cannot be used for Room S). Empty dots represent the training microphones, consisting of two arrays T_b ($b = 1, \ldots, 2$) of 4 microphones each. The first 3 validation arrays and the training arrays form a $1\,\text{m} \times 1\,\text{m}$ area, that will be referred at as *sweet-spot*, whereas the fourth validation array is the *isolated* array. In both training and validation, two different WN input signals are fed to two loudspeakers X and Y, indicated with a full square. The microphone noise was set very low (SNR = 60 dB) so not to influence the performance of the identification. It is assumed that the loudspeaker-microphone distances are known, so that the acoustic delay parameter d_r can be set in the filter for each r^{th} acoustic channel[3].

[3] in a practical implementation, the d_r-samples delays are applied to the intermediate signals before they are fed to the multi-channel linear coefficients.

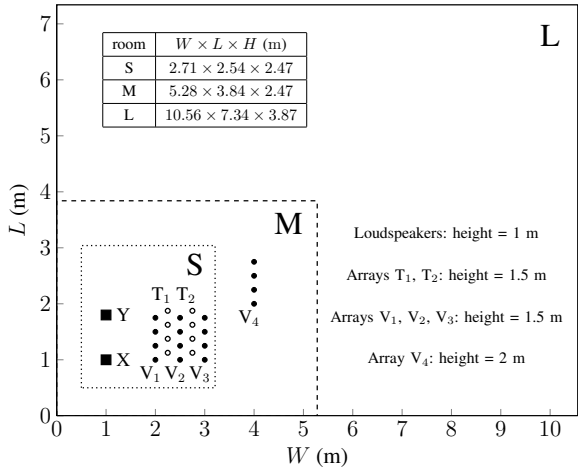

Figure 5.8: Schematics of the three simulated rooms considered, with the relative position of loudspeakers (■), and training (○) and validation (•) microphones.

In the training, a set $\boldsymbol{p}_m^{\mathcal{A}}$ of common poles of the multi-channel OBF adaptive filter is identified from the 8 microphone signals of arrays T_b using the SB-OBF-GMP algorithm with $T_s = 0.5\,\mathrm{s}$ and variables defined as in Section 5.3.2. When the identification is performed from two (or more) loudspeakers, the two loudspeaker signals have to be filtered individually by two SIMO OBF filters in parallel. If the input signals are uncorrelated, as for different realizations of WN, the MIMO system can be correctly identified from the microphone signals, and a set of poles common to all the MIMO acoustic channels considered can be obtained with the pole selection strategy of the SB-OBF-GMP algorithm with very minor modifications. The identification from correlated signals is discussed in Section 5.5.1 in the framework of stereophonic acoustic echo cancellation (SAEC). Even though the pole grid could be chosen according to some prior knowledge of the room and its RT, here the same grid has been used in all conditions. The pole grid has $L = 4000$ pole pairs, with 400 angles uniformly placed from 1 Hz to 399 Hz (giving approximately 1 Hz of resolution), and 10 radii logarithmically distributed from 0.85 to 0.9925. These limits for the radius ρ are chosen to include the minimum and the maximum values of the RT T_{60} considered, according to the formula [1]

$$\rho = 10^{\frac{-3}{T_{60} f_s}} \tag{5.47}$$

The lower bound ($\rho_{\min} = 0.85$), corresponding to relatively wide resonances with a decay time of 50 ms, is meant to allow for the approximation of the superposition of multiple resonances.

In the validation, the estimated common poles $\boldsymbol{p}_m^{\mathcal{A}}$ are fixed in the multi-channel OBF filter and only the linear coefficients are adapted from the microphone signals of both training and validation arrays. The adaptation on the training and validation arrays using the same signals is performed also with FIR adaptive filters, whose performance is compared to that of the OBF adaptive filters with the same number $M = 2m$ of adaptive linear coefficients. For both filters, the NLMS algorithm in (5.20) is used, as the case of very low orders requiring the OBF-NLMS algorithm is not considered here. The validation on each array for a specific value of M is repeated 3 times, each time using 2 different realizations of WN (30 s long, one per each loudspeaker) as the input signals.

The NM, as defined in (5.31), is computed for each acoustic channel for the loudspeaker-microphone pairs in the specific configuration (2 loudspeakers, 1 array), and then averaged over the number of acoustic channels and the 3 repetitions,

$$\overline{\mathrm{NM}}_{\mathrm{Z}}(\infty) = 10\log_{10}\left(\frac{1}{3R_{\mathrm{Z}}}\sum_{r=1}^{R_{\mathrm{Z}}}\sum_{k=1}^{3}\frac{\|\boldsymbol{h}_r^k - \hat{\boldsymbol{h}}_r^k(\infty, \boldsymbol{p}_m^{\mathcal{A}}, \hat{\boldsymbol{\theta}}_r)\|_2^2}{\|\boldsymbol{h}_r^k\|_2^2}\right)\ \mathrm{dB}, \tag{5.48}$$

where $\hat{\boldsymbol{h}}_r(\infty, \boldsymbol{p}_m^{\mathcal{A}}, \hat{\boldsymbol{\theta}}_r)$ indicates the RIR estimated at steady-state at the end of the k^{th} repetition for a set of poles $\boldsymbol{p}_m^{\mathcal{A}}$ identified during training, and Z indicates the array used in the validation (e.g. T_b or V_a) with a total of R_{Z} channels. In practice, the NM at steady-state is computed by averaging the misalignment obtained on the last 10 s of the validation signal, when the parameters have normally converged, as

$$\frac{\|\boldsymbol{h}_r^k - \hat{\boldsymbol{h}}_r^k(\infty, \boldsymbol{p}_m^{\mathcal{A}}, \hat{\boldsymbol{\theta}}_r)\|_2^2}{\|\boldsymbol{h}_r^k\|_2^2} \approx \frac{1}{10f_s}\sum_{n=20f_s}^{30f_s}\frac{\|\boldsymbol{h}_r^k - \hat{\boldsymbol{h}}_r^k(n, \boldsymbol{p}_m^{\mathcal{A}}, \hat{\boldsymbol{\theta}}_r)\|_2^2}{\|\boldsymbol{h}_r^k\|_2^2}. \tag{5.49}$$

Also the average convergence time $\overline{\mathrm{CT}}_{\mathrm{Z}}$ is computed per each array Z as the time instant at which the average NM reaches +2 dB above $\overline{\mathrm{NM}}_{\mathrm{Z}}(\infty)$. These quantities are illustrated in Figure 5.9.

In the following analysis, four measures are considered:

(i) the average NM on the training sets (speakers X and Y, and arrays T_b, $R_{T_b} = 8$) calculated for the validation microphone signals, using OBF and FIR adaptive filters.

$$\overline{\mathrm{NM}}_{\mathrm{T}}(\infty) = \frac{1}{2}\sum_{b=1}^{2}\overline{\mathrm{NM}}_{\mathrm{T}_b}(\infty), \tag{5.50}$$

(ii) the distance in dB between $\overline{\mathrm{NM}}_{\mathrm{T}}(\infty)$ and the average NM using OBF adaptive filters on the validation sets,

$$\overline{\Delta\mathrm{NM}}_{\mathrm{TV}}(\infty) = \frac{1}{N_a}\sum_{a=1}^{N_a}\Big[\overline{\mathrm{NM}}_{\mathrm{T}}(\infty) - \overline{\mathrm{NM}}_{\mathrm{V}_a}(\infty)\Big]\ \mathrm{dB}, \tag{5.51}$$

with $\overline{\mathrm{NM}}_{\mathrm{V}_a}(\infty)$ computed using (5.48) on set a (speakers X and Y, and array V_a, $R_{V_a} = 8$) and N_a the number of validation sets considered in the average,

(iii) the distance in dB between the average NM for the validation sets obtained using OBF adaptive filters and FIR adaptive filters,

$$\overline{\Delta\mathrm{NM}}_{\mathrm{VF}}(\infty) = \frac{1}{N_a}\sum_{a=1}^{N_a}\Big[\overline{\mathrm{NM}}_{\mathrm{V}_a}(\infty) - \overline{\mathrm{NM}}_{\mathrm{F}_a}(\infty)\Big]\ \mathrm{dB}, \tag{5.52}$$

with $\overline{\mathrm{NM}}_{\mathrm{F}_a}(\infty)$ computed using (5.48) for FIR adaptive filters on set a (speakers X and Y, and array V_a),

(iv) the average convergence rate, defined as the ratio between the NM at +2 dB above the steady-state and the convergence time CT, averaged over multiple arrays,

$$\overline{\mathrm{CR}} = \frac{1}{N_a}\sum_{a=1}^{N_a}\left[\frac{\overline{\mathrm{NM}}_{\mathrm{Z}_a}(\infty) + 2}{\overline{\mathrm{CT}}_{\mathrm{Z}_a}}\right]\ \mathrm{dB/s}, \tag{5.53}$$

calculated on the validation sets for OBF and FIR adaptive filters ($\mathrm{Z}_a = \{\mathrm{V}_a, \mathrm{F}_a\}$).

For all these averaged measures, also the minimum and maximum values are computed and plotted as vertical bars in the plots that follow. The analysis is performed on the 9 room configurations (3 rooms, 3 RTs each), first on the arrays in the sweet-spot, and then separately on the isolated array.

Evaluation at the sweet-spot

The results for the measures, computed as in (5.50)–(5.53) for the training and the first 3 validation sets, are given in Figure 5.10, where each row corresponds to a measure and each column to one room. The first row, instead, shows the magnitude responses (loudspeaker X, first microphone of array T_2) for the 3 rooms with 3 different values of the RT, with the vertical lines indicating the corresponding Schroeder frequency.

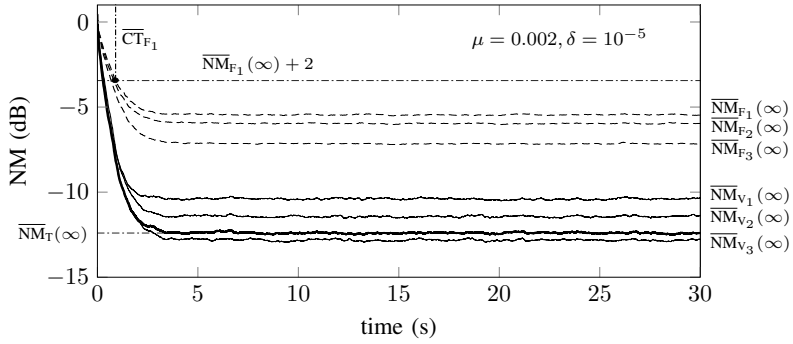

Figure 5.9: Illustration of the quantities used in the analysis measures: the NM computed on the training microphones (thick), and on the validation arrays using OBF (solid) and FIR (dashed) filters. Room M, $T_{60} = 0.75$ s, $M = 60$.

***(i)* Misalignment at training positions** Looking at the curves in the second row of Figure 5.10, the first thing that can be noticed is that the average NM reduces consistently for increasing numbers M of the linear coefficients. This is of course expected, since the common poles are identified at the training positions based on the minimization of the average power of the residuals, which is also the reason for the small differences between the NM at the two training arrays (see the 'short' bars in the plots).

Also, the absolute reduction in the NM is dependent on the RT: a shorter RT implies wider resonances and a more prominent modal overlap, so that a smaller number of poles is required to approximate the RTF (e.g. notice in the top-left plot the reduced number of spectral peaks for $T_{60} = 0.15\,\text{s}$ compared to $T_{60} = 0.5\,\text{s}$). However, a shorter RT also implies a faster time decay of the RIR, which means that an FIR filter as well requires less parameters to attain a given NM. The NM thus reduces with the RT both for OBF and FIR filters for increasing M, with the former showing a stronger reduction, which will be analyzed further on the validation sets.

Another factor influencing the absolute reduction in the NM is the volume of the room, and more precisely the number of modes $N_{f_s/2}$ in the RTF. Indeed, large rooms have a large number of modes in a fixed frequency range and thus a higher modal density. However, they normally also have a long RT, so that the overlap between modal resonances is only partial (at least at low frequencies), as can be noticed in the top-right plot. This means that, for a given RT and a given number of poles, the absolute NM that is achievable depends on the room dimensions, with larger rooms having higher NM (e.g. compare the three rooms for $T_{60} = 0.5\,\text{s}$ and $M = 120$).

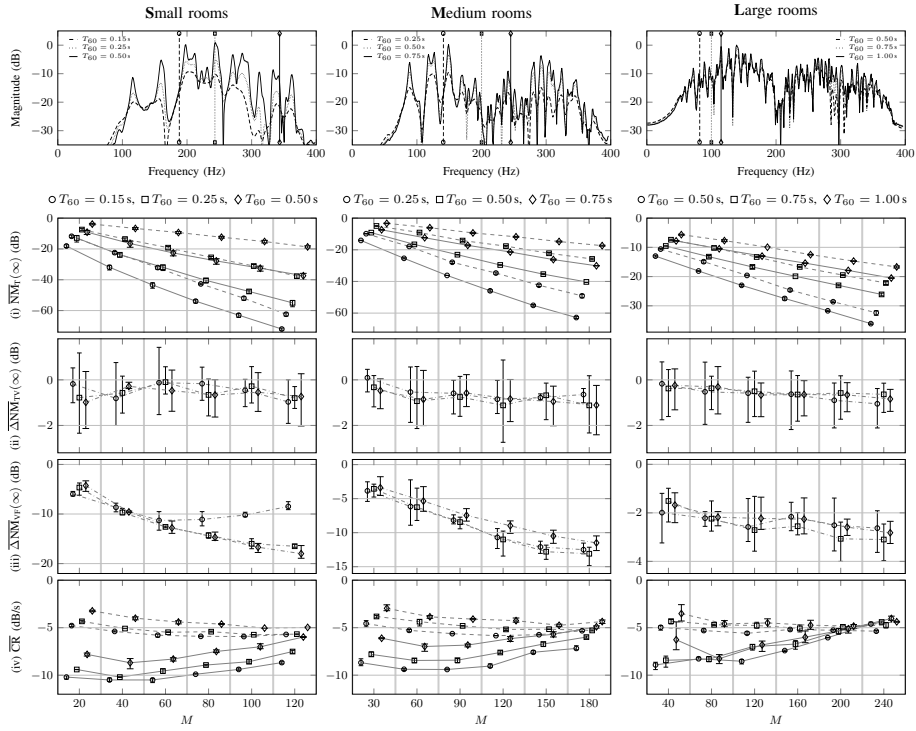

Figure 5.10: *Sweet-spot*: identification results for measures in (5.50-5.53), using OBF (solid) and FIR (dashed) filters with different orders M, and for the 9 cases considered (3 rooms, 3 RTs). Top row: example magnitude responses and corresponding f_{Sch} (vertical lines).

An alternative way of quantifying the difference in performance of OBF and FIR filters is to compare the number of coefficients that is required for them to achieve a predefined value of the NM. As an example, consider the medium room with $T_{60} = 0.75\,\mathrm{s}$ at $\overline{\mathrm{NM}}_{\mathrm{T}}(\infty) = -18\,\mathrm{dB}$: in this case, OBF filters require $M = 90$ linear coefficients, while FIR filters $M = 180$.

***(ii)* Misalignment at validation positions** The assessment of the NM at the training positions is useful to analyze the performance of OBF filters for the identification algorithm used. In practice, however, the receiver may change position in space, so that it is of interest analyzing the degradation in performance at other positions when the poles are kept fixed. The variations in the NM in positions others than those at which the poles were computed

is quantified here by means of the measure defined in (5.51). The third row in Figure 5.10 shows that the degradation introduced by moving the receivers within the sweet-spot area is limited to 1 dB (with ± 1 dB variability), irregardless of the room characteristics. This is a good indication that the estimated poles can be considered common, at least in the neighborhood of the training positions.

The validation of the identification results at a position away from the training positions is considered later on. Moreover, the tracking performance, i.e. the ability of the adaptive algorithm to track variations in the RTF due, for instance, to the receiver changing position during adaptation, is analyzed in the AEC example in Section 5.5.1.

***(iii)* Comparison between OBF and FIR filters** The third measure considered is intended to verify the actual advantage that is given by OBF filters over FIR filters at validation positions within the sweet-spot, which is the actual situation in case poles are estimated at some training positions and kept fixed when used at other positions. Starting from small rooms, it can be seen that the gap between OBF and FIR filters, defined as in (5.52), increases for increasing M. However, it can be noticed that, for short RT, this gap is reduced above a certain value of M, the reason being that all the main resonances of the RTF have been already modeled, so that adding one more pole pair reduces the NM less than adding two more taps to an FIR filter.

The same trend is noticed for medium rooms, which have a larger number of resonant modes in the frequency range considered. As already mentioned, a longer RT implies sharper resonances and the need of a larger number of poles to achieve a small NM. For this reason the absolute reduction in the NM is less than in small rooms (as is the case at the training positions). It can be also seen that for longer RT (see $T_{60} = 0.75$ s), the gap increases more slowly for increasing M. This can be ascribed to the lower degree of modal overlap and the consequent increase in the number of resonances that have to be modeled. This fact is even more noticeable for large rooms, where the number of modes becomes even larger. Even though the number of required coefficients for FIR filters increases (given the longer decay time of the RIRs), a large number of coefficients is required for OBF filters as well, so that the difference between the two types of filters becomes less significant (which is the main argument against the use of IIR filters in the literature [240, 239]).

Another factor influencing the gap between the performances of OBF and FIR filters is the finite resolution of the pole grid of the SB-OBF-GMP algorithm used in the training phase. Even though the grid-based approach is effective in finding a good low-order approximation of the RTFs, it happens that a mode

is not efficiently modeled by a single pole pair in the OBF filter because the true pole of the system has values for the angle and/or radius in-between the values used in the grid. Thus, this lack of resolution may lead to a reduced gap between the performances of OBF and FIR filters, especially when the modal resonances are sharper and more numerous, as happens in large rooms.

***(iv)* Convergence rate** The convergence rate (CR) measure in (5.53) is meant to assess the relation between the NM at steady-state and the time it takes for the adaptation algorithm to attain it. The CR is clearly dependent on the step size μ, so that a faster CR is achieved for a larger μ, at the expense of a larger steady-state error (see Section 5.2.2). As shown in Section 5.2.2, the estimation error for OBF adaptive filters between two successive iterations is governed by the term $\gamma_M(\omega, \boldsymbol{p})$ in (5.24), with $\gamma_M(\omega, \boldsymbol{p}) = M$ for FIR filters. In these simulations, the step size in the NLMS adaptation rule in (5.20) was kept fixed at $\mu = 0.002$, so that $\tilde{\mu} = \mu M$ increases with increasing values of M. It is worth recalling that the loudspeaker signals are in this case unit-variance WN sequences (i.e. $S_u(\omega) = 1$, $\forall\omega$), and that the microphone noise variance is quite low ($\sigma_v^2 \approx 10^{-6}$).

From the last row of Figure 5.10 it can be seen that the CR is basically constant for FIR filters, with possibly a slightly lower rate for small model orders (the first samples of a downsampled RIR correspond to early reflections, which are large in absolute value and so more iterations are needed for the coefficients to converge from zero). Since $\tilde{\mu}$ increases for increasing M, the first term of the error in (5.24) must decrease, whereas the second term increases. It follows that, for the CR to be constant, the two terms must compensate each other.

The faster convergence rate for OBF filters at low values of M can then be partially explained by the fact that $\gamma_M(\omega, \boldsymbol{p}) < M$ for most values of the frequency, so that the second term in (5.24) remains small. By adding more poles, the NM reduces, but $\gamma_M(\omega, \boldsymbol{p})$ increases, at least at those frequencies where the pole density is high. The other reason can be found in how the NM decreases for increasing M. Whereas, apart from early reflections, the NM for FIR filters tends to decrease linearly (on a logarithmic scale), OBF filters show a decreasing exponential trend of the NM (see [113, 156] for more convincing examples than in Figure 5.10). This is a consequence of the fact that with OBF filters, and especially when iterative procedures are used, the most dominant room modes are modeled first, so that the relative reduction in the NM decreases with increasing M, and consequently the CR.

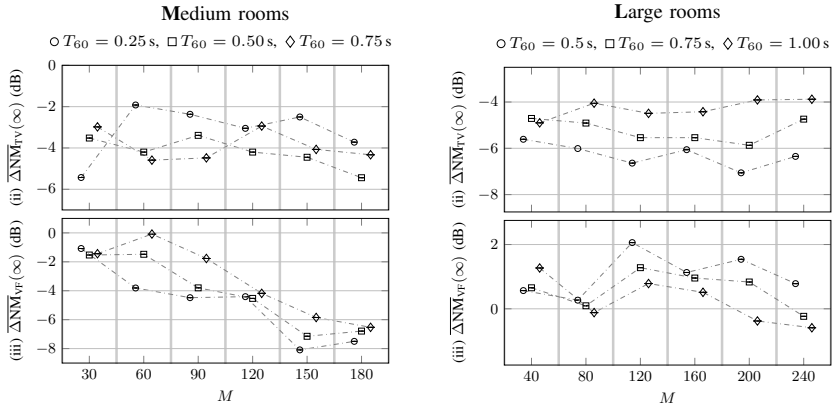

Figure 5.11: *Isolated array*: identification results for measures in (5.51-5.52), using OBF and FIR filters with different orders M, and for the 6 cases considered (2 rooms, 3 RTs).

Evaluation at the isolated array

The above analysis focused on receivers relatively close to the training positions. Here, validation is performed on an array of microphones V_4 placed at roughly 2 m from the training microphones. The results are shown in Figure 5.11. Even in this case, the difference in the NM between training and validation are basically independent on the room characteristics, but larger than in the sweet-spot. The degradation is even more prominent for larger rooms (around 5.5 dB), where the fact of having more and sharper resonances increases the chances of a higher variability in the spatial distribution of the modes in the room (i.e. the finite number of common poles selected at the sweet-spot may not correspond, at least in part, to the set of dominant modes in the isolated array). The analysis of the comparison between OBF and FIR filters at validation positions done above is still valid in this case, with the difference that the gap is smaller due to the higher NM at V_4. The trends seen in Figure 5.10 (fourth row) are recognizable also in this case, with the gap between OBF and FIR filters getting larger for increasing M in the medium room case, and no substantial difference for large rooms, with FIR filters slightly outperforming OBF filters. It follows that common poles estimated in a restricted area of a room do not correspond to all the poles of the system, so that more microphones should be placed around the room in order to extend the sweet-spot area. This is not a surprising result, whose implications are discussed later on in Section 5.6.

5.4.2 Real room (SMARD database)

The same experiment and the same analysis described above in the case of simulated RIRs is repeated here for RIRs measured in a real rectangular room of dimension 7.34 m × 8.09 m × 2.87 m with a measured RT below 400 Hz of $T_{60} = 0.3\,\text{s}$ ($V = 209\,\text{m}^3$, $N_{f_s/2} = 1387$, $f_{\text{Sch}} = 93\,\text{Hz}$, $N_{f_{\text{Sch}}} = 17$) [50]. Two loudspeakers and two orthogonal arrays were used in the simulations, the training set consisting of two loudspeakers and 8 microphones from the first array (inter-microphone distance ≈15 cm) and the 2 validation sets consisting of the same two loudspeakers and 4 microphones from the second array[4]. The same room and configurations were also simulated using the RIM, so to compare the results with respect to the real room. Already from the comparison of the magnitude response at the top of Figure 5.12 (microphone 4Y, first array), it is clear that the simulated response does not match the actual response very closely, the main reason being the fact that in the simulated scenario the same homogeneous broadband reflection coefficient was used for all the surfaces. The result is that the real response shows some strong and sharp resonances (probably due to a lack of absorption at low frequency) and a high degree of modal overlap at higher frequency, whereas the simulated response has sharp resonances also above 200 Hz.

The same kind of analysis as above is performed also in this case. The NM at training positions shows characteristics similar to previous simulation results, with a distance between OBF and FIR filters around -6 dB. For the simulated room, the NM decreases more slowly as a results of the larger number of spectral peaks to be modeled, as already discussed. The degradation at validation positions is around 3 dB, but with a larger variability if compared to the simulated room. The OBF-FIR gap (fourth row) increases only slightly for increasing M; for the real room, the reason being that, apart from few slowly decaying modes below 100 Hz, the response decays quite fast, so that it can be approximated with a relatively small number of coefficients of an FIR filter as well. For the simulated room the gap is only 1 dB, because the large number of spectral peaks makes the advantage of OBF over FIR filters less significant, as it was the case for the large rooms in the simulated environment discussed above. Finally, the CR results are in accordance with previous analysis, with OBF filters showing high efficiency especially for small M, i.e. when the most prominent modes are approximated with a small number of poles, such that $\gamma_M(\omega, \boldsymbol{p})$ has small values.

[4] more specifically, configurations 1000 and 1100 for the training set, and 1001 and 1101 for the validation sets (see http://www.smard.es.aau.dk/).

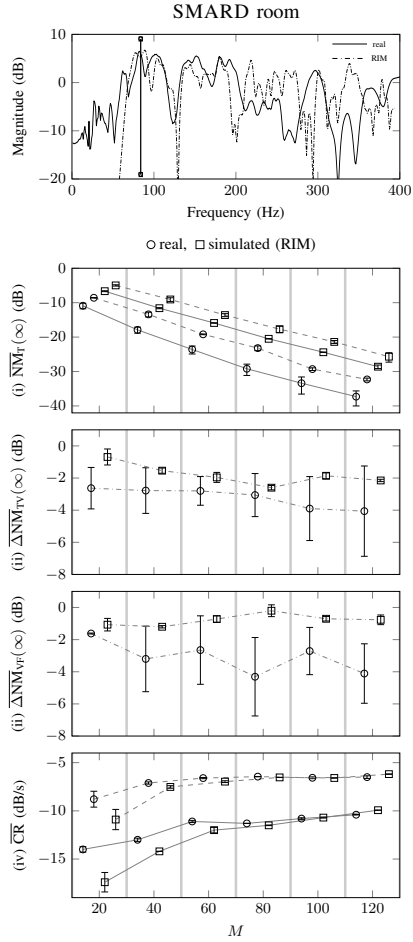

Figure 5.12: *SMARD*: identification results for measures in (5.50-5.53), using OBF (solid) and FIR (dashed) filters with different orders M, and for the measured and simulated responses of the SMARD room. Top: example magnitude response and corresponding f_{Sch} (vertical line).

5.5 Applications in acoustic signal enhancement

In the previous section it has been shown that the use of OBF adaptive filters can bring an advantage compared to FIR filters in terms of identification accuracy and convergence behavior, depending on the characteristics of the room, and that common poles estimated at given positions inside a room can be considered

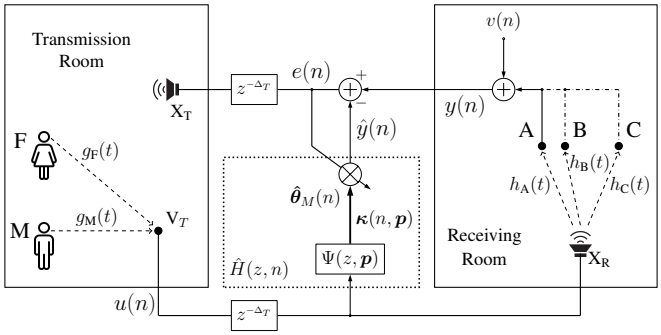

Figure 5.13: Schematics of the AEC scenario using an OBF filter.

valid at other positions in their vicinity. The aim of the current section is to show, by means of examples, how OBF adaptive filters can be practically used in RASE applications. It has been discussed in Section 5.2 how OBF filters with fixed poles can be seen as a generalization of FIR filters, also possessing similar properties in terms of coefficient adaptation. The only difference consists in the location of the fixed poles, which influences the estimation accuracy, but also convergence and the variability of the coefficients at steady-state. It follows that most of the tasks encountered in RASE applications can be tackled with the same strategies usually employed with FIR filters.

Even though the application examples presented in this section, as well as the previous analysis, focus on low frequencies, the extension to higher frequencies is straightforward, especially if a subband processing approach is used. The discussion about the use of OBF filters at higher frequencies is postponed to Section 5.6. In the following, examples are provided for two RASE tasks, namely AEC and RRE.

5.5.1 Acoustic echo cancellation (AEC)

The first scenario, depicted in Figure 5.13, considers a simple monophonic acoustic echo canceler. A male speaker M is talking in the transmission room (small RIM room, $T = 0.15\,\mathrm{s}$, position X, see Figure 5.8), and his voice signal $s_\mathrm{M}(t)$ is convolved with RIR $g_\mathrm{M}(t)$ and picked up by microphone $\mathrm{V_T}$ (first microphone, array $\mathrm{V_3}$). The resulting (sampled and delayed by Δ_T samples) signal $u(n-\Delta_T)$ is reproduced in the receiving room (SMARD room, configurations 1000 and 1001) through loudspeaker $\mathrm{X_R}$. The reproduced speech signal is then convolved with RIR $h_\mathrm{A}(t)$, picked up by the microphone at position A (corresponding to microphone 1Y in the first orthogonal array) and, finally,

the (sampled and delayed) microphone signal $y(n)$ is transmitted back through loudspeaker $\mathrm{X_T}$ to the transmission room, where it is perceived as echo. At a certain time t_1, the male speaker stops talking, whereas the female speaker F starts talking (position Y). After some time, at t_2, while F is still speaking, the microphone in the receiving room moves from position A to position B (microphone 3Y, first array) and then, at time t_3, to position C (microphone 7X, second array), so that the acoustic path in the receiving rooms first becomes $h_{\mathrm{B}}(t)$, and then $h_{\mathrm{C}}(t)$. Finally, at time t_4, the female voice stops, and the male speaker resumes talking, with the microphone in the receiving room still at position C. An external noise source is present in the receiving room, which produces a WN signal $v(t)$, assumed uncorrelated to the speech signal $u(t)$ and distributed in space (i.e. not localized), with $= 30\,\mathrm{dB}$ SNR.

The aim of AEC is to identify the acoustic path in the receiving room so as to allow the removal of the echo from the microphone signal $y(n)$. For this purpose, the loudspeaker signal $u(n)$ is processed by a filter $\hat{H}(z,n)$, either an FIR or an OBF adaptive filter, whose linear coefficients are adapted using NLMS, thus producing the signal $\hat{y}(t)$. The poles of the OBF adaptive filters are first identified using the SB-OBF-GMP algorithm from the male speech sequence [291] with voice activity detection (VAD) activated and using the configuration described in Section 5.4.2 (only one sound source, maximum radius in the grid $\rho = 0.975$, and stage duration $T_s = 1\,\mathrm{s}$). A set of 30 pole pairs is identified ($M = 60$) and then fixed in the OBF echo canceler, counting $m = {}^{M}/_{2}$ resonator sections having TFs $\Psi(z, \boldsymbol{p}_m^{\mathcal{A}}) = \{\Psi_1^{\pm}(z, \tilde{\boldsymbol{p}}_1^{\mathcal{A}}), \ldots, \Psi_m^{\pm}(z, \tilde{\boldsymbol{p}}_m^{\mathcal{A}})\}$. Two cases are considered for the order of the FIR echo canceler, $M_F = 60$ and $M_F = 80$, with the latter determined such that the steady-state NM at the training positions is roughly equal to the one obtained by the OBF filters in the identification (i.e. -17 dB). The step size is set to $\mu = 0.002$ for both the OBF and the FIR echo canceler, giving comparable convergence rate for both.

One drawback of using OBF filters in AEC consists in the necessity of estimating the acoustic delay parameter d (cfr. Figure 5.1). Different methods are available in the literature, from single-channel algorithms based on (generalized) cross-correlation or on adaptive filters, to more sophisticated multi-channel algorithms (see [292] for an overview). For simplicity, the acoustic delays of the acoustic paths used in the following examples (selected from the SMARD database) are assumed to be known and applied to both OBF and FIR filters.

In addition to the difficulties described in Section 5.3.2 about identifying the room acoustic system from speech signals, such as non-stationarity and the high-pass characteristics of speech, an extra difficulty is represented here by the presence of reverberation in the loudspeaker signal $u(n)$, which has a negative impact on the accuracy of the identification of the poles, compared to the case of an 'anechoic' loudspeaker signal, as can be seen in Figure 5.14. Reverberation in

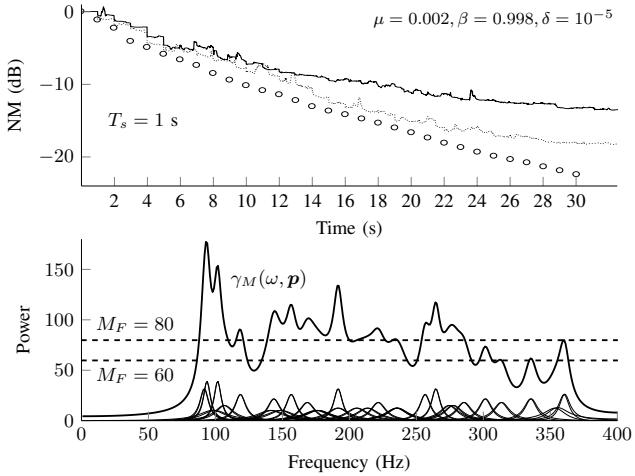

Figure 5.14: Top: the NM for the OBF-GMP algorithm (○) and for the SB-OBF-GMP algorithm with 'anechoic' (dotted) and reverberated (solid) speech input signal (EBU-SQAM) in the AEC scenario. Bottom: The power responses of the 30 pairs of basis functions generated from the estimated pole set $\boldsymbol{p}$, and the resulting $\gamma_M(\omega, \boldsymbol{p})$ and $\gamma_{M_F}(\omega) = M_F$.

the transmission room introduces additional coloration in the loudspeaker signal, thus increasing the condition number C of the signal correlation matrix, which leads to slower convergence of the parameters (see Section 5.2.2). Moreover, the spectral properties of the RTF of the transmission room have an influence on the excitation characteristics of the loudspeaker signal, so that an RTF in the receiving room may not be fully identifiable. In the case considered, the magnitude response of the Small simulated room (see top-left plot of Figure 5.10, $T = 0.15$ s) presents small energy below 100 Hz and above 370 Hz, plus a number of deep anti-resonances. It follows that the system is not excited enough in those regions, thus making the identification more difficult. This is confirmed by the lack of OBFs in some parts of the spectrum, as seen in the bottom plot of Figure 5.14 showing the power responses of the OBFs built from the estimated poles $\boldsymbol{p}$ and the resulting shape of the $\gamma_M(\omega, \boldsymbol{p})$ factor in (5.12) compared to the two constant factors $\gamma_{M_F}(\omega) = M_F$ considered for the FIR echo cancelers. Nevertheless, as long as the characteristics of the speech signal are similar in terms of excitation to those of the speech signal used in the pole identification, the set of poles estimated in the training phase allows to achieve good cancellation performances, as shown in the following.

Apart from the NM in (5.31), a common measure to evaluate the effectiveness of

an acoustic echo canceler is the echo return loss enhancement (ERLE), defined as the ratio of the power of the microphone signal in the receiving room and the power of the residual signal after cancellation,

$$\mathrm{ERLE}(n) = 10\log_{10}\left(\frac{S_y(n)}{S_e(n)}\right)\mathrm{dB}, \tag{5.54}$$

where estimates of $S_y(n)$ and $S_e(n)$ of the two signals are obtained by low-pass filtering their instantaneous power. Figure 5.15 shows the NM and the ERLE obtained in the scenario described above using the standard NLMS for adapting the coefficients of both types of canceler. It can be seen that, when $M_F = M = 60$, the FIR canceler initially converges quickly to low values of NM, but then it starts diverging with the passing of time. This is a consequence of the order M_F of the FIR canceler being much shorter than the effective length of the RIR. Indeed, FIR filters using standard adaptive algorithms provide a biased estimate with a large variance when the system is undermodeled [160] (i.e. for RIRs truncated to a small number of samples), so that the unmodeled part of the RIR (i.e. its 'tail') contributes to an increase in the NM [161] and to a reduction of the ERLE. Increasing the order of the FIR canceler to $M_F = 80$ reduces this effect and a performance comparable to that of the OBF canceler with $M = 60$ is achieved. The OBF canceler, due to its IIR, is indeed less prone to misalignment problems related to a low filter order.

To test the robustness to undermodeling in the AEC case, the order of both cancelers is reduced to $M = M_F = 40$ (for the OBF filter only the first 20 estimated pole pairs are used). The results are shown in Figure 5.16, showing on one hand the aggravated misalignment problem for the FIR canceler, and, on the other hand, the well-behaved dynamic properties and the good misalignment performance of the OBF canceler even at a lower order. Also, it is worth noticing the difference in the ERLE compared to the previous case. It follows that OBF filters represent a more robust choice than FIR filters in AEC, especially when the number of adaptive coefficients is required to be low.

Moreover, as discussed in Section 5.2, OBF filters can be regarded as a generalization of FIR filters, so that most of the algorithms developed for the latter can be modified to work with the former as well (in most of the cases just substituting the expression of the gradient vector, as seen in Section 5.2.2). The problem of slow convergence, normally originating from the poor excitation properties of the speech signals, can be addressed in different ways, such as using VSS algorithms (see [285, 286, 287] for an overview and examples), regularization techniques [185], or more complex adaptation algorithms. As an example, the scenario above is considered to evaluate the use of the APA[284] with projection order $Q = 8$ ($\mu = 0.005$, $\delta = 10^{-4}$). It is shown in Figure 5.17 how the increase in complexity improves the speed of convergence for both

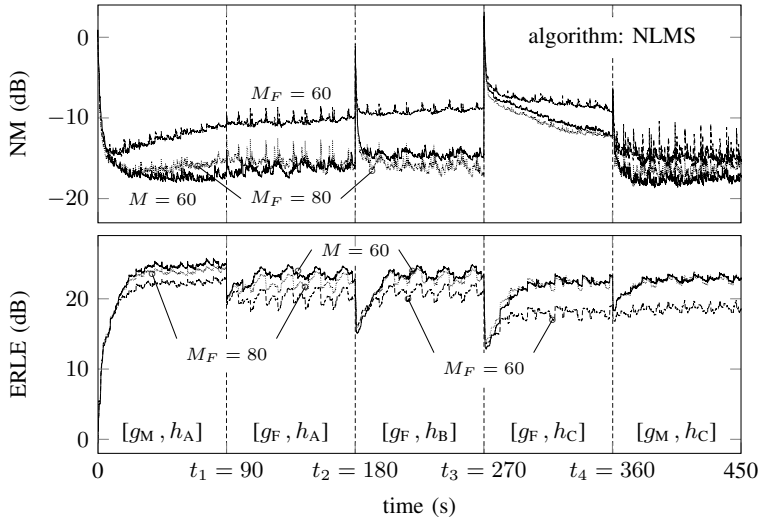

Figure 5.15: The NM (top) and the ERLE (bottom) for the AEC scenario, using an OBF filter of order M and FIR filters of order M_F.

cancelers compared to when NLMS is used (see Figure 5.15), but all the other considerations still apply.

As a final note, the behavior of the echo cancellation algorithms when abrupt changes occur in the scenario is analyzed. First, notice at t_1, when the male speaker stops and the female speaker starts talking in the transmission room, that the degradation in both NM and ERLE is limited, with the coefficients of the canceler quickly adapting to the varied excitation characteristics of the speech signal. Modified conditions in the receiving room, instead, degrade the NM and the ERLE more dramatically, especially in the case (at t_3) when the receiver position moves to the second array. This is mostly due to the fact that, being at low frequencies, the RTF at position C presents more differences compared to position A and B, which belong to the same array. Nevertheless, the common poles estimated on the first array are still valid at position C, as confirmed by the low NM in the last period, when the male speaker resumes talking.

It follows that a monophonic acoustic echo canceler needs to reconverge especially when large changes in the receiving room occur, whereas less problems are encountered when the speaker and/or its location varies in the transmission room. This is not true in SAEC [236, 161], where two microphones are located in the transmission room and two loudspeakers in the receiving room, in which case

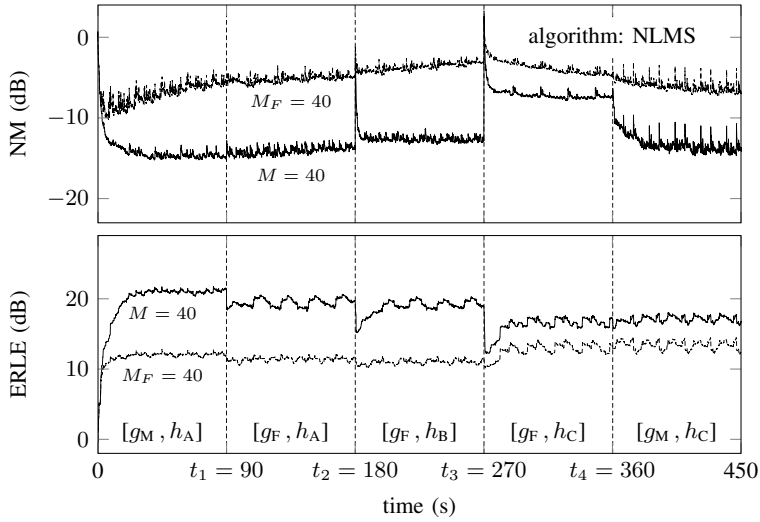

Figure 5.16: The NM (top) and the ERLE (bottom) for the AEC scenario, using an OBF filter and an FIR filter, both of order $M = M_F = 40$.

other issues have to be considered. The fact that the two loudspeaker signals $u_1(t)$ and $u_2(t)$ originate from a common source (the speech signal), convolved with two RIRs g_{M_1} and g_{M_2} in the transmission room, results in a high degree of their cross-correlation. In turn, this produces a very ill-conditioned covariance matrix, which leads to high misalignment and slow convergence [161, 293]. The misalignment can be reduced by using higher orders of the echo canceler, at the expense of slower convergence rate and higher ill-conditioning. Conversely, ill-conditioning is reduced if the order of the cancelers is much lower than the effective length of the RIRs in the transmission room, at the expense of a higher misalignment.

The IIR nature of OBF filters could help in achieving lower misalignment with a lower number of adaptive coefficients, thus meeting the second requirement more easily. However, it is often the case that using an IIR echo canceler does not solve the ill-conditioning problem to an acceptable degree. Moreover, the robustness to coloration of OBF filters and their good numerical conditioning properties discussed in Section 5.2 only refer to the monophonic case, whereas the ill-conditioning still remains in the stereophonic case. Thus, it is necessary to introduce a preprocessing stage with the aim of reducing the cross-correlation of the stereo signals. Common approaches achieve partial decorrelation by introducing psychoacoustically-masked noise [236], by applying a non-linear function, such as the one proposed in [161], or by preprocessing the input

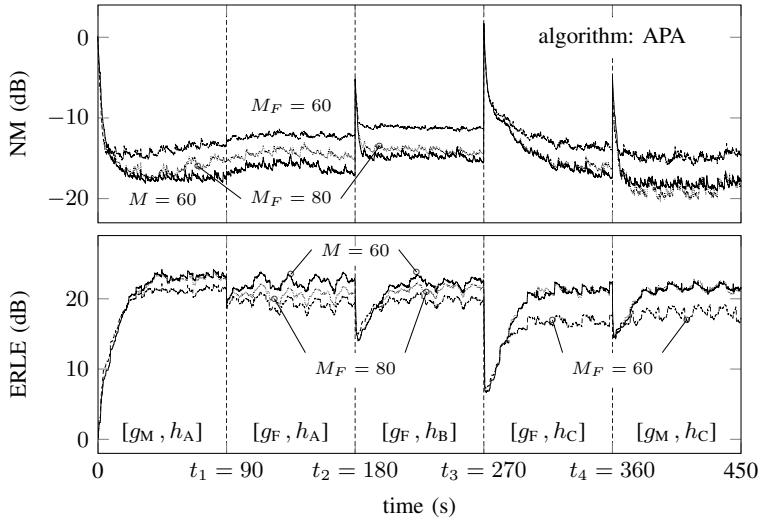

Figure 5.17: The NM (top) and the ERLE (bottom) for the AEC scenario, using an OBF filter of order M and FIR filters of order M_F, both using the APA.

signals with a pair of time-varying all-pass filters [237]. This last approach, with parameters as proposed in [294] for low frequency subbands (chosen in such a way not to introduce perceptible speech degradation), has proven to be an effective decorrelation method for the low-frequency case considered here. When the two loudspeaker signals are successfully decorrelated, poles of a multi-channel OBF adaptive filter can be estimated similarly to the cases presented above, and the echo cancellation performance is comparable to the monophonic AEC.

5.5.2 Room response equalization (RRE)

The other RASE application considered is the equalization of RTFs (see [9] for an overview). In general terms, the aim of RRE is to correct the RTF for deviations from a desired target response, so as to improve the quality of the sound reproduced in the room or of the signal captured by a microphone. Digital filters are used to modify the frequency spectrum of the source signal before it is sent to the loudspeaker or after it is captured by a microphone, such that the spectrum of the equalized microphone signal is as close as possible to the source signal spectrum.

The use of OBF filters for minimum and nonminimum-phase equalization of room responses was proposed in [28]. It was suggested that the possibility of fixing the poles of the equalizer allows to obtain a desired frequency resolution and potentially reduce the number of filter parameters necessary for a given degree of equalization. The advantage of fixing the poles is the ability to control the resolution of the equalizer not only in terms of frequency distribution (the angle of the poles), but also in terms of frequency selectivity (the radius of the poles), so that an equalizer can be designed, for instance, to correct for sharp resonances and notches in low frequencies (by having dense poles with large radius) and to correct for the overall envelope of the magnitude response at higher frequencies. Another possibility is to limit the angle of the poles, with the pole with the lowest angle determining the high-pass cut-off frequency of the equalized response. Different pole placement strategies were proposed in [28], from the simple distribution of poles on the unit disc based on a desired frequency resolution, to the poles obtained from modeling an estimate of the inverse RTF. When the equalizer is designed from the minimum-phase inverse of a measured single-point RTF, the equalizer can only compensate for the magnitude room response. Moreover, a measured RTF has to be available or at least identified in advance. In [277], a blind equalization method is presented in which the RTF is first estimated as an FIR filter using a method relying on higher-order statistics. Then, the poles of the equalizer are placed in correspondence with the zeros of the RTF which are closest to the unit circle or obtained by an FIR-to-IIR conversion of the equalizer impulse response.

The example presented in the following focuses on adaptive single-input/single-output (SISO) equalization. The two scenarios considered, depicted in Figure 5.18, are those in which the signal to be processed by the equalization filter is either the microphone signal $y(n)$ (top scheme) or, alternatively, the source signal $s(n)$ before it is sent to the loudspeaker (bottom scheme). The coefficients of the equalizer with TF $\hat{F}(z,n)$ are adapted based on the residual error signal $\varepsilon(n)$ between the equalized microphone signal $\tilde{y}(n)$ and, assuming a flat target response, the delayed source signal. The equalization problem can be formalized as

$$\text{minimize } \mathbb{E}\{\varepsilon^2(n)\} = \mathbb{E}\{\big(s(n-\Delta) - \tilde{y}(n)\big)^2\}, \tag{5.55}$$

where Δ is a modeling delay and $\tilde{y}(n) = H(q,n)\hat{F}(q,n)s(n)$ is the equalized microphone signal, i.e. the source signal filtered by the equalizer and convolved with the RTF at time n, $H(z,n)$. In case the equalizer is an OBF filter with a predefined set of poles $\boldsymbol{p}$, having TF $\hat{F}(z,n)$ defined as in (5.4), the minimization is performed by adapting the linear coefficients $\hat{\boldsymbol{\theta}}_i(n)$ (see Figure 5.18).

Here we propose two direct equalization filter design methods (i.e. not requiring a priori inversion of a RTF) which directly estimate the poles of the equalizer.

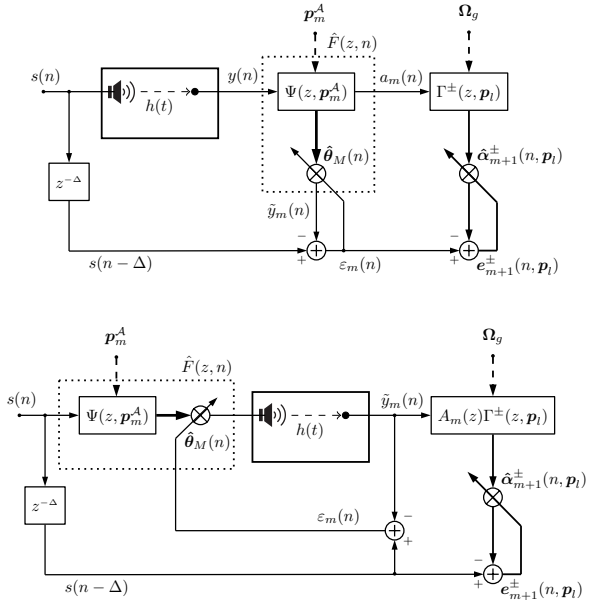

Figure 5.18: The simplified schematics of the off-line (top) and the on-line (bottom) methods for pole estimation of an OBF equalizer.

Both methods use a modified version of the SB-OBF-GMP algorithm. The first method is an off-line design procedure that aims at estimating the poles by post-equalizing the microphone signal $y(n)$. The setting is typical of equalization intended to dereverberate the microphone signal, in which case equalization is not performed in the room during the estimation. The SB-OBF-GMP pole estimation algorithm can be used in this case, with the microphone signal $y(n)$ entering the OBF filter instead of the loudspeaker signal (compare the top schematics of Figure 5.18 with Figure 5.4). The pole selection strategy is unmodified, given that in this case the coefficients $\hat{\alpha}_{m+1}^{\pm}(n,\boldsymbol{p}_l)$ represent an estimate of the correlation between the residual error signal $\varepsilon_m(n)$ produced by the current OBF equalizer with m resonator sections having TFs $\Psi(z,\boldsymbol{p}_m^{\mathcal{A}}) = \{\Psi_1^{\pm}(z,\tilde{\boldsymbol{p}}_1^{\mathcal{A}}),\ldots,\Psi_m^{\pm}(z,\tilde{\boldsymbol{p}}_m^{\mathcal{A}})\}$ and the output of the series of m all-pass filters processed by the candidate OBFs $\Gamma_i^{\pm}(z,\boldsymbol{p}_l)$ in the dictionary. If desired, once the poles are estimated, the equalizer can be moved in front of the loudspeaker, in a pre-equalization setup.

The second method is instead an on-line design procedure in which the equalizer built from the set of active (already selected) poles is placed in front of the loudspeaker, thus performing the actual equalization task in the room already

during the pole estimation, while the pole selection is still performed at the microphones side (see right schematics of Figure 5.18). However, in this case, the regression vector required for the adaptation of the linear coefficients of the OBF filter (which would correspond to the intermediate signals convolved with the RIR) is not available, as they cannot be retrieved from the microphone signal. An option, to be verified, would be to compute the regression vector using an estimate of the RIR obtained at the previous iteration of the algorithm. Another drawback is related to the first stages of the algorithm, when the OBF equalizer has only few poles. In this case, the equalizer response being a linear combination of sparse resonant responses, the microphone signal, instead of being equalized, will present a resonant response, with a resulting degradation of the reproduced sound quality. A more useful application of the on-line method then consists in starting with a pre-equalizer built from a predefined set of poles (either chosen arbitrarily or pre-estimated with the off-line method), with the possibility of adding new poles while equalization is performed in the room.

Apart from the possibility of determining the order of the equalizer (by interrupting the algorithm when the desired equalization is achieved), the advantage of estimating the poles compared to distributing them in a fixed configuration consists in having frequency resolution determined by the optimization process and not by an initial choice. This is useful in particular when the order of the equalization filter is constrained to a small value (not too small, as said above), so that at each stage the one pole pair (among those available in the grid) for which the equalization error reduces the most is selected. When the number of poles increases, the algorithm tends to distribute them more evenly in the whole frequency range, so that estimating the poles may not bring a significant advantage over a fixed configuration.

As an example, equalization results are shown in the low frequency range of a RTF taken from the SMARD database (microphone 7Y, first array) [50]. The grid counts $L = 2000$ poles with 400 different angles distributed uniformly between 80 Hz and 399 Hz. The radius of the poles decreases exponentially at the increase of the angle [28, 113] according to $\rho_i = \varrho^{\frac{\vartheta_i}{\pi}}$, where 5 values for ϱ (the radius defined at the Nyquist frequency) were distributed logarithmically between 0.7 and 0.925, such that pole density toward the unit circle is increased (see shaded area in top-right plot of Figure 5.19). In this way, the equalizer will be able to correct for sharper deviations at lower frequencies and provide a smoother correction at higher frequencies. The equalization results of the OBF equalizer with $m = 20$ pole pairs estimated with the off-line method ($T_s = 0.5$ s) are compared with a fixed-pole OBF equalizer with $m = 20$ pole pairs distributed in the same way, but only for $\varrho = 0.85$ (bottom-right plot of Figure 5.19). The modeling delay was set to 35 samples and both equalizers are adapted with NLMS ($\mu = 0.002$, $\delta = 10^{-5}$).

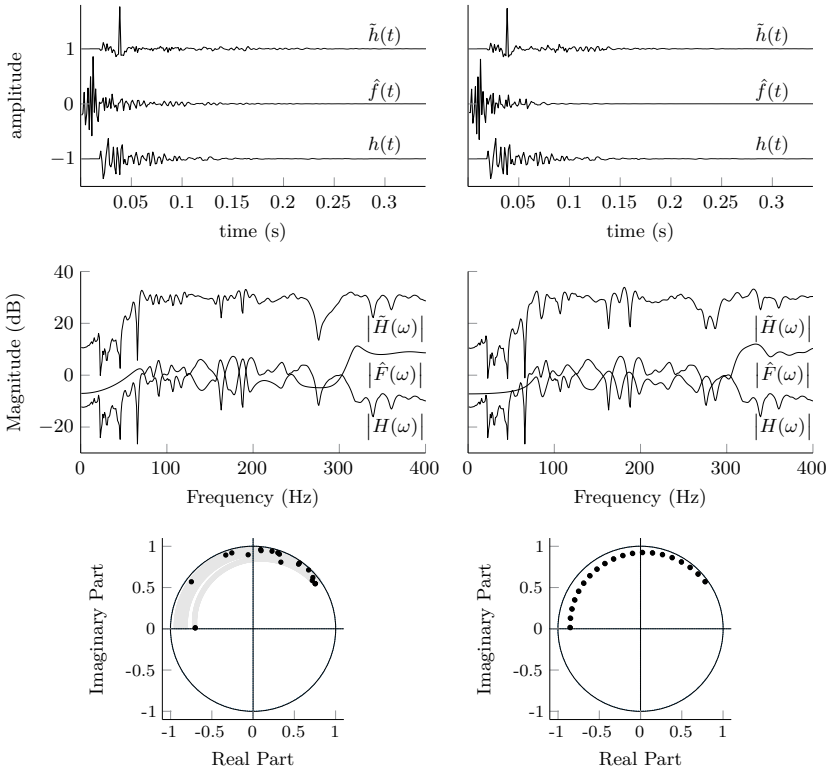

Figure 5.19: Equalization example results using poles estimated with the proposed off-line method (left column) and using a fixed configuration of 20 pole pairs (right column). The plots show (top) the RIR $h(t)$, the equalizer impulse response $\hat{f}(t)$, the equalized RIR $\tilde{h}(t)$, and (center) their respective magnitude frequency responses $|H(\omega)|$, $|\hat{F}(\omega)|$, and $|\tilde{H}(\omega)|$ (the latter shifted by 20 dB).

From Figure 5.19 it can be seen that the resolution below 200 Hz is higher for the equalizer with estimated poles, with sharper peaks and dips in the equalizer magnitude response $|\hat{F}(\omega)|$. The result is a flatter equalized response $|\tilde{H}(\omega)|$, also with reduced depth of the anti-resonances. At higher frequencies, only 2 poles are estimated, against 9 in the fixed configuration, which explains the lower resolution in the former above 250 Hz. Also notice in both cases that the response is not equalized below 80 Hz, corresponding to the lowest frequency allowed for the poles. Concerning the time domain, the presence of some estimated poles with large radius translates to a longer response of the equalization filter $\hat{f}(t)$ and a different distribution of the error in the equalized

response $\tilde{h}(t)$ after the main pulse; for the case with estimated poles, this post-ringing is longer but with lower amplitude compared to the case with predefined poles. Also notice in both cases the presence of a significant pre-ringing effect, due to the low order of the equalizer and the short modeling delay. Both pre-ringing and post-ringing can be reduced by employing higher filter orders and longer modeling delays.

The extension to adaptive multi-point (SIMO) equalization [208], is straightforward for poles already fixed in the equalizer. The estimation of the poles using the two methods suggested is still possible, but presents limitations, mainly of a complexity nature. In the off-line configuration, poles could be estimated for each acoustic channel independently, but this would imply running the full algorithm in parallel for each microphone signal. For the on-line setup, a set of poles common to all channels could be computed, but still each microphone signal would need to be processed by the all-pass filter $A_m(z)$ and by the $2L$ candidate OBFs $\Gamma^{\pm}(z, \boldsymbol{p}_l)$. Moreover, the idea of having an equalizer with common poles would be motivated by the debatable assumption that not only resonances, but also anti-resonances are a characteristics of the room, independent of the source and receiver positions.

5.6 Discussion

The main idea of representing a RTF by means of a pole-zero model is that a better approximation can be obtained by reducing the distance between the model poles and the true poles of the room acoustic system. From a different viewpoint, employing an IIR filter may help reducing the number of filter coefficients compared to the sampled truncated RIR representation inherent to modeling with FIR filters. The applicability of IIR filters in place of FIR filters in RASE applications has always been a matter of debate. A number of works [240, 239] concluded that the advantages are normally out-weighted by the increased difficulties encountered in the estimation and adaptation of the filter parameters, such as convergence to local minima or instability. It has been discussed in the first part of this chapter that IIR adaptive filters based on OBFs make the problem of instability easy to keep under control (the information about the pole radius being readily available) and, by virtue of orthogonality, provide a well-conditioned estimation problem under a wide range of conditions. As a result, OBF adaptive filters show the fastest convergence among fixed-poles IIR filters. Moreover, OBF adaptive filters can be seen as a generalization of FIR filters, so that, when poles are fixed, most filter adaptation algorithms developed for FIR filters can be easily applied to the OBF case as well.

When approaching a RASE problem, the first thing to be considered is whether OBF adaptive filters can bring an advantage compared to the common FIR filters. Since stability and convergence are not an issue, the choice should be made with respect to different criteria, based on the specific application scenario. First, the characteristics of the room play an important role in defining the expected performance in the identification of the room acoustic system using OBF filters. To summarize the findings from Section 5.4, an OBF filter can provide a significant advantage when it is able to approximate the RTF with a small number of filter parameters, i.e. when the RTF is characterized by either a small number of sparse resonances (such as at low frequency in small rooms) or a significant modal overlap (such as in rooms with high absorption). As the volume of the room increases, and with that the RT, the number of relatively sharp modes increases as well, and consequently the number of spectral peaks that have to be modeled. These results were observed in the low-frequency band considered (0-400 Hz), for which the modal density can be high, but possibly not high enough to be able to approximate the superposition of multiple resonances with a small number of poles. Moreover, real rooms present non-homogeneous characteristics at different frequency regions, as noticed in the SMARD database example. Thus, simulation results presented should be used just as an indication, as the performance in real scenarios may differ due to a number of different reasons.

Second, the potentially expected reduction in the filter order and the use of a common set of poles may not only bring computational savings, but also help in overcome some of the problems encountered in RASE applications, such as undermodeling, as suggested in the AEC example. It has been shown that common poles estimated at low frequencies in a limited area of the room provide a good approximation also at locations in the surrounding area, but less so in other parts of the room. It follows that the estimation of the poles might have to be performed in a wider area, with ensuing issues pertaining to the spatial sampling of the microphone setup. This may not be too much of an issue at higher frequency, where the higher modal overlap reduces the variability of the modal distribution in space, as suggested above. Third, the possibility of fixing the poles in the filter allows to achieve a desired frequency resolution, either by fixed configurations of the poles or by estimation algorithms, as suggested in the RRE example. This is a feature not encountered in FIR filters, unless warping techniques are used.

However, the use of OBF filters is beneficial only provided that the poles are well estimated, or at least fixed in a meaningful way according to some prior knowledge or some desired properties of the filter. The estimation of the poles is then possibly the most critical issue for a wide-spread adoption of OBF filters in RASE applications. Whereas stable and effective modeling algorithms

are already available, also for common poles (see [113] and Section 5.1), the identification from input-output signals is more challenging. The identification algorithm proposed in this work proved to be effective in estimating a common set of poles for SIMO and MIMO systems, also from speech signals. Nonetheless, the identification is a time and resource-consuming task, not only with the proposed method, but also when recursive algorithms are applied.

It follows that in some situations, it may not be possible to identify the poles from input-output signals during the actual RASE task. In these cases, one option is to estimate common poles in advance, and then keep them fixed during the task. Another possibility is to first obtain a solution for the problem at hand using already available methods for FIR filters, and then convert, with one of the modeling methods cited, the obtained filter to an OBF filter with reduced order, whose linear coefficients can be adapted to track RTF variations.

Future work should then address the identification problem. Gradient-based recursive algorithms, as discussed in Section 5.2.4, should be investigated further. The proposed algorithm could be used to obtain good initial values for the pole parameters, so that only few iterations of the recursive algorithm should suffice. This would probably overcome the limitations imposed by the discrete resolution of the pole grid and attain improved performance, at least in some situations (e.g. when modes are sharp and only moderately overlapping). The actual improvement achievable by refining the pole parameters should be verified, also taking into account the necessity of adapting the pole parameters to track possible changes in the room acoustic system.

Modeling and identification at higher frequency bands should be also assessed. It can be expected at higher frequencies that the increased modal density and the increased absorption would result in favorable conditions for the applicability of OBF filters. Subband modeling seems to be a promising direction [32], but more in-depth analysis is required to understand if the potential advantages over FIR filters are consistent enough to justify the additional processing.

5.7 Conclusion

In this chapter, the use of OBF adaptive filters in room acoustic system identification and RASE applications has been considered. Since some of the problems typical of IIR filters are not encountered in the OBF case, the choice between an OBF and an FIR filter should be made based on application requirements and the characteristics of the application scenario. One aim of this chapter was to provide the reader with the knowledge and some of the necessary tools to make an informed decision in this regard.

The main properties of OBF adaptive filters have been reviewed, with a focus on the error performance and dynamic behavior of filter adaptation algorithms; in this context, a modified version of the NLMS algorithm (analogous to the adaptation rule in TD algorithms and named OBF-NLMS) has been suggested to deal with issues at very low model orders. An identification algorithm has been proposed, capable of estimating a set of common poles for a MIMO room acoustic system, from both WN and speech signals. The algorithm has been used to identify the RTFs at low frequencies in different scenarios for real and simulated rooms, highlighting the relation between the characteristics of the room, such as its volume and it reverberation time, and the expected performance of OBF and FIR filters. Although the analysis has been performed at low frequencies, which is already of interest in applications such as RRE or SAEC, the methods and algorithms presented are applicable at higher frequency as well, especially if a subband approach is adopted.

Finally, examples in the context of two RASE applications were given. In the AEC example, it has been shown that OBF filters can provide good identification and cancellation performances already for small model orders, for which FIR filters may encounter undermodeling problems. In the RRE example, two methods were presented to directly estimate the poles of an equalizer in a SISO scenario, useful to allocate frequency resolution where it is more necessary, and to keep the order of the equalizer as low as possible. To conclude, OBF adaptive filters represent a useful and flexible tool for approaching most of the problems encountered in RASE applications, whose adoption should be considered based on each specific case, and whose possibilities may still have to be explored.

Part IV

Equalization

Chapter 6

Loudspeaker and room equalization with IIR parametric filters

An automatic design procedure for low-order IIR parametric equalizers

Giacomo Vairetti, Enzo De Sena, Michael Catrysse, Søren Holdt Jensen, Marc Moonen, and Toon van Waterschoot

Submitted for publication to *J. Audio Eng. Soc.*, Apr. 2018.

The candidate's contributions as first author include: literature study, co-development of the presented algorithms, software implementation and computer simulations, co-design of the evaluation experiments, co-formulation of the conclusions, text redaction and editing.

Abstract

Parametric equalization of an acoustic system aims to compensate for the deviations of its response from a desired target response using parametric digital filters. An optimization procedure is presented for the automatic design of a low-order equalizer using parametric infinite impulse response (IIR) filters, specifically second-order peaking filters and first-order shelving filters. The proposed procedure minimizes the sum of square errors (SSE) between the system and the target complex frequency responses, instead of the commonly used difference in magnitudes, and exploits a previously unexplored orthogonality property of one particular type of parametric filter. This brings a series of advantages over the state-of-the-art procedures, such as an improved mathematical tractability of the equalization problem, with the possibility of computing analytical expressions for the gradients, an improved initialization of the parameters, including the global gain of the equalizer, the incorporation of shelving filters in the optimization procedure, and a more accentuated focus on the equalization of the more perceptually relevant frequency peaks. Examples of loudspeaker and room equalization are provided, as well as a note about extending the procedure to multi-point equalization and transfer function modeling.

6.1 Introduction

Parametric equalization of an acoustic system aims to compensate for the deviations of its response from a target response using parametric digital filters. The general purpose is to improve the perceived audio quality by correcting for linear distortions introduced by the system [295, 7, 8, 296]. Linear distortions, usually perceived as spectral coloration (i.e. timbre modifications) [297, 298], are related to changes in the magnitude and phase of the complex frequency response with respect to a target response. Even though phase distortions are perceivable in some conditions [49], their effect is usually small compared to large variations in the magnitude of the frequency response [299]. Consequently, a low-order equalizer should focus on correcting the magnitude response of the system, rather than its phase response.

Parametric equalizers using cascaded infinite impulse response (IIR) filter sections consisting of peaking and shelving filters are commonly used [300, 301, 302, 214], especially when a low-order equalizer is required. Indeed, the possibility of adjusting gain, central frequency and bandwidth of each section of the equalizer results in a greater flexibility and, if the values of the parameters are well-chosen, in a reduced number of equalizer parameters w.r.t. ,

for instance, a graphic equalizer with fixed central frequencies and bandwidths, or a finite impulse response (FIR) filter. However, since manually adjusting the values of the control parameters, as often done, can be difficult or may lead to unsatisfactory results, the availability of automatic design procedures is beneficial.

For a parametric equalizer design procedure to be fully automatic, various relevant aspects should be considered, such as the number of filter sections available, typically fixed between 3 and 30 based on the application, and the structure of the filter sections, which can have different characteristics and be parametrized in different ways, especially in terms of the bandwidth parameter [300]. Other design choices pertain the definition of a target response, based on a prototype or defined by the user, and its 0-dB line, relative to which the global gain of the equalizer will be set, as well as preprocessing operations, such as smoothing of the system frequency response. Once all these aspects are determined, an automatic design procedure requires the definition of an optimization criterion (or cost function), typically in terms of a distance between the equalized system magnitude response and the target magnitude response, as well as the choice of an optimization algorithm for the estimation of the parameter values of the filter sections. The focus of this chapter is on automatic parametric equalizer design procedures operating in a sequential way, optimizing one filter section at a time, starting with the one that reduces the cost function the most, i.e. in order of importance in the equalization [215, 216]. The idea is to select an initial filter section, to search for better parameter values by minimizing the cost function using an iterative optimization algorithm, and then move to the initialization and optimization of the next filter section.

The choice of the cost function has a fundamental role in determining the final performance of the design procedure. The characteristics of the first- and second-order peaking and shelving filters used in minimum-phase low-order parametric equalizers are well suited for the equalization of the magnitude response and have only a small influence on the phase response. As a consequence, the cost function generally chosen uses the difference between the magnitudes of the equalized response and the target response, discarding the phase response. The procedure described by Ramos et al. [215] uses a cost function which is the average absolute difference between the equalized magnitude response and the target magnitude response, computed on a logarithmic scale. More recently, Behrends et al. [216] proposed a series of modifications to the aforementioned procedure, including the evaluation of the cost function on a linear scale. Such a choice is meant to favor the equalization of frequency peaks, which are known to be more audible than dips [303]. This is a desirable feature, especially for low-order equalizers, which also limits the selection of filters producing a sharp boost in the response that may cause clipping in the audio system.

In the proposed procedure, the focus on equalizing peaks is even more prominent. The cost function employed uses the sum of squared errors (SSE) between the equalized and the target complex frequency responses. Minimizing the SSE does not explicitly aim at maximizing the 'flatness' of the equalized magnitude response, as for the procedures cited above, but rather at compensating for the deviations of the equalized response by putting more emphasis in the equalization of energetic frequency peaks over dips. Even though the use of the SSE may be a less intuitive way of defining the equalization problem, it brings some advantages over using the magnitude response error. Specifically, the SSE gives the possibility of computing analytical expressions for the gradients of the cost function w.r.t. the parameters of the filter sections, such that efficient line search optimization algorithms can be used, and of estimating the global gain of the equalizer (i.e. the 0-dB line). Moreover, if only the linear-in-the-gain (LIG) structure of the parametric filters [300, 301] is used, the gain parameters can be estimated in closed form using least squares (LS), thus enabling the use of a grid search procedure for the initialization of the other filter parameters, as well as the inclusion of first-order shelving filters in the optimization procedure. It follows that most of the design aspects to be considered are based on the minimization of the cost function and not on arbitrary choices or assumptions regarding the magnitude response to be equalized, as in the procedures in [215] and [216], briefly described in Section 6.2.

The present chapter is organized as follows: Section 6.2 gives an overview of the state-of-the-art procedures for automatic equalizer design using parametric IIR filters. Section 6.3 formalizes and discusses the equalization problem defined in terms of the SSE. In Section 6.4, LIG parametric IIR filters are described and the closed-form expression for the gain parameter is derived. The proposed automatic procedure for parameter estimation of a low-order parametric equalizer is detailed in Section 6.5. In Section 6.6, results of the equalization of a loudspeaker response are evaluated using different error-based objective measures [296], as well as objective measures of perceived audio quality [304, 298, 305]. In Section 6.7, application to room response equalization is also considered. The modification to the proposed procedure for multi-point equalization and transfer function modeling is briefly discussed in Section 6.8. Section 6.9 concludes the chapter.

Terminology

The following terms and conventions are defined and used throughout the chapter. The term *system response* $H_0(k)$ indicates the frequency response to be equalized, which could be either a loudspeaker response, a room response, or a joint loudspeaker-room response. The radial frequency index k refers to the

evaluation of the transfer function on the unit circle at the k^{th} radial frequency bin ω_k (k is short for $e^{j\omega_k/f_s}$, with f_s the sampling frequency). The *equalized response* $H_s(k)$ is defined as the system response filtered by the parametric equalizer having s filter sections. The term *parametric equalizer* refers to the cascade of S parametric filters, while the term *parametric filter* refers to either a *peaking* filter with filter order $m = 2$ or a *shelving* filter with filter order $m = 1$. A parametric filter has two possible implementation forms: a LIG form, typically used in the literature with a positive gain (in dB) to generate a *boost* in the filter response, and a nonlinear-in-the-gain (NLIG) form, typically used with a negative gain (in dB) to generate a *cut* in the filter response (see Section 6.4).

6.2 State-of-the-art procedures

The purpose of parametric equalization is to compensate for the deviations of the system frequency response $H_0(k)$ from a user-defined target frequency response $T(k)$ using a parametric equalizer of order M with overall response $F_M(k)$. In other words, the purpose is to filter $H_0(k)$ with the equalizer $F_M(k)$ in order to approximate the target response as closely as possible, based on the following error:

$$E_M(k) = W(k)\big\{H_0(k) \cdot F_M(k) - T(k)\big\}. \tag{6.1}$$

with $W(k)$ a weighting function used to give more or less importance to the error at certain frequencies.

Different cost functions are possible. In the procedure proposed by Ramos et al. [215], the mean absolute error between the magnitudes in dB of the equalized response and the target response, computed on a logarithmic frequency scale, was chosen to account for the 'double logarithmic behavior of the ear',

$$\epsilon_M^{\text{dB}} = \frac{20}{N}\sum_k \Big| W(k)\Big\{\log_{10}|H_0(k)\cdot F_M(k)| - \log_{10}|T(k)|\Big\}\Big|, \tag{6.2}$$

with N the number of frequencies included in the frequency range of interest. The system magnitude response $|H_0(k)|$, as commonly done in low-order parametric equalization, is smoothed by a certain fractional-octave factor (usually $1/8^{\text{th}}$ or $1/12^{\text{th}}$) in order to remove narrow peaks and dips that are less audible [303] and to facilitate the search for the optimal parameter values. For each filter section, the procedure in [215] uses a heuristic algorithm to optimize the parameters. The procedure was extended in [306] to include second-order shelving and high-pass (HP) and low-pass (LP) filters in the equalizer design.

The decision of including shelving filters has to be made by analyzing the error areas above and below the target magnitude response at the beginning and at the end of the frequency range of interest. A shelving (or HP/LP) filter is then included if the error area is larger than a predefined threshold, with the values of the filter parameters optimized using the same heuristic algorithm. Another extension proposed in [215] adds the possibility of reducing the order of the parametric equalizer by removing the peaking filters that are correcting for inaudible peaks and dips, according to psychoacoustic considerations [303].

In Behrends et al. [216], the higher perceptual relevance of spectral peaks is directly taken into account in the definition of the cost function by considering the error on a linear magnitude scale, instead of a logarithmic scale, i.e.

$$\epsilon_M^{\text{lin}} = \frac{1}{N} \sum_k \Big| W(k) \Big\{ |H_0(k) \cdot F_M(k)| - |T(k)| \Big\} \Big|. \tag{6.3}$$

While the cost function used in Eq. (6.2) equally weights the error produced by deviations of the equalized magnitude response above and below the target response, the evaluation of the cost function on a linear scale as in Eq. (6.3), gives more importance to the portions of the equalized magnitude response that lie above the target, thus favoring the removal of frequency peaks, rather than the boosting of the dips, In [216], Behrends et al. also suggest to employ a derivative-free algorithm, called the Rosenbrock method [307], which offers a gradient-like behavior, and thus faster convergence.

A critical aspect of the procedures by Ramos et al. [215] and Behrends et al. [216] is the selection of the initial values of the parameters of each new parametric filter. The selection is done by computing the areas of the magnitude response above and below the target, using either (6.2) or (6.3). The largest area becomes the one to be equalized, with the half-way point between the two zero-crossing points and the negation of its level (in dB) defining the central frequency and gain of the filter section, respectively, and the -3 dB points defining the bandwidth (or Q-value). This approach assumes that the system magnitude response is a combination of peaks and dips above and below the target magnitude response. The problem with such an assumption is that, in case of highly irregular system magnitude responses, the initial filter placement approach may provide initial values quite distant from a local minimizer. In this case, the reduction in the cost function provided by the initial filter may even be quite limited. Furthermore, the placement of the 0-dB line becomes an important aspect of the procedure, for which a clear solution was not provided.

An example system magnitude response, similar to an example in [216], is given in Figure 6.1, also showing the filter responses for the initial values computed with different procedures. Between 100 Hz and 16 kHz, there are

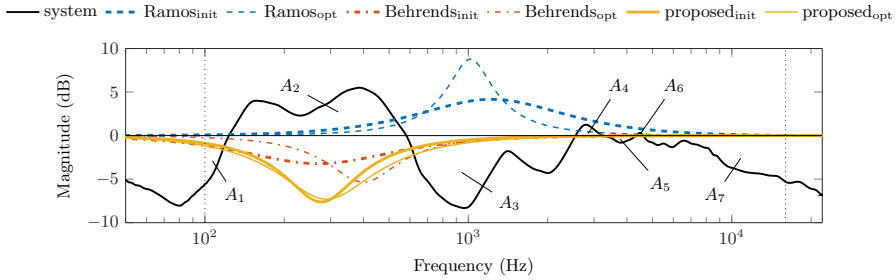

Figure 6.1: Initialized (thick lines) and optimized (thin lines) responses of a single filter section using different procedures.

seven error areas $A_1 - A_7$ above and below the predefined flat target magnitude response. The procedure by Ramos et al. [215] places the initial filter based on the largest error area computed according to (6.2), which is A_3 in the example; the parameters of the initial filter are chosen as described above; the irregularity of the system magnitude response makes the selection based on the half-way point of the area far from optimal, with the initial filter far from the optimal solution (also shown in the figure). The largest error area for the procedure in [216], computed according to (6.3), is instead A_2. As shown in the figure, using the same approach as in [215] leads to similar problems. A peak finding approach, as also suggested in [216], may provide a better initialization in this particular example, but it may not be effective in general and introduces the problem of defining the initial value for the bandwidth. The initial filter obtained with the proposed procedure is also shown in the figure. The initialization, which will be described in Section 6.5, is not based on the largest error area approach, but on a grid search with optimal gain (in LS sense) computed w.r.t. the SSE. It can be seen that initial parameters are found quite close to the optimal ones.

Other examples of automatic parametric equalizer design can be found in [308], where nonlinear optimization is used to find the parameters of a parametric equalizer starting from initial values selected using peak finding; in [309], where the gains of a parametric equalizer with fixed frequencies and bandwidths are estimated in closed form exploiting a self-similarity property of the peaking filters on a logarithmic scale; and in [310], where a gradient-based optimization of the parameters of an equalizer is proposed, which uses filters parametrized in terms of the numerator and denominator coefficients of the transfer function and not a constrained form defined in terms of gain, central frequency and bandwidth, as the one considered in this chapter.

6.3 Equalization based on the SSE

In this chapter, a cost function is used based on the SSE between the frequency responses, i.e.

$$\epsilon_M^{\text{SSE}} = \frac{1}{N}\sum_k \big(W(k)\big[H_0(k)\cdot F_M(k) - T(k)\big]\big)^2. \tag{6.4}$$

Such formulation, even though less intuitive than (6.2) and (6.3), brings some advantages, as will be detailed later on: (i) it provides an improved mathematical tractability of the equalization problem, with the possibility of computing analytical expressions for the gradients w.r.t. the filter parameters; (ii) when the parametric filter is in the LIG implementation form, it leads to a closed-form expression for the gain parameters (see Section 6.4), which simplifies the automatic design procedure; (iii) it provides a better way to initialize a parametric filter prior to optimization; (iv) it allows to include first-order shelving filters, and (v) to estimate the global constant gain in closed-form; and (vi) it focuses on the equalization of the more perceptually relevant frequency peaks rather than the dips.

The parametric equalizer considered, comprising a cascade of minimum-phase parametric filters, has a minimum-phase response. An interesting property of a minimum-phase response is that its frequency response $H(\omega)$ is completely determined by its magnitude response. The phase $\phi_H(\omega)$ is, indeed, given by the inverse Hilbert transform $\mathcal{H}^{-1}\{\cdot\} = -\mathcal{H}\{\cdot\}$ of the natural logarithm of the magnitude [45, 311]:

$$\begin{aligned} &H(\omega) = |H(\omega)|e^{j\phi_H(\omega)}, \\ &\text{with } \phi_H(\omega) = -\mathcal{H}\{\ln|H(\omega)|\}. \end{aligned} \tag{6.5}$$

This is a consequence of the fact that the log frequency response is an *analytic signal* in the frequency domain

$$\ln H(\omega) = \ln|H(\omega)| + j\phi_H(\omega), \tag{6.6}$$

whose time-domain counterpart is the so-called *cepstrum* [45]. In the digital domain, the phase response of the minimum-phase frequency response $H(k)$ can be obtained as the imaginary part $\mathcal{I}$ of the DFT of the folded real periodic cepstrum $\hat{h}(n) = \text{IDFT}\{\ln|H(k)|\}$

$$\phi_H(k) = \mathcal{I}\{\text{DFT}\{\text{fold}\{\hat{h}(n)\}\}\} \tag{6.7}$$

where the DFT and IDFT operators indicate the discrete Fourier transform and its inverse, and the *fold* operation has the effect of folding the anti-causal part

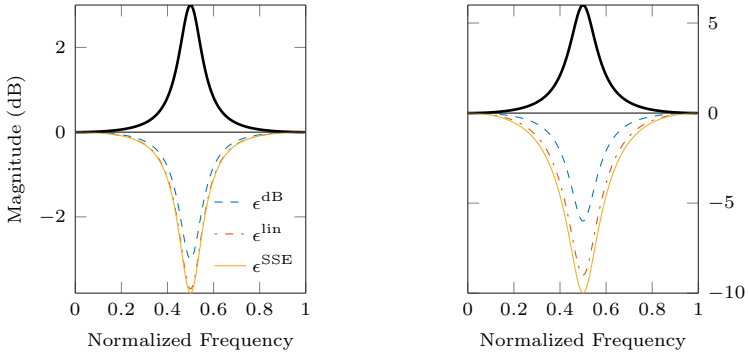

Figure 6.2: Two peaking filters with gains $G = 3\,\mathrm{dB}$ and $G = 6\,\mathrm{dB}$ (thick lines), and the corresponding cut filter responses (thin lines) with gain optimized to give equal error using different cost functions.

of $\hat{h}(n)$ onto its causal part. More details can be found in [312] or [67]. Thus, given the relation between the magnitude and the phase of a minimum-phase frequency response as given in (6.5), minimizing the cost function in (6.4), remarkably, still corresponds to a magnitude-only equalization.

The use of the SSE in (6.4) compared to the linear function in (6.3) puts more emphasis on the error generated by strong peaks, as described in more detail in Appendix B.1. Here an intuitive interpretation is given as follows. In Figure 6.2, the boost magnitude response of two peaking filters with positive gains $G = 3\,\mathrm{dB}$ and $G = 6\,\mathrm{dB}$ is considered. A cut in the filter magnitude response, having the same central frequency and bandwidth, is obtained using a negative gain. The negative gain parameter is optimized such that the error w.r.t. the 0-dB line computed with the cost functions in (6.2), (6.3) and (6.4), is equal to the one obtained for the boost response. For the cost function in (6.2), the cut filter response is obviously specular to the boost filter response on a logarithmic scale (the gain is $-G$), whereas for the cost functions in (6.3) and (6.4) it is not. This is the consequence of the fact that the evaluation of the error on a linear scale puts more weight on values above the 0-dB line. Whereas for the $G = 3\,\mathrm{dB}$ gain case (left plot) the cost function in (6.3) and (6.4) produce almost the same error, for higher gains (see right plot for $G = 6\,\mathrm{dB}$) the SSE gives more emphasis to errors above the 0-dB line.

6.4 Linear-in-the-gain parametric filters

Digital IIR filters used in parametric equalizers are first- and second-order IIR filters, with constraints on the filter magnitude response defined at the zero frequency, at the Nyquist frequency, and, for peaking filters, at the central frequency. Different parameterizations satisfying these constraints are possible. However, even though the various parameterizations have different definitions for the bandwidth parameter, all parameterizations satisfying the same constraints are equivalent [300].

Among different possibilities, the structure of first- and second-order parametric filters originally proposed by Regalia and Mitra [301] is chosen here. This structure, shown in Figure 6.3, comprises an all-pass (AP) filter $A_m(z)$ of order m and a feed-forward path. If the AP filter is independent from the gain parameter V, the parametric filter has a transfer function $F_m(z)$ which is linear in the gain V,

$$F_m(z) = \frac{1}{2}[(1+V) + (1-V)A_m(z)] \tag{6.8}$$

$$= \frac{1}{2}[(1 + A_m(z)) + V(1 - A_m(z))], \tag{6.9}$$

where expression (6.9), corresponding to the equivalent filter structure in Figure 6.3b, highlights this linear dependency [302, 214]. Given that for $V > 0$ the filter response is minimum-phase, whereas for $V < 0$ it is maximum-phase [301], only filters with positive linear gain will be considered.

Another characteristic of this filter structure, which is exploited in the proposed procedure, follows from the energy preservation property [313] of the AP filter: since the energy of the output signal of the AP filter is equal to the energy of its input signal, the signals $y^{\eta}(n) = x(n) + z(n)$, corresponding to a notch, and $y^{\beta}(n) = x(n) - z(n)$, corresponding to a resonance, are found to be orthogonal to each other. An intuitive proof is provided in Appendix B.2. If follows that, when the gain parameter V does not appear in the AP filter transfer function, the gain V is only acting on the resonant response $y^{\beta}(n)$, whereas the notch response $y^{\eta}(n)$ is not changed when V is modified. This can be seen in Figure 6.4, showing the magnitude response of two shelving filters (left) and two peaking (right) filters in LIG form with gains $V = 2$ and $V = 0.5$, together with the corresponding notch and resonance responses. If should be noticed that the LIG filter structure is able to produce both a boost and a cut in the response, even though the cut response tend to have a reduced bandwidth [301], as discussed below.

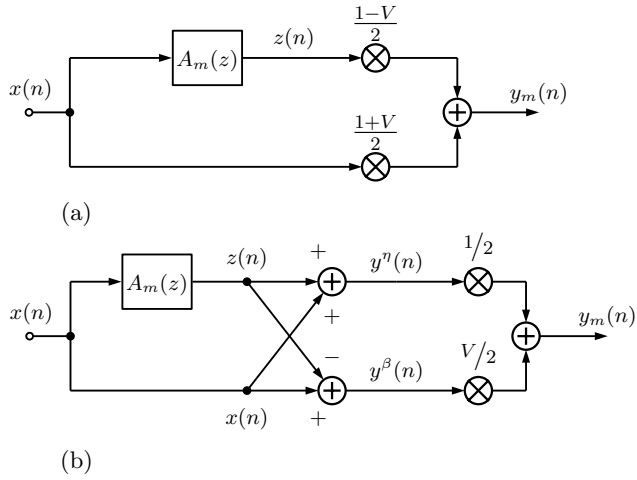

Figure 6.3: The Regalia-Mitra parametric filter

6.4.1 First-order shelving filters

A shelving filter is used whenever the lowest or highest portion of the system frequency response has to be enhanced or reduced. Shelving filters are described by a set of two parameters, namely the gain V and the transition frequency f_c, defined as the -3 dB notch bandwidth. By using the filter structure in (6.8) or (6.9), a first-order shelving filter at low frequencies (LFs) or at high frequencies (HFs), respectively, is obtained by defining a first-order AP filter as

$$A_1^{\mathrm{LF}}(z) = \frac{a_{\mathrm{LF}} - z^{-1}}{1 - a_{\mathrm{LF}} z^{-1}}, \quad A_1^{\mathrm{HF}}(z) = \frac{a_{\mathrm{HF}} + z^{-1}}{1 + a_{\mathrm{HF}} z^{-1}}\,. \tag{6.10}$$

The LIG form is obtained by defining the parameter a in terms of the transition frequency f_c and the sampling frequency f_s as

$$a_{\mathrm{LF}}^b = \frac{1 - \tan(\pi f_c/f_s)}{1 + \tan(\pi f_c/f_s)}, \quad a_{\mathrm{HF}}^b = \frac{\tan(\pi f_c/f_s) - 1}{\tan(\pi f_c/f_s) + 1}. \tag{6.11}$$

As a consequence, the AP filter does not depend on the gain V. However, for $0 < V < 1$, when the filter represents a cut, the effective transition frequency of the filter response tends towards lower (or higher for the HF case) frequencies (see left plot of Figure 6.4 or [301]). To obtain a cut response, for $0 < V < 1$, with response specular to the one obtained with the LIG form when V is replaced by $1/V$, the parameter a has to be modified to be dependent on the gain [214],

$$a_{\mathrm{LF}}^c = \frac{V - \tan(\pi f_c/f_s)}{V + \tan(\pi f_c/f_s)}, \quad a_{\mathrm{HF}}^c = \frac{\tan(\pi f_c/f_s) - V}{\tan(\pi f_c/f_s) + V} \tag{6.12}$$

which yields the NLIG form of a shelving filter.

Another option would be to redefine the parameter a in order to obtain a single expression that provides specular responses for a boost with gain V and a cut with gain $1/V$ [314, 300, 295]. However, the resulting filter structure of the *proportional* shelving filter is nonlinear in the gain parameter.

Finally, it should be noticed, also from the left plot of Figure 6.4, that the notch response $y^{\eta}(n)$ of the LF shelving filter corresponds to a first-order HP filter (i.e. when $V = 0$). The same is true also for the notch response of the HF shelving filter, which corresponds to a first-order LP filter.

6.4.2 Second-order peaking filters

Peaking filters are used to compensate for peaks or dips in the system magnitude response. As for first-order shelving filters, second-order peaking filters can be implemented with the filter structure in (6.8) by defining a second-order AP filter as

$$A_2(z) = \frac{a + d(1+a)z^{-1} + z^{-2}}{1 + d(1+a)z^{-1} + az^{-2}}, \tag{6.13}$$

with $d = -\cos(2\pi f_0/fs)$, where f_0 is the central frequency of the peaking filter. The LIG form is obtained by defining the bandwidth parameter a as

$$a^b = -\frac{\tan(\pi f_b/f_s) - 1}{\tan(\pi f_b/f_s) + 1}, \tag{6.14}$$

with f_b defined as the -3 dB notch bandwidth obtained for $V = 0$ [301, 300]. Similar to first-order shelving filters, peaking filters do not show a specular response when replacing V by $1/V$ (see right plot of Figure 6.4 or [301]). In order to obtain symmetric boost and cut responses, either the NLIG form [214] for $0 < V < 1$, with

$$a^c = -\frac{\tan(\pi f_b/f_s) - V}{\tan(\pi f_b/f_s) + V}, \tag{6.15}$$

or the proportional filters in [314, 300, 295] could be used. In both cases, the linear dependency w.r.t the gain parameter is lost. Only the LIG form is used in the proposed automatic equalization procedure. It is possible in any case to convert the parameters of a filter, either shelving or peaking, from the LIG form to the NLIG or the proportional form.

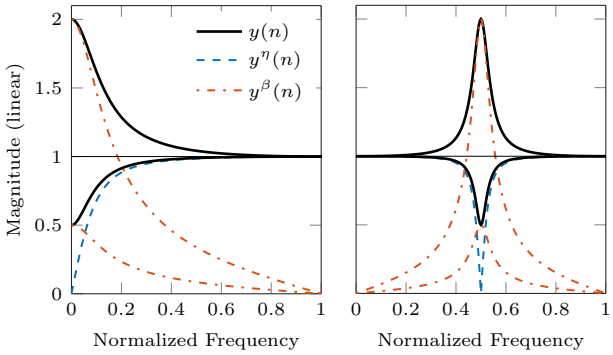

Figure 6.4: Shelving and peaking filters in LIG form

6.4.3 LS solution for the gain parameter

The advantage of the LIG form is that the linearity and orthogonality properties described above enable a closed-form solution for the estimation problem of the gain parameter. When the equalizer is made of only one parametric filter, the cost function in (6.4) can be written as

$$\epsilon_m^{\mathrm{SSE}} = \frac{1}{N}\sum_k \big(W(k)\{\frac{1}{2}H_0(k)[F_m^{\eta}(k) + VF_m^{\beta}(k)] - T(k)\}\big)^2, \tag{6.16}$$

where $F_m^{\eta}(k) = 1 + A_m(k)$ and $F_m^{\beta}(k) = 1 - A_m(k)$, respectively, and $k = 1, \ldots, N$. The minimization of the cost function is performed by setting to zero the first-order partial derivative of $\epsilon_M^{\mathrm{SSE}}$ w.r.t. V. The LS solution is obtained by

$$\hat{V} = \frac{\sum_k |W(k)|^2 F_m^{\beta *}(k) H_0^*(k) T(k)}{\sum_k |W(k)|^2 |H_0(k)|^2 |F_m^{\beta}(k)|^2} \tag{6.17}$$

with $\{\cdot\}^*$ indicating complex conjugation, which is independent from $F_m^{\eta}(k)$ because of the orthogonality between $F_m^{\eta}(k)$ and $F_m^{\beta}(k)$ (see details in Appendices B.2 and B.3). This feature will be also used in the parameter initialization, as described in Section 6.5. Indeed, if the equalizer is designed one parametric filter at a time, the optimal value $\hat{V}_s$ of the gain parameter of the s^{th} filter section, is obtained by substituting the system frequency response $H_0(z)$ in (6.17) with the equalized response $H_{s-1}(z)$.

6.5 Proposed design procedure

The aim of the proposed procedure is to design a parametric equalizer of order M as a cascade of S filter sections, each consisting of a parametric filter of order $m_s = 1$ (shelving) or $m_s = 2$ (peaking) having frequency response $F_{m_s}(k)$ defined as in (6.8-6.9), i.e.

$$F_M(k) = C \prod_{s=1}^{S} F_{m_s}(k), \text{ with } M = \sum_{s=1}^{S} m_s, \tag{6.18}$$

where s indicates the filter section index and C a global gain. The parameter values of the s^{th} filter section are optimized so as to minimize the cost function $\mathcal{F}(a_s, d_s, V_s)$, defined as

$$\epsilon^{\text{SSE}}_{s,m_s} = \frac{1}{N} \sum_{k} \Big(W(k) \Big\{ H_{s-1}(k) F_{m_s}(k) - T(k) \Big\} \Big)^2, \tag{6.19}$$

with H_{s-1} the system response filtered by the equalizer comprising the previous $s-1$ filter sections.

The proposed design procedure consists of the steps depicted in Figure 6.5 and detailed in the rest of the section. A preliminary step is to define a target response $T(k)$ and a minimum-phase preprocessed version of the system response $H_0(k)$. Optionally, the value of the global gain C can be estimated in closed-form using LS. The design of each new filter section can be divided into two stages. The first stage provides initial parameter values by means of a grid search, in which the optimal gain parameter for predefined discrete values of the central frequency and bandwidth is estimated as described above. The second stage consists of a line search optimization, which is intended to iteratively refine the initial parameter values and reach a local minimum of the cost function.

6.5.1 Spectral preprocessing

The spectral preprocessing of the system frequency response follows the steps outlined in [312]: first, the system magnitude response $|H_0(k)|$ is smoothed according to the Bark frequency scale, in order to approximate the critical bands of the ear, using a moving-average (MA) filter with bandwidth increasing with frequency. Apart for frequencies below 500 Hz, at which the smoothing is performed over a fixed 100 Hz interval, the bandwidth of the filter is set to an interval equal to 20% of the frequency. The amount of smoothing can then be controlled by the length of the window of the MA filter; either fractional

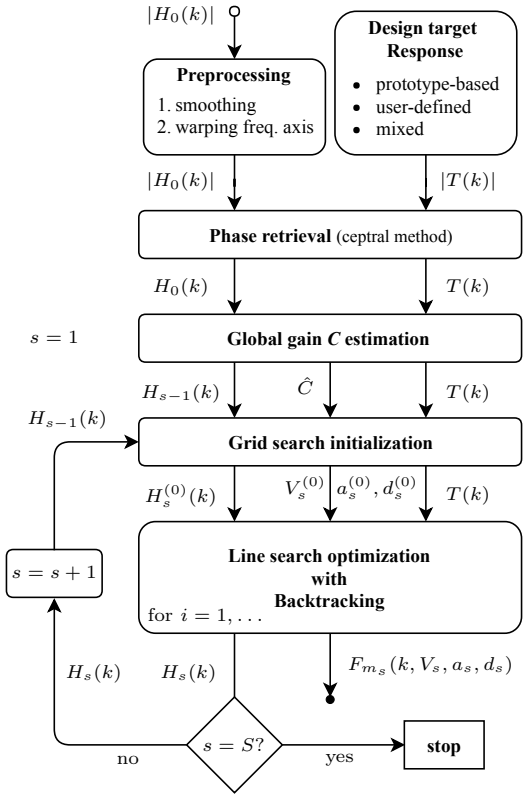

Figure 6.5: Schematics of the proposed design procedure.

critical bandwidth smoothing or fractional-octave smoothing can be easily used instead.

The second (and optional) step of the spectral preprocessing in [312] is to warp the frequency axis in order to approximate the Bark frequency scale, i.e. to allocate a higher resolution to the LFs. An alternative, also adopted in this chapter, is to resample the frequency axis from linear to logarithmic, by defining a logarithmically spaced axis, e.g. with $1/48^{\text{th}}$-octave resolution as in [215], and thus evaluating the magnitude response at those frequency points (e.g. using Horner's method [45], after the phase retrieval step explained below). Yet another way of favoring the equalization of a given frequency range, which can be used in conjunction with the strategies above, is to tune the weighting function $W(k)$ in (6.19) accordingly.

Finally, the cost function in (6.19) requires the minimum-phase response $H_0(k)$ to be retrieved from the preprocessed system magnitude response. A common solution, also suggested in [312], to create a minimum-phase frequency response is by means of the cepstral method [45, 312], where the smoothed (and/or warped) magnitude response is used to retrieve the corresponding phase response, as given in (6.5-6.7). Notice that, in order to avoid time-aliasing given by deep notches that can remain in the magnitude response after smoothing (e.g. towards 0 Hz), it is advisable to increase the FFT size to a high power of two and to clip the response, as suggested in [67].

6.5.2 Target response

Although the choice of the target response is arbitrary, it should be made cautiously. If the target response is too distant from the system frequency response, the equalization will be more difficult to be realized. For instance, if the lower cut-off frequency of the target response is below the lower cut-off frequency of the loudspeaker, the equalizer would contain a parametric filter with positive gain, which would move the loudspeaker driver outside its working range.

There is no complete agreement on the optimal target response for loudspeaker/room response equalization, and no single target for all sound reproduction purposes and all listeners can be defined [315]. It is out of the scope of this chapter to discuss the characteristics of an optimal, according to some criterion, target response for different sound reproduction systems and situations. Here only a brief overview of different approaches and guidelines is given. The target response can be defined in its magnitude and then its phase can be retrieved with the cepstral method.

Prototype-based: A prototype target magnitude response can be defined as, e.g., a band-pass filter transfer function or the magnitude response of a different loudspeaker. In this case, particular attention should be given to matching the cut-off frequencies of the system magnitude response and of the prototype target response, in order to avoid overloading of the loudspeaker driver. Another option is to use a strongly smoothed version of the system magnitude response, such as the one-octave smoothed response [308] or smoothing based on power averaged sound pressure [316], which eliminates peaks and dips, while preserving the coarse spectral envelope of the system response.

User-defined: A target magnitude response can be obtained as an interpolation of a set of points defined w.r.t. the system magnitude response [216]. In this way, it is easy to match the cut-off frequencies of the system magnitude response and to determine any desired characteristic of the response in the pass-band.

Mixed strategies: A combination of the two approaches can be used. For instance, the target magnitude response may be obtained by smoothing the system magnitude response in the LFs and in the HFs, whereas the response in the middle range may be defined by the user, e.g. a flat response or a boost at LFs.

6.5.3 Optimal global gain

Another aspect to consider is the optimization of the global gain C of the parametric equalizer, or, equivalently, the setting of the 0-dB line. Indeed, this has an influence on the characteristics of the filters selected by the design procedure. Centering a loudspeaker response around 0 dB would most likely avoid the selection of wide-band filters. However, in case of a room response, it is more difficult to determine the level at which the response should be centered, so that wide-band filters, with possibly high gains, are more likely to be selected, especially if the target response is not chosen carefully.

As described in Section 6.2, the placement of the 0-dB line is a critical aspect in the procedures proposed in [215] and [216]; the requirement for the system magnitude response to be centered around the 0-dB line of the target response in order to create error areas to be equalized is somewhat arbitrary. A possibility would be to place the 0-dB line by visual inspection or as the mean of the magnitude response of the system within a frequency range of interest (e.g. mid frequencies). This solution is not guaranteed to be an optimal one.

The use of the cost function based on the SSE, instead, allows the estimation of a global gain using LS, similarly to the estimation of the linear gain described in Section 6.4.3; by replacing the parametric equalizer $F_M(k)$ in (6.4) by a constant C, an estimate for the global gain $\hat{C}$ is given as

$$\hat{C} = \frac{\sum_k |W(k)|^2 \, H_0^*(k) T(k)}{\sum_k |W(k)|^2 |H_0(k)|^2} \tag{6.20}$$

This global gain C can be regarded as a scaling factor that centers the system response around the 0-dB line that minimizes the cost function in (6.4). Since the SSE puts more emphasis on the peaks (see Section 6.3), the system

magnitude response will tend to have dips that are more prominent than the peaks w.r.t. the target response. This may not be desirable, as the design procedure may favor the boost of spectral dips rather then the cut of spectral peaks. If desired, this may be avoided by adding an offset of a few dB to the global gain in order to restore the emphasis on the equalization of peaks over dips.

6.5.4 Grid search initialization and constraints

The initialization of the parameters of each new parametric filter in the cascade, as well as the selection of either a peaking or a shelving filter, is performed in an automatic way by means of a grid search using a discrete set of possible frequency and bandwidth values. A pole grid is defined, similarly to [113], where the radius and angle of complex poles determine respectively the bandwidth f_b and central frequency f_0 of the peaking filters. The radius of the real poles defines the transition frequencies f_c of LF (positive real poles) and HF (negative real poles) shelving filters. The gain for the filters built using each pole p in the grid is defined by LS estimation as described in Section 6.4.3, and the parameters of the filter that reduces the SSE the most are selected as initial parameter values of the current filter section. The gain can be limited based on hardware specifications, by defining a minimum (e.g. $V_{\min} = 0.25$) and a maximum value (e.g. $V_{\max} = 4$). Note that, being the system response minimum-phase, the gain V will always be positive [301].

Given the critical-band smoothing and the logarithmic resolution of the frequency axis, the angle $\sigma = 2\pi f_0/f_s$ of the complex poles, which define peaking filters, can be discretized according to a logarithmic or a Bark-scale distribution, with minimum and maximum angles defined, for instance, by the frequency limits of the equalization. The radius $\rho = \sqrt{a}$ of the complex poles $p = \rho e^{j\sigma}$ can be defined between a lower and an upper limit determined by the constraints imposed on the gain and bandwidth parameters for the different values of σ. It is common to define constraints in terms of the Q-factor, which provides an indication of the filter bandwidth relative to its central frequency [214]. The parameter a can be converted into the corresponding Q-factor in closed form, but the two cases of $V > 1$ and $V < 1$ must be addressed separately. Filters in the LIG form defined in terms of the parameters a and d (see Section 6.4.2), can be converted in the corresponding LIG *boost* form and NLIG *cut* forms defined in terms of Q and the auxiliary variable $K = \tan(\pi f_0/f_s)$ as in [214],

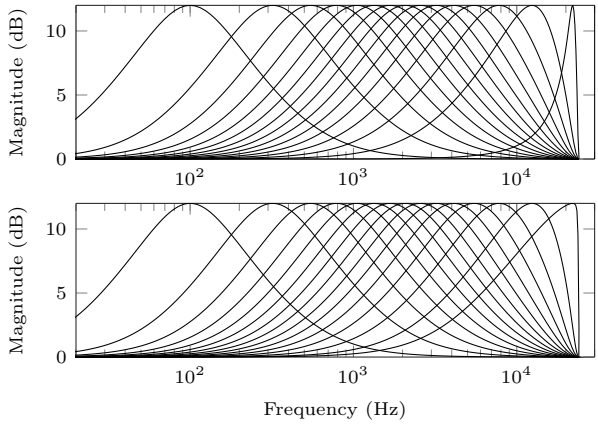

Figure 6.6: Magnitude response of constant-Q (top) and constant relative bandwidth (bottom) peaking filters.

respectively with

$$Q^b = \frac{\sin(2\pi f_0/f_s)}{2\tan(\pi f_b/f_s)} = \frac{\sin(\sigma)}{2\frac{1-a^b}{1+a^b}} \qquad \text{if } V > 1\,, \tag{6.21}$$

$$Q^c = \frac{\sin(2\pi f_0/f_s)}{2V\tan(\pi f_b/f_s)} = \frac{\sin(\sigma)}{2V\frac{1-a^b}{1+a^b}} \qquad \text{if } V < 1\,. \tag{6.22}$$

The Q-factor can be limited as well in order to avoid filters too narrow-band (e.g. $Q_{\max} = 10$) or too wide-band (e.g. $Q_{\min} = 0.5$).

However, for given fixed values of Q and V, the actual bandwidth (in octaves) of a peaking filter reduces for increasing frequencies and the filter response on a logarithmic scale becomes asymmetric when f_0 approaches $f_s/2$ (top plot of Figure 6.6). In order to keep the relative bandwidth approximately constant over the whole frequency range (bottom plot of Figure 6.6), the radius ρ of the complex poles is set to decrease exponentially with increasing angle σ, according to $\rho = R^{\frac{\sigma}{\pi}}$, with R the value of the radius defined at the Nyquist frequency [28]. The value for R can be computed to match the response of a filter defined in terms of a given Q [214] at a given angular frequency σ_q. The parameter a_q is

computed from (6.21-6.22) as

$$a_q^b = \frac{2Q^b - \sin(\sigma_q)}{2Q^b + \sin(\sigma_q)} \qquad \text{if } V > 1, \tag{6.23}$$

$$a_q^c = \frac{2VQ^c - \sin(\sigma_q)}{2VQ^c + \sin(\sigma_q)} \qquad \text{if } V < 1, \tag{6.24}$$

from which the corresponding $R = a_q^{\frac{\pi}{2\sigma_q}}$ is obtained. The limits for R are computed the same way inserting the constraints in (6.23-6.24). The minimum and maximum radius at the Nyquist frequency for $V > 1$ ($R^b_{\min}$, $R^b_{\max}$) are computed from (6.23) for $Q = Q^b_{\min}$ and $Q = Q^b_{\max}$, whereas for $V < 1$, $R^c_{\min}$ and $R^c_{\max}$ are computed from (6.24) for $Q = Q^c_{\min}$ and $Q = Q^c_{\max}$, with $V = V_{\min}$. This results in two partially overlapping allowed areas of the unit disc, one valid when $V > 1$ and the other when $V < 1$, , where generally $R^c_{\min} < R^b_{\min}$ and $R^c_{\max} < R^b_{\max}$.

In general, the bandwidth constraints for filter with $V > 1$ (Q^b) and filters with $V < 1$ (Q^c) can be chosen to be different, with the limitation dictated by the requirement of having a positive value for a (and thus ρ real). From (6.24) with $\sigma_q = \pi/2$, it is required that $Q^c_{\min} > 1/2V_{\min}$, thus trading-off between sharp cut filters with high gain and broader cut filters with limited gain. Also, it is required from (6.23) that $Q^b_{\min} > 0.5$ (which is anyway quite wide, approximately 2.5 octaves). Notice that for very large bandwidths, the filter responses tend to skew towards the Nyquist frequency, but less dramatically than for the filters with fixed Q (see Figure 6.6). A unique allowed area could be found by setting $Q^c_{\min} = Q^b_{\min}/V_{\min}$ and $Q^c_{\max} = Q^b_{\max}/V_{\min}$, but this would lead to filters with cut responses ($V < 1$) much narrower compared to boost responses ($V > 1$).

Regarding the values for R between $R^c_{\min}$ and $R^b_{\max}$, it is suggested in [113] to set the desired number of radii (for each angle) and distribute them logarithmically in order to increase density towards the unit circle (obtaining the so-called Bark-exp grid [113]) and thus to increase the resolution of narrow peaking filters. If the allowed areas do not coincide, the complex poles with smaller radius are valid only for $\hat{V} < 1$ (i.e. cut responses), whereas they would produce too wide boost responses for $\hat{V} > 1$. On the other hand, complex poles very close to the unit circle, valid for $\hat{V} > 1$, would produce too narrow cut responses for $\hat{V} < 1$. It is then necessary to check the constraints after the estimation of the optimal gains $\hat{V}$, and select the initial filter as the one that minimizes the cost function within the constraints. This can be done by checking that the parameter $a_s = \rho_s^2$ of the selected complex pole $p_s = \rho_s e^{j\sigma_s}$ satisfies $a^b_{\min} \leq a_s \leq a^b_{\max}$ or $a^c_{\min} \leq a_s \leq a^c_{\max}$, where $a^b_{\min}$ and $a^b_{\max}$ are computed from (6.23) for $Q = Q^b_{\min}$ and $Q = Q^b_{\max}$, and $a^c_{\min}$ and $a^c_{\max}$ from (6.24) for $Q = Q^c_{\min}$ and $Q = Q^c_{\max}$, with $V = V_{\min}$, where σ_q is replaced by σ_s.

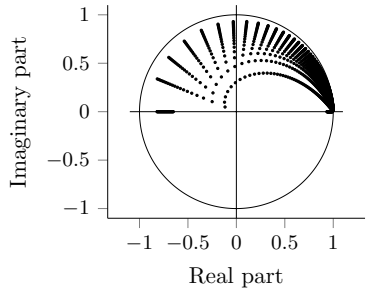

Figure 6.7: A Bark-exp pole grid for the grid-search.

Finally, the radius of the real poles, determining the transition frequency f_c of the shelving filters, may be set arbitrarily within the range of equalization. The effective transition frequency corresponding to ρ can be easily computed from (6.11) and (6.12), for $V > 1$ and $V < 1$, respectively. An upper and a lower limit for the radius of real poles can be imposed using (6.11) for $V > 1$ and using (6.12) with $V = V_{\min}$ for $V < 1$. It is also possible to include first-order HP/LP filters in the grid search by forcing the gain of the shelving filters to zero, effectively using only their notch responses, as mentioned in Section 6.4.

An example Bark-exp pole grid is shown in Figure 6.7, with $Q^c_{\min} = Q^b_{\min} = 0.75$ and $Q^c_{\max} = Q^b_{\max} = 10$, and with $V_{\max} = 1/V_{\min} = 4$, where a^b_q and a^c_q in (6.23) and (6.24) are evaluated at $\sigma_q = \pi/4$, giving a good balance between narrow and wide band filters. The central frequencies f_0 are distributed between 100 Hz and 21 kHz, with poles having 75 possible angles, and 20 possible radii. The cut-off frequencies f_c of the candidate shelving and HP/LP filters are linearly distributed between 100 Hz and 1 kHz, and between 18 kHz and 21 kHz.

6.5.5 Line search optimization

Once the pole $p_s = \rho_s e^{j\sigma_s}$ corresponding to the optimal parametric filter in the grid search is selected, the parameters $d_s^{(0)} = -\cos(\sigma_s)$ and $a_s^{(0)} = \rho_s^2$ are used as the initial conditions of a line search optimization [288], meant to refine their value and reduce the cost function $\mathcal{F}(a_s, d_s, V_s)$ further. In the optimization, $\sigma_s^{(0)}$ is used instead of $d_s^{(0)}$ to take into account its cosinusoidal nature, important in the computation of the search direction. The cost function in (6.19), indeed, allows the computation of the gradients w.r.t. to the filter parameters, thus enabling the use of gradient-based algorithms, such as steepest descent (SD), quasi-Newton or Gauss-Newton (GN) algorithms, which guarantee fast convergence to a local minimum, provided that the initial values are chosen

properly. The assumption that the initial filter parameters obtained with the grid search are sufficiently close to a local minimum is reasonable, as long as the density of the poles in the grid is sufficiently high. The same assumption is required also for the derivative-free algorithms in [215] and in [216], in order to guarantee convergence in a relatively small number of iterations, with the exception that in those cases the initial filter parameters are obtained by an indirect minimization of the cost function, without verifying if the initial values provide a good starting point for the equalization.

The parameter vector, initialized as $\boldsymbol{\theta}^{(0)} = [a_s^{(0)}, \sigma_s^{(0)}]^T$ for a complex pole (peaking filter), or $\boldsymbol{\theta}_s^{(0)} = a_s^{(0)}$ for a real pole (shelving filter), is updated at each iteration $i = 0, 1, 2, \ldots$ as

$$\boldsymbol{\theta}_s^{(i+1)} = \boldsymbol{\theta}_s^{(i)} + \mu^{(i)} \boldsymbol{p}^{(i)}, \tag{6.25}$$

where $\mu^{(i)}$ indicates the step size, and $\boldsymbol{p}^{(i)}$ the search direction along which the step is taken in order to reduce the cost function in (6.19), such that

$$\mathcal{F}(\boldsymbol{\theta}_s^{(i)} + \mu^{(i)} \boldsymbol{p}^{(i)}, V_s^{(i)}) < \mathcal{F}(\boldsymbol{\theta}_s^{(i)}, V_s^{(i)}), \tag{6.26}$$

where $V_s^{(0)}$ is the gain estimated in the grid search, which is updated by LS estimation at each evaluation of the cost function. In other words, the search direction $\boldsymbol{p}^{(i)}$ has to be a descent direction, i.e. $\boldsymbol{p}^{(i)T} \boldsymbol{\nabla} \mathcal{F}_s^{(i)} < 0$ with $\boldsymbol{\nabla} \mathcal{F}_s^{(i)} = \boldsymbol{\nabla} \mathcal{F}(\boldsymbol{\theta}_s^{(i)}, V_s^{(i)})$ the gradient of the cost function (i.e. the vector of its first-order partial derivatives) w.r.t. the parameters in $\boldsymbol{\theta}_s^{(i)}$,

$$\boldsymbol{\nabla} \mathcal{F}_s^{(i)} = {}^{\partial \mathcal{F}_s^{(i)}}/_{\partial \boldsymbol{\theta}_s^{(i)}} = [{}^{\partial \mathcal{F}_s^{(i)}}/_{\partial a_s^{(i)}}, {}^{\partial \mathcal{F}_s^{(i)}}/_{\partial \sigma_s^{(i)}}]^T, \tag{6.27}$$

with $\{\cdot\}^T$ indicating the vector transpose. The analytic expressions for the gradient are given in Appendix B.4.

The search direction generally has the form

$$\boldsymbol{p}^{(i)} = -\{\boldsymbol{B}^{(i)}\}^{-1} \boldsymbol{\nabla} \mathcal{F}_s^{(i)}, \tag{6.28}$$

where $\boldsymbol{B}^{(i)}$ is a symmetric and nonsingular matrix, whose form differentiates the different methods. When $\boldsymbol{B}^{(i)}$ is an identity matrix, $\boldsymbol{p}^{(i)}$ is the SD and (6.28) corresponds to the SD method. When $\boldsymbol{B}^{(i)}$ is the exact Hessian $\boldsymbol{\nabla}^2 \mathcal{F}_s^{(i)}$ (i.e. the matrix of second-order partial derivatives), (6.28) corresponds to the Newton method. The Hessian can be approximated at each iteration without the need for computing the second-order partial derivatives, leading to quasi-Newton methods, such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. The GN method, instead, computes the search direction by expressing the

derivatives of $\mathcal{F}_s^{(i)}$ in terms of the Jacobians $\boldsymbol{\nabla e}_s^{(i)}$, as

$$\begin{aligned}
\boldsymbol{p}^{(i)} &= -\Big(\boldsymbol{\nabla e}^{(i)H}\boldsymbol{\nabla e}^{(i)}\Big)^{-1}\boldsymbol{\nabla e}^{(i)H}\boldsymbol{e}^{(i)}, \quad \text{with} \\
\boldsymbol{\nabla e}^{(i)} &= {}^{\partial \boldsymbol{e}^{(i)}}\!/_{\partial \boldsymbol{\theta}_s^{(i)}} = \big[{}^{\partial \boldsymbol{e}^{(i)}}\!/_{\partial a_s^{(i)}}, {}^{\partial \boldsymbol{e}^{(i)}}\!/_{\partial \boldsymbol{\sigma}_s^{(i)}}\big]^T, \\
\boldsymbol{e}^{(i)} &= [e(1, \boldsymbol{\theta}_s^{(i)}, V_s^{(i)}), \ldots, e(N, \boldsymbol{\theta}_s^{(i)}, V_s^{(i)})]^T, \\
e(k, \boldsymbol{\theta}_s^{(i)}, V_s^{(i)}) &= W(k)\Big\{\frac{1}{2}H_{s-1}(k)F_{m_s}^{(i)}(k) - T(k)\Big\}, \\
F_{m_s}^{(i)}(k) &= F_{m_s}(k, \boldsymbol{\theta}_s^{(i)}, V_s^{(i)})
\end{aligned} \tag{6.29}$$

with $\{\cdot\}^H$ indicating Hermitian transpose, where the Jacobians are obtained as an intermediate step in the calculation of the gradients (see Appendix B.4). The GN method approximates the Hessian with $\boldsymbol{\nabla e}^{(i)H}\boldsymbol{\nabla e}^{(i)}$, thus having convergence rate similar to the Newton method, i.e. faster than the SD method.

The convergence rate of line search algorithms also depends on the choice of the step size $\mu^{(i)}$. In order to select a value of $\mu^{(i)}$ that achieves a significant reduction of $\mathcal{F}_s^{(i)}$ without the need to optimize for $\mu^{(i)}$, backtracking with the Armijo's sufficient decrease condition [288] is used. The backtracking strategy consists in starting with a large step size $\mu^{(i)} < 1$ ($\mu^{(i)} = 1$ for Newton and quasi-Newton methods) and iteratively reducing it by means of a contraction factor $\kappa \in (0, 1)$, such that $\mu^{(i)} \leftarrow \kappa\mu^{(i)}$. At each repetition of the backtracking, a sufficient decrease condition is evaluated to ensure that the algorithm gives reasonable descent along $\boldsymbol{p}^{(i)}$. The condition in (6.26) is however not sufficient to ensure convergence to a local minimum. A different condition is then required, such as the commonly used Armijo's sufficient decrease condition

$$\mathcal{F}(\boldsymbol{\theta}_s^{(i)} + \mu^{(i)}\boldsymbol{p}^{(i)}, V_s^{(i)}) \leq \gamma\mu^{(i)}\boldsymbol{p}^{(i)T}\boldsymbol{\nabla}\mathcal{F}(\boldsymbol{\theta}_s^{(i)}, V_s^{(i)}) \tag{6.30}$$

with $\gamma \in (0, 1)$, which states that a decrease in $\mathcal{F}_s^{(i)}$ is sufficient if proportional to both $\mu^{(i)}$ and $\boldsymbol{p}^{(i)T}\boldsymbol{\nabla}\mathcal{F}_s^{(i)}$. A final value for $\mu^{(i)}$ is obtained when the Armijo's condition is fulfilled, or when it becomes smaller than a predefined value $\mu_{\min}$. Also, the parameters in $\boldsymbol{\theta}_s^{(i)} + \mu^{(i)}\boldsymbol{p}^{(i)}$ should be checked to ensure that $a_s^{(i)}$ and $\sigma_s^{(i)}$ still satisfy the constraints described in the previous section. Stability is guaranteed by $a_{\max} < 1$.

The line search for the current stage terminates when $\boldsymbol{p}^{(i)T}\boldsymbol{\nabla}\mathcal{F}_s^{(i)} \leq \tau$, with τ a specified tolerance or when a maximum number of iterations I is reached. It should be mentioned that it is possible to include a closed-form expression of V in terms of a_s and d_s in the filter transfer function $F_{m_s}(k)$ in (6.19), at the

expense of more complicated analytic expressions for the gradients. Another alternative is to include the gain V in the vector of parameters $\boldsymbol{\theta}_i$ and perform the line-search without updating the gain parameter between two iterations. However, experimental results showed that the speed of convergence and the final result of these two alternatives are comparable to the results of the line-search algorithm described above.

6.6 Loudspeaker equalization example

In this section, an example of parametric equalization of a loudspeaker response is presented. The aim is to show the performance of the proposed procedure described above, in comparison to the state-of-the-art procedures presented in Section 6.2. In an attempt to keep the comparison as fair as possible, the same target response, the same range of equalization 100 Hz-21 kHz, and the same preprocessing (logarithmic frequency axis, Bark-scale smoothing, etc.) is used for the three procedures considered. The target response is built to match the pass-band characteristics of the loudspeaker response, using second-order high-pass and low-pass Butterworth filters with cut-off frequency of 250 Hz and 22 kHz, respectively. The loudspeaker response is scaled so that the 0-dB line of the target response corresponds to the response mean value between 400 Hz and 6 kHz, which satisfies the requirement of the state-of-the-art procedures of having peaks and dips to be equalized (see Figure 6.8). The same termination conditions are used for all procedures; the algorithm moves to the next filter section whenever either a maximum number of iterations (e.g. $I = 100$) is reached, or the step size gets smaller than a given value (e.g. $\mu_{\min} = 10^{-4}$), or the reduction in the cost function in a number of previous iterations (e.g. 10) is less than a predefined tolerance value (e.g. $\tau = 10^{-8}$). The Rosenbrock method [307] is applied for both the state-of-the-art procedures, using a step expansion factor $\alpha = 1.5$ and a step contraction factor $\zeta = 0.75$, starting from an initial variation of 0.5% of the value of the initial filter parameters (see [216]). In the procedure by Ramos et al. (**R**) [215], the Q-factor of the filter is initialized based on the bandwidth of the selected error area, while in the one by Behrends et al. (**B**) [216] it is set to $Q_0 = 2$.

The Bark-exp grid used in the proposed procedure (**P**) is the one in Figure 6.7. In the example, the GN algorithm is used in the line search, which provides very similar results as SD in a much smaller number of iterations. The initial step size is set to $\mu^{(i)} = 0.9$, the contraction factor for the backtracking to $\kappa = 0.8$, and the Armijo's condition constant to $\gamma = 0.05$. The global gain C is estimated as explained in Section 6.5.

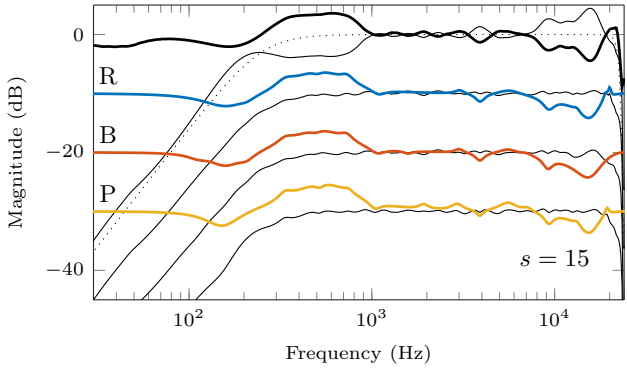

Figure 6.8: Loudspeaker equalization. Top: the unequalized response (solid) with the target response (dotted) and the ideal high-order FIR equalizer (thick); From top to bottom (10 dB offset): the equalized response (solid) and the corresponding equalizer (thick) using procedures R, B and P.

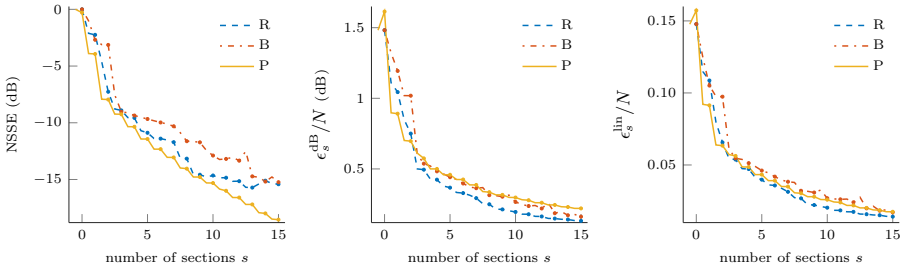

Figure 6.9: The error produced by the different procedures at each stage according to the different cost functions.

The error produced by the different procedures with increasing number of filter sections s is shown in Figure 6.9. As expected, the proposed procedure (P) performs best in minimizing the normalized SSE (NSSE), i.e. the error in (6.19) normalized w.r.t. the error in (6.4) computed without equalizer ($F_M(k) = 1$) and converted to decibels; the procedure by Ramos et al. (R), with cost function as in (6.2), outperforms the other procedures in minimizing the logarithmic error, whereas the procedure by Behrends et al. (B) fails to minimize the linear cost function in (6.3) more than the other procedures (at least in this example). Procedure P is the one that, for all cost functions considered, is able to achieve the largest error reduction in the first two stages. Also, the error for procedure P exhibits a staircase-like behavior, which is due to the vicinity of the initial

parameter values to a local minimum and the subsequent small improvement given by the line search. In general, the different procedures for an increasing number of stages are not too different from each other in terms of equalization performance, all capable of attaining the target response to a certain degree, as can be seen in Figure 6.8 for $s = 15$. A difference is found in the total number of iterations (n_i), with procedure P using the GN algorithm having an order of magnitude less than the other procedures, including the backtracking (see Table 6.1), due to the efficiency of both the initialization and the GN algorithm. However, the grid search and each iteration of the line search are computationally more demanding than the iterations of the Rosenbrock algorithm, eventually obtaining similar execution times for the different procedures.

Apart from the performance evaluation based on the different cost functions themselves, other measures are considered, namely the spectral flatness measure (SFM) and the spectral distance measure (SDM) described in [296]. The SFM is the ratio between the geometric mean and the arithmetic mean of the power spectrum (on a linear frequency scale). The target response not necessarily being flat in the range of equalization, the measure is computed using the power spectrum of the equalized system response divided by the target response $\tilde{H}_s(k) = {}^{H_s(k)}/_{T(k)}$

$$\mathrm{SFM}_s = N \frac{\sqrt[N]{\prod_k |\tilde{H}_s(k)|^2}}{\sum_k |\tilde{H}_s(k)|^2}, \tag{6.31}$$

so that the ideal high-order FIR equalizer defined as $D(k) = {}^{T(k)}/_{H_0(k)}$ has SFM=1. The SDM is also based on the power spectrum of the responses, and it is given by

$$\mathrm{SDM}_s = \sqrt{\sum_k \left| \frac{|\bar{H}_s(k)|^2 - |\bar{T}(k)|^2}{N} \right|^2}, \tag{6.32}$$

where in this case $\bar{H}_s(k)$ and $\bar{T}(k)$ are the loudspeaker and target responses resampled on a logarithmic frequency scale with ${}^1/_5$ octave resolution [296]. Results for these two measures are shown in Table 6.1 for the different procedures using equalizers with 5, 10 and 15 parametric filters. For both measures, the greatest improvement is achieved with 5 filters only, with smaller improvements for increasing number of filters. It is interesting to notice that, even though not specifically designed to maximize the SFM of the loudspeaker response as for procedures B and R, the proposed procedure (P) achieves a good level of flatness. Regarding the SDM, procedure P achieves a performance close to that of procedure R. Also notice that the optimization of the global gain (at $s = 0$) already contributes to a reduction of the SDM.

From a subjective point of view, it is commonly accepted that a flat (in the pass-band) frequency response is perceived as more natural, and that deviations

measure	procedure	$s = 0$	$s = 5$	$s = 10$	$s = 15$
n_i	R	0	232	552	824
	B	0	260	593	885
	P	0	29	53	70
SFM_s	R	0.922	**0.991**	**0.996**	**0.998**
	B	0.922	0.986	0.990	0.993
	P	0.922	0.983	0.991	0.995
SDM_s	R	0.435	**0.100**	**0.045**	**0.030**
	B	0.435	0.118	0.078	0.051
	P	0.375	0.101	0.064	0.035

Table 6.1: Error-based objective measures

from this are perceived as spectral coloration. Perceptual objective measures based on these assumptions are used here for speech and music stimuli. These are the average log-spectral difference measure (LSDM) [304], and a measure based on a linear distortion auditory model [305], referred to here as perceptual linear distortion measure (PLDM).

The LSDM is the square difference between the logarithm of the magnitude responses of a clean speech segment convolved with the target response $S_T(k)$ (reference) and the same speech segment convolved with the equalized loudspeaker response $S_{H_s}(k)$,

$$\mathrm{LSDM}_s = \sqrt{\frac{1}{N}\sum_k \big[\log(S_{H_s}(k)) - \log(S_T(k))\big]^2} \tag{6.33}$$

The average LSDM is then computed for those segments (in this case of 25 ms with 15 ms overlap) where active speech is detected. The PLDM is a measure of the perceived subjective naturalness of speech or music w.r.t. linear distortions, represented by spectral ripples and tilts in the magnitude response (see [305] for detailed information).

Results obtained using a male voice speech signal [291] and a music signal consisting of the instrumental introduction of a rock song [317] (having a wide-band spectrum) are shown in Table 6.2. For both measures considered, a strong improvement for the low-order equalizers ($s = 5$) is shown, with procedure R slightly better than procedure P, except for the LSDM using the music signal. For the LSDM, the use of higher-order equalizers improves the scores only slightly, whereas a more consistent improvement is still visible for the PLDM

measure	procedure	$s = 0$	$s = 5$	$s = 10$	$s = 15$
$\mathrm{LSDM}_{\mathrm{speech}}$	R	0.350	**0.144**	**0.118**	**0.099**
	B	0.350	0.158	0.143	0.135
	P	0.350	0.160	0.142	0.118
$\mathrm{LSDM}_{\mathrm{music}}$	R	0.308	0.165	0.143	0.127
	B	0.308	0.171	0.163	0.159
	P	0.296	**0.146**	**0.136**	**0.110**
$\mathrm{PLDM}_{\mathrm{speech}}$	R	0.785	**0.349**	**0.222**	0.178
	B	0.785	0.366	0.258	**0.174**
	P	0.785	0.399	0.259	0.217
$\mathrm{PLDM}_{\mathrm{music}}$	R	0.960	**0.380**	**0.267**	**0.224**
	B	0.960	0.414	0.314	0.257
	P	0.960	0.405	0.278	0.238

Table 6.2: Perception-based objective measures

with $s = 10$. Notice, however, that small differences in the score values will most likely not be perceived as a difference in sound quality.

6.7 Room equalization example

The proposed procedure can be applied to the equalization of the combined loudspeaker/room response without major modifications. Differently from loudspeaker equalization, the purpose of room transfer function (RTF) equalization is not only to obtain a more balanced response w.r.t. a target response, but also to compensate (as much as possible) for strong resonances at LFs. Thus, the smoothing should be less prominent, with fractional-octave smoothing (e.g. $1/6^{\mathrm{th}}$) preferred over Bark-scale smoothing, which has constant resolution below 500 Hz. Based on the amount of smoothing, which determines the level of detail in the spectral envelope of the response, the number of parametric filters required to attain the target response with a certain accuracy may vary.

The definition of the target response is a critical issue. RTFs have a more irregular frequency structure than loudspeaker responses, which cannot be easily recognized as deviations from a flat response. Moreover, spectral complexity combined with a less aggressive smoothing result in less smooth error surfaces produced by the cost functions, presenting a large number of local minima. In order to obtain a target response that produces peaks and dips in the RTF to

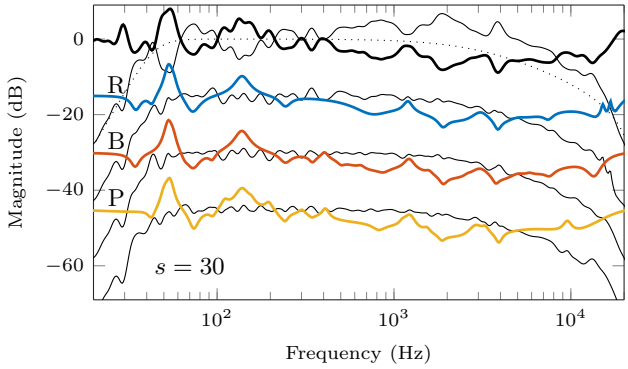

Figure 6.10: Room equalization. Top: the unequalized response (solid) with the target response (dotted) and the ideal high-order FIR equalizer (thick); From top to bottom (15 dB offset): the equalized response (solid) and the corresponding equalizer (thick) using procedures R, B and P.

be equalized, as required by the state-of-the-art procedures, one could use a strongly smoothed (e.g. one-octave resolution) version of the response as the target, which however may not provide a desired response. These procedures can be still used in general, but are more likely to incur into problems.

The proposed procedure is instead less sensitive to the selection of the target response, and will always start optimizing a new filter section from an initial point reasonably close to a useful local minimum, provided that the density of the poles in the grid is proportional to the amount of smoothing applied. In this example, the number of possible angles was increased to 300. This, however, makes the grid search at each stage more computationally demanding. If efficiency is an issue, an option is to start from a dense grid and effectively use only a subset of the grid points, different for every filter section (e.g., by taking one every $n = 4$ angles and shift over one angle for the following $n - 1$ filter sections). Simulation results show that this solution leads to very similar results to those obtained with the solution based on the full grid.

In the example shown here, the magnitude of a RTF measured in the parliament hall of the Provinciehuis Oost-Vlaanderen in Ghent, Belgium, having a reverberation time of 1.5 s, has been smoothed with 1/6 octave resolution and the equalization is evaluated in the range 30 Hz-18 kHz. The target response used is a combination of a forth-order high-pass Butterworth filter with cut-off frequency of 45 Hz and a first-order low-pass Butterworth filter with cut-off frequency of 3 kHz, which produces a slight roll-off at higher frequencies (see Figure 6.10). Table 6.3 provides results of the different error functions and

measure	procedure	$s = 0$	$s = 10$	$s = 20$	$s = 30$
n_i	R	0	642	1150	1584
	B	0	668	1353	2026
	P	0	231	595	792
NSSE_s	R	0	**-12.6**	-14.6	-14.6
	B	0	-9.4	-13.9	-16.5
	P	0	-11.9	**-15.7**	**-18.3**
$\epsilon_s^{\mathrm{dB}}/N$ (6.2)	R	3.91	**0.80**	**0.54**	0.54
	B	3.91	0.96	0.57	**0.42**
	P	3.91	0.94	0.65	0.51
$\epsilon_s^{\mathrm{lin}}/N$ (6.3)	R	0.402	**0.077**	0.051	0.051
	B	0.402	0.082	0.052	**0.037**
	P	0.402	**0.077**	**0.048**	**0.037**
SFM_s (6.31)	R	0.915	0.966	0.979	0.978
	B	0.915	**0.986**	**0.992**	**0.995**
	P	0.915	0.960	0.966	0.979
SDM_s (6.32)	R	1.146	0.176	0.106	0.104
	B	1.146	0.186	0.106	0.066
	P	1.146	**0.146**	**0.078**	**0.049**

Table 6.3: Error-based objective measures (RTF)

other error measures produced by the different procedures. Given the more complicated error function surfaces, the GN algorithm in the proposed procedure (P) now needs more iterations than in the previous example, but still two or three times fewer than the other two procedures. As in the loudspeaker example, all procedures are able to achieve equalization with a comparable accuracy (see also Figure 6.10 for $s = 30$), with the strongest improvement obtained in the first 10 stages. The proposed procedure achieves better results in terms of NSSE and of SDM, and slightly worse performances in the error measures related to the flatness of the equalized response. Also in this example, procedure B is slightly outperformed by the other procedures in its own cost function, but has better performance in terms of the SFM. It should also be noted that procedure R do not achieve any improvement with the inclusion of the last 10 filter sections ($s = 30$). In this particular case, the Rosenbrock method gets stuck into a local minimum and is unable to correct for the largest error area, which is then selected at each new filter initialization, thus failing to produce any further performance improvement.

6.8 Note on multi-point equalization and transfer function modeling

When the aim is to improve the response of a loudspeaker at multiple listening angles or the room response at multiple positions, a single equalizer can be designed based on a prototype response which contains the common acoustic features of the multiple responses. Averaging and smoothing the magnitude responses was proven to offer an effective solution [318], which also increases robustness to spatial variations. The proposed procedure can then be extended to multi-point equalization by including an averaging operation in the prepocessing step.

A very similar procedure to the one described for automatic design of a parametric equalizer can be applied to the problem of transfer function modeling. This idea can be useful, for instance, to model the ideal FIR equalizer $D(k) = {}^{T(k)}/_{H_0(k)}$ using a low order parametric filter, the response of a graphic equalizer [309], or more generally any minimum-phase transfer function. For the modeling problem, the cost function becomes

$$\epsilon_M^{\mathrm{SSE}} = \frac{1}{N} \sum_k \big(W(k)\big[D(k) - F_M(k)\big]\big)^2. \tag{6.34}$$

Also in this case, LIG parametric filters can be used, with the possibility of computing the gradients w.r.t. the other filter parameters.

6.9 Conclusion and future work

An automatic procedure for the design of a low-order parametric equalizer has been proposed, which uses a series of second-order peaking filters and first-order shelving filters. The proposed procedure minimizes the SSE between the system response and the target responses, instead of the commonly used difference in the magnitude responses, bringing some advantages, such as an improved mathematical tractability of the equalization problem, with the possibility of computing analytical expressions for the gradients w.r.t. the filter parameters and a closed-form solution for the estimation of the gain parameters, an improved parameter initialization, the inclusion of shelving filters in the optimization procedure, and a more accentuated focus on the equalization of frequency peaks over dips. Examples of loudspeaker and room response equalization have shown that effective equalization using a small number of parametric filters can be achieved. The proposed procedure can be extended

to multi-point equalization, by means of a prototype average response, and to transfer function modeling.

Acknowledgments

The authors would like to thank Televic N.V. for the use of their equipment and their premises, especially Vincent Soubry and Frederik Naessens for the useful collaboration in the early stage of this work. The authors would also like to thank Martin Møller (Bang & Olufsen) for providing the loudspeaker responses and Rainer Huber (Uni Oldenburg) for the fruitful discussions and the code for the PLDM. The scientific responsibility of this work is assumed by its authors.

Chapter 7

Multi-channel equalization of car cabin acoustics

Automatic calibration of car cabin acoustics in a multi-channel equalization framework

Giacomo Vairetti, Thomas Dietzen, David Pelegrin Garcia, Marc Moonen, and Toon van Waterschoot

Final report for the IWT (Agency for Innovation by Science and Technology) project *RAVENNA: Proof-of-concept of a Rationed Architecture for Vehicle Entertainment and NVH Next-generation Acoustics*, in collaboration with Premium Sound Solutions N.V.

The candidate's contributions as first author include: literature study, co-development of the presented modeling algorithms, software implementation and computer simulations, co-design of the evaluation experiments, co-formulation of the conclusions, text redaction and editing.

Abstract

This chapter is based on the final report for the IWT project RAVENNA: *Proof-of-concept of a Rationed Architecture for Vehicle Entertainment and NVH Next-generation Acoustics*. The report deals with the problem of equalization of an acoustic system, specifically an audio reproduction system inside a car cabin. The objective is the design of a multiple-input/multiple-output (MIMO) equalizer filter able to correct for minimum-phase as well as nonminimum-phase distortions of the acoustic paths of one primary speaker at different receivers within a given listening region, with the help of a number of support loudspeakers. The approach adopted falls into a polynomial-based control system framework, which allows the inclusion of other functionalities, such as equalization at higher frequencies, channel similarity using two or more primary speakers, multi-zone equalization, and listening enhancement in noisy conditions. In this chapter, the theoretical solution to the equalization problem is carefully analyzed and an interpretation is given. Methods for efficient transfer function modeling are presented, based on an existing method, which has been modified to reduce numerical problems through regularization, to estimate a single set of denominator coefficients common to all transfer functions, and to accurately estimate the all-pass excess-phase component of the transfer functions.

7.1 Introduction

The present chapter deals with the problem of equalization of an acoustic system, specifically an audio reproduction system inside a car cabin. In the single-input/single-output (SISO) case, the problem corresponds to the design of a single pre-filter intended to correct for the deviations of the system transfer function (TF) with respect to an ideal response. When several loudspeakers are used to equalize the response at one receiver position, an individual filter is designed for each loudspeaker independently. However, designing an equalizer based on the responses measured at only one receiver position will produce errors at other positions, due to the high spatial variability of the TFs.

A common solution to multipoint equalization is to use minimum-phase filters to compensate for the minimum-phase part of the acoustic impulse responses (AIRs) in a given listening region. In this case, common distortion components in the TF magnitude responses are compensated for. The minimum-phase compensating filter, however, is not sufficient to correct for all deviations of the time-domain AIRs, which are typically mixed-phase.

For this reason, in order to correct the responses at multiple receiver positions, mixed-phase filters should be used. The problem with mixed-phase filters is that they generally introduce errors in the equalized response, known as *pre-ringing* or pre-echo. It is then important to keep the level of this type of artifact under control so that it is inaudible when the equalizer is applied. In order to do so, only the nonminimum-phase distortions that are common to all receiver positions should be corrected [224].

Different approaches for the multiple-input/multiple-output (MIMO) equalization problem are possible, ignoring methods in which the position of the loudspeakers is optimized. A common approach is based on multi-channel filter design, which is based on the exact inversion of TF matrices either in the time domain or in the frequency domain (e.g. [219, 208, 220]). These methods, however, are not robust to TF variations, and require the number of receivers to be smaller than the number of loudspeakers.

The work presented here uses a different approach [225, 226]: the idea is to select a *primary* loudspeaker to be equalized, with a number of *support* loudspeakers intended to help the primary loudspeaker to attain the desired target response. In this chapter, a detailed analysis and implementation of the method in [225] is presented. This method performs a partial channel inversion and sound field superposition in order to reach the desired target response. In other words, after individual phase compensation of all loudspeakers, the support loudspeakers are used to equalize one or more primary loudspeakers. It has been shown in [225] that a significant reduction of the mean square error (MSE) and of the spatial variability is obtained, particularly at low frequencies (LFs), and that the performance of the equalization increases with the number of loudspeakers used. Also, the solution is robust to modeling errors and provides a way to control the level of the pre-ringing introduced.

The reason of choosing the method in [225] is not only dictated by the results shown for LF equalization. The method makes use of a polynomial-based control system framework, which can be used to jointly address other problems: equalization at high frequencies (HFs), with limitations determined by the spacing between microphones in the listening region, balanced equalization of two primary loudspeakers (stereo staging) [319], i.e. the minimization of the differences between two TFs after equalization, and of multiple listening positions [320], i.e. staging, as well as sound field control [321, 322] and active noise control [323]. The method and its extensions are already used in commercial systems by Dirac Research AB, which go under the commercial name of Dirac Unison [324]. A number of patents has been filed in relation to this technology [325, 326, 327, 328].

The present chapter is structured as follows: Section 7.2 introduces the problem

statement, and analyzes the framework for equalization of single-input/multiple-output (SIMO) and MIMO systems; the objective of this section is to concisely put together the work in [224, 225, 329][1], and to provide an interpretation of the solution proposed therein. Section 7.3 describes the modeling of the TFs using infinite impulse response (IIR) filters. The algorithm for TF modeling, known as BU method [73], is used. Modifications are proposed for the modeling of TFs with a common denominator, derived independently from [32], and of their all-pass (AP) excess-phase components, as required for the design of an AP filter meant to remove phase distortions common to all positions in the listening region. Moreover, regularization is introduced to avoid problems of rank deficiency in the estimation of the model parameters. Section 7.4 presents some simulation results of the acoustic modeling for the in-car LF equalization. Section 7.5 concludes the chapter.

Notation and conventions

The same notation and terminology from [225] is used throughout the chapter[2]. *Scalar- and vector-valued discrete-time signals* are denoted by normal and boldface italic letters, such as $s(n)$ and $\boldsymbol{s}(n)$, respectively. Discrete-time filters and transfer functions are represented by polynomial and rational matrices in the backward shift operator q^{-1}, defined by $q^{-1}s(n) = s(n-1)$, where q^{-1} corresponds to z^{-1} or $e^{-j\omega}$ in the frequency domain.

Constant matrices are denoted by boldface capital letters as, for example, $\mathbf{C}$. The element at row i and column j of a matrix $\mathbf{C}$ is denoted $\mathbf{C}_{(i,j)}$, whereas the ith row and jth column of a matrix are denoted $\mathbf{C}_{(i,:)}$ and $\mathbf{C}_{(:,j)}$, respectively. *Scalar polynomials* are denoted by italic capital letters as $C(q^{-1}) = c_0 + c_1 q^{-1} + c_2 q^{-2} + \cdots + c_P q^{-P}$, where the nonnegative integer P is called the *degree* of $C(q^{-1})$. *Polynomial matrices* are denoted by bold italic capital letters as $\boldsymbol{C}(q^{-1}) = \mathbf{C}_0 + \mathbf{C}_1 q^{-1} + \cdots + \mathbf{C}_P q^{-P}$, which define a polynomial matrix as a polynomial whose coefficients consist of matrices. An alternative, but equivalent, definition of a polynomial matrix is that of a matrix whose elements consist of polynomials.

Rational matrices are denoted by bold calligraphic letters as $\boldsymbol{\mathcal{G}}(q^{-1})$, and are represented on right matrix fraction description (MFD) form as $\boldsymbol{\mathcal{G}}(q^{-1}) = \boldsymbol{C}(q^{-1})\boldsymbol{D}^{-1}(q^{-1})$, which for SIMO systems is equivalent to the common denominator form $\boldsymbol{\mathcal{G}}(q^{-1}) = \boldsymbol{C}(q^{-1})/D(q^{-1})$, where $\boldsymbol{C}(q^{-1})$ is a polynomial matrix and the scalar monic polynomial $D(q^{-1})$ is the least common denominator of all rational elements in $\boldsymbol{\mathcal{G}}(q^{-1})$. For scalar rational functions, normal calligraphic

[1] notice that, in some cases, the rephrasing from the reference material may be limited.
[2] the notation is defined almost exactly as in [225].

letters are used, like $\mathcal{G}(q^{-1})$. The arguments q^{-1}, q, z^{-1}, z, etc. will often be omitted if there is no risk of misunderstanding. All signals and polynomial coefficients are assumed to be real valued.

For any polynomial matrix $\boldsymbol{C}(q^{-1})$, or scalar polynomial $C(q)$, their respective *conjugates* are defined as $\boldsymbol{C}_*(q^{-1}) = \mathbf{C}_0 + \mathbf{C}_1 q + \cdots + \mathbf{C}_P q^P$ and $C(q) = c_0 + c_1 q + c_2 q^2 + \cdots + c_P q^P$. The *reciprocals* of $\boldsymbol{C}(q^{-1})$ and $C(q^{-1})$ are defined as $\bar{\boldsymbol{C}}(q^{-1}) = q^{-P}\boldsymbol{C}_*(q)$ and $\bar{C}(q^{-1}) = q^{-P}C_*(q)$, respectively. The symbol $\odot$ indicates element-wise multiplication (Hadamard product). The notation $diag(\boldsymbol{v})$, for a column vector $\boldsymbol{v}$, represents a diagonal matrix with the elements of $\boldsymbol{v}$ along the diagonal. A filter or a TF having l inputs and p outputs is said to be of dimension $p|l$.

7.2 Theoretical solution to the equalization problem

In this section, the equalization problem statement is introduced, and the theoretical solution to the equalization problem for the SIMO case and the MIMO case are discussed.

7.2.1 Problem statement

The aim of equalization of an acoustic system is to compensate for the deviations of the system response from an ideal behavior. In other words, the scope is to design a MIMO equalizer that, applied before the loudspeakers, would precompensate for deviations of the response measured at different receiver positions.

In this work, a MIMO system comprising l loudspeakers placed outside a single listening region $\Omega \in \mathbb{R}^3$, and p microphones distributed inside Ω is considered. If the input signals to the l loudspeakers are represented by a signal vector $\boldsymbol{u}(n) = [u_1(n), \ldots, u_l(n)]$ of dimension $l|1$ and the output signals from the p microphones by $\boldsymbol{y}(n) = [y_1(n), \ldots, y_p(n)]$ of dimension $p|1$, the input-output relation for the MIMO system is described by

$$\boldsymbol{y}(n) = \boldsymbol{\mathcal{H}}(q^{-1})\boldsymbol{u}(n), \tag{7.1}$$

where the rational matrix $\boldsymbol{\mathcal{H}}$ of dimension $p|l$ contains $p \times l$ rational scalar TFs of the form $\mathcal{H}_{ij}(q^{-1}) = B_{ij}(q^{-1})/A_{ij}(q^{-1})$, with $B_{ij}(q^{-1})$ and $A_{ij}(q^{-1})$ scalar polynomials of arbitrary degree and with $A_{ij}(q^{-1})$ having its roots inside the unit circle ($i = 1, \ldots, p$ and $j = 1, \ldots, l$). The roots of $A_{ij}(q^{-1})$, i.e. the poles

of the TF $\mathcal{H}_{ij}(q^{-1})$ are known to be common to all loudspeaker and microphone positions, so that a common denominator polynomial $A(q^{-1})$ could be used to model each TF. However, in order to account for differences in the response of the various loudspeakers, a common denominator $A_j(q^{-1})$ will be used to model the TFs describing the acoustic path between loudspeaker j and each of the p microphones in Ω, resulting in the following representation

$$\begin{aligned}\boldsymbol{\mathcal{H}}(q^{-1}) &= \boldsymbol{B}(q^{-1})\boldsymbol{A}^{-1}(q^{-1}) \\ &= \begin{bmatrix} B_{11}(q^{-1}) & \dots & B_{1l}(q^{-1}) \\ \vdots & & \vdots \\ B_{p1}(q^{-1}) & \dots & B_{pl}(q^{-1}) \end{bmatrix} \times \begin{bmatrix} A_1(q^{-1}) & & 0 \\ & \ddots & \\ 0 & & A_l(q^{-1}) \end{bmatrix}^{-1}.\end{aligned} \tag{7.2}$$

In order to obtain a MIMO equalizer for a wide listening region Ω, a very large number of microphones should be distributed inside Ω. In order to obtain a robust MIMO equalizer designed from relatively sparse measurements, a probabilistic modeling technique has been used in order to represent the variability of the TF. Each modeled TF $\mathcal{H}_{ij}(q^{-1})$ is decomposed into two parts, a *nominal part* $\mathcal{H}_{0ij}(q^{-1})$, which represents components of the TF that vary slowly with space (e.g. low frequencies and the first-order reflections), and a stochastic *uncertainty part* $\Delta\mathcal{H}_{ij}(q^{-1})$ which is intended to capture the variability in space of the TFs within Ω, especially at higher frequencies and in the 'late' reflections. By dropping the shift operator argument q^{-1} for a moment, and describing the two parts as pole-zero (PZ) models, the individual TFs can be written as

$$\begin{aligned}\mathcal{H}_{ij} &= \mathcal{H}_{0ij} + \Delta\mathcal{H}_{ij} = \frac{B_{0ij}}{A_{0j}} + \Delta B_{ij}\frac{B_{1j}}{A_{1j}} \\ &= \frac{B_{0ij}A_{1j} + \Delta B_{ij}B_{1j}A_{0j}}{A_{0j}A_{1j}} = \frac{\hat{B}_{0ij} + \Delta B_{ij}\hat{B}_{1j}}{A_{0j}A_{1j}} \triangleq \frac{B_{ij}}{A_j}\end{aligned} \tag{7.3}$$

where ΔB_{ij} is a polynomial with zero-mean random variables as coefficients, scaled so that $\mathbb{E}\{|\Delta B_{ij}(e^{-j\omega})|^2\} = 1$, and B_{1j}/A_{1j} is a filter, common to TFs model related to each loudspeaker, for shaping the spectral distribution of the uncertainty part.

In polynomial matrix form, the above representation becomes

$$\begin{aligned}\boldsymbol{\mathcal{H}} &= \boldsymbol{\mathcal{H}}_0 + \boldsymbol{\Delta\mathcal{H}} = \boldsymbol{B}_0\boldsymbol{A}_0^{-1} + \Delta\boldsymbol{B}\boldsymbol{B}_1\boldsymbol{A}_1^{-1} \\ &= (\boldsymbol{B}_0\boldsymbol{A}_1 + \Delta\boldsymbol{B}\boldsymbol{B}_1\boldsymbol{A}_0)(\boldsymbol{A}_0\boldsymbol{A}_1)^{-1} = (\hat{\boldsymbol{B}}_0 + \Delta\boldsymbol{B}\hat{\boldsymbol{B}}_1)(\boldsymbol{A}_0\boldsymbol{A}_1)^{-1} \triangleq \boldsymbol{B}\boldsymbol{A}^{-1},\end{aligned} \tag{7.4}$$

with $\hat{\boldsymbol{B}}_0 = \boldsymbol{B}_0\boldsymbol{A}_1$, $\hat{\boldsymbol{B}}_1 = \boldsymbol{B}_1\boldsymbol{A}_0$, $\boldsymbol{B} = (\hat{\boldsymbol{B}}_0 + \Delta\boldsymbol{B}\hat{\boldsymbol{B}}_1)$ and $\boldsymbol{A} = \boldsymbol{A}_0\boldsymbol{A}_1$. The matrices $\boldsymbol{B}$, $\boldsymbol{B}_0$ and $\Delta\boldsymbol{B}$ have dimension $p|l$, whereas matrices $\boldsymbol{B}_1$, $\boldsymbol{A}$, $\boldsymbol{A}_0$, and $\boldsymbol{A}_1$ are diagonal matrices of dimension $l|l$, where the jth element of the diagonal is common to all TFs related to the jth loudspeaker. This modeling technique is described in detail in Section 7.3.

7.2.2 SIMO equalizer

The case when only one loudspeaker is considered [224] is discussed first. The objective in this case is to design a single equalizer filter $\mathcal{R}(q^{-1})$ so that the measured TFs $\boldsymbol{\mathcal{H}}(q^{-1})$ from the loudspeaker to the multiple receivers in the listening region are corrected and closer to the predefined target TFs $\boldsymbol{\mathcal{D}}(q^{-1})$ (see Figure 7.1). If both the rational matrices $\boldsymbol{\mathcal{H}}(q^{-1})$ and $\boldsymbol{\mathcal{D}}(q^{-1})$ are modeled as PZ models with common denominators, then the error signal of a SIMO system can be written as

$$\begin{aligned}\boldsymbol{y}(n) &= \boldsymbol{\mathcal{D}}(q^{-1})w(n) - \boldsymbol{\mathcal{H}}(q^{-1})\mathcal{R}(q^{-1},q)w(n)\\ &= \frac{\boldsymbol{D}(q^{-1})}{E(q^{-1})}w(n) - \frac{\boldsymbol{B}(q^{-1})}{A(q^{-1})}\mathcal{R}(q^{-1},q)w(n)\end{aligned} \tag{7.5}$$

where

$$\begin{aligned}\boldsymbol{\mathcal{D}}(q^{-1}) &= \boldsymbol{D}(q^{-1})/E(q^{-1}) = [D_1(q^{-1})\dots D_p(q^{-1})]^T/E(q^{-1}),\\ \boldsymbol{\mathcal{H}}(q^{-1}) &= \boldsymbol{B}(q^{-1})/A(q^{-1}) = [B_1(q^{-1})\dots B_p(q^{-1})]^T/A(q^{-1}),\end{aligned} \tag{7.6}$$

$A(q^{-1})$ and $E(q^{-1})$ are stable monic polynomials, and $w(n)$ is a scalar-valued input signal defined as a zero-mean unit-variance white noise.

The objective in the SIMO case is to design the equalizer filter $\mathcal{R}(q^{-1},q)$ so that the sum of the power of the p components in the vector-valued error signal $\boldsymbol{y}(n)$ is minimized, i.e. the MSE cost function becomes

$$J = \mathbb{E}\{||\boldsymbol{y}(n)||_2^2\} = \mathbb{E}\{\mathrm{tr}(\boldsymbol{y}(n)\boldsymbol{y}^T(n))\}. \tag{7.7}$$

It was shown in [224] that the MSE-optimal, stable, possibly noncausal, mixed-phase SIMO precompensator $\mathcal{R}(q^{-1},q)$ is required to have the structure

$$\mathcal{R}(q^{-1},q) = q^{-d_0}\mathcal{F}_*(q)\mathcal{R}_1(q^{-1}) = q^{-d_0}\frac{\bar{F}_*(q)}{F_*(q)}\mathcal{R}_1(q^{-1}) = q^{-d_0}\frac{B_*^c(q)}{\beta_*^c(q)}\mathcal{R}_1(q^{-1}). \tag{7.8}$$

The polynomial $\bar{F}(q^{-1})$ is such that the zeros of $\bar{F}(z^{-1})$ are the common excess-phase zeros (the zeros outside the unit circle $|z| = 1$) of the elements in

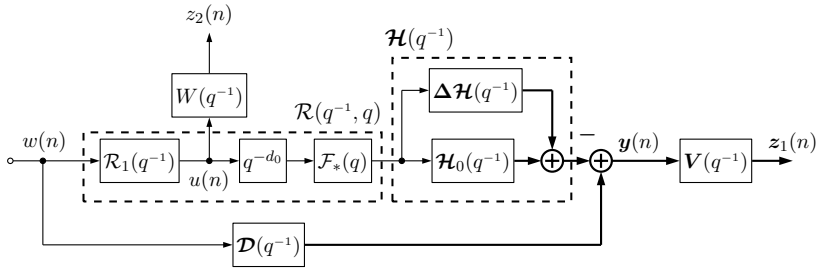

Figure 7.1: Block diagram of the robust SIMO feedforward control problem. The thin lines represent scalar signals, and the thick lines represent vector-valued signals of dimension $p|1$ (adapted from [225]).

$\boldsymbol{B}(q^{-1})$. This is expressed in the rightmost side of the equation above, with $B^c(q^{-1})$ indicating the numerator polynomial common to all the modeled TFs, and $\beta^c(q^{-1})$ its minimum-phase part, so that $\tilde{B}^c(q^{-1}) = {}^{B^c(q^{-1})}\!/_{\beta^c(q^{-1})}$ is the common AP excess-phase (nonminimum-phase) part of $B^c(q^{-1})$. It follows that $\mathcal{F}(q^{-1}) = \bar{F}(q^{-1})/F(q^{-1})$ is a causal stable AP filter and its conjugate $\mathcal{F}_*(q)$ is a stable noncausal AP filter. The response of $\mathcal{F}_*(q)$ is noncausal, which is the reason why a delay of d_0 samples is applied to shift the response in the causal domain. The role of this AP filter is to remove group delay distortions common to all positions in the listening region.

In practice, there are no exactly common excess-phase zeros among the different TFs. However, nearly common excess-phase zeros can be found by clustering zeros together, where a cluster contains exactly one zero for each TF. An AP filter can then be designed from the zeros chosen as the center of these clusters. When the compensator is applied, pre-ringing is introduced in each equalized TF, with a level dependent on the distance between the center of the clusters and the actual excess-phase zeros of the TF. By controlling the allowed maximum dimension of the clusters, the level of pre-ringing introduced can be predicted and thus controlled, as will be explained in Section 7.3. In practical applications, the finite impulse response (FIR) approximation of the noncausal response of $\mathcal{F}_*(q)$ is used, and a modeling delay d_0 is chosen large enough so that (part of) the noncausal response becomes causal. The same modeling delay has to be applied to all the TFs in the target rational polynomial matrix

$$\begin{aligned}\boldsymbol{\mathcal{D}}(q^{-1}) = \boldsymbol{D}(q^{-1})/E(q^{-1}) &= [D_1(q^{-1}) \dots D_p(q^{-1})]^T/E(q^{-1}) \\ &= [q^{-d_0}\tilde{D}_1(q^{-1}) \dots q^{-d_0}\tilde{D}_p(q^{-1})]^T/E(q^{-1}).\end{aligned} \tag{7.9}$$

Finally, $\mathcal{R}_1(q^{-1})$ represents a causal stable filter, whose structure derives from

the unique solution to the bilateral Diophantine equation (see [224, 225] and references within). To simplify the derivation of $\mathcal{R}_1(q^{-1})$, and given that $\mathcal{F}(q^{-1})$ can be computed in advance, an augmented system incorporating the delayed AP response can be defined as

$$\tilde{\boldsymbol{\mathcal{H}}}(q^{-1}) = \boldsymbol{\mathcal{H}}(q^{-1})q^{-d_0}\mathcal{F}_*(q) = \frac{\boldsymbol{B}(q^{-1})q^{-d_0}\mathcal{F}_*(q)}{A(q^{-1})} = \frac{\tilde{\boldsymbol{B}}(q^{-1})}{A(q^{-1})} \tag{7.10}$$

The causal stable filter $\mathcal{R}_1(q^{-1})$ thus assumes the structure

$$\mathcal{R}_1(q^{-1}) = A(q^{-1})\beta^{-1}(q^{-1})\Big\{\beta_*^{-1}(q)\tilde{\boldsymbol{B}}_*(q)\boldsymbol{D}(q^{-1})/E(q^{-1})\Big\}_+, \tag{7.11}$$

where $\beta(q^{-1})$ is the root mean square (RMS) spatial average model defined as the minimum-phase spectral factor of

$$\beta_*(q)\beta(q^{-1}) = \tilde{\boldsymbol{B}}_*(q)\tilde{\boldsymbol{B}}(q^{-1}) = \boldsymbol{B}_*(q)\boldsymbol{B}(q^{-1}) = \sum_{i=1}^{p} B_{i*}(q)B_i(q^{-1}), \tag{7.12}$$

and the operator $\{\cdot\}_+$ represents the causal portion of the response of its argument. Here an interpretation of the rather complicated expression for the MSE-optimal precompensator reported in (7.8) and (7.11) is given (more details are provided in Appendix C.1). For this scope, the notation is slightly simplified and the two equations combined into

$$\mathcal{R} = q^{-d_0}\mathcal{F}_*\frac{A}{\beta}\Big\{q^{d_0}\mathcal{F}\frac{\boldsymbol{B}_*}{\beta_*}\frac{\boldsymbol{D}}{E}\Big\}_+ = q^{-d_0}\mathcal{F}_*\frac{A}{\beta}\Big\{\frac{\tilde{\boldsymbol{B}}_*}{\beta_*}\frac{\boldsymbol{D}}{E}\Big\}_+. \tag{7.13}$$

The term $q^{-d_0}\mathcal{F}_*$ in the first part of the right-hand side (RHS) of the equation is the excess-phase equalizer, which corresponds to the time-reversed and delayed FIR approximation of the common AP filter $\mathcal{F}$. The next term A/β is the minimum-phase average equalizer, which corresponds to the inverse of the average minimum-phase model response β/A. The argument of the causal operator is meant to shape the equalizer w.r.t. the desired target TFs and to compensate for an average of the non-common excess-phase part of the modeled TFs.

A common case is when the target TFs at all receiver positions are the same, so that the numerator polynomial matrix $\boldsymbol{B}(q^{-1})$ can be replaced by their complex spatial average model $B_0(q^{-1}) = \sum_{i=1}^{p} B_i(q^{-1})$. The SISO case is instead a special case of the SIMO case presented here. In this case, the AP filter contains all the excess-phase zeros of the modeled TF $B(q^{-1})$, and the optimal SISO compensator becomes

$$\mathcal{R} = q^{-d_0}\mathcal{F}_*\frac{A}{\beta}\frac{D}{E}. \tag{7.14}$$

One problem in this approach is the fact that the number p of TFs is limited, so that the RMS average β is a good representative of the minimum-phase response at the receiver positions, but not at other position within Ω. To improve robustness, in [224] spectral smoothing of β with fractional-octave (e.g. 1/6 octave) resolution has been suggested. Another problem is related to the bandlimited frequency range of the loudspeaker. In order for the compensator to avoid correcting for the responses outside the working frequency range of the loudspeaker, a weighting polynomial $W(q^{-1})$ is applied to the control signal $u(n) = \mathcal{R}_1(q^{-1})w(n)$. Another modification is the introduction of a diagonal weighting matrix $\boldsymbol{V}(q^{-1})$ of dimension $p|p$ for the error signal $\boldsymbol{y}(n)$, so that the error in different frequency regions can be controlled. The introduction of these weighting matrices leads to a modification of the expression in (7.12)

$$\beta_*\beta = \boldsymbol{B}_*\boldsymbol{V}_*\boldsymbol{V}\boldsymbol{B} + A_*W_*WA, \tag{7.15}$$

and of the cost function in (7.7), which becomes

$$J = \mathbb{E}\Big\{||\boldsymbol{V}(q^{-1})\boldsymbol{y}(n)||_2^2 + ||W(q^{-1})u(n)||_2^2\Big\}. \tag{7.16}$$

Notice that, when the probabilistic model given in (7.4) is used, the numerator polynomial matrix of the augmented system in (7.10) becomes

$$\tilde{\boldsymbol{B}} = \boldsymbol{B}q^{-d_0}\mathcal{F}_* = (\hat{\boldsymbol{B}}_0 + \Delta\boldsymbol{B}\hat{\boldsymbol{B}}_1)q^{-d_0}\mathcal{F}_* = \check{\boldsymbol{B}}_0 + \Delta\boldsymbol{B}\check{\boldsymbol{B}}_1, \tag{7.17}$$

with $\check{\boldsymbol{B}}_0 = \hat{\boldsymbol{B}}_0 q^{-d_0}\mathcal{F}_*$ and $\check{\boldsymbol{B}}_1 = \hat{\boldsymbol{B}}_1 q^{-d_0}\mathcal{F}_*$. As a consequence, the expression in (7.15) is modified as

$$\beta_*\beta = \check{\boldsymbol{B}}_{0*}\boldsymbol{V}_*\boldsymbol{V}\check{\boldsymbol{B}}_0 + A_*W_*WA + \check{\boldsymbol{B}}_{1*}\mathbb{E}\{\Delta\boldsymbol{B}_*\boldsymbol{V}_*\boldsymbol{V}\Delta\boldsymbol{B}\}\check{\boldsymbol{B}}_1, \tag{7.18}$$

with $A = A_0A_1$, which, given the scaling of the variance of $\Delta\boldsymbol{B}$ to unit, can be simplified as

$$\beta_*\beta = \check{\boldsymbol{B}}_{0*}\boldsymbol{V}_*\boldsymbol{V}\check{\boldsymbol{B}}_0 + A_*W_*WA + \check{\boldsymbol{B}}_{1*}\boldsymbol{I}_l\,\mathrm{tr}(\boldsymbol{V}_*\boldsymbol{V})\check{\boldsymbol{B}}_1. \tag{7.19}$$

The expression in (7.13) for the final SIMO compensator then becomes

$$\mathcal{R} = q^{-d_0}\mathcal{F}_*\frac{A}{\beta}\left\{\frac{\check{\boldsymbol{B}}_{0*}\boldsymbol{V}_*\boldsymbol{V}\boldsymbol{D}}{\beta_* E}\right\}_+. \tag{7.20}$$

A solution to obtain β and its inverse is described in a section of the original report, not included here, for the MIMO case, which is a generalization of the SIMO case, as explained below.

7.2.3 MIMO equalizer

In the MIMO case, the objective is to design a set of l compensators $\boldsymbol{\mathcal{R}}(q^{-1},q) = [\mathcal{R}_1(q^{-1},q)\ \dots\ \mathcal{R}_l(q^{-1},q)]^T$ in such a way that $l-1$ support loudspeakers help attaining the target TFs defined for a primary loudspeaker. It was shown in [225] that a sufficient condition for obtaining a solution in the MIMO case is to apply noncausal phase compensation filters $\mathcal{F}_{1*}(q),\dots,\mathcal{F}_{l*}(q)$, designed similarly as in the SIMO case, to each of the loudspeakers, and then design a full causal and stable MIMO compensator $\boldsymbol{\mathcal{R}}_1(q^{-1})$ such that the target TF of the primary loudspeaker is attained with minimum error. The role of the filters $\mathcal{F}_{1*}(q),\dots,\mathcal{F}_{l*}(q)$ is to remove group delay distortions that are common and systematic throughout Ω for each loudspeaker. The role of $\boldsymbol{\mathcal{R}}_1(q^{-1})$, instead, is to use all the individual phase-corrected loudspeaker responses in an optimal way, so to obtain an overall sum response that is closer to the target TF than it would be in the SIMO case.

The p-dimensional error signal in the MIMO case is defined similarly to (7.5)

$$\begin{aligned}\boldsymbol{y}(n) &= \boldsymbol{\mathcal{D}}(q^{-1})w(n) - \boldsymbol{\mathcal{H}}(q^{-1})\boldsymbol{\mathcal{R}}(q^{-1},q)w(n)\\ &= \frac{\boldsymbol{D}(q^{-1})}{E(q^{-1})}w(n) - \boldsymbol{B}(q^{-1})\boldsymbol{A}^{-1}(q^{-1})\boldsymbol{\mathcal{R}}(q^{-1},q)w(n)\end{aligned} \tag{7.21}$$

where $\boldsymbol{\mathcal{D}}$ contains the modeling delay d_0 as in (7.9), and $\boldsymbol{\mathcal{H}}$ is defined as in (7.2). The MSE-optimal MIMO-compensator is given, analogously to (7.8), by

$$\boldsymbol{\mathcal{R}}(q^{-1},q) = \tilde{\boldsymbol{\Delta}}(q^{-1})\boldsymbol{\mathcal{F}}_*(q)\boldsymbol{\mathcal{R}}_1(q^{-1}), \tag{7.22}$$

where

$$\begin{aligned}\tilde{\boldsymbol{\Delta}}(q^{-1}) &= diag\Big(\big[q^{-(d_0-d_1)}\ \dots\ q^{-(d_0-d_l)}\big]^T\Big),\\ \boldsymbol{\mathcal{F}}(q^{-1}) &= diag\Big(\big[\mathcal{F}_1(q^{-1})\ \dots\ \mathcal{F}_l(q^{-1})\big]^T\Big),\\ \boldsymbol{\mathcal{R}}_1(q^{-1}) &= \big[\mathcal{R}_{11}(q^{-1})\ \dots\ \mathcal{R}_{1l}(q^{-1})\big]^T,\end{aligned} \tag{7.23}$$

where $d_j, j = 1,\dots,l$ are individual delays used to include individual deviations in distances between the listening region and each loudspeaker, whereas the elements in $\boldsymbol{\mathcal{F}}(q^{-1})$ are computed as in the SIMO case for each individual loudspeaker.

The matrices $\tilde{\boldsymbol{\Delta}}(q^{-1})$ and $\boldsymbol{\mathcal{F}}(q^{-1})$ can be incorporated into the augmented system

$$\tilde{\boldsymbol{\mathcal{H}}}(q^{-1}) = \boldsymbol{\mathcal{H}}(q^{-1})\tilde{\boldsymbol{\Delta}}(q^{-1})\boldsymbol{\mathcal{F}}_*(q) = \boldsymbol{B}(q^{-1})\tilde{\boldsymbol{\Delta}}(q^{-1})\boldsymbol{\mathcal{F}}_*(q)\boldsymbol{A}^{-1}(q^{-1}) = \tilde{\boldsymbol{B}}(q^{-1})\boldsymbol{A}^{-1}(q^{-1}), \tag{7.24}$$

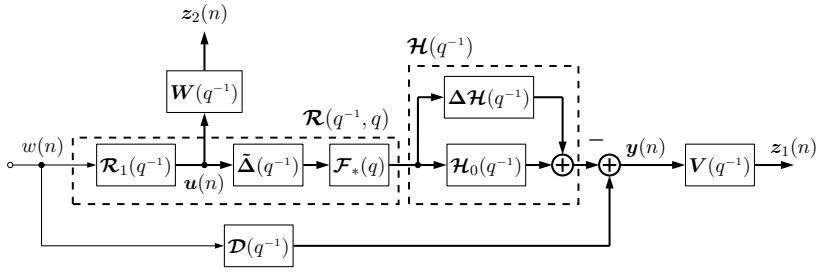

Figure 7.2: Block diagram of the constrained MIMO equalizer design. The thin lines represent scalar signals, and the thick lines represent vector-valued signals of dimension $p|1$ (adapted from [225]).

where, in case the probabilistic modeling defined in (7.4) is used

$$\tilde{\boldsymbol{B}} = \boldsymbol{B}\tilde{\boldsymbol{\Delta}}\boldsymbol{\mathcal{F}}_* = (\hat{\boldsymbol{B}}_0 + \Delta\boldsymbol{B}\hat{\boldsymbol{B}}_1)\tilde{\boldsymbol{\Delta}}\boldsymbol{\mathcal{F}}_* = \check{\boldsymbol{B}}_0 + \Delta\boldsymbol{B}\check{\boldsymbol{B}}_1, \tag{7.25}$$

with $\check{\boldsymbol{B}}_0 = \hat{\boldsymbol{B}}_0\tilde{\boldsymbol{\Delta}}\boldsymbol{\mathcal{F}}_*$, $\check{\boldsymbol{B}}_1 = \hat{\boldsymbol{B}}_1\tilde{\boldsymbol{\Delta}}\boldsymbol{\mathcal{F}}_*$, and $\boldsymbol{A} = \boldsymbol{A}_0\boldsymbol{A}_1$.

In order to attain the target TF of the primary loudspeaker, the objective is to find the optimal causal and stable MIMO compensator that minimizes the cost function defined, with reference to Figure 7.2, as

$$J = \bar{\mathbb{E}}\Big\{\mathbb{E}\big\{||\boldsymbol{V}(q^{-1})\boldsymbol{y}(n)||_2^2\big\} + \mathbb{E}\big\{||\boldsymbol{W}(q^{-1})\boldsymbol{u}(n)||_2^2\big\}\Big\}. \tag{7.26}$$

with $\boldsymbol{y}(n)$ and $\boldsymbol{u}(n)$ the error and control signals, respectively, and $\boldsymbol{W}(q^{-1})$ a $l|l$ weighting matrix applied to the control signal.

The final causal, stable, MSE-optimal MIMO compensator is then given by

$$\boldsymbol{\mathcal{R}} = \tilde{\boldsymbol{\Delta}}\boldsymbol{\mathcal{F}}_*\boldsymbol{A}\boldsymbol{\beta}^{-1}\Big\{\boldsymbol{\beta}_*^{-1}\,\check{\boldsymbol{B}}_{0*}\boldsymbol{V}_*\boldsymbol{V}\boldsymbol{D}/E\Big\}_+, \tag{7.27}$$

where the $l|l$ unique minimum-phase spectral factor $\boldsymbol{\beta}$ is obtain by factorizing

$$\boldsymbol{\beta}_*\boldsymbol{\beta} = \check{\boldsymbol{B}}_{0*}\boldsymbol{V}_*\boldsymbol{V}\check{\boldsymbol{B}}_0 + \boldsymbol{A}_*\boldsymbol{W}_*\boldsymbol{W}\boldsymbol{A} + \check{\boldsymbol{B}}_{1*}\boldsymbol{I}_l\,\mathrm{tr}(\boldsymbol{V}_*\boldsymbol{V})\check{\boldsymbol{B}}_1. \tag{7.28}$$

The computation of $\boldsymbol{\beta}$ and its inverse $\boldsymbol{\beta}^{-1}$ from the RHS of (7.28) is discussed in a section of the original report, not included here.

Checklist

Here a summary of the equations necessary to practically implement the MIMO controller is given, so to identify the elements that need to be modeled, and the

design choices to be made.

$$\boldsymbol{\mathcal{R}} = \tilde{\boldsymbol{\Delta}}\boldsymbol{\mathcal{F}}_*\boldsymbol{A}\beta^{-1}\Big\{\beta_*^{-1}\,\check{\boldsymbol{B}}_{0*}\boldsymbol{V}_*\boldsymbol{V}\boldsymbol{D}/E\Big\}_+ \tag{7.29}$$

$$\tilde{\boldsymbol{\Delta}} = diag\Big(\big[q^{-(d_0-d_1)}\ \ldots\ q^{-(d_0-d_l)}\big]^T\Big) \tag{7.30}$$

$$\begin{aligned}\boldsymbol{\mathcal{F}} &= diag\Big(\big[\mathcal{F}_1(q^{-1})\ \ldots\ \mathcal{F}_l(q^{-1})\big]^T\Big)\\ &= diag\Big(\Big[\frac{\bar{F}_1(q^{-1})}{F_1(q^{-1})}\ \ldots\ \frac{\bar{F}_l(q^{-1})}{F_l(q^{-1})}\Big]^T\Big)\\ &= diag\Big(\Big[\frac{B_1^c(q^{-1})}{\beta_1^c(q^{-1})}\ \ldots\ \frac{B_l^c(q^{-1})}{\beta_l^c(q^{-1})}\Big]^T\Big)\end{aligned} \tag{7.31}$$

$$\beta_*\boldsymbol{\beta} = \check{\boldsymbol{B}}_{0*}\boldsymbol{V}_*\boldsymbol{V}\check{\boldsymbol{B}}_0 + \boldsymbol{A}_*\boldsymbol{W}_*\boldsymbol{W}\boldsymbol{A} + \check{\boldsymbol{B}}_{1*}\boldsymbol{I}_l\,\mathrm{tr}(\boldsymbol{V}_*\boldsymbol{V})\check{\boldsymbol{B}}_1 \tag{7.32}$$

$$\check{\boldsymbol{B}}_0 = \hat{\boldsymbol{B}}_0\tilde{\boldsymbol{\Delta}}\boldsymbol{\mathcal{F}}_* \tag{7.33}$$

$$\check{\boldsymbol{B}}_1 = \hat{\boldsymbol{B}}_1\tilde{\boldsymbol{\Delta}}\boldsymbol{\mathcal{F}}_* \tag{7.34}$$

$$\hat{\boldsymbol{B}}_0 = \boldsymbol{B}_0\boldsymbol{A}_1 \tag{7.35}$$

$$\hat{\boldsymbol{B}}_1 = \boldsymbol{B}_1\boldsymbol{A}_0 \tag{7.36}$$

$$\boldsymbol{B} = (\hat{\boldsymbol{B}}_0 + \Delta\boldsymbol{B}\hat{\boldsymbol{B}}_1) \tag{7.37}$$

$$\boldsymbol{A} = \boldsymbol{A}_0\boldsymbol{A}_1 \tag{7.38}$$

The optimal MIMO compensator has dimension $l|1$, the target TFs matrix $\boldsymbol{\mathcal{D}} = {}^{\boldsymbol{D}}/_E$ (and thus $\boldsymbol{D}$) has dimension $p|1$, the minimum-phase spectral factor $\boldsymbol{\beta}$ has dimension $l|l$, whereas the weighting matrices $\boldsymbol{W}$ and $\boldsymbol{V}$ have dimension $l|l$ and $p|p$ respectively. The polynomial matrices $\check{\boldsymbol{B}}_0$, $\check{\boldsymbol{B}}_1$, $\hat{\boldsymbol{B}}_0$, $\hat{\boldsymbol{B}}_1$, $\boldsymbol{B}_0$, $\boldsymbol{B}_1$, and $\boldsymbol{\Delta B}$ have dimensions $p|l$, while $\boldsymbol{A}$, $\boldsymbol{A}_0$, $\boldsymbol{B}_1$, $\boldsymbol{A}_1$, $\tilde{\boldsymbol{\Delta}}$, and $\boldsymbol{\mathcal{F}}$ are diagonal matrices of dimensions $l|l$. The modeling of the TFs using the probabilistic modeling technique and of the identification of the nearly-common excess-phase zeros is presented in Section 7.3.

Design choices pertain to the selection of the target TFs $\boldsymbol{\mathcal{D}} = {}^{\boldsymbol{D}}/_E$, the modeling delay d_0, the definition of the weighting matrices $\boldsymbol{V}$ and $\boldsymbol{W}$, and the selection of the parameters to control the pre-ringing level in the design of $\boldsymbol{\mathcal{F}}$. Apart from the latter, shortly described in Section 7.4, the other aspects are not discussed in this chapter, but can be found in the final report.

7.3 Acoustic modeling

The first step in the design of a MIMO controller, is actually the modeling of the measured TFs and the design of the AP filters from the clustering of nearly-common excess-phase zeros. In the following, all the TFs are modeled as IIR filters, which are preferred to FIR filters (which however are still an option) because they provide more flexibility and lower model orders for the same modeling accuracy. In the practical equalizer design, the FIR approximation of the IIR filter responses will be used in some cases. As a preliminary step, the acoustic delays of each individual AIR were computed (using a peak detection algorithm), stored for use in $\tilde{\boldsymbol{\Delta}}$ in the controller design, and removed from the AIRs prior to modeling. The probabilistic modeling (Sections 7.3.1 and 7.3.2) is performed on the full-band TFs, whereas the TF modeling (Sections 7.3.3, 7.3.4 and 7.3.5) is performed on the LF (probabilistic) TFs, i.e. after resampling to $f'_S = {}^{f_S}/_{D_s}$, with $D_s \in \mathbb{Z}^+$.

7.3.1 Probabilistic modeling

As briefly mentioned in the introduction, the MIMO controller is designed starting from a number of AIRs measured within a given listening area Ω. The number of microphones in $\Omega \in \mathbb{R}^3$ should be ideally large, with the actual number determined by the dimensions of Ω, and the spacing between microphones limiting the highest frequency that can be effectively equalized. When an insufficient number of microphones is available to design the equalizer, a certain amount of overfitting at the microphone positions has to be expected. In order to obtain a more spatially robust solution from a limited number of sparse microphones, a probabilistic model is used to describe the TF variability in the listening area [329].

The measured TFs in $\boldsymbol{\mathcal{H}}(q^{-1})$ can be decomposed as in (7.4) into two parts, a *nominal part* $\boldsymbol{\mathcal{H}}_0(q^{-1})$ and a stochastic *uncertainty part* $\boldsymbol{\Delta\mathcal{H}}(q^{-1})$

$$\boldsymbol{\mathcal{H}}(q^{-1}) = \boldsymbol{\mathcal{H}}_0(q^{-1}) + \boldsymbol{\Delta\mathcal{H}}(q^{-1}) \tag{7.39}$$

where $\boldsymbol{\Delta\mathcal{H}}(q^{-1})$ is parameterized by zero-mean random variables, assumed to be independent from any signal in the system, and describes possible deviations from the nominal part. The nominal part contains the direct path and the LF parts of the TFs, which are the components that vary slowly with space. In the region of modal frequencies, i.e. below the Schroeder's frequency, the sound field is less diffuse than at higher frequencies. Moreover, when the wavelength is much larger than the microphone spacing, the sound pressure in-between adjacent microphones can be interpolated linearly from the sound pressure

at the microphones. The 'complementary' uncertainty part $\boldsymbol{\Delta\mathcal{H}}(q^{-1})$ models the reverberant part of the AIRs and the sound field at higher frequencies, where the spatial variability in-between microphones is more prominent. When dealing with equalization at LFs, the use of this probabilistic model is less important, since the LF part of the TFs is included in the nominal part only. In the implementation described below, a certain degree of variability at lower frequencies was allowed, in order to improve robustness of the solution (and to help overcome the lack of measurements in the listening area).

The two parts of the TF matrix can be described in the right MFD notation as

$$\begin{aligned}\boldsymbol{\mathcal{H}} &= \boldsymbol{\mathcal{H}}_0 + \boldsymbol{\Delta\mathcal{H}} = \boldsymbol{\mathcal{H}}_0 + \boldsymbol{\Delta B\mathcal{H}}_1 = \boldsymbol{B}_0\boldsymbol{A}_0^{-1} + \Delta\boldsymbol{B}\boldsymbol{B}_1\boldsymbol{A}_1^{-1} \\ &= (\boldsymbol{B}_0\boldsymbol{A}_1 + \Delta\boldsymbol{B}\boldsymbol{B}_1\boldsymbol{A}_0)(\boldsymbol{A}_0\boldsymbol{A}_1)^{-1} = (\hat{\boldsymbol{B}}_0 + \Delta\boldsymbol{B}\hat{\boldsymbol{B}}_1)(\boldsymbol{A}_0\boldsymbol{A}_1)^{-1} \triangleq \boldsymbol{B}\boldsymbol{A}^{-1},\end{aligned} \tag{7.40}$$

with $\hat{\boldsymbol{B}}_0 = \boldsymbol{B}_0\boldsymbol{A}_1$, $\hat{\boldsymbol{B}}_1 = \boldsymbol{B}_1\boldsymbol{A}_0$, $\boldsymbol{B} = (\hat{\boldsymbol{B}}_0 + \Delta\boldsymbol{B}\hat{\boldsymbol{B}}_1)$ and $\boldsymbol{A} = \boldsymbol{A}_0\boldsymbol{A}_1$. The matrices $\boldsymbol{B}$, $\boldsymbol{B}_0$ and $\Delta\boldsymbol{B}$ have dimension $p|l$, whereas matrices $\boldsymbol{B}_1$, $\boldsymbol{A}$, $\boldsymbol{A}_0$, and $\boldsymbol{A}_1$ are diagonal matrices of dimension $l|l$, where the jth element of the diagonal is common to all TFs related to the jth loudspeaker. Each element of $\Delta\boldsymbol{B}$ is a polynomial with zero-mean random variables as coefficients, scaled so that $\mathbb{E}\{|\Delta B_{ij}(e^{-j\omega})|^2\} = 1$. Each element of the diagonal of $\boldsymbol{\mathcal{H}}_1 = \boldsymbol{B}_1\boldsymbol{A}_1^{-1}$ is an IIR filter, common to all modeled TFs related to one loudspeaker, used for shaping the spectral distribution of the uncertainty part. The use of FIR filters is an option, which can be easily obtained by setting $\boldsymbol{A}_0 = 1$ and $\boldsymbol{A}_1 = 1$.

The decomposition of the TF matrix into two parts is performed, as suggested in [329], by applying a variable low-pass (LP) filter to each TF to obtain the nominal part H_{0ij} of the AIR, while the complementary variable high-pass (HP) filter is used to obtain the corresponding reverberant part H_{Rij}. The cut-off frequency of the LP/HP filter pair decreases from the Nyquist frequency to a certain frequency f_C, over a given time frame, starting some time t_{start} (e.g. 1 ms) after the direct path of the AIR, and ending at t_{stop} (e.g. 7 ms). The variable LP/HP filter pair was implemented frame-by-frame, by using overlap-and-add with non-overlapping rectangular windows, where the frame length L is determined by t_{start} (for $f_S = 44.1$kHz and $t_{\text{start}} = 1$ ms, $L = 45$). A variable linear phase FIR filter design technique was used to design the LP/HP filters [330, 331]. A prototype LP filter h_{M_0} is first obtained, with a desired transition bandwidth and cut-off frequency at -6 dB defined as $\omega_{c_0} = 2\pi(0.25 + {}^1\!/_{2L})$ (for L odd). Here a window-based design technique was used (see e.g. [45]), with a Hamming window $w(n)$ of length L (which for $L = 45$ has a stop-band attenuation of almost -60 dB), so that $h_{M_0} = w(n)\,\omega_{c_0}\text{sinc}(\omega_{c_0}n)$ (with $n = -{}^{(L-1)}\!/_2 : {}^{(L-1)}\!/_2$). A LP filter with arbitrary cut-off frequency ω_c

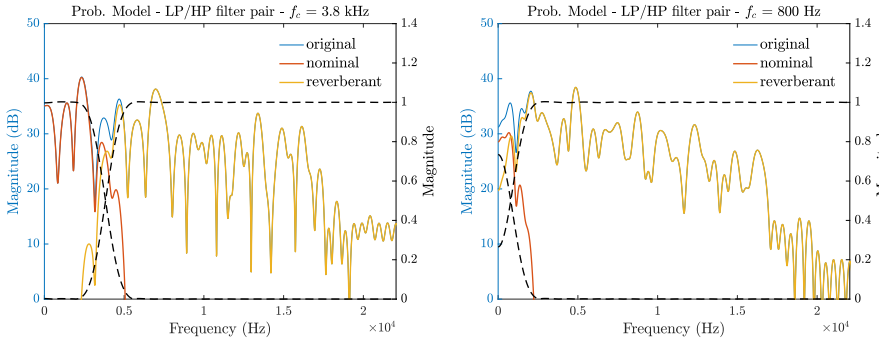

Figure 7.3: Variable HP/LP filters (Listener 1, FL loudspeaker, L microphone).

can be obtained from the prototype filter as

$$h_c^{LP}(n) = \begin{cases} c(n)\,\omega_c & \text{for } n = 0 \\ c(n)\,\sin(\omega_c n) & \text{for } 1 \leq |n| \leq {}^{(L-1)}/_{2} \end{cases}, \tag{7.41}$$

where $c(n) = h_{M_0}(n)/\sin(\omega_{c_0} n)$. At each frame, starting from the second, the cut-off frequency is reduced, so that at t_{stop} the final cut-off frequency f_C is reached. The complementary HP filter is simply obtained as

$$h_c^{HP}(n) = \begin{cases} 1 - h_c^{LP}(n) & \text{for } n = 0 \\ -h_c^{LP}(n) & \text{for } 1 \leq |n| \leq {}^{(L-1)}/_{2} \end{cases}. \tag{7.42}$$

The cut-off frequency of the LP should not be less than half the transition bandwidth, otherwise the LP filter does not reach unit level in the pass-band. In the implementation described, the final cut-off frequency f_C becomes slightly smaller than the transition bandwidth, so that some low frequencies are included in the reverberant uncertainty part. This can be seen in the left plot of Figure 7.3. If this feature is not desirable, a prototype LP filter with shorter transition bandwidth should be designed. Figure 7.4 shows the original frequency response of one TF and the resulting nominal and reverberant parts. Notice that, for what discussed above, the stochastic reverberant part includes some low frequencies.

The second step in the probabilistic model technique consists in the design of the shaping filter $\boldsymbol{\mathcal{H}}_1 = \boldsymbol{B}_1 \boldsymbol{A}_1^{-1}$ intended to model the spectral envelope of the reverberant parts H_{Rij} for $i = 1, \ldots, p$. For this purpose, for each loudspeaker j, an FIR shaping filter H_{1j} of order N_{tr} is constructed from H_{Rij} by using their average periodogram $\Phi = \frac{1}{p}\sum_{i=1}^{p} H_{Rij*} H_{Rij}$. A triangular window of length $2N_{\text{tr}} - 1$ is applied symmetrically over the polynomial coefficients of

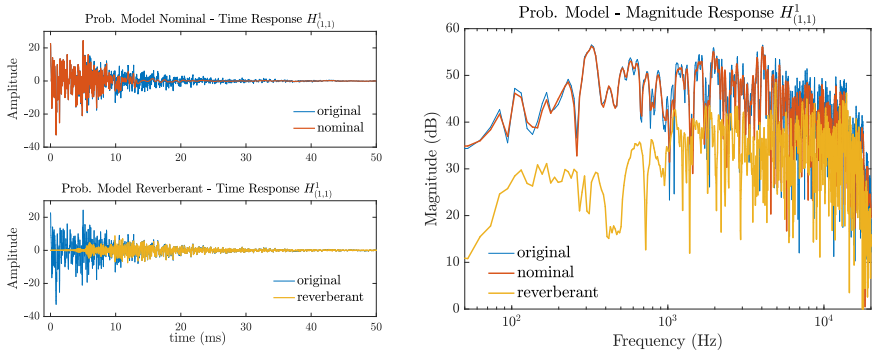

Figure 7.4: Time and magnitude responses of the nominal and reverberant parts of one TF (Listener 1, FL loudspeaker, L microphone).

Φ, yielding an average Blackman-Tukey spectral estimate $\hat{\Phi}$ (Welch method for spectral estimation) [332]. The shaping filter H_{1j} is then obtained as the minimum-phase spectral factor of $\hat{\Phi}$ (e.g. using the cepstral method [45, 67] – see `rceps` MATLAB function).

7.3.2 Virtual receivers

The probabilistic modeling technique described above has been used as a method to generate a number of virtual receivers in order to compensate for the lack of receivers in the listening region Ω. The idea is to start from the p_r actual measurements in Ω recorded for the jth loudspeaker, and for each of them generate p_v virtual measurements as variations of the actual measurements using the probabilistic modeling technique, so to obtain $p = p_r(p_v + 1)$ receivers inside Ω. After the probabilistic model has been computed from the p_r actual measurements $H_{ij}(q^{-1})$, thus obtaining p_v nominal parts $H_{0ij}(q^{-1})$ (with $i = 1, \ldots, p_r$) and one shaping filter $H_{1j}(q^{-1})$, the recording $\breve{H}_{vj}(q^{-1})$ at the vth virtual receiver is generated by applying a random variation to the nominal part and a temporal envelope to the uncertainty part, so to obtain $\breve{H}_{vj}(q^{-1}) = \breve{H}_{0vj}(q^{-1}) + \Delta\breve{H}_{vj}(q^{-1})$.

The nominal part $\breve{H}_{0vj}(q^{-1})$ of the TF $\breve{H}_{vj}(q^{-1})$ of a virtual receiver starting from the nominal part $H_{0ij}(q^{-1})$ of the ith real TF is obtained by applying a low-passed random gaussian noise $\breve{w}(n)$ distributed as $\mathcal{N} \sim (0, \sigma_v)$ (e.g. $\sigma_v = 0.005$), where the low-pass filter $\breve{h}^{\mathrm{LP}}(n)$ (e.g. with normalized cut-off frequency $\omega_c = 0.4$) is intended to limit the variability at high frequency, which is later introduced in the uncertainty part. The variance σ_v is scaled with respect to

the norm of the TF, so that the nominal part for a virtual receiver becomes

$$\breve{H}_{0vj}(q^{-1}) = H_{0ij}(q^{-1})\Big\{1 + \big[\breve{h}^{\mathrm{LP}}(q^{-1}) \cdot \{\breve{w}(n) \cdot ||H_{0ij}(q^{-1})||_2\}\big]\Big\}. \tag{7.43}$$

The second step is to add a temporal envelope to the uncertainty part. The probabilistic model uses a random sequence to obtain the coefficients of the polynomials $\Delta B_{ij}(q^{-1})$, which does not decay in time. In order to obtain a realistic time-response, the temporal envelope $h_{ij}^{\mathrm{env}}(q^{-1})$ of the reverberant part H_{Rij}, calculated as the upper RMS envelope determined using a sliding window (see MATLAB function `envelope.m`), is applied to a random polynomial $\Delta\breve{B}_{vj}(q^{-1})$, which is then filtered by the shaping filter H_{1j} obtained from the probabilistic modeling,

$$\Delta\breve{H}_{vj}(q^{-1}) = \big[\Delta\breve{B}_{vj}(q^{-1}) \odot h_{ij}^{\mathrm{env}}(q^{-1})\big] \cdot H_{1j}(q^{-1}). \tag{7.44}$$

7.3.3 Transfer function modeling (BU method)

Once the nominal and uncertainty parts of the TFs are obtained, the next step consists in modeling the numerator and denominator polynomials of $\boldsymbol{\mathcal{H}}_0 = \boldsymbol{B}_0\boldsymbol{A}_0^{-1}$ and $\boldsymbol{\mathcal{H}}_1 = \boldsymbol{B}_1\boldsymbol{A}_1^{-1}$. We will start with $\boldsymbol{\mathcal{H}}_1 = \boldsymbol{B}_1\boldsymbol{A}_1^{-1}$, by modeling the elements of the polynomial matrices $\boldsymbol{B}_1$, and $\boldsymbol{A}_1$.

As described in previous sections, each FIR shaping filter $H_{1j}(q^{-1})$ is designed as the average spectral envelope of the reverberant part of the TF at p receiver positions for a given loudspeaker, so that one shaping filter $\mathcal{H}_{1j}(q^{-1})$ per loudspeaker has to be modeled. An output-error method for pole-zero modeling of TFs was used, called the BU method [73], which outperforms better-known methods, such as the Steiglitz-McBride method [70], in terms of stability and efficiency of the resulting polynomials (i.e. the polynomials have lower degree). The only limitation is that the degree of the numerator and denominator polynomials has to be the same, and it has to be defined a priori. Regularization is introduced to avoid ill-conditioning problems in the estimation of the model parameters.

The BU method first estimates the denominator polynomial of the PZ model as the denominator of an AP filter, while the numerator polynomial is obtained in closed form. The objective of the BU method is to obtain a least squares (LS) IIR approximation of an FIR filter. In this case, the aim is to obtain an approximation of the FIR shaping filter $H_{1j}(q^{-1})$ with an IIR of the form $\mathcal{H}_{1j}(q^{-1}) = {}^{B_{1j}(q^{-1})}\!/_{A_{1j}(q^{-1})}$. The cost function to be minimized is then the

ℓ_2-norm of the approximation error $E_{1j}(q^{-1})$,

$$J_{\mathrm{BU}} = ||E_{1j}(q^{-1})||_2^2 = ||H_{1j}(q^{-1}) - \mathcal{H}_{1j}(q^{-1})||_2^2 = \left|\left|H_{1j}(q^{-1}) - \frac{B_{1j}(q^{-1})}{A_{1j}(q^{-1})}\right|\right|_2^2. \tag{7.45}$$

In other words, the LS approximation of the FIR TF $H_{1j}(q^{-1}) = \sum_{\lambda=0}^{K} h_\lambda q^{-\lambda}$ of order K by an IIR filter $\mathcal{H}_{1j}(q^{-1}) = {}^{B(q^{-1})}/_{A(q^{-1})}$ of order $M < K$, with $B_{1j}(q^{-1}) = \sum_{\mu=0}^{M} b_\mu q^{-\mu}$ and $A_{1j}(q^{-1}) = \sum_{\mu=0}^{M} a_\mu q^{-\mu}$ (with $a_0 = 1$), consists in the estimation of $M+1$ numerator coefficients b_μ ($\mu = 0, 1, \ldots, M$) and M denominator coefficients a_μ ($\mu = 1, 2, \ldots, M$), such that the cost function in (7.45) is minimized. In the rest of this section we will temporarily drop the subscripts to be more general and improve readability.

The BU method relies on the Walsh theorem [75], which states that for a given denominator polynomial $A(q^{-1})$ with roots (poles) α_μ, the best approximation of $B(q^{-1})$ in LS sense is given by the unique function that interpolates $H(q^{-1})$ at points ${}^1/_{\alpha_\mu^*}$ and at infinity. This theorem tells us that the estimation of the denominator coefficients a_μ can be decoupled from the estimation of the numerator coefficients b_μ, so that once the coefficients a_μ are found, the coefficients b_μ are obtained as the solution of an interpolation problem. Another consequence is that the approximation error $E(q^{-1})$ has $M+1$ zeros located at ${}^1/_{\alpha_\mu^*}$ and at infinity, so that it can be written as

$$E(q^{-1}) = H(q^{-1}) - \frac{B(q^{-1})}{A(q^{-1})} = \frac{q^{-(M+1)}A(q)}{A(q^{-1})}R(q^{-1}) = \frac{q^{-M}A(q)}{A(q^{-1})}q^{-1}R(q^{-1}),$$

$$\text{with } R(q^{-1}) = q^{-K}H(q)\frac{q^{-M}A(q)}{A(q^{-1})} = \sum_{\lambda=0}^{K-1} r_\lambda q^{-\lambda}. \tag{7.46}$$

an unknown FIR TF of order $K-1$. From the equation above, for known polynomials $A(q^{-1})$ and $R(q^{-1})$, the numerator polynomial $B(q^{-1})$ can be obtained in closed form as

$$B(q^{-1}) = H(q^{-1})A(q^{-1}) - q^{-(M+1)}A(q)R(q^{-1}). \tag{7.47}$$

The estimation of the denominator polynomial, instead, uses the concept of 'complementary' signal [74], for which, if the reciprocal of $H(q^{-1})$ (which correspond to a time-reversal of the FIR coefficients h_λ) is fed to an AP filter $\mathcal{A}(q^{-1})$, the energy of the resulting polynomial $U(q^{-1}) = q^{-K}H(q) \cdot \mathcal{A}(q^{-1})$ is equal to the energy of $H(q^{-1})$, but it is partitioned in the following way [29, 31]

$$\sum_{\lambda=0}^{\infty} |H(q^{-1})|^2 = \sum_{\lambda=0}^{\infty} |U(q^{-1})|^2 = \sum_{\lambda=0}^{K} |U(q^{-1})|^2 + \sum_{\lambda=K+1}^{\infty} |U(q^{-1})|^2, \tag{7.48}$$

where the first term of the RHS of the equation is the approximation error energy, which is what the algorithm will try to minimize, whereas the second term is the energy of the approximation. An outline (adapted from [31]) of the algorithm is provided here (more details can be found in [73]):

- The algorithm is based on the approximation of an AP filter of order M

$$\mathcal{A}^{(\kappa)}(q^{-1}) = \frac{q^{-M}A^{(\kappa)}(q)}{A^{(\kappa-1)}(q^{-1})} \tag{7.49}$$

  where a new monic polynomial $A^{(\kappa)}(q^{-1})$ is estimated at each iteration κ (with $\{1, A^{(1)}(q^{-1}), A^{(2)}(q^{-1}), \dots\}$), and is restricted to the form

$$A^{(\kappa)}(q^{-1}) = 1 + \sum_{\mu=1}^{M} a_{\mu}^{(\kappa)} q^{-\mu} = 1 + \tilde{A}^{(\kappa)}(q^{-1}), \quad \kappa = 1, 2, \dots. \tag{7.50}$$

- The polynomial ratio $\mathcal{A}(q^{-1})$ converges to an AP function if $||A^{(\kappa)}(q^{-1}) - A^{(\kappa-1)}(q^{-1})||_2 \to 0$.
- The objective is to minimize the energy of the error polynomial $U^{(\kappa)}(q^{-1}) = \mathcal{A}^{(\kappa)}(q^{-1})X(q^{-1})$, where $X(q^{-1}) = q^{-K}H(q)$ is the reciprocal of $H(q^{-1})$.
- Define $Y^{(\kappa)}(q^{-1}) = X(q^{-1})/A^{(\kappa-1)}(q^{-1}) = \sum_{\lambda=0}^{K-1} y_{\lambda}^{(\kappa)} q^{-\lambda}$, so that $U^{(\kappa)}(q^{-1}) = q^{-M}A^{(\kappa)}(q)Y^{(\kappa)}(q^{-1}) = \sum_{\lambda=0}^{K-1} u_{\lambda}^{(\kappa)} q^{-\lambda}$.
- Use (7.50) and rearrange, obtaining

$$Y^{(\kappa)}(q^{-1})\left[q^{-(M-1)}\tilde{A}^{(\kappa)}(q)\right] = U^{(\kappa)}(q^{-1}) - q^{-M}Y^{(\kappa)}(q^{-1}). \tag{7.51}$$

- Collect the coefficients of equal order on both sides of the equation above, obtaining the following vectors

$$\begin{aligned} \boldsymbol{a}^{(\kappa)} &= [a_M^{(\kappa)}, a_{M-1}^{(\kappa)}, \dots, a_1^{(\kappa)}]^T \\ \boldsymbol{u}^{(\kappa)} &= [u_0^{(\kappa)}, u_1^{(\kappa)}, \dots, u_{K-1}^{(\kappa)}]^T \\ \boldsymbol{y}^{(\kappa)} &= -[0, \dots, 0, y_0^{(\kappa)}, y_1^{(\kappa)}, \dots, y_{K-M-1}^{(\kappa)}]^T \end{aligned} \tag{7.52}$$

and the $K \times M$ design matrix

$$\boldsymbol{Y}^{(\kappa)} = \begin{bmatrix} y_0^{(\kappa)} & 0 & \dots & 0 \\ y_1^{(\kappa)} & y_0^{(\kappa)} & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ y_{M-1}^{(\kappa)} & \dots & & y_0^{(\kappa)} \\ \vdots & & & \vdots \\ y_{K-1}^{(\kappa)} & \dots & & y_{K-M}^{(\kappa)} \end{bmatrix} \tag{7.53}$$

which corresponds to an overdetermined set of linear equations $\boldsymbol{Y}^{(\kappa)}\boldsymbol{a}^{(\kappa)} = \boldsymbol{u}^{(\kappa)} + \boldsymbol{y}^{(\kappa)}$, with $\boldsymbol{a}^{(\kappa)}$ and $\boldsymbol{u}^{(\kappa)}$ unknown.

- At each iteration κ, the LS solution of $\boldsymbol{Y}^{(\kappa)}\boldsymbol{a}^{(\kappa)} = \boldsymbol{y}^{(\kappa)}$, which minimizes the energy of $\boldsymbol{u}^{(\kappa)} = \boldsymbol{Y}^{(\kappa)}\boldsymbol{a}^{(\kappa)} - \boldsymbol{y}^{(\kappa)}$, is computed (e.g. using QR decomposition), thus obtaining the coefficients $a_\mu^{(\kappa)}$ of $\tilde{A}^{(\kappa)}(q)$, and consequently of $A(q^{-1})$.

- The matrix $\boldsymbol{Y}^{(\kappa)}$, as the number of iterations increases, can become rank deficient. In order to avoid that, some regularization can be introduced, with the set of equations becoming $(\boldsymbol{Y}^{T(\kappa)}\boldsymbol{Y}^{(\kappa)} + \lambda \boldsymbol{I}_M)\boldsymbol{a}^{(\kappa)} = \boldsymbol{Y}^{T(\kappa)}\boldsymbol{y}^{(\kappa)}$, with λ a regularization parameter (e.g. $\lambda = 10^{-8}$).

- After a number of iterations, or when a specified error threshold is reached, an estimate of the numerator polynomial $B(q^{-1})$ is computed as

$$\hat{B}(q^{-1}) = H(q^{-1})\hat{A}(q^{-1}) - q^{-(M+1)}\hat{A}(q)R(q^{-1}), \tag{7.54}$$

where $\hat{A}(q^{-1})$ is the polynomial with minimum LS error among all the estimates computed in the iterative procedure, and

$$R(q^{-1}) = q^{-K}H(q)\frac{q^{-M}\hat{A}(q)}{\hat{A}(q^{-1})} \tag{7.55}$$

is the reciprocal of the polynomial $\hat{U}(q^{-1})$ with minimum energy.

7.3.4 Common-denominator TF modeling (CD-BU method)

The next step is to model the numerator and denominator polynomials of $\boldsymbol{\mathcal{H}}_0 = \boldsymbol{B}_0\boldsymbol{A}_0^{-1}$, i.e. the elements of the polynomial matrices $\boldsymbol{B}_0$, and $\boldsymbol{A}_0$. As mention already in Section 7.1, the polynomials $A_{ij}(q^{-1})$ (for $i = 1, \dots, p$) do not depend on the spatial position of the receivers within Ω, based on

the assumption that the modal resonances excited by the jth loudspeaker are present, with different levels, at each receiver position (i.e. the poles of $A_{ij}(z^{-1})$ are common to all p TFs). It follows that a common denominator $A_j(q^{-1})$ to model the TFs linked to the jth loudspeaker has to be found ($\boldsymbol{A}_0$ is a diagonal polynomial matrix of dimensions $l|l$).

The BU method described above can be easily extended to the multi-channel case[3]. The extension, called here Common-Denominator BU (CD-BU) method, is quite trivial, but provides more accurate results than other methods, such as the equation-error method in [103]. The cost function is just the sum of the approximation errors of the p TFs (for the jth loudspeaker),

$$\begin{aligned} J_{\text{CD-BU}} &= \sum_{i=1}^{p} ||E_{0ij}(q^{-1})||_2^2 = \sum_{i=1}^{p} ||H_{0ij}(q^{-1}) - \mathcal{H}_{0ij}(q^{-1})||_2^2 \\ &= \sum_{i=1}^{p} \left|\left| H_{0ij}(q^{-1}) - \frac{B_{0ij}(q^{-1})}{A_{0j}(q^{-1})} \right|\right|_2^2 \end{aligned} \tag{7.56}$$

which has to be minimized w.r.t. $p(M+1)$ numerator coefficients $b_{\mu,i}$ ($\mu = 0, 1, \ldots, M$ and $i = 1, \ldots, p$) and M denominator coefficients a_μ ($\mu = 1, 2, \ldots, M$).

The estimation of the denominator polynomial $A_{0j}(q^{-1})$ is performed by minimizing the sum of the energy of the error polynomials $U_{0ij}^{(\kappa)}(q^{-1}) = \mathcal{A}_{0j}(q^{-1}) X_{0ij}(q^{-1})$, where $X_{0ij}(q^{-1}) = q^{-K} H_{0ij}(q)$ is the reciprocal of $H_{0ij}(q^{-1})$ and $\mathcal{A}_{0j}(q^{-1})$ is the AP filter built from $A_{0j}(q^{-1})$ defined as in (7.49), leading to p equations of the type in (7.51),

$$Y_{0ij}^{(\kappa)}(q^{-1}) \left[q^{-(M-1)} \tilde{A}_{0j}^{(\kappa)}(q) \right] = U_{0ij}^{(\kappa)}(q^{-1}) - q^{-M} Y_{0ij}^{(\kappa)}(q^{-1}). \tag{7.57}$$

with $Y_{0ij}^{(\kappa)}(q^{-1}) = q^{-K} H_{0ij}(q) / A_{0j}^{(\kappa-1)}(q^{-1})$ and $A_{0j}^{(\kappa)}(q^{-1}) = 1 + \sum_{\mu=1}^{M} a_{\mu j}^{(\kappa)} q^{-\mu} = 1 + \tilde{A}_{0j}^{(\kappa)}(q^{-1})$. These p equations can be put in vector form as

$$\begin{aligned} \boldsymbol{a}_{0j}^{(\kappa)} &= [a_{Mj}^{(\kappa)}, a_{(M-1)j}^{(\kappa)}, \ldots, a_{1j}^{(\kappa)}]^T \\ \boldsymbol{u}_{0j}^{(\kappa)} &= [\boldsymbol{u}_{01j}^{(\kappa)}, \boldsymbol{u}_{02j}^{(\kappa)}, \ldots, \boldsymbol{u}_{0pj}^{(\kappa)}]^T \\ \boldsymbol{y}_{0j}^{(\kappa)} &= [\boldsymbol{y}_{01j}^{(\kappa)}, \boldsymbol{y}_{02j}^{(\kappa)}, \ldots, \boldsymbol{y}_{0pj}^{(\kappa)}]^T \\ \boldsymbol{Y}_{0j}^{(\kappa)} &= [\boldsymbol{Y}_{01j}^{(\kappa)}, \boldsymbol{Y}_{02j}^{(\kappa)}, \ldots, \boldsymbol{Y}_{0pj}^{(\kappa)}]^T \end{aligned} \tag{7.58}$$

[3] this simple extension has been derived independently from the work published in 2017 [32].

where the vectors $\boldsymbol{u}_{0ij}^{(\kappa)}$ and $\boldsymbol{y}_{0ij}^{(\kappa)}$ and the matrices $\boldsymbol{Y}_{0ij}^{(\kappa)}$ (for $1 = 1, \ldots, p$) are defined as in (7.52) and (7.53). Some regularization to the possibly rank-deficient square matrix $(\boldsymbol{Y}_{0j}^{T(\kappa)}\boldsymbol{Y}_{0j}^{(\kappa)})$ can be applied also in this case to avoid ill-conditioning in the LS solution of the overdetermined set of equations $\boldsymbol{Y}_{0j}^{(\kappa)}\boldsymbol{a}_{0j}^{(\kappa)} = \boldsymbol{y}_{0j}^{(\kappa)}$.

After a number of iterations, the estimated coefficients $\hat{\boldsymbol{a}}_{0j}^{(\hat{\kappa})}$ for which the energy of $\boldsymbol{u}_j^{(\hat{\kappa})}$ is the lowest are chosen as the coefficients of the polynomial $\hat{A}_{0j}(q^{-1})$. The estimates $\hat{B}_{0ij}(q^{-1})$ of the numerator polynomials are then readily obtained as

$$\hat{B}_{0ij}(q^{-1}) = H_{0ij}(q^{-1})\hat{A}_{0j}(q^{-1}) - q^{-(M+1)}\hat{A}_{0j}(q)R_{0ij}(q^{-1}), \tag{7.59}$$

with $R_{0ij}(q^{-1})$ defined as in (7.55).

7.3.5 Nearly-common excess-phase zeros modeling

The last step required is the modeling of the diagonal matrix $\boldsymbol{\mathcal{F}}$ in (7.31), whose elements are the AP functions built from excess-phase zeros common to all (probabilistic) TFs $\boldsymbol{\mathcal{H}}_{(:,j)}(q^{-1}) = \boldsymbol{B}(q^{-1})\boldsymbol{A}^{-1}(q^{-1})$, for the jth loudspeaker. The excess-phase zeros of a TF $\mathcal{H}_{(i,j)}(z^{-1}) = {}^{B_{ij}(z^{-1})}/_{A_j(z^{-1})}$ are the zeros of the AP excess-phase TF $\tilde{B}_{ij}(z^{-1}) = {}^{B_{ij}(z^{-1})}/_{\beta_{ij}(z^{-1})}$, with $\beta_{ij}(z^{-1})$ the minimum-phase spectral factor of $B_{ij}(z^{-1})$ which can be obtained using, e.g., the cepstral method [45]; the vector containing the coefficients of $B_{ij}(q^{-1})$ has to be padded with a large number of zeros in order to obtain a reliable factorization, with $\beta_{ij}(z^{-1})$ of the same order as $B_{ij}(z^{-1})$ and the excess-phase polynomial $\tilde{B}_{ij}(z^{-1})$ truncated (for practical reasons) to a certain order.

Excess-phase zeros modeling (AP-BU method)

The zeros of each excess-phase TF $\tilde{B}_{ij}(q^{-1}) = {}^{B_{ij}(q^{-1})}/_{\beta_{ij}(q^{-1})}$ can be estimated using a slightly modified version of the iterative part of the BU method (called here AP-BU) and computing the roots of the numerator of the estimated AP TF (or equivalently the reciprocal of the roots (poles) of the denominator). The cost function for the proposed modification is given by

$$\begin{aligned} J_{\text{AP-BU}} &= \sum_{i=1}^{p} ||\tilde{E}_{ij}(q^{-1})||_2^2 = \sum_{i=1}^{p} ||\tilde{B}_{ij}(q^{-1}) - \hat{\tilde{\mathcal{B}}}_{ij}(q^{-1})||_2^2 \\ &= \sum_{i=1}^{p} \left|\left| \tilde{B}_{ij}(q^{-1}) - \frac{q^{-\tilde{M}}Q_{ij}(q)}{Q_{ij}(q^{-1})} \right|\right|_2^2. \end{aligned} \tag{7.60}$$

The estimation of the denominator polynomial $Q_{ij}(q^{-1})$ of order $\tilde{M}$ is performed as in the BU method, with the only difference that the set of linear equations has a slightly different form, due to the fact that the numerator polynomial is the reciprocal of $Q_{ij}(q^{-1})$

$$Y_{ij}^{(\kappa)}(q^{-1})\left[q^{-(\tilde{M}-1)}\tilde{Q}_{ij}^{(\kappa)}(q)\right] = U_{ij}^{(\kappa)}(q^{-1}) - q^{-\tilde{M}}Y_{0ij}^{(\kappa)}(q^{-1}) + q^{-K}, \tag{7.61}$$

with $Y_{ij}^{(\kappa)}(q^{-1}) = q^{-K}\tilde{B}_{ij}(q)/Q_{ij}^{(\kappa-1)}(q^{-1})$ and $Q_{ij}^{(\kappa)}(q^{-1}) = 1 + \sum_{\mu=1}^{M} a_{\mu ij}^{(\kappa)} q^{-\mu} = 1 + \tilde{Q}_{ij}^{(\kappa)}(q^{-1})$.

The order $\tilde{M}$ can be estimated a priori by evaluating the unwrapped phase response $\phi_{\tilde{B}}(\omega)$ of the AP function $\tilde{B}_{ij}(q^{-1})$ at π as

$$\tilde{M} = -\frac{\phi_{\tilde{B}}(\pi)}{\pi} \tag{7.62}$$

This result comes from the properties of the sum of two AP functions [333], for which the phase of an AP TF is given by the following relation

$$\phi_{\tilde{B}}(\omega) = -\tilde{M}\omega - 2\sum_{m=1}^{\infty}\frac{\mathcal{S}_m^{\tilde{B}}\sin(m\omega)}{m}, \tag{7.63}$$

with $\mathcal{S}_m^{\tilde{B}}$ the first-order root moments of $\tilde{B}_{ij}(q^{-1})$ ($m \in \mathbb{Z}$), which gives the model order of the AP TF in (7.62) when evaluated at π.

In practice, the AP model order is increased to $\tilde{M}_2 = \tilde{M} + 2$ to account for real zeros at 0 Hz and at the Nyquist frequency $f_s/2$. A way to control how the identification of the zeros performs, is to evaluate the equivalent modeling error $\tilde{E}_{ij}(q^{-1}) = \left[\tilde{B}_{ij}(q^{-1}) - 1\right] - \left[\tilde{\mathcal{B}}_{ij}(q^{-1}) - 1\right]$ and compare the magnitude responses of $\left[\tilde{B}_{ij}(q^{-1}) - 1\right]$ and $\left[\tilde{\mathcal{B}}_{ij}(q^{-1}) - 1\right]$, which have the same poles as $\tilde{B}_{ij}(q^{-1})$ and $\tilde{\mathcal{B}}_{ij}(q^{-1})$. An example is given in Figure 7.5. Notice that poles very close to the origin, which comes from the fact that the order is increased w.r.t. to the actual order, can be safely removed, since it would have almost no contribution to the AP response. Also notice that, when limiting the equalization to LFs (e.g. $f_S' = 441$ Hz), the model order $\tilde{M}$ is usually quite small (e.g. between 0 and 5), with the extreme case in which $B_{ij}(q^{-1})$ is minimum-phase (i.e. $\tilde{B}_{ij}(z^{-1}) = 1$). To account for this case, when the energy of $\tilde{E}_{ij}(q^{-1})$ is above a certain threshold, the AP-BU algorithm is run again with $\tilde{M}_2 \leftarrow \tilde{M}_2 - 1$.

Nearly-common excess-phase zeros clustering

When the zeros of each excess-phase TF $\tilde{B}_{ij}(q^{-1})$ are correctly identified, the zeros that are common to all TFs have to be determined for each jth loudspeaker

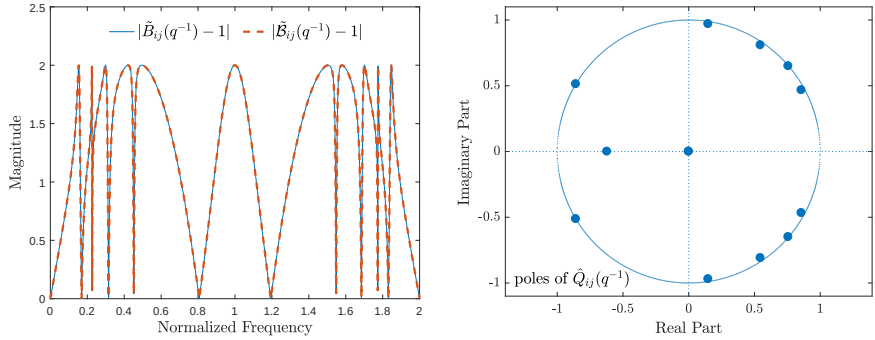

Figure 7.5: Magnitude responses of $\left[\tilde{B}_{ij}(q^{-1})-1\right]$ and $\left[\tilde{\mathcal{B}}_{ij}(q^{-1})-1\right]$ (which they perfectly overlap) and the corresponding poles (Listener 1-left, FL loud., $f'_S = 881$ Hz).

(with $i = 1,\ldots,p_s$ and p_s the number of receivers selected for the design of the compensator). In practice, there are no actual common zeros, so that the excess-phase compensators have to be built from sets of nearly-common zeros. The problem is that the compensator will introduce pre-ringing in the equalized responses, with level depending on the distance between the zeros introduced in the compensator and the actual excess-phase zeros of the TFs.

A strategy to cluster nearly-common zeros was introduced in [334, 224] in which clusters are considered to be invertible if the level of preringing that is introduced is kept below a predefined threshold. The center of the clusters, i.e. the candidate zeros eligible to be included in $\boldsymbol{\mathcal{F}}_*$, are chosen to be the excess-phase zeros $z_{om} = r_{om}e^{j\omega_{om}} \in \boldsymbol{z}_o$ of the complex average $\tilde{B_{oj}}(q^{-1})$ of the p_s excess-phase TF (estimated as above), which were proved [224] to represent the weighted average of the zeros z_{ij} of the individual excess-phase TF $\tilde{B}_{ij}(q^{-1})$. It was also shown in [334, 224] how the preringing level introduced for each equalized TF can be quantified based on the distance between each zero $z_{om} \in \boldsymbol{z}_o$ (with $m = 1,\ldots,\tilde{M}_o$) and the actual zeros $z_{im} \in \boldsymbol{z}_i$ (with $m = 1,\ldots,\tilde{M}_i \geq \tilde{M}_o$) belonging to the corresponding clusters $\mathcal{C}_m$. The main result is that the level of pre-ringing, introduced by a compensator built from an excess-phase zero $z_{om} = r_{om}e^{j\omega_{om}} \in \boldsymbol{z}_o$ at receiver position i (for loudspeaker j), measured at a given time instant t_r before the direct sound, can be quantified as

$$L^m_{\mathrm{dB}} = 20\log_{10}(C_{im}r_{om}^{-\kappa_r}) \tag{7.64}$$

with $\kappa_r = \lceil t_r f_S \rceil$, and

$$C_{im} = \frac{|z_{om} + \epsilon|^2 - |z_{om}|^2}{|z_{om}|^2 \cos \Phi}$$

$$\text{with} \quad \Phi = \arctan\left(\frac{\frac{2\Re(\epsilon)|z_{om}|^2}{|z_{om}+\epsilon|^2-|z_{om}|^2} - \Re(z_{om})}{\Im(z_{om})} \right), \tag{7.65}$$

$$\epsilon = z_{om} - z_{im}\,.$$

A constraint on the admissible pre-ringing level at time $t < t_r$ can be easily imposed by setting $L^m_{\text{dB}} < L_{\max}$.

The clustering algorithm proposed in [224] is reported in Appendix C.2. The algorithm is run separately from the other modeling tasks, because the common zeros should be determined based on the loudspeaker and microphone selected for the desgin of the equalizer. What differentiate this algorithm from conventional clustering algorithms is the fact that exactly one zero from each TF should be included in each cluster. It was shown in [224], that the pre-ringing constraint (with given κ_r and $L_{\max}$) defines a cluster $\mathcal{C}_m$ whose dimensions are determined by the radius of its central zero z_{om}; the cluster gets larger when z_{om} moves away from the unit circle. This means that is it more likely to find 'invertible' clusters not too close to the unit circle. Another method, alternative to the clustering approach, would be to use a common-denominator AP-BU method, possibly using different values for the model order $\tilde{M}$, and then check for the pre-ringing constraint on the estimated common zeros.

In practice, it is likely that the zeros in some of the clusters do not respect the constraint, with the result of the cluster not being 'invertible'. In order to obtain a useful excess-phase compensator in this case, two options are contemplated. The first option consists of increasing the radius r_{om} of the zero at the center of the cluster that does not respect the constraint in order to reduce the level of pre-ringing in (7.64), as suggested in [335]. By doing so, the residual pre-ringing is compressed in time, but amplified in level. In order to compensate for this amplification, another AP has to be included in the compensator, where the excess-phase zero has a larger radius than r_{om}, but the same angle. The radius can be optimized in such a way that the pre-ringing constraint is satisfied (see [335]).

Here, a second option is suggested, which consists of relaxing the constraint and making it frequency-dependent, according to perceptual consideration about pre-ringing at different frequencies. The pre-ringing can be controlled defining the parameters $L_{\max}$ and t_r, which determine the level of the pre-ringing at a particular time before the desired component. This way, the pre-ringing

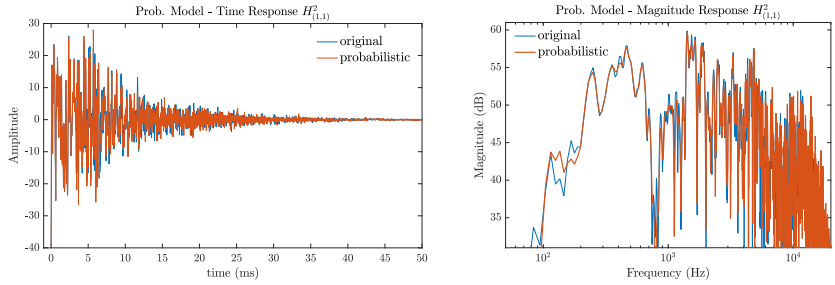

Figure 7.6: Time and magnitude responses of the original and probabilistic TF.

introduced by the controller can be traded off against the equalization quality. It follows that the audibility of the pre-ringing artifacts in the equalized responses should be evaluated.

7.4 Simulation results

In this section some examples are provided to give an idea of the results that can be expected in the modeling stage. Simulation results for the equalization are described in the original report, not included here.

7.4.1 Modeling results

In the examples shown here, the modeling of the response for the front-left loudspeaker and the right microphone of the dummy head at listening position 1 is considered.

Probabilistic modeling and virtual receivers

The probabilistic modeling uses a variable LP filter as described in Section 7.3.1, using frames of 1 ms, with the cut-off frequency reducing linearly from the Nyquist frequency to $f_c = 800Hz$ in 7 ms. The length of the spectral envelope window was set to $N_{\text{tr}} = 1500$. Figure 7.6 shows an example of the result of the probabilistic modeling, showing the variations in the time and magnitude response. Notice that at LFs these variations are of the order of just a few dBs, which seems reasonable to represent the spatial variations around the microphone positions.

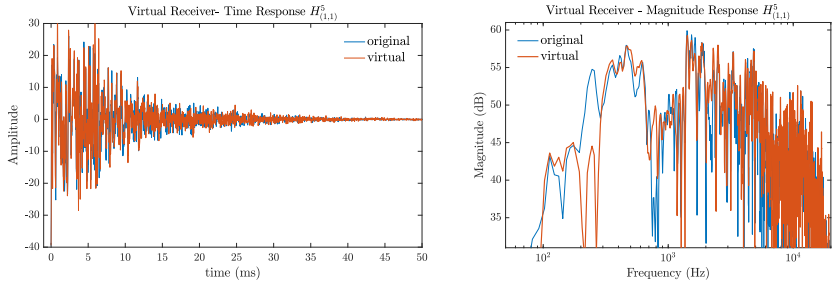

Figure 7.7: Time and magnitude responses of the original TF and a virtual receiver TF.

Figure 7.7 gives an example response of a virtual receiver. Here, a variation $\sigma_v = 0.003$ applied to the nominal part of the actual TF was used. It is quite difficult to assess how representative these virtual responses are of actual responses at points close in space to the actual receivers. Ideally, the number of measurement positions would be large enough not to require the use of virtual receivers.

TF modeling

In the following, some examples of the resulting TF models are shown. The model order used are $M = 25$ for the non-probabilistic modeling $\boldsymbol{\mathcal{H}} = \boldsymbol{B}\boldsymbol{A}^{-1}$ (which is not being used in the design stage, but is put here for reference), $M_0 = 18$ for the modeling of the nominal part of the TF $\boldsymbol{\mathcal{H}}_0 = \boldsymbol{B}_0\boldsymbol{A}_0^{-1}$, and $M_1 = 9$ for the modeling of the spectral shaping filters $\boldsymbol{\mathcal{H}}_1 = \boldsymbol{B}_1\boldsymbol{A}_1^{-1}$. These model orders proved to be a good trade-off between model accuracy and efficiency. The regularization parameter for the modeling algorithms was chosen as $\lambda = 10^{-12}$, that seems large enough to avoid rank deficiency problems of the design matrix in the BU and CD-BU algorithms.

Figure 7.8 shows the common denominator modeling without the probabilistic model $\mathcal{H}_{ij} = B_{ij}/A_j$, showing almost perfect approximation of the TF. In Figure 7.9, the corresponding nominal model $\mathcal{H}_{0ij} = B_{0ij}/A_{0j}$ obtained with the CD-BU method is given for $M_0 = 18$; a lower model order was used in order to facilitate the design stage, especially when multiple loudspeakers are used. Some modeling error is present, which is however acceptable, especially considering that a certain degree of variation was introduced in the nominal part also at LF in the previous step. Figure 7.10 shows the modeling result of the BU method for the shaping filter $\mathcal{H}_{1j} = B_{1j}/A_{1j}$ with good approximation; $\mathcal{H}_{1j}(q^{-1})$ has degree depending on the length N_{tr} of the triangular window used

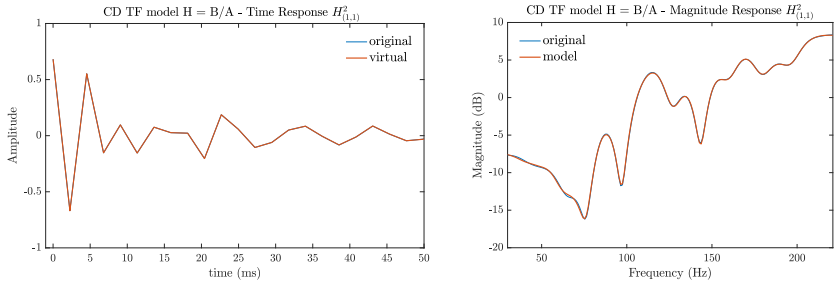

Figure 7.8: Time and magnitude responses of the original resampled TF and the (non probabilistic) modeled TF. Model order $M = 25$.

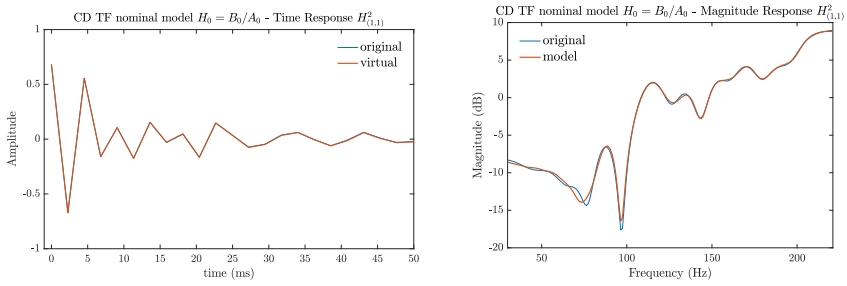

Figure 7.9: Time and magnitude responses of the nominal TF and the nominal modeled TF. Model order $M_0 = 18$.

in the probabilistic modeling, which is chosen to be lower than the degree of $\mathcal{H}_{0ij}(q^{-1})$. Also in this case, some modeling error is acceptable, so that the use of a lower order M_1 is possible. Notice the dynamic range in the magnitude response of the nominal model and of the shaping filter. The difference is due to the fact that the energy of the shaping filter at very LFs is very small (and it would have been zero if an LP/HP filter with shorter transition bandwidth would have been used). In practice, this mean that the contribution of the shaping filter model $\mathcal{H}_{1j}$ is rather limited at LFs.

Regarding the modeling of the excess-phase zeros, the AP-BU method is able to provide perfect modeling of the AP TF $\tilde{B}_{ij}$, when the latter is not truncated to a too low degree, with the model order $\tilde{M}$ set as in (7.62). Being the number of excess-phase zeros very small in the narrow-band range considered, using high-order TFs (i.e. long polynomials $\tilde{B}_{ij}(q^{-1})$ after the minimum-phase/AP decomposition) does not lead to computational issues in the AP-BU algorithm. This is true especially in the LFs, where the TF are decimated to a very low sample frequency. Zeros with very large radius, corresponding to poles close to

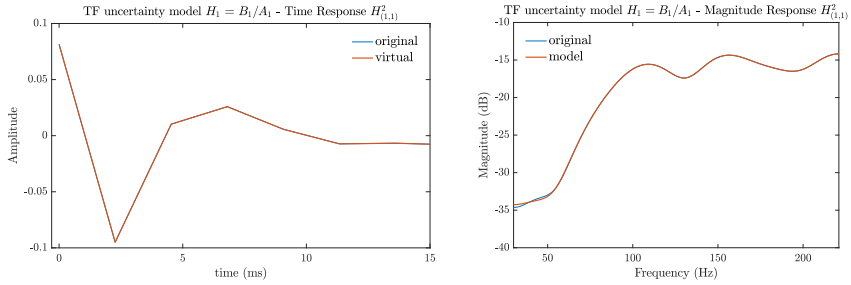

Figure 7.10: Time and magnitude responses of the resampled shaping filter and the modeled one. Model order $M_1 = 9$.

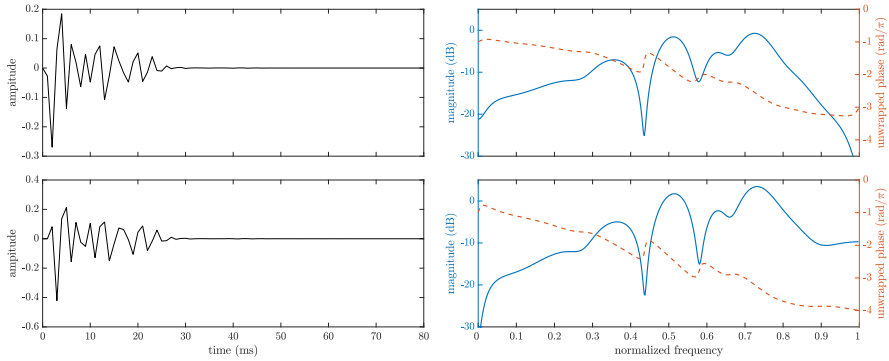

Figure 7.11: Modeled primary loudspeaker AIR (left) and TF (right) at receiver position 1 (top) and at receiver position 2 (bottom).

the origin, which are the result of the absence of zeros at the zero frequency and/or at the Nyquist frequency, can be safely removed in the construction of the AP TFs $\mathcal{F}_j$. An example was already given in Figure 7.5.

7.4.2 Pre-ringing control

In order to illustrate the effect of controlling the pre-ringing, a simulation has been carried out with two speakers and two microphones. The modeled TF of the primary speaker is shown in Fig. 7.11. The equalizer is computed for two different choices regarding the pre-ringing: first for $t_r = 50\,\text{ms}$, and second for $t_r = 5\,\text{ms}$. The level $L_{\max}$ is set to $-60\,\text{dB}$ in both cases.

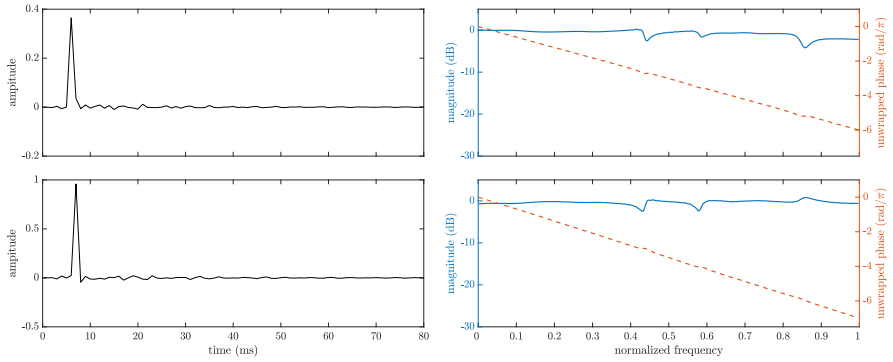

Figure 7.12: Equalized primary loudspeaker AIR (left) and TF (right) at receiver position 1 (top) and at receiver position 2 (bottom). One complex pair of nearly-common zeros included in $\boldsymbol{\mathcal{F}}(q^{-1})$.

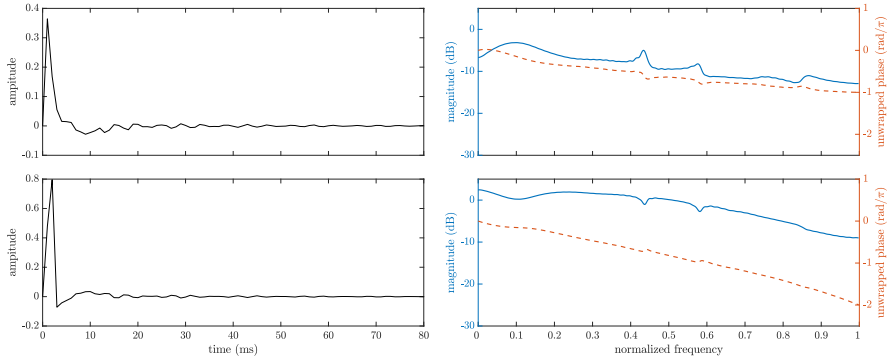

Figure 7.13: Equalized primary loudspeaker AIR (left) and TF (right) at receiver position 1 (top) and at receiver position 2 (bottom). No nearly-common zeros included in $\boldsymbol{\mathcal{F}}(q^{-1})$.

In the first case, the clustering algorithm finds one complex nearly-common zero pair for the primary speaker, while in the second case, no nearly-common zeros are found, such that $\boldsymbol{\mathcal{F}}(q^{-1})$ reduces to an identity. The target TF of the equalized system was set to be one. The equalized TFs and AIRs represented by $\boldsymbol{\mathcal{H}}^{\text{eq}} \triangleq \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{R}}$ are shown for both cases in Fig. 7.12 and Fig. 7.13, respectively. As can be seen, the TF equalization works significantly better in the first case, i.e. when nearly-common zeros are found, than in the second case, when no

nearly-common zeros are found. From the analysis of the AIR in Figure 7.12, however, it can be seen that some pre-ringing is introduced, whose effect should be assessed from a perceptual point of view. The tuning of the parameters $L_{\max}$ and t_r is therefore an important aspect to be considered.

7.5 Conclusion and future work

The focus of this chapter was on the equalization at LFs of a MIMO sound reproduction system. The solution, mainly adopted from [225], is based on a polynomial-based framework, which finds its roots in the control theory field. A careful and comprehensive analysis of this solution to the design of a MSE-optimal equalizer was the scope of this chapter, which comes together with a working implementation of the equalization design procedure. The basic idea of the adopted solution is to equalize the response of a primary loudspeaker measured at different positions inside a given listening region. A given number of loudspeakers are used in support of the primary loudspeaker in order to reduce the deviations from the desired target response.

A summary of the theoretical solution to the equalization problem, both for SIMO and MIMO systems, was provided, with insight of a more intuitive interpretation of such solution. The final equalizer consists of three parts: a minimum-phase part intended to equalize an average of the minimum-phase part of the TF, an excess-phase part meant to compensate for nonminimum-phase distortions common to all TFs (associated to one loudspeaker), and a residual part intended to equalize an average of the residual unequalized, non-common responses and to attain the desired target response.

One of the issues is the fact that the excess-phase part of the equalizer requires a set of excess-phase zeros common to all TFs. Only nearly common zeros can be found in practice, with the distance between zeros related to different TFs determining the level of pre-ringing (or pre-echo) in the equalized response. A nearly-common excess-phase zeros clustering algorithm based on a constraint of maximum admissible level of pre-ringing was implemented, with the trade-off between equalization performance and pre-ringing artifacts easily controllable by means of a small number of intuitive parameters.

Another issue is robustness of the equalizer; the fact that the equalizer is designed starting from a limited number of spatially sparse measurement points may produce a solution which is not effective in all points inside the listening region. A solution based on a probabilistic model was used to introduce a certain level of variability in the TFs, especially at mid/high frequencies and for late reverberation.

Detailed instructions on how to model the TFs were given; algorithms for efficient TF probabilistic modeling, both with individual or common denominators, were presented, as well as an algorithm to accurately estimate the excess-phase zeros of the TFs. A way of trading-off between the levels of pre-ringing and equalization has been discussed. Practical aspects of the equalizer design, such as the approximation of the minimum-phase spectral factor and its inversion, or the design of the target response and the weighting matrices, and simulation results pertaining the performance of the equalizer with varying number of microphones and loudspeakers can be found in the original report.

Chapter 8

Conclusion

This thesis presented research spanning a wide range of topics in room acoustic signal processing, from measuring and analyzing room impulse responses (RIRs), through their modeling and identification, to the application of signal enhancement algorithms aimed at improving the sound quality of acoustic signals in rooms. Such a wide scope is the result of the belief that, to be able to design effective algorithms for room acoustic signal enhancement (RASE) applications, it is important to first analyze and understand the characteristics of the acoustic response of a room, so that efficient models and identification algorithms can be developed. The focus of our investigation has been directed mainly, but not exclusively, to the low-frequency region in small rooms, where strong modal resonances are sparse and unevenly distributed in space and frequency, resulting in large variations of sound pressure level, which have a detrimental effect on sound quality and represent one of the main problems to be tackled by algorithms for RASE.

To summarize, the major objectives were the following: (i) to characterize and analyze the room acoustics in the critical region of modal frequencies; (ii) to develop efficient parametric models and the relative parameter estimation algorithms for modeling room responses; (iii) to apply the developed models and algorithms in a system identification framework using adaptive filters; and, finally, (iv) to design and implement effective low-complexity solutions for some of the problems encountered in RASE applications, with a special focus on digital equalization.

Room impulse response measurements and analysis at low frequencies

Concerning the first objective, a new database of RIRs has been introduced in **Chapter 2**, measured in a rectangular room using subwoofers as sound sources. The SUBRIR database aims to provide reliable acoustic measurements within the frequency region of modal resonances, which are not offered by databases available, and it is expected to find application in the testing of RASE algorithms intended for music reproduction and in the validation of physical models for room acoustics. The challenges in measuring RIRs at low frequencies (LFs), related to high levels of ambient noise and the strongly nonlinear behavior of the subwoofer, have been addressed. The exponential sine-sweep (ESS) method was chosen for its desirable properties in terms of robustness to noise and room transfer function (RTF) variations, and, most of all, for its ability to reject part of the regular and irregular nonlinear artifacts. Unfortunately, the ESS method is not immune to all kind of artifacts, as odd-order harmonic nonlinearities and impulsive distortions partially overlap with the causal response of the room. A careful calibration of the measuring equipment is then necessary to prevent these artifacts to occur in the first place. Limiting the loudspeaker level, however, results in poor signal-to-noise ratio (SNR) in the recorded signals. Even though the SNR can be improved in postprocessing by synchronous averaging of multiple RIRs, the level of the so-called noise floor is likely to be too high to be able to obtain reliable estimates of the frequency-dependent reverberation time (RT) at very LFs using standard procedures. Thus, a procedure has been suggested which uses a fixed-bandwidth cosine-modulated filterbank to reduce the influence of the band-pass filters, and estimates the RT from a noiseless approximation of the RIRs obtained with the models and algorithms described in Part II. The analysis of the retrieved RIRs and of the estimated RT revealed the presence of strong room modes, with modal frequencies in good accordance with their theoretical values. Of particular interest is the very energetic first axial mode with a decay of almost 1.5 s, which is almost three times the RT at other frequencies, and the partial presence of the cavity mode response, even though the latter may not be excited by commercial subwoofers with a high-pass filter built in.

Our contributions in the context of room acoustic measurements can be summarized as follows:

- A novel database of RIRs measured in the modal frequency region.
- Best practice recommendations for acoustic measurements at very LFs.
- A reliable procedure for estimating reverberation time at very LFs.

Parametric modeling of room acoustics using OBF models

The second major objective regarded room acoustic modeling by means of parametric models. A consistent part of thesis has been dedicated to the investigation of parametric models based on orthonormal basis functions (OBFs) and to the development of identification algorithms for the estimation of the model parameters. These scalable iterative algorithms are based on the desirable properties of OBF models and on their previously unexplored interpretation as an approximation of the RTF resulting from a superposition of a finite number of resonant responses. The main idea of employing OBF models is that a good approximation of the RTF can be obtained by reducing the distance between the model poles and the true poles of the room acoustic system.

In **Chapter 3**, the main properties of OBF models have been discussed, the most important being orthogonality, which provides a numerically well-conditioned estimation problem, with estimates less prone to numerical inaccuracies. The nonlinear problem of estimating the pole parameters was overcome by using a matching pursuit (MP) approach, which performs an iterative grid search on a set of complex-conjugate pairs of stable poles distributed on the unit disc in a way that favors the modeling of sharp resonances at LFs. Orthogonality assures that the pole estimation is numerically well-conditioned. The OBF-MP algorithm developed has been compared to the state-of-the-art method, the warped BU (wBU) method, to all-zero (AZ) and pole-zero (PZ) models in terms of their modeling performance. Simulation results have shown that OBF models, with parameters estimated either with OBF-MP or wBU, are able to achieve a reduction in the approximation error compared to AZ and PZ models, even when the increase in the filter complexity is taken into account, and without the instability problems encountered with PZ models. The same modeling accuracy achieved with conventional models can be provided by OBF models with a reduction in the number of model parameters of roughly 50% in full-band and up to 75% in the low and mid frequencies. The features of the OBF-MP algorithm that offer an advantage compared to the wBU method are the following: (i) *scalability*, given by the iterative pole selection, (ii) *stability*, enforced by the fixed pole grid, which also avoids polynomial factorizations, (iii) *flexibility* in the allocation of the spectral resolution, provided by the freedom of positioning poles on the unit disc based on a desired resolution or prior knowledge, only limited in number by considerations concerning the algorithmic complexity.

In **Chapter 4**, the OBF-MP algorithm has been extended to the common-denominator estimation of multiple RTFs. The simple modification that leads to the OBF-GMP algorithm is intended to reduce the number of parameters required to model the RTFs by estimating a set of poles common

to all source/receiver positions, with position-dependent linear coefficients. Simulation results performed on the SUBRIR database have shown that a further reduction up to 50% in the total number of model parameters can be obtained for the same modeling accuracy achieved with the OBF-MP algorithm.

Our contributions in the context of parametric modeling of room acoustics can be summarized as follows:

- A flexible algorithm delivering efficient, scalable, and stable OBF model estimates from one or multiple RTFs.
- A means of allocating frequency resolution arbitrarily based on a desired resolution or prior knowledge, not available in state-of-the-art methods.
- A reduction in the number of model parameters of 50% compared to AZ and PZ models in full-band, and up to 75% in the low and mid frequencies.
- A further reduction up to 50% in the total number of model parameters in the modeling of multiple RTFs at LFs when a common set of poles is estimated.

Identification of room acoustic systems with OBF adaptive filters

The third major objective was to investigate the applicability of the developed models and algorithms for the identification of room acoustic systems from input-output data. The identification of RIRs, as required by most RASE algorithms, is typically performed using adaptive filters.

Chapter 5, which has been structured to serve as a tutorial on the topic, provides a review of the most important properties of OBF adaptive filters in terms of their error performance and the dynamic behavior of the adaptation. Orthogonality, indeed, allows to develop analysis tools enabling a comparison between the performance of OBF filters, other fixed-poles adaptive filters (FPAFs), and finite impulse response (FIR) filters. In short, orthogonality provides a well-conditioned identification problem under a wide range of conditions, which translates to faster convergence and lower variability compared to FPAFs with the same pole set. Regarding the comparison with FIR filters, the difference in performance is regulated to a large extent by the position of the poles in the OBF filter, which determines not only the estimation accuracy based on the distance from the true poles of the system, but also the rate of convergence and the variability of the adaptive coefficients with respect to the spectral characteristics of the noise and of the input signal.

An iterative scalable algorithm has been introduced. The stage-based SB-OBF-GMP algorithm, which exploits the grid-search idea of the OBF-GMP algorithm, identifies a common set of poles of a multiple-input/multiple-output (MIMO) room acoustic system, both from white noise and speech signals. The filter coefficients of the multi-channel OBF filter are adapted with a modified version of the NLMS algorithm, meant to deal with issues at very low model orders, whereas the standard NLMS is used to track the instantaneous correlation between the current residual signals and the output of candidate new sections of the OBF filter, built from poles in the grid. The algorithm has been used to identify the RTFs at LFs in different scenarios for real and simulated rooms. Experimental results have shown that a significant improvement in terms of accuracy and convergence compared to FIR filters and good robustness with respect to position changes within a relatively large area are achieved especially in small or damped rooms, as OBF filters are particularly efficient in approximating room responses with either isolated resonances or highly overlapping ones. The potential reduction in the filter order and the use of a common set of poles may not only bring computational savings, but also help in addressing some of the problems encountered in RASE applications. An example in an acoustic echo cancellation scenario has shown that OBF filters can provide good identification and cancellation performances already for small model orders, for which FIR filters may encounter problems related to undermodeling. Moreover, the possibility of fixing the poles in the filter allows to achieve a desired frequency resolution, either by fixed configurations of the poles or by estimation algorithms, which is a feature not available in standard FIR filters. An example in this regard was given in the context of digital equalization, showing how the poles of the inverse filter can be estimated directly from input-output data, thus allocating resolution where needed, while keeping the order of the equalizer as low as possible.

Our contributions in the context of identification of room acoustic systems can be summarized as follows:

- A review of OBF adaptive filters and of the properties of the most common adaptation algorithms.
- A scalable algorithm capable of identifying a common set of poles of a multi-channel room acoustic system, both from white noise and speech signals.
- An analysis of identification results in different scenarios at LFs with respect to the characteristics of the room, both real and simulated.

- Improvements in terms of identification accuracy and convergence compared to FIR filters, as well as robustness with respect to changes in the microphone positions, especially in small or damped rooms.
- A discussion on the applicability of OBF adaptive filters to RASE applications, not only in terms of efficiency, but also with respect to the possibility of addressing some of the most common problems.
- An example in the context of acoustic echo cancellation (AEC) at LFs, showing good identification and cancellation performances already for small model orders, for which FIR filters may encounter problems related to undermodeling.
- An example in the context of room response equalization (RRE) exemplifying how the poles of a low-order equalizer can be estimated directly from input-output data using an OBF adaptive filter, thus allocating resolution where needed.

Equalization of loudspeaker, room and car cabin responses

The fourth and last main objective of this thesis was that of designing and implementing effective low-complexity solutions for digital equalization applications.

Chapter 6 introduces an iterative procedure for designing a low-order equalizer using parametric infinite impulse response (IIR) filters, specifically peaking and shelving filters, to be used in the compensation of loudspeaker and room magnitude responses. Despite the fact that these filters can only perform a minimum-phase equalization, the proposed procedure minimizes the sum of square errors between the system and the target complex responses. Moreover, the previously unexplored orthogonality property of a particular implementation form of these parametric filters allows to compute the least squares optimal value for the gain parameter in closed-form. This property has been exploited in the initialization of the filter parameters, which are then refined by a line search optimization. Other advantages of this procedure, compared to state-of-the-art methods, are an improved mathematical tractability of the equalization problem, with the possibility of computing analytical expressions of the gradients, an improved initialization of the parameters, including the global gain of the equalizer, the incorporation of shelving filters in the optimization procedure, robustness to local minima, and a more accentuated focus on the equalization of the more perceptually relevant spectral peaks. Examples of loudspeaker and room response equalization have proved that an effective low-order equalizer can be designed, with good performances with respect to the state-of-the-art

methods for a number of different error functions and perceptual objective measures. Moreover, the proposed design procedure can be easily extended to multi-point equalization, by means of a prototype average response, and to minimum-phase transfer function modeling.

Chapter 7 describes the practical implementation of an existing solution for the design of a robust MIMO equalizer meant to correct for nonminimum-phase distortions in low-frequency acoustic responses of a car cabin. The approach aims at equalizing the response of a primary loudspeaker in the listening area by partial response inversion and by sound field superposition provided by a number of support loudspeakers. The implementation of the adopted solution, which is based on a polynomial-based control system framework, requires to model different aspects of the acoustic transfer functions (ATFs). Most of the details regarding modeling were not described in the original work, which required further investigation into modeling techniques. In order to improve robustness to ATF variations within the listening area, a probabilistic modeling approach is applied, which uses a variable low-pass filter to decompose an ATF into a deterministic low-frequency component, and a stochastic high-frequency component. The same idea has been also used to generate 'virtual' ATFs to compensate for the limited number of measurements available in the listening area. The common-denominator modeling of the ATFs was also recommended. For this purpose, the common-denominator BU method has been derived, with the inclusion of a regularization parameter to mitigate rank deficiency problems. Another requirement is the estimation of approximately common excess-phase zeros for the design of a stable noncausal all-pass (AP) filter, whose role is to remove phase distortions common to all positions in the listening area. The idea is to first carry out a minimum-phase/all-pass decomposition of the ATFs, and then model the AP components. A modified version of the BU method has been suggested to find the zeros of the AP responses, which can then be clustered together. It has been found that the model order can be estimated a priori from the unwrapped phase response of the AP responses. It turns out that the order of the AP responses is rather low, typically between 0 and 5, suggesting that the ATF at LFs is indeed approximately minimum-phase. The modified algorithm proved to be able to model the AP responses with high accuracy.

Our contributions in the context of equalization can be summarized as follows:

- A novel low-order equalizer automatic design procedure using parametric IIR filters, specifically peaking and shelving filters.
- An improved mathematical tractability of the minimum-phase equalization problem using IIR parametric filters.

- An improved automatic initialization of the filter parameters and a stronger focus on the equalization of spectral peaks.
- An interpretation and implementation guidelines of an existing solution for the equalization of nonminimum-phase distortions in low-frequency acoustic responses of a car cabin.
- Efficient methods for common-poles PZ modeling and AP modeling for nonminimum-phase equalization.

Industrial relevance of the research work

Most RASE applications, such as equalization, dereverberation, acoustic echo and feedback cancellation among others, are nowadays commonly found in most communication devices, such as teleconference systems, and audio equipment. In all these applications, the desired signal enhancement task has to be performed in real-time, which imposes limitations in terms of computational complexity and latency. Thus, the efficiency of the digital filters employed in RASE applications is an important aspect to consider. The wide-spread use of FIR filters is motivated by their simplicity and the large availability of effective solutions, but it may not be the best choice in terms of efficiency.

The core topic of this thesis, namely room acoustic modeling and identification using OBF models and OBF adaptive filters (Chapters 3 to 5), presented an alternative to FIR filters. Especially when its poles are fixed, an OBF filter possesses similar characteristics to an FIR filter, with the added benefits of having an IIR and the possibility of arbitrarily allocating spectral resolution, potentially yielding higher accuracy with lower complexity. The actual advantage of using OBF adaptive filters is however conditional to the position of the poles and their distance from the actual poles of the RTF (or its inverse). This means that in practical RASE applications, a calibration phase would be necessary to first estimate a set of, possibly common, poles from input-output data, which can than be kept fixed during the signal processing task, or slowly adapted to keep track of variations in time of the room acoustics (if computational requirements allow). A calibration phase is already employed in existing teleconference applications, such as Skype™, suggesting that the pre-estimation of the poles is a viable idea. Therefore, we believe that our work could encourage the adoption of fixed-poles IIR filters in a number of audio signal processing applications.

As a remark, we also believe that the room acoustic modeling algorithms developed in this thesis could find application in the context of artificial reverberation. Methods for synthesizing reverberation by means of a parallel

of resonating filters have recently come on the market[1], which could benefit from our estimation algorithms to parametrize and then synthesize the acoustic response of real rooms.

The rest of this thesis (Chapters 2, 6, and 7) dealt with different aspects of loudspeaker and room equalization, and is the result of a collaboration with three different companies operating in the field of high-tech communication systems and consumer audio products, namely Bang & Olufsen A/S (Denmark), Televic N.V. (Belgium), and Premium Sound Solutions N.V. (Belgium).

In recent years, digital room correction (DRC) solutions appeared as a complementary tool to many high-end sound reproduction systems. The aim of DRC is to equalize the loudspeaker/room response when two or more loudspeakers are placed inside a listening room. The idea of DRC is first to perform a series of RIR measurements at different microphone positions for fixed loudspeaker positions, e.g. using the ESS method, and then automatically design an equalizer capable of compensating for the deviations of the measured responses from a desired response. An example of such automatic equalizer design for magnitude response equalization was given in Chapter 6. The proposed procedure uses possibly the most common type of equalizer found in commercial products, such that the estimated parameters for the equalizer could be directly used in practice. A possible refinement of the proposed design procedure could consist in the inclusion of psychoacoustical criteria in the optimization process, so as to obtain a perceptually better equalization. An equalizer also able to correct for nonminimum-phase distortions can be designed when multiple loudspeakers are available, e.g. when listening to stereophonic content from a 5.1 surround system. An analysis and guidelines were presented in Chapter 7 for the implementation of an existing solution for the automatic design of such an equalizer, which is already used in commercial DRC systems by Dirac Research AB (Sweden).

Another requirement of DRC systems is the availability of measured RIRs. The recommendations for performing acoustic measurements and the reliable procedure for estimating reverberation time at LFs presented in Chapter 2 could be put into practice as analysis tools in this context. Possible improvements to make the measurement procedure fully automatic concern the calibration of the loudspeaker output level to mitigate the effects of the nonlinear behavior of the loudspeaker, which could also be complemented by a method for modeling and possibly control these nonlinearities.

[1] for instance, the Moodal© spectral resonator plugin developed by Tritik (France) consists, most likely, of a parallel of resonators.

Suggestions for future research

A consistent part of this thesis has been devoted to modeling and identification of room acoustics using OBF models and their related adaptive filters. Even though the theory of OBF models has been largely treated in the system identification literature and previously applied in the field of audio and acoustic signal processing, our work, to the best of our knowledge, is the first investigation of the applicability of multi-pole OBF adaptive filters to RASE applications. We believe, indeed, that the properties of OBF adaptive filters, described in Chapter 5, and the fact that most algorithms developed for FIR filters are easy to extend to OBF filters, make them good candidates to tackle past and future issues in room acoustic signal processing.

The most important problem, which remains partially unsolved, is the identification of the pole parameters from input-output data, especially in the case of non-stationary and non-white signals. Indeed, the advantages of adopting OBF filters are conditional to the fact that the poles of the filter are close enough to the system poles. Regarding the approach adopted in this work, some issues are still to be solved, mostly related to the algorithmic computational complexity and the suboptimal solution given by the discrete nature of the set of candidate poles. Although useful to avoid the use of nonlinear optimization algorithms and to arbitrarily allocate frequency resolution, the use of a grid search may prevent to fully exploit the modeling capabilities of OBF models. One option would be to use the proposed algorithms employing a coarse grid as a way of initializing the values of the pole parameters, which can be then refined by nonlinear optimization techniques or recursive algorithms [170]. Simplified expressions for the gradients, as suggested in [171], could be used, although possible convergence to local minima has to be considered.

An alternative would be to combine FIR and OBF filters together, similarly to what has been proposed in [265] for the parallel filter model. The difference, in this case, is that the resulting combined filter will have an infinite impulse response (IIR) and will be orthonormal, thus keeping the desirable numerical properties of OBF filters[2]. Different strategies could be adopted in this case. For instance, an FIR filter with order equal to the maximum expected acoustic delay could be placed in front of an OBF filter; this way, the FIR filter would take care of modeling the acoustic delay and, possibly, the early reflections of the RIR, whereas the OBF filter, with fixed or adaptive poles, would model the reverberation tail efficiently.

The computational issues could be addressed by means of approaches in the frequency domain [336] and/or by subband techniques [337]. For instance, a

[2]after all, an FIR filter can be interpreted as an OBF filter with poles in the origin.

subband modeling algorithm could be developed using the so-called frequency zooming (FZ) ARMA method [207], where the complex-valued subband responses could be then modeled using OBF models with complex poles only, with the real-valued responses obtained only at synthesis by complex-conjugation. Subband modeling would be useful also to investigate the modeling and identification capabilities of OBF models in different regions of the spectrum. It can be expected that the increased modal density and the increased absorption would result in favorable conditions for the applicability of OBF filters. More absorption also implies shorter responses, such that the actual advantage over FIR filters should be assessed also at higher frequencies.

Another open question, only partially addressed in this work, is related to the identification of common poles. Although the idea proved to be useful to reduce the total number of parameters, we did not address the issue related to spatial sampling [338], i. e. related to the minimum number of microphones and their relative distance necessary to accurately estimate the room modes at LFs. Providing an answer to this question would open new possibilities, enabling the development of RASE algorithms robust to RTF variations, and possibly interpolation methods of RTFs at LFs [197]. Results in this direction would be also useful for other applications, such as for simulating moving sources or microphones in artificial reverberation, especially in the case of the modal reverberator [83] described in the introduction of this thesis.

One of the assumptions commonly made in room acoustic signal processing, including our work, is the linearity of the room acoustic system. However, as discussed in Chapter 2, nonlinearities may appear in the acoustic response when a loudspeaker is driven at high levels, as often happens with subwoofers or loudspeakers in small devices. Modeling nonlinearities, such as the harmonic distortions in a loudspeaker, may then provide ways to control or cancel them. Nonlinear OBF filters, specifically those derived from Legendre polynomials [339], have been suggested for nonlinear modeling of loudspeaker responses, showing improved performances with respect to the better-known Hammerstein and Volterra models [340]. These filters are also orthoghonal and linear in the filter coefficients, such that adaptive filters can be applied similarly to other linear OBF filters. An alternative, suggested very recently [341], is to estimate the Volterra kernels through regularized single-pole OBF models. The combination of a filter of this kind with a linear OBF filter, and possibly an FIR filter, would generate an orthonormal filter with whom to model and identify both the linear and nonlinear components of a loudspeaker/room acoustic system.

Appendix A

Appendix to Chapter 5

A.1 Gradient expressions for the poles

Here the full and approximated expressions for the gradients with respect to the k^{th} pole pair defined as $\boldsymbol{p}_k = [\zeta_k, \eta_k] = [-2\rho_k \cos\vartheta_k, \rho_k^2]$ are given. After simple calculations and some reordering (and omitting the time index n and the q operator, s.t. $D_k \equiv D_k(n,q)$ and $\chi_k \equiv \chi_k(n) = \{\zeta_k, \eta_k\}$), the partial derivatives appearing in the second and third terms of equation (5.35) are given by

$$\frac{\partial \tilde{y}_k(n)}{\partial \chi_k} = \left(\theta_k^+ \frac{\partial N_k^+}{\partial \chi_k} + \theta_k^- \frac{\partial N_k^-}{\partial \chi_k}\right) x_k(n) - \frac{1}{D_k}\frac{\partial D_k}{\partial \chi_k}\tilde{y}_k(n)\,, \tag{A.1}$$

$$\frac{\partial \tilde{y}_i(n)}{\partial \chi_k} = \left(\theta_i^+ N_i^+ + \theta_i^- N_i^-\right) \frac{\prod_{j=k+1}^{i-1} A_j}{D_i} \frac{\partial \bar{D}_k}{\partial \chi_k} x_k(n) - \frac{1}{D_k}\frac{\partial D_k}{\partial \chi_k}\tilde{y}_i(n)\,, \tag{A.2}$$

where $x_k(n) = P_k \prod_{j=1}^{k-1} A_j u(n)$ (the output of the k^{th} resonator) and $\tilde{y}_i(n), \forall i$, are readily available signals. By noticing in (5.3) that $|1 \pm p_k| = \sqrt{1 \mp \zeta_k + \eta_k}$,

the partial derivatives in (A.1) and (A.2) with respect to ζ_k are given as

$$\frac{\partial N_k^+}{\partial \zeta_k} = -\frac{\sqrt{c}}{2\sqrt{a}}(z^{-1}-1)\,,$$

$$\frac{\partial N_k^-}{\partial \zeta_k} = \frac{\sqrt{c}}{2\sqrt{b}}(z^{-1}+1)\,,$$

$$\frac{\partial D_k}{\partial \zeta_k} = z^{-1}\,,\quad \frac{\partial \bar{D}_k}{\partial \zeta_k} = z^{-1}\,, \tag{A.3}$$

with $a = 1-\zeta_k+\eta_k$, $b = 1+\zeta_k+\eta_k$ and $c = {}^{(1-\eta_k)}/_2$, whereas the same expressions with respect to η_k are given as

$$\frac{\partial N_k^+}{\partial \eta_k} = \frac{1}{2}\left[\frac{\sqrt{c}}{\sqrt{a}} - \frac{\sqrt{a}}{2\sqrt{c}}\right](z^{-1}-1)\,,$$

$$\frac{\partial N_k^-}{\partial \eta_k} = \frac{1}{2}\left[\frac{\sqrt{c}}{\sqrt{b}} - \frac{\sqrt{b}}{2\sqrt{c}}\right](z^{-1}+1)\,,$$

$$\frac{\partial D_k}{\partial \eta_k} = z^{-2}\,,\quad \frac{\partial \bar{D}_k}{\partial \eta_k} = 1\,. \tag{A.4}$$

The first term of expression (A.2), however, can be computationally expensive, especially for small k. For this reason, a simplification has been proposed in [171] assuming slow convergence of the parameters and poles close to the unit circle, which allows to approximate the gradient only by the term in (A.1).

A.2 The BB-OBF-GMP identification algorithm

Here, the BB-OBF-GMP identification algorithm, introduced in [154], is described. Similarly to the SB-OBF-GMP algorithm, the purpose of the BB-OBF-GMP algorithm is to estimate the poles of a multi-channel OBF adaptive filter from single-input/multiple-output (SIMO) data. The algorithm aims to build a SIMO adaptive OBF filter including one common pole pair in the set of active poles $\boldsymbol{p}_m^{\mathcal{A}}$ at a time, so that the sum of instantaneous squared errors of the R acoustic channels in (5.38) is minimized. The algorithm is designed for stationary white noise input signals, and, for this reason, it uses the LMS algorithm for adapting the linear coefficients of the multi-channel OBF adaptive filter, according to

$$\hat{\boldsymbol{\Theta}}_M(n+1) = \hat{\boldsymbol{\Theta}}_M(n) + \mu\boldsymbol{\kappa}(n,\boldsymbol{p}_m^{\mathcal{A}})\boldsymbol{\epsilon}_m(n)\,. \tag{A.5}$$

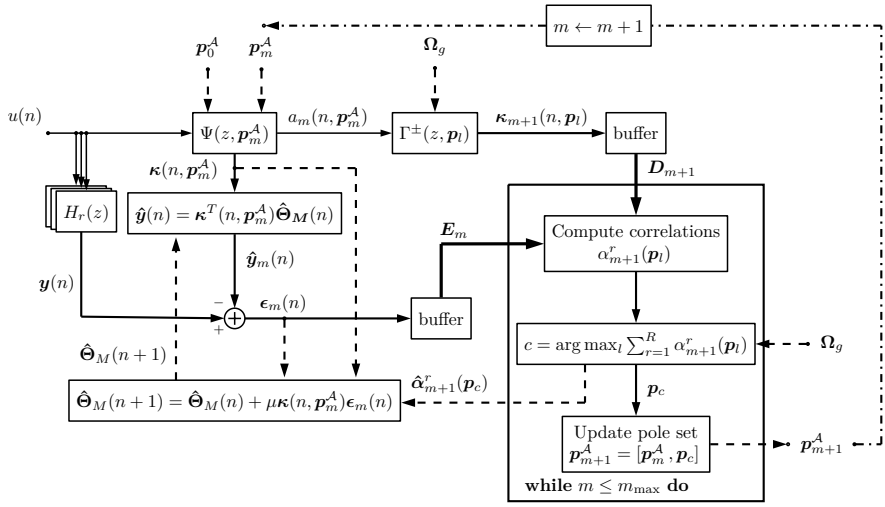

Figure A.1: The schematics of the BB-OBF-GMP identification algorithm. Inbound dashed lines represent initial conditions and inputs, while outbound dashed lines represent outputs.

Also in this algorithm, the poles of the adaptive OBF filter are estimated using a grid-based matching pursuit algorithm, which is depicted in Figure A.1 with a slightly simplified notation. The main difference compared to the SB-OBF-GMP algorithm is that the correlation coefficients between the candidate intermediate signals and the prediction error signals are not tracked in time, but obtained by least squares (LS) estimation on blocks of data of length N_b samples. In each block (i.e. every N_b samples), one pole pair is selected from the grid of L candidate poles pairs $\boldsymbol{p}_l \in \boldsymbol{\Omega}_g$ as the one that produces the pair of candidate intermediate signals $\boldsymbol{\kappa}_{m+1}(n, \boldsymbol{p}_l) = [\kappa_{m+1}^{+}(n, \boldsymbol{p}_l), \kappa_{m+1}^{-}(n, \boldsymbol{p}_l)]$ that is mostly correlated with the last N_b samples of the prediction error signal produced in each acoustic channel considered. The vector $\boldsymbol{\kappa}_{m+1}(n, \boldsymbol{\Omega}_g) = [\boldsymbol{\kappa}_{m+1}(n, \boldsymbol{p}_1), \ldots, \boldsymbol{\kappa}_{m+1}(n, \boldsymbol{p}_L)]^T$ of the intermediate signals computed for all the pole pairs $\boldsymbol{p}_l \in \boldsymbol{\Omega}_g$ is then collected for N_b samples and stacked to build the dictionary $\boldsymbol{D}_{m+1}$, which is a $N_b \times 2L$ matrix, whose columns $\boldsymbol{d}_{m+1}^{\pm}(\boldsymbol{p}_l)$ are the last N_b samples of the L pairs of intermediate signals $\boldsymbol{\kappa}_{m+1}(n, \boldsymbol{p}_l)$. At each block, a pole pair is selected based on the correlation of the pairs of intermediate signal vectors $\boldsymbol{d}_{m+1}^{\pm}(\boldsymbol{p}_l)$ with the last N_b samples of the prediction error vector $\boldsymbol{\epsilon}_m(n)$, stacked to form the $N_b \times R$ matrix $\boldsymbol{E}_m$, whose columns $\boldsymbol{\epsilon}_m^r$ contain the last N_b samples of the prediction error signals $\epsilon_m^r(n)$. The correlation of each pair of intermediate signal vectors $\boldsymbol{d}_{m+1}^{\pm}(\boldsymbol{p}_l)$ with the vector $\boldsymbol{\epsilon}_m^r$ for the r-th

channel is computed as

$$\alpha_{m+1}^{r}(\boldsymbol{p}_l) = \sqrt{\left[\alpha_{m+1}^{r+}(\boldsymbol{p}_l)\right]^2 + \left[\alpha_{m+1}^{r-}(\boldsymbol{p}_l)\right]^2}$$
$$= \sqrt{\left(\left[\boldsymbol{d}_{m+1}^{+}(\boldsymbol{p}_l)\right]^T \boldsymbol{\epsilon}_m^r\right)^2 + \left(\left[\boldsymbol{d}_{m+1}^{-}(\boldsymbol{p}_l)\right]^T \boldsymbol{\epsilon}_m^r\right)^2}, \tag{A.6}$$

where the correlation coefficients $\alpha_{m+1}^{r\pm}(\boldsymbol{p}_l)$ can be obtained as the elements of the $2L \times R$ matrix $\boldsymbol{\Lambda}_{m+1} = \boldsymbol{D}_{m+1}^T \boldsymbol{E}_m$.

The pair of candidate intermediate signal vectors in the dictionary having maximum correlation with the prediction error matrix $\boldsymbol{E}_m$ is selected according to $c = \arg\max_l \sum_{r=1}^{R} \alpha_{m+1}^{r}(\boldsymbol{p}_l)$ and the corresponding pole pair $\boldsymbol{p}_c \in \boldsymbol{\Omega}_g$ is added to the active pole set $\boldsymbol{p}_{m+1}^{\mathcal{A}}$ and included in the multi-channel OBF adaptive filter. The linear filter coefficients $\hat{\boldsymbol{\theta}}_{m+1}^{r}(n) = [\hat{\theta}_{m+1}^{r+}(n), \hat{\theta}_{m+1}^{r-}(n)]$ are set equal to the correlation coefficients $\hat{\boldsymbol{\alpha}}_{m+1}^{r}(\boldsymbol{p}_c) = [\hat{\alpha}_{m+1}^{r+}(\boldsymbol{p}_c), \hat{\alpha}_{m+1}^{r-}(\boldsymbol{p}_c)]$ (with $r = 1, \ldots, R$), normalized w.r.t. the norm of $\boldsymbol{d}_{m+1}^{\pm}(\boldsymbol{p}_c)$. In this way, the linear filter parameters are already close to their optimal value, so that a small value for μ can be used in order to achieve better accuracy with the LMS algorithm. Finally, the algorithm moves to the next block ($m \leftarrow m + 1$) where another pole pair is estimated from the last N_b samples of the prediction error signals and of the $(m+1)$-th candidate intermediate signals as described above, until a desired number of pole pairs $m_{\max}$ has been reached or some other stopping criterion based on the error in (5.38) is satisfied.

Notice that the pole selection criterion, based on the maximum of the average correlation over the R channels, is only valid under the assumption that the input signal has a flat spectrum. In case of a non-white input signal, the correlation coefficients should be extracted from the matrix $\boldsymbol{\Lambda}_{m+1} = (\boldsymbol{D}_{m+1}^T \boldsymbol{D}_{m+1})^{-1} \boldsymbol{D}_{m+1}^T \boldsymbol{E}_m$. However, the matrix inversion, or alternatively the computation of the pseudo-inverse, increases the computational complexity and is prone to numerical inaccuracies, which is one of the reasons that led to the SB-OBF-GMP algorithm.

Appendix B

Appendix to Chapter 6

B.1 SSE minimum-phase cost function

Here the sum of squared errors (SSE) cost function in (6.4) for the minimum-phase equalization problem is analyzed, using the relation by which the frequency response of a minimum-phase transfer function $H(k)$ can be written as

$$H(k) = |H(k)|e^{-j\mathcal{H}\{\ln|H(k)|\}} \tag{B.1}$$

For simplicity, the weighting matrix in (6.4) is set to $W(k) = 1$ and the notation is simplified. The cost function in (6.4) can be elaborated in terms of magnitude and phase of the frequency responses involved, as shown in (B.2), where the Euler's rule and the linear property of the Hilbert transform have been used ($\{\cdot\}^*$ indicates complex conjugation). It can be noticed that the optimal equalizer, for which $E^2(k) = 0$, is defined as $F(k) = \frac{T(k)}{H(k)} = \frac{|T(k)|}{|H(k)|}e^{j(\phi_T(k)-\phi_H(k))}$.

This cost function has a quadratic form, which assumes large values whenever the power of the equalized magnitude response is significantly larger than the power of the target response, and whenever the difference between their magnitude responses on a natural logarithmic scale is large (i.e. when the value

of cos is far from one).

$$\begin{aligned}
E^2(k) =&(H(k)F(k) - T(k))^*(H(k)F(k) - T(k)) \\
=&(F^*(k)H^*(k) - T^*(k))(H(k)F(k) - T(k)) \\
=&|F(k)H(k)|^2 + |T(k)|^2 \\
&- |F(k)H(k)T(k)|\big(e^{-j(\phi_F(k)+\phi_H(k)-\phi_T(k))} - e^{j(\phi_F(k)+\phi_H(k)-\phi_T(k))}\big) \\
=&|F(k)H(k)|^2 + |T(k)|^2 - 2|F(k)H(k)T(k)|\cos(\phi_F(k) + \phi_H(k) - \phi_T(k)) \\
=&|F(k)H(k)|^2 + |T(k)|^2 - 2|F(k)H(k)T(k)| \\
&\cos(-j\mathcal{H}\{\ln|F(k)|\} - j\mathcal{H}\{\ln|H(k)| + j\mathcal{H}\{\ln|T(k)|\}) \\
=&|F(k)H(k)|^2 + |T(k)|^2 \\
&- 2|F(k)H(k)T(k)|\cos(-j\mathcal{H}\{\ln|F(k)| + \ln|H(k)| - \ln|T(k)|\}) \\
=&|\tilde{H}(k)|^2 + |T(k)|^2 - 2|\tilde{H}(k)T(k)|\cos(-j\mathcal{H}\{\ln|\tilde{H}(k)| - \ln|T(k)|\})
\end{aligned} \tag{B.2}$$

To simplify the analysis even further, a zero-phase, flat target response ($T(k) = 1$) is considered, for which the cost function assumes the form

$$E^2(k) = |\tilde{H}(k)|^2 + 1 - 2|\tilde{H}(k)|\cos(-j\mathcal{H}\{\ln|\tilde{H}(k)|\}) \tag{B.3}$$

In this case, it is easy to see that the error is larger when the equalized magnitude response of $\tilde{H}(k)$ has values larger than one, with the error increasing more than linearly for increasing magnitude, which explains the focus on the equalization of strong peaks.

B.2 The orthogonality property of the Regalia-Mitra parametric filters

A brief explanation of the property introduced in Section 6.4 is provided here. Define a vector $\boldsymbol{x}$ containing N samples of the input signal $x(n)$ and the vector $\boldsymbol{z}$ containing N samples of the output signal $z(n)$ of the all-pass filter $A_m(z)$. A known property of an all-pass filter is the preservation of the energy, such that the energy of the input signal is equal to the energy of the output signal

$$\sum_{n=-\infty}^{\infty} |z(n)|^2 = \sum_{n=-\infty}^{\infty} |x(n)|^2 \tag{B.4}$$

or, in terms of vector inner products, $\langle \boldsymbol{z}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{x} \rangle$. With reference to Figure 6.3b, the output of the filter $y_m(n)$ is formed from the weighted summation of two signals, $y^{\eta}(n) = x(n) + z(n)$ and $y^{\beta}(n) = x(n) - z(n)$. The orthogonality of these two signals can be assessed from their inner product,

$$\langle \boldsymbol{y}^n, \boldsymbol{y}^r \rangle = \langle \boldsymbol{x} + \boldsymbol{z}, \boldsymbol{x} - \boldsymbol{z} \rangle \tag{B.5}$$

$$= \langle \boldsymbol{x}, \boldsymbol{x} \rangle - \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{z}, \boldsymbol{x} \rangle - \langle \boldsymbol{z}, \boldsymbol{z} \rangle = 0, \tag{B.6}$$

which follows from the equality stated above, $\langle \boldsymbol{z}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{x} \rangle$, and from $\langle \boldsymbol{x}, \boldsymbol{z} \rangle = \langle \boldsymbol{z}, \boldsymbol{x} \rangle$.

B.3 Gain LS estimation

The first-order partial derivative of the cost function in (6.19) w.r.t. the gain parameter V is given by equation (B.7).

$$\begin{aligned}
\frac{\partial \epsilon_m^{\mathrm{SSE}}}{\partial V} =& \frac{1}{N} \sum_k \Big(\frac{1}{2} W(k) H_s(k) F_m^{\beta}(k) \Big)^* \Big(W(k) \big[H_s(k) \cdot F_m(k) - T(k) \big] \Big) \\
=& \frac{1}{N} \sum_k \Big(\frac{1}{2} F_m^{\beta *}(k) H_s^*(k) W^*(k) \Big) \Big(\frac{1}{2} W(k) H_s(k) F_m^{\eta} \Big) \\
&+ \frac{1}{N} \sum_k \Big(\frac{1}{2} F_m^{\beta *}(k) H_s^*(k) W^*(k) \Big) \Big(\frac{V}{2} W(k) H_s(k) F_m^{\beta} \Big) \\
&- \frac{1}{N} \sum_k \Big(\frac{1}{2} F_m^{\beta *}(k) H_s^*(k) W^*(k) \Big) \Big(W(k) T(k) \Big).
\end{aligned} \tag{B.7}$$

Since orthogonality in the time domain (see Appendix B.2) implies orthogonality also in the frequency domain, the first summation in the equation equals zero ($F_m^{\eta}(k)$ and $F_m^{\beta}(k)$ are orthogonal). By setting $\partial \epsilon_m^{\mathrm{SSE}} / \partial V = 0$, the following equation is obtained

$$\begin{aligned}
V \sum_k \Big(\frac{1}{2} F_m^{\beta *}(k) H_s^*(k) W^*(k) \Big) \Big(\frac{1}{2} W(k) H_s(k) F_m^{\beta} \Big) \\
= \sum_k \Big(\frac{1}{2} F_m^{\beta *}(k) H_s^*(k) W^*(k) \Big) \Big(W(k) T(k) \Big).
\end{aligned} \tag{B.8}$$

which provides an estimate for the gain as

$$\hat{V} = \frac{\sum_k \big[F_m^{\beta *}(k) H_s^*(k) W^*(k) \big] \big[W(k) T(k) \big]}{\sum_k \big[F_m^{\beta *}(k) H_s^*(k) W^*(k) \big] \big[W(k) H_s(k) F_m^{\beta}(k) \big]} \tag{B.9}$$

which is equivalent to the expression in (6.17).

B.4 Gradients and Jacobians expressions

Based on the method chosen to perform the line search at each stage, the search direction $\boldsymbol{p}_i$ requires the computation of the gradients $\boldsymbol{\nabla}\mathcal{F}_s^{(i)} = {}^{\partial\mathcal{F}_s^{(i)}}/_{\partial\boldsymbol{\theta}_s^{(i)}}$ (in the SD and quasi-Newton methods) or of the Jacobians $\boldsymbol{\nabla}\boldsymbol{e}_s^{(i)} = {}^{\partial\boldsymbol{e}_s^{(i)}}/_{\partial\boldsymbol{\theta}_s^{(i)}}$ (in the GN method), where[1]

$$\mathcal{F}_s^{(i)} = \frac{1}{N}\sum_k e(k,\boldsymbol{\theta}_s^{(i)},V_s^{(i)})^2 = \frac{1}{N}\,\boldsymbol{e}_s^{(i)H}\boldsymbol{e}_s^{(i)} \tag{B.10}$$

The gradient of the cost function can be written as

$$\boldsymbol{\nabla}\mathcal{F}_s^{(i)} = \frac{2}{N}\sum_k \frac{\partial e_s^{(i)}(k)}{\partial\boldsymbol{\theta}_s^{(i)}} e_s^{(i)}(k) = \frac{2}{N}\boldsymbol{\nabla}\boldsymbol{e}_s^{(i)H}\boldsymbol{e}_s^{(i)}, \tag{B.11}$$

where the Jacobian is given by, for $k = 1,\ldots,N$,

$$\frac{\partial e_s^{(i)}(k)}{\partial\boldsymbol{\theta}_s^{(i)}} = \frac{1}{2}W(k)H_{s-1}(k)\frac{\partial F_{m_s}(k,\boldsymbol{\theta}_s^{(i)},V_s^{(i)})}{\partial\boldsymbol{\theta}_s^{(i)}}, \tag{B.12}$$

so that the partial derivatives $\frac{\partial F_{m_s}^{(i)}(k)}{\partial\boldsymbol{\theta}_s^{(i)}}$ for peaking and shelving filters are required. In order to use the Newton method, the exact Hessian $\boldsymbol{\nabla}^2\mathcal{F}_s^{(i)} = {}^{\partial^2\mathcal{F}_s^{(i)}}/_{\partial\boldsymbol{\theta}_s^{(i)2}}$ should be computed. Analytic expressions for the second-order partial derivatives can be obtained, but the advantages of using the Newton method are outweighed by a higher complexity.

Peaking filters

The frequency response of peaking filters in the linear-in-the-gain (LIG) form can be written, substituting (6.13) in (6.9), as

$$F_2(k) = \frac{(1+a)(1+2dk+k^2)+V(1-a)(1-k^2)}{2(1+d(1+a)k+ak^2)}, \tag{B.13}$$

[1]For convenience here k stands for $e^{-j\omega_k/f_s}$, so that the transfer functions are evaluated at $z = e^{j\omega_k/f_s}$, and z^{-1} can be substituted by k. $\{\cdot\}^H$ indicates Hermitian transpose.

and its first-order partial derivatives w.r.t. the parameters a and $\sigma = \cos^{-1}(-d)$ as

$$\frac{\partial F_2(k)}{\partial a} = \frac{(1-V)(1-k^2)(1+2dk+k^2)}{2(1+d(1+a)k+ak^2)^2} \tag{B.14}$$

$$\frac{\partial F_2(k)}{\partial \sigma} = \frac{\sin(\sigma)(1-V)(1-k^2)(1-a^2)k}{2(1+d(1+a)k+ak^2)^2}. \tag{B.15}$$

Shelving and high-pass (HP)/low-pass (LP) filters

The frequency response of LF and high frequency (HF) shelving filters in the LIG form can be written, substituting (6.10) in (6.9), respectively as

$$F_1^{\mathrm{LF}}(k) = \frac{(1+a)(1-k)+V(1-a)(1+k)}{2(1-ak)} \tag{B.16}$$

$$F_1^{\mathrm{HF}}(k) = \frac{(1+a)(1+k)+V(1-a)(1-k)}{2(1+ak)}, \tag{B.17}$$

and its partial derivatives w.r.t. the parameter a as

$$\frac{\partial F_1^{\mathrm{LF}}(k)}{\partial a} = \frac{(1-V)(1-k^2)}{2(1-ak)^2} \tag{B.18}$$

$$\frac{\partial F_1^{\mathrm{HF}}(k)}{\partial a} = \frac{(1-V)(1-k^2)}{2(1+ak)^2}. \tag{B.19}$$

The frequency response for HP and LP filters is obtained by setting $V = 0$ in (B.16) and (B.17), respectively, and their partial derivatives w.r.t. a by setting $V = 0$ in (B.18) and (B.19).

Appendix C

Appendix to Chapter 7

C.1 SIMO MSE-optimal equalizer

Here more details about the interpretation of the simplified expression in (7.13) for the SIMO MSE-optimal equalizer, reported here for convenience,

$$\mathcal{R} = q^{-d_0}\mathcal{F}_* \frac{A}{\beta}\Big\{q^{d_0}\mathcal{F}\,\frac{\boldsymbol{B}_*}{\beta_*}\frac{\boldsymbol{D}}{E}\Big\}_+,$$

are provided. The polynomial matrix $\boldsymbol{B}$ of the model numerator polynomials B_i can be approximated as a combination of two parts, one common to all B_i's, here called B^c and a non-common part $\boldsymbol{B}^n$, so that each numerator polynomial can be written as $B_i \approx B^c B_i^n$. Also, the RMS average can be divided in a common and a non-common part as $\beta \approx \beta^c\beta^n$, where β^c is the minimum-phase equivalent of B^c, which together define the all-pass filter $\mathcal{F} = B^c/\beta^c$ in (7.8). The expression above can then be written as

$$\mathcal{R} \approx q^{-d_0}\frac{B^c_*}{\beta^c_*}\frac{A}{\beta^c\beta^n}\Big\{q^{d_0}\frac{B^c}{\beta^c}\,\frac{B^c_*\boldsymbol{B}^n_*}{\beta^c_*\beta^n_*}\frac{\boldsymbol{D}}{E}\Big\}_+,$$

And simplified further, noticing that $B^c_*B^c = \beta^c_*\beta^c$,

$$\mathcal{R} \approx q^{-d_0}\frac{B^c_*}{\beta^c_*}\frac{A}{\beta^c\beta^n}\Big\{q^{d_0}\frac{\boldsymbol{B}^n_*}{\beta^n_*}\frac{\boldsymbol{D}}{E}\Big\}_+.$$

If now the same decomposition is used, the modeled TFs can be written as

$$\mathcal{H} \approx \frac{B^c\boldsymbol{B}^n}{A}.$$

We first apply the average minimum-phase equalizer A/β, thus obtaining

$$\hat{\mathcal{H}} = \mathcal{H}\frac{A}{\beta} \approx \frac{B^c\boldsymbol{B}^n}{A}\frac{A}{\beta^c\beta^n} = \frac{B^c\boldsymbol{B}^n}{\beta^c\beta^n}.$$

Then we apply the time-reversed and delayed all-pass response $q^{-d_0}\mathcal{F}_* = q^{-d_0}B_*^c/\beta_*^c$, obtaining the non-common excess-phase all-pass TFs

$$\grave{\mathcal{H}} = \hat{\mathcal{H}}q^{-d_0}\mathcal{F}_* \approx \frac{B^c\boldsymbol{B}^n}{\beta^c\beta^n}q^{-d_0}\frac{B_*^c}{\beta_*^c} = q^{-d_0}\frac{\boldsymbol{B}^n}{\beta^n}.$$

The previous equation holds exactly if the decomposition $B_i = B^cB_i^n$ actually exists, while pre-ringing is introduced in any other case (i.e. if $\mathcal{F}$ uses nearly-common excess phase zeros instead of truly common excess phase zeros). It is clear now that the argument of the causal operator $\{\cdot\}_+$ has the purpose of implicitly equalizing the non-common excess-phase part of the TFs by means of their average, weighted by the target TFs

$$\acute{\mathcal{H}} = \grave{\mathcal{H}}\Big\{q^{d_0}\frac{\boldsymbol{B}_*^n}{\beta_*^n}\frac{\boldsymbol{D}}{E}\Big\}_+ = q^{-d_0}\frac{\boldsymbol{B}^n}{\beta^n}\Big\{q^{d_0}\frac{\boldsymbol{B}_*^n}{\beta_*^n}\frac{\boldsymbol{D}}{E}\Big\}_+ = q^{-d_0}\frac{\boldsymbol{B}^n}{\beta^n}\Big\{\frac{\boldsymbol{B}_*^n}{\beta_*^n}\frac{\tilde{\boldsymbol{D}}}{E}\Big\}_+,$$

where $\boldsymbol{D} = q^{-d_0}\tilde{\boldsymbol{D}}$ was used in the RHS of the last equation. Note that any acoustic delay in $\boldsymbol{B}_*^n$ is compensated for by the delays in the target $\boldsymbol{D}$, whereas the delay in the first term of the RHS of the equation is the equalization delay. Also notice that, in the SISO case, the non-common terms B^n and β^n are equal to one, so that the compensator has the form (7.14) and the equalized response is given by

$$\acute{\mathcal{H}} = \mathcal{H}\mathcal{R} = \frac{B}{A}\,q^{-d_0}\frac{B_*}{\beta_*}\frac{A}{\beta}\frac{\tilde{D}}{E} = q^{-d_0}\frac{B}{\beta}\frac{B_*}{\beta_*}\frac{\tilde{D}}{E} = \frac{D}{E}, \tag{C.1}$$

i.e. in theory we perform perfect equalization up to the delay d_0. However, note that d_0 is theoretically infinitely long, since it is defined as the length of the all-pass impulse response of $\mathcal{F}$ (i.e. truly perfect equalization requires an acausal filter, which is not possible in practice).

The above interpretation can be extended to the case in which the probabilistic model is used and to the MIMO case. However, the introduction of the probabilistic model and the inclusion of the matrices $\boldsymbol{V}$ and $\boldsymbol{W}$ in the definition of the spectral factor (see e.g. (7.28)), make the interpretation more involved.

C.2 Zero-clustering algorithm

Here, the zero clustering algorithm suggested in [224] is reported, with a slightly different notation. The scope of the algorithm is to classify the excess-phase

zeros of $B_{1j}, \ldots, B_{p_s j}$, identified using the AP-BU method, into separated clusters centered around the excess phase zeros of the complex average TF B_{oj}. The requirement is that each cluster must contain exactly one zero from each TF B_{ij}.

Suppose that B_{oj} contains $\tilde{M}_o$ zeros z_{om} (with $m = 1, \ldots, \tilde{M}_o$) outside the unit circle in the upper half plane, and that each B_{ij} contains $\tilde{M}_i$ zeros $z_{im} \in \boldsymbol{z}_i$ (with $m = 1, \ldots, \tilde{M}_i$ and $i = 1, \ldots, p_s$). In [224], it was assumed that $\tilde{M}_o \leq \tilde{M}_i$; the case in which $\tilde{M}_o > \tilde{M}_i$ can be included anyway, with the difference that not every $z_{om} \in \boldsymbol{z}_o$ will be the center of a cluster. The idea is that each zero $z_{om} \in \boldsymbol{z}_o$ is associated with one zero $z_{im} \in \boldsymbol{z}_i$ for each ith TF, so that $\hat{M} = \min\{\tilde{M}_o, \min_i\{\tilde{M}_i\}\}$ clusters $\mathcal{C}_m$ are being formed, defined as

$$\mathcal{C}_m = \{z_{1k_m^1}, z_{2k_m^2}, \ldots, z_{p_s k_m^{p_s}}\} \tag{C.2}$$

where the indexes k_m^i determine which of the zeros $z_{i1}, \ldots, z_{i\tilde{M}_i} \in \boldsymbol{z}_i$ is to be associated with a certain nominal zero $z_{om} \in \boldsymbol{z}_o$. Another set $\mathcal{Z}_o$ is being used, along with an index set μ, defined as

$$\mu = \{\mu_1, \ldots, \mu_{\bar{M}}\} \subset \{1, \ldots, M_o\} \tag{C.3}$$

$$\mathcal{Z}_o = \{z_{o\mu_1}, \ldots, z_{o\mu_{\bar{M}}}\} \subset \{z_{o1}, \ldots, z_{o\tilde{M}_o}\} \tag{C.4}$$

where μ is always ordered, i.e., $\mu_j < \mu_{j+1}$, $j = 1, ..., \bar{M} - 1$.

Note that $\bar{M}$ is the number of elements in μ and $\mathcal{Z}_o$, which varies between different passes through the algorithm. The algorithm is greedy in the sense that, by a principle of 'mutually nearest neighbors', it prioritizes dense and well separated clusters instead of minimizing a global criterion based on average distances, as is often the case with other clustering algorithms. The algorithm is described in pseudo code below.

Algorithm 4 Zero-clustering algorithm [224]

for $m = 1$ to $\tilde{M}_o$ **do**
  $\mathcal{C}_m \leftarrow \emptyset$
end for
for $i = 1$ to p_s **do**
  $\mathcal{Z}_o \leftarrow \boldsymbol{z}_o$; $\mathcal{X}_0 \leftarrow \emptyset$; $\mu \leftarrow \{1, \ldots, \tilde{M}_o\}$; $\xi \leftarrow \emptyset$;
  repeat
    for j=1 to $\bar{M}$ **do**
      $m \leftarrow \mu_j$
      Let $z_{ik_m^i}$ be the zero $\in \boldsymbol{z}_i$ closest to z_{om};
      Let $z_{ok_m^i}$ be the zero $\in \mathcal{Z}_o$ closest to $z_{ik_m^i}$;
      if $z_{ok_m^i} = z_{om}$ **then**
        Add $z_{ik_m^i}$ to $\mathcal{C}_m$: $\mathcal{C}_m \leftarrow \mathcal{C}_m \cup \{z_{ik_m^i}\}$;
        Remove $z_{ik_m^i}$ from $\boldsymbol{z}_i$: $\boldsymbol{z}_i \setminus \{z_{ik_m^i}\}$;
      else if
        thenAdd z_{om} to $\mathcal{X}_0$: $\mathcal{X}_0 \leftarrow \mathcal{X}_0 \cup \{z_{om}\}$;
        Add m to ξ: $\xi \leftarrow \xi \cup \{m\}$;
      end if
    end for
    $\mathcal{Z}_o \leftarrow \mathcal{X}_0$;
    $\mu \leftarrow \xi$;
  until $\mathcal{Z}_o = \emptyset$ or $\bar{M} < \tilde{M}_o - \tilde{M}_i$
end for

Bibliography

[1] H. Kuttruff, *Room acoustics*. Spon Press, 2009.

[2] L. Beranek, *Concert halls and opera houses: music, acoustics, and architecture*. Springer, 2012.

[3] P. Naylor and N. Gaubitch, *Speech Dereverberation*. Signals and Communication Technology, Springer, 2010.

[4] T. J. Cox, P. D'Antonio, and M. R. Avis, "Room sizing and optimization at low frequencies," *J. Audio Eng. Soc.*, vol. 52, no. 6, pp. 640–651, 2004.

[5] T. J. Cox and P. D'Antonio, *Acoustic absorbers and diffusers: theory, design and application*. CRC Press, 2016.

[6] J. N. Mourjopoulos, "Digital equalization of room acoustics," *J. Audio Eng. Soc.*, vol. 42, no. 11, pp. 884–900, 1994.

[7] L. D. Fielder, "Analysis of traditional and reverberation-reducing methods of room equalization," *J. Audio Eng. Soc.*, vol. 51, no. 1/2, pp. 3–261, 2003.

[8] M. Karjalainen, T. Paatero, J. N. Mourjopoulos, and P. D. Hatziantoniou, "About room response equalization and dereverberation," in *Proc. 2005 IEEE Workshop Appls. Signal Process. Audio Acoust. (WASPAA '05)*, pp. 183–186, IEEE, 2005.

[9] S. Cecchi, A. Carini, and S. Spors, "Room response equalization—a review," *Appl. Sci.*, vol. 8, no. 1, p. 16, 2017.

[10] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 20, no. 5, pp. 1421–1448, 2012.

[11] V. Välimäki, J. Parker, L. Savioja, J. O. Smith, and J. Abel, "More than 50 years of artificial reverberation," in *Proc. AES 60th Int. Conf.: DREAMS (Dereverb. Reverb. Audio Music Speech)*, (Leuven, Belgium), AES, 2016.

[12] D. Griesinger, "Improving room acoustics through time-variant synthetic reverberation," in *Preprints Audio Eng. Soc. Conv. 90*, AES, 1991.

[13] T. Lokki, J. Pätynen, T. Peltonen, and O. Salmensaari, "A rehearsal hall with virtual acoustics for symphony orchestras," in *Preprints Audio Eng. Soc. Conv. 126*, AES, 2009.

[14] M. Vorländer, *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. Springer, 2007.

[15] D. R. Begault, *3D Sound for Virtual Reality and Multimedia*. San Diego, CA, USA: Academic Press Professional, Inc., 1994.

[16] E. De Sena, H. Hacıhabiboğlu, and Z. Cvetković, "Scattering delay network: An interactive reverberator for computer games," in *Proc. AES 41st Int. Conf.: Audio for Games*, AES, 2011.

[17] M. Vorländer, D. Schröder, S. Pelzer, and F. Wefers, "Virtual reality for architectural acoustics," *J. Build. Perform. Simul.*, vol. 8, no. 1, pp. 15–25, 2015.

[18] P. Zahorik, "Assessing auditory distance perception using virtual acoustics," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1832–1846, 2002.

[19] Y. Seki and T. Sato, "A training system of orientation and mobility for blind people using acoustic virtual reality," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 1, pp. 95–104, 2011.

[20] D. Pelegrín-García and J. Brunskog, "Speakers' comfort and voice level variation in classrooms: Laboratory research," *J. Acoust. Soc. Am.*, vol. 132, no. 1, pp. 249–260, 2012.

[21] D. Pelegrín-García, E. De Sena, T. van Waterschoot, M. Rychtáriková, and C. Glorieux, "Localization of a virtual wall by means of active echolocation by untrained sighted persons," *Appl. Acoust.*, vol. 139, pp. 82–92, 2018.

[22] C. Breining, P. Dreiscitel, E. Hansler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, "Acoustic echo control. an application of very-high-order adaptive filters," *IEEE Signal Process. Mag.*, vol. 16, no. 4, pp. 42–69, 1999.

[23] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, S. L. Gay, *et al.*, *Advances in network and acoustic echo cancellation*. Springer, 2001.

[24] T. van Waterschoot and M. Moonen, "Fifty years of acoustic feedback control: State of the art and future challenges," *Proc. IEEE*, vol. 99, no. 2, pp. 288–327, 2011.

[25] S. Müller and P. Massarani, "Transfer-function measurement with sweeps," *J. Audio Eng. Soc.*, vol. 49, no. 6, pp. 443–471, 2001.

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

[27] P. Heuberger, P. van den Hof, and B. Wahlberg, *Modelling and Identification with Rational Orthogonal Basis Functions*. Springer, 2005.

[28] M. Karjalainen and T. Paatero, "Equalization of loudspeaker and room responses using Kautz filters: Direct least squares design," *EURASIP J. Adv. Signal Process.*, vol. 2007, no. 1, pp. 185–185, 2007.

[29] T. Paatero and M. Karjalainen, "Kautz filters and generalized frequency resolution: Theory and audio applications," *J. Audio Eng. Soc.*, vol. 51, no. 1/2, pp. 27–44, 2003.

[30] T. Paatero, "Modeling of long and complex responses using Kautz filters and time-domain partitions," in *Proc. 12th Eur. Signal Process. Conf. (EUSIPCO '04)*, pp. 313–316, 2004.

[31] T. Paatero *et al.*, *Generalized linear-in-parameter models: theory and audio signal processing applications.* PhD thesis, Helsinki University of Technology, 2005.

[32] S. Hashemgeloogerdi and M. Bocko, "High precision robust modeling of long room responses using wavelet transform," in *Proc. 2017 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '17)*, pp. 356–360, IEEE, 2017.

[33] M. Karjalainen and T. Paatero, "Generalized source-filter structures for speech synthesis," in *Proc. 7th Eur. Conf. Speech Comm. Tech.*, 2001.

[34] L. S. Ngia, "Recursive identification of acoustic echo systems using orthonormal basis functions," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 3, pp. 278–293, 2003.

[35] S. Hashemgeloogerdi and M. Bocko, "Inherently stable weighted least-squares estimation of common acoustical poles with the application in feedback path modeling utilizing a Kautz filter," *IEEE Signal Process. Lett.*, vol. 25, no. 3, pp. 368–372, 2018.

[36] J. Zeng and R. de Callafon, "Filters parametrized by orthonormal basis functions for active noise control," in *Proc. ASME 2005 Int. Mech. Eng. Cong. Expo.*, pp. 201–208, ASME, 2005.

[37] F. Jacobsen and P. Juhl, *Fundamentals of general linear acoustics.* John Wiley & Sons Ltd, 2013.

[38] J. Mourjopoulos, "On the variation and invertibility of room impulse response functions," *J. Sound Vib.*, vol. 102, no. 2, pp. 217–228, 1985.

[39] M. Schroeder and K. Kuttruff, "On frequency response curves in rooms. comparison of experimental, theoretical, and Monte Carlo results for the average frequency spacing between maxima," *J. Acoust. Soc. Am.*, vol. 34, no. 1, pp. 76–80, 1962.

[40] M. Stephenson, *Assessing the quality of low frequency audio reproduction in critical listening spaces.* PhD thesis, University of Salford, 2012.

[41] B. Fazenda, *Perception of room modes in critical listening spaces.* PhD thesis, University of Salford, 2004.

[42] P. Rubak and L. G. Johansen, "Coloration in natural and artificial room impulse responses," in *Proc. AES 23rd Int. Conf.: Signal Process. Audio Record. Reprod.*, AES, 2003.

[43] S. Bech, "Timbral aspects of reproduced sound in small rooms. I," *J. Acoust. Soc. Am.*, vol. 97, no. 3, pp. 1717–1726, 1995.

[44] N. Kaplanis, S. Bech, S. H. Jensen, and T. van Waterschoot, "Perception of reverberation in small rooms: a literature study," in *Proc. AES 55st Int. Conf.: Spatial Audio*, AES, 2014.

[45] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-time Signal Processing (2nd Ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1999.

[46] R. P. Genereux, "Adaptive loudspeaker systems: Correcting for the acoustic environment," in *Proc. 8th Int. Conf. Audio Eng. Soc.: The Sound of Audio*, Audio Engineering Society, 1990.

[47] P. Rubak and L. G. Johansen, "Design and evaluation of digital filters applied to loudspeaker/room equalization," in *Preprints Audio Eng. Soc. Conv. 108*, AES, 2000.

[48] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Am.*, vol. 66, no. 1, pp. 165–169, 1979.

[49] L. G. Johansen and P. Rubak, "The excess phase in loudspeaker/room transfer functions: Can it be ignored in equalization tasks?," in *Preprints Audio Eng. Soc. Conv. 100*, 1996.

[50] J. K. Nielsen, J. R. Jensen, S. H. Jensen, and M. G. Christensen, "The single- and multichannel audio recordings database (SMARD)," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC 2014)*, (Antibes-Juan Les Pins, France), pp. 40–44, 2014.

[51] J. Bradley, H. Sato, and M. Picard, "On the importance of early reflections for speech in rooms," *J. Acoust. Soc. Am.*, vol. 113, no. 6, pp. 3233–3244, 2003.

[52] I. Arweiler and J. M. Buchholz, "The influence of spectral characteristics of early reflections on speech intelligibility," *J. Acoust. Soc. Am.*, vol. 130, no. 2, pp. 996–1005, 2011.

[53] A. Lindau, L. Kosanke, and S. Weinzierl, "Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses," in *Preprints Audio Eng. Soc. Conv. 128*, 2010.

[54] G. Defrance, L. Daudet, and J.-D. Polack, "Using matching pursuit for estimating mixing time within room impulse responses," *Acta Acust. united Ac.*, vol. 95, no. 6, pp. 1071–1081, 2009.

[55] W. Klippel, "Tutorial: Loudspeaker nonlinearities - causes, parameters, symptoms," *J. Audio Eng. Soc.*, vol. 54, no. 10, pp. 907–939, 2006.

[56] A. Lundeby, T. E. Vigran, H. Bietz, and M. Vorländer, "Uncertainties of measurements in room acoustics," *Acta Acust. united Ac.*, vol. 81, no. 4, pp. 344–355, 1995.

[57] G.-B. Stan, J.-J. Embrechts, and D. Archambeau, "Comparison of different impulse response measurement techniques," *J. Audio Eng. Soc.*, vol. 50, no. 4, pp. 249–262, 2002.

[58] A. Torras-Rosell and F. Jacobsen, "Measuring long impulse responses with pseudorandom sequences and sweep signals," in *Proc. 39th Int. Congr. Noise Control Eng. (INTER-NOISE 2010)*, (Lisbon, Portugal), 2010.

[59] D. D. Rife and J. Vanderkooy, "Transfer-function measurement with maximum-length sequences," *J. Audio Eng. Soc.*, vol. 37, no. 6, pp. 419–444, 1989.

[60] A. J. Berkhout, D. de Vries, and M. M. Boone, "A new method to acquire impulse responses in concert halls," *J. Acous. Soc. Am.*, vol. 68, no. 1, pp. 179–183, 1980.

[61] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Preprints Audio Eng. Soc. Conv. 108*, (Paris, France), 2000.

[62] A. Torras-Rosell and F. Jacobsen, "A new interpretation of distortion artifacts in sweep measurements," *J. Audio Eng. Soc.*, vol. 59, no. 5, pp. 283–289, 2011.

[63] J. S. Abel and P. Huang, "A simple, robust measure of reverberation echo density," in *Preprints Audio Eng. Soc. Conv. 121*, AES, 2006.

[64] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am.*, vol. 37, no. 3, pp. 409–412, 1965.

[65] M. Karjalainen, P. Ansalo, A. Mäkivirta, T. Peltonen, and V. Välimäki, "Estimation of modal decay parameters from noisy response measurements," *J. Audio Eng. Soc.*, vol. 50, no. 11, pp. 867–878, 2002.

[66] H. W. Löllmann and P. Vary, "Estimation of the frequency dependent reverberation time by means of warped filter-banks," in *Proc. 2011 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '11)*, pp. 309–312, IEEE, 2011.

[67] J. O. Smith, *Introduction to Digital Filters with Audio Applications*. online book, 2007 edition, available: http://ccrma.stanford.edu/∼jos/filters/, accessed Oct. 2014.

[68] G. Long, D. Shwed, and D. Falconer, "Study of a pole-zero adaptive echo canceller," *IEEE Trans. Circuits Syst.*, vol. 34, no. 7, pp. 765–769, 1987.

[69] Y. Tomita, A. A. Damen, and P. M. Van Den Hof, "Equation error versus output error methods," *Ergonomics*, vol. 35, no. 5-6, pp. 551–564, 1992.

[70] K. Steiglitz and L. McBride, "A technique for the identification of linear systems," *IEEE Trans. Autom. Control*, vol. 10, pp. 461–464, 1965.

[71] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[72] C. S. Burrus and T. W. Parks, "Time domain design of recursive digital filters," *IEEE Trans. Audio Electroacoust.*, vol. 18, no. 2, pp. 137–141, 1970.

[73] H. Brandenstein and R. Unbehauen, "Least-squares approximation of FIR by IIR digital filters," *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 21–30, 1998.

[74] T. Young and W. Huggins, "'complementary' signals and orthogonalized exponentials," *IRE Trans. Circuit Theory*, vol. 9, no. 4, pp. 362–370, 1962.

[75] J. L. Walsh, *Interpolation and approximation by rational functions in the complex domain*, vol. 20. Am. Math. Soc., 1935.

[76] V. Y. Pan, "Solving a polynomial equation: some history and recent progress," *SIAM review*, vol. 39, no. 2, pp. 187–220, 1997.

[77] G. A. Sitton, C. S. Burrus, J. W. Fox, and S. Treitel, "Factoring very-high-degree polynomials," *IEEE Signal Process. Mag.*, vol. 20, no. 6, pp. 27–42, 2003.

[78] J. O. Smith and J. S. Abel, "Bark and ERB bilinear transforms," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 6, pp. 697–708, 1999.

[79] J. A. Belloch, F. J. Alventosa, P. Alonso, E. S. Quintana-Ortí, and A. M. Vidal, "Accelerating multi-channel filtering of audio signal on ARM processors," *J. Supercomput.*, vol. 73, no. 1, pp. 203–214, 2017.

[80] B. Bank, "Perceptually motivated audio equalization using fixed-pole parallel second-order filters," *IEEE Signal Process. Lett.*, vol. 15, pp. 477–480, 2008.

[81] B. Bank, "Audio equalization with fixed-pole parallel filters: An efficient alternative to complex smoothing," in *Preprints Audio Eng. Soc. Conv. 128*, 2010.

[82] J. Rämö, V. Välimäki, and B. Bank, "High-precision parallel graphic equalizer," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 12, pp. 1894–1904, 2014.

[83] J. S. Abel, S. Coffin, and K. Spratt, "A modal architecture for artificial reverberation with application to room acoustics modeling," in *Preprints Audio Eng. Soc. Conv. 137*, AES, 2014.

[84] M. Guldenschuh, "Least-mean-square weighted parallel IIR filters in active-noise-control headphones," in *Proc. 22nd Eur. Signal Process. Conf. (EUSIPCO '14)*, pp. 1367–1371, 2014.

[85] G. Ramos, M. Cobos, B. Bank, and J. A. Belloch, "A parallel approach to HRTF approximation and interpolation based on a parametric filter model," *IEEE Signal Process. Lett.*, vol. 24, no. 10, pp. 1507–1511, 2017.

[86] B. Bank, "Computationally efficient nonlinear Chebyshev models using common-pole parallel filters with the application to loudspeaker modeling," in *Preprints Audio Eng. Soc. Conv. 130*, 2011.

[87] B. Bank, "Warped IIR filter design with custom warping profiles and its application to room response modeling and equalization," in *Preprints Audio Eng. Soc. Conv. 130*, AES, 2011.

[88] B. Bank, S. Zambon, and F. Fontana, "A modal-based real-time piano synthesizer," *IEEE Trans. Audio Speech Lang. Process.*, vol. 18, no. 4, pp. 809–821, 2010.

[89] B. Bank and G. Ramos, "Improved pole positioning for parallel filters based on spectral smoothing and multiband warping," *IEEE Signal Process. Lett.*, vol. 18, no. 5, pp. 299–302, 2011.

[90] A. Bjorck, *Numerical methods for least squares problems.* SIAM, 1996.

[91] A. Härmä, M. Karjalainen, L. Savioja, V. Välimäki, U. K. Laine, and J. Huopaniemi, "Frequency-warped signal processing for audio applications," *J. Audio Eng. Soc.*, vol. 48, no. 11, pp. 1011–1031, 2000.

[92] A. Härmä and U. K. Laine, "A comparison of warped and conventional linear predictive coding," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 579–588, 2001.

[93] M. Karjalainen, E. Piirilä, and A. Järvinen, "Loudspeaker response equalisation using warped digital filters," in *Proc. NorSig-96*, pp. 367–370, 1996.

[94] M. Karjalainen, V. Välimäki, H. Penttinen, and H. Saastamoinen, "DSP equalization of electret film pickup for the acoustic guitar," *J. Audio Eng. Soc*, vol. 48, no. 12, pp. 1183–1193, 2000.

[95] M. Tyril, J. A. Pedersen, and P. Rubak, "Digital filters for low-frequency equalization," *J. Audio Eng. Soc.*, vol. 49, no. 1/2, pp. 36–43, 2001.

[96] O. Kirkeby, P. Rubak, L. G. Johansen, and P. A. Nelson, "Implementation of cross-talk cancellation networks using warped FIR filters," in *Proc. AES 16th Int. Conf.: Spatial Sound Reprod.*, 1999.

[97] T. van Waterschoot and M. Moonen, "Adaptive feedback cancellation for audio signals using a warped all-pole near-end signal model," in *Proc. 2008 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '08)*, (Las Vegas, NV, USA), pp. 269–272, Apr. 2008.

[98] L. Ausiello, T. van Waterschoot, and M. Moonen, "A first-order frequency-warped sigma delta modulator with improved signal-to-noise ratio," in *Proc. 16th Eur. Signal Process. Conf. (EUSIPCO '08)*, (Lausanne, Switzerland), Aug. 2008.

[99] G. Ramos, J. J. López, and B. Pueo, "Cascaded warped-FIR and FIR filter structure for loudspeaker equalization with low computational cost requirements," *Digital Signal Process.*, vol. 19, no. 3, pp. 393–409, 2009.

[100] P. Gil-Cacho, T. van Waterschoot, M. Moonen, and S. H. Jensen, "Multi-microphone acoustic echo cancellation using multi-channel warped linear prediction of common acoustical poles," in *Proc. 18th Eur. Signal Process. Conf. (EUSIPCO '10)*, pp. 2121–2125, 2010.

[101] M. Karjalainen, A. Härmä, and U. K. Laine, "Realizable warped IIR filters and their properties," in *Proc. 1997 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '97)*, vol. 3, pp. 2205–2208, IEEE, 1997.

[102] A. Härmä, "Implementation of frequency-warped recursive filters," *Signal Process.*, vol. 80, no. 3, pp. 543–548, 2000.

[103] Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 2, pp. 320–328, 1994.

[104] Y. Haneda, S. Makino, and Y. Kaneda, "Multiple-point equalization of room transfer functions by using common acoustical poles," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 4, pp. 325–333, 1997.

[105] F. Fontana, L. Gibin, D. Rocchesso, and O. Ballan, "Common pole equalization of small rooms using a two-step real-time digital equalizer," in *Proc. 1999 IEEE Workshop Appls. Signal Process. Audio Acoust. (WASPAA '99)*, pp. 195–198, IEEE, 1999.

[106] Y. Haneda, S. Makino, Y. Kaneda, and N. Kitawaki, "Common-acoustical-pole and zero modeling of head-related transfer functions," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 2, pp. 188–196, 1999.

[107] H. Schepker and S. Doclo, "Least-squares estimation of the common pole-zero filter of acoustic feedback paths in hearing aids," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 24, no. 8, pp. 1334–1347, 2016.

[108] G. Bunkheila, M. Scarpiniti, R. Parisi, and A. Uncini, "Stereo acoustical echo cancellation based on common poles," in *Proc. 2009 16th Int. Conf. Digital Signal Process.*, pp. 1–6, IEEE, 2009.

[109] Y. Haneda, Y. Kaneda, and N. Kitawaki, "Common-acoustical-pole and residue model and its application to spatial interpolation and extrapolation of a room transfer function," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 6, pp. 709–717, 1999.

[110] Y. Liu, P. Yi, and Y. Wu, "Efficient implementation of FIR type time domain equalizers for MIMO wireless channels via M-LESQ," in *Proc. 2009 IEEE 20th Int. Symp. Pers. Indoor Mobile Radio Commun.*, pp. 286–290, IEEE, 2009.

[111] C.-U. Lei and N. Wong, "WISE: Warped impulse structure estimation for time-domain linear macromodeling," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 2, no. 1, pp. 131–139, 2012.

[112] B. Ninness and F. Gustafsson, "A unifying construction of orthonormal bases for system identification," *IEEE Trans. Autom. Control*, vol. 42, no. 4, pp. 515–521, 1997.

[113] G. Vairetti, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. van Waterschoot, "A scalable algorithm for physically motivated and sparse approximation of room impulse responses with orthonormal basis functions," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 7, pp. 1547–1561, 2017.

[114] S. Takenaka, "On the orthogonal functions and a new formula of interpolation," in *Jpn. J. Math.*, vol. 2, pp. 129–145, Math. Soc. Jpn., 1925.

[115] F. Malmquist, "Sur la détermination d'une classe de fonctions analytiques par leurs valeurs dans un ensemble donné de points," *Comptes Rendus du Sixiemme Congress des mathematiciens scandinaves. Kopenhagen, Denmark*, p. 253, 1925.

[116] N. Wiener, *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. MIT press Cambridge, 1949.

[117] W. H. Kautz, "Transient synthesis in the time domain," *Trans. IRE Prof. Group Circuit Theory*, no. 3, pp. 29–39, 1954.

[118] P. W. Broome, "Discrete orthonormal sequences," *J. Assoc. Comput. Mach.*, vol. 12, no. 2, pp. 151–168, 1965.

[119] B. Epstein, *Orthogonal families of analytic functions*. New York: Macmillan, 1965.

[120] P. J. Davis, *Interpolation and approximation*. Courier Corp., 1975.

[121] I. M. Horowitz, *Synthesis of feedback systems.* New York: Academic Press, 1963.

[122] R. King and P. Paraskevopoulos, "Digital Laguerre filters," *Int. J. Circuit Theory Appls.*, vol. 5, no. 1, pp. 81–91, 1977.

[123] B. Wahlberg, "System identification using Laguerre models," *IEEE Trans. Autom. Control*, vol. 36, no. 5, pp. 551–562, 1991.

[124] B. Wahlberg, "System identification using Kautz models," *IEEE Trans. Autom. Control*, vol. 39, no. 6, pp. 1276–1282, 1994.

[125] P. Bodin and B. Wahlberg, "Thresholding in high order transfer function estimation," in *Proc. 33rd IEEE Conf. Decision Control*, vol. 4, pp. 3400–3405, IEEE, 1994.

[126] P. S. C. Heuberger, P. M. J. Van den Hof, and O. Bosgra, "A generalized orthonormal basis for linear dynamical systems," *IEEE Trans. Autom. Control*, vol. 40, no. 3, pp. 451–465, 1995.

[127] B. Ninness and F. Gustafsson, "Orthonormal bases for system identification," in *Proc. 3rd Eur. Control Conf.*, vol. 1, pp. 13–18, 1995.

[128] B. Wahlberg and P. Mäkilä, "On approximation of stable linear dynamical systems using Laguerre and Kautz functions," *Automatica*, vol. 32, no. 5, pp. 693–708, 1996.

[129] H. Akçay and B. Ninness, "Rational basis functions for robust identification from frequency and time-domain measurements," *Automatica*, vol. 34, no. 9, pp. 1101–1117, 1998.

[130] B. Ninness and J. Gómez, "Frequency domain analysis of tracking and noise performance of adaptive algorithms," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1314–1332, 1998.

[131] G. W. Davidson and D. D. Falconer, "Reduced complexity echo cancellation using orthonormal functions," *IEEE Trans. Circuits Syst.*, vol. 38, no. 1, pp. 20–28, 1991.

[132] H. Perez and S. Tsujii, "A system identification algorithm using orthogonal functions," *IEEE Trans. Signal Process.*, vol. 39, no. 3, pp. 752–755, 1991.

[133] A. C. den Brinker, "Laguerre-domain adaptive filters," *IEEE Trans. Signal Process.*, vol. 42, no. 4, pp. 953–956, 1994.

[134] T. O. e Silva, "On the determination of the optimal pole position of Laguerre filters," *IEEE Trans. Signal Process.*, vol. 43, no. 9, pp. 2079–2087, 1995.

[135] A. C. den Brinker, F. Benders, and T. O. e Silva, "Optimality conditions for truncated Kautz series," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 43, no. 2, pp. 117–122, 1996.

[136] P. M. Mäkilä, "Approximation of stable systems by Laguerre filters," *Automatica*, vol. 26, no. 2, pp. 333–345, 1990.

[137] G. A. Dumont and C. C. Zervos, "Adaptive control based on orthonormal series representation," in *Adaptive Systems in Control and Signal Processing*, pp. 105–113, Elsevier, 1987.

[138] R. Tóth, *Modeling and identification of linear parameter-varying systems*, vol. 403. Springer, 2010.

[139] L. D. Tufa, M. Ramasamy, and M. Shuhaimi, "Improved method for development of parsimonious orthonormal basis filter models," *J. Process Control*, vol. 21, no. 1, pp. 36–45, 2011.

[140] W. Mi and T. Qian, "Frequency-domain identification: an algorithm based on an adaptive rational orthogonal system," *Automatica*, vol. 48, no. 6, pp. 1154–1162, 2012.

[141] K. Tiels and J. Schoukens, "Wiener system identification with generalized orthonormal basis functions," *Automatica*, vol. 50, no. 12, pp. 3147–3154, 2014.

[142] T. Chen and L. Ljung, "Regularized system identification using orthonormal basis functions," in *Proc. 2015 Eur. Control Conf. (ECC)*, pp. 1291–1296, IEEE, 2015.

[143] D. Mayer, "Orthonormal filters for identification in active control systems," *Smart Mat. Struct.*, vol. 24, no. 12, p. 125037, 2015.

[144] A. Khouaja and H. Messaoud, "Iterative selection of GOB poles in the context of system modeling," *Int. J. Autom. Comput.*, pp. 1–10, 2016.

[145] R. Schumacher and G. H. Oliveira, "Unifying method to construct rational basis functions for linear and nonlinear systems," *Circuits Syst. Signal Process.*, pp. 1–19, 2017.

[146] T. Kondo, S. Yamaoka, and Y. Ohta, "A hyper-parameter estimation algorithm in bayesian system identification using OBFs-based kernels," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 14162–14167, 2017.

[147] M. A. H. Darwish, G. Pillonetto, and R. Tóth, "The quest for the right kernel in Bayesian impulse response identification: The use of OBFs," *Automatica*, vol. 87, pp. 318–329, 2018.

[148] T. Paatero, M. Karjalainen, and A. Härmä, "Modeling and equalization of audio systems using Kautz filters," in *Proc. 2001 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '01)*, vol. 5, pp. 3313–3316, IEEE, 2001.

[149] T. Paatero and M. Karjalainen, "New digital filter techniques for room response modeling," in *Proc. AES 21st Int. Conf.: Architectural Acoustics and Sound Reinforcement*, AES, 2002.

[150] T. Paatero, "An audio motivated hybrid of warping and Kautz filter techniques," in *Proc. 11th Eur. Signal Process. Conf. (EUSIPCO '02)*, pp. 627–630, 2002.

[151] L. S. Ngia, "Separable nonlinear least-squares methods for efficient off-line and on-line modeling of systems using Kautz and Laguerre filters," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 48, no. 6, pp. 562–579, 2001.

[152] D. Friedman, "On approximating an FIR filter using discrete orthonormal exponentials," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 29, no. 4, pp. 923–926, 1981.

[153] B. Sarroukh, S. J. Van Eijndhoven, and A. C. Den Brinker, "An iterative solution for the optimal poles in a Kautz series," in *Proc. 2001 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '01)*, vol. 6, pp. 3949–3952, IEEE, 2001.

[154] G. Vairetti, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. van Waterschoot, "Multichannel identification of room acoustic systems with adaptive IIR filters based on orthonormal basis functions," in *Proc. 2016 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '16)*, (Shanghai, China), pp. 16–20, IEEE, 2016.

[155] H. Brandenstein and R. Unbehauen, "Weighted least-squares approximation of FIR by IIR digital filters," *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 558–568, 2001.

[156] G. Vairetti, E. De Sena, T. van Waterschoot, M. Moonen, M. Catrysse, N. Kaplanis, and S. H. Jensen, "A physically motivated parametric model for compact representation of room impulse responses based on orthonormal basis functions," in *Proc. 10th Eur. Congr. Expo. Noise Control Eng. (EURONOISE '15)*, (Maastricht, The Netherlands), pp. 149–154, 2015.

[157] S. Haykin, *Adaptive filter theory*. Pearson Education India, 2008.

[158] P. S. Diniz, *Adaptive filtering*. Springer, 1997.

[159] C. Paleologu, S. Ciochina, and J. Benesty, "Variable step-size NLMS algorithm for under-modeling acoustic echo cancellation," *IEEE Signal Process. Lett.*, vol. 15, pp. 5–8, 2008.

[160] G. Rombouts, T. van Waterschoot, K. Struyve, P. Verhoeve, and M. Moonen, "Identification of undermodelled room impulse responses," in *Proc. 2005 Int. Workshop Acoust. Echo Noise Control (IWAENC '05)*, 2005.

[161] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 2, pp. 156–165, 1998.

[162] R. D. Poltmann, "Stochastic gradient algorithm for system identification using adaptive FIR-filters with too low number of coefficients," *IEEE Trans. Circuits Syst.*, vol. 35, no. 2, pp. 247–250, 1988.

[163] J. J. Shynk, "Adaptive IIR filtering," *IEEE ASSP Mag.*, vol. 6, no. 2, pp. 4–21, 1989.

[164] J. J. Shynk, "Adaptive IIR filtering using parallel-form realizations," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 37, no. 4, pp. 519–533, 1989.

[165] G. Williamson and S. Zimmermann, "Globally convergent adaptive IIR filters based on fixed pole locations," *IEEE Trans. Signal Process.*, vol. 44, no. 6, pp. 1418–1427, 1996.

[166] H. J. W. Belt, *Orthonormal bases for adaptive filtering*. PhD thesis, Technische Universiteit Eindhoven, 1997.

[167] B. Ninness, H. Hjalmarsson, and F. Gustafsson, "The fundamental role of general orthonormal bases in system identification," *IEEE Trans. Autom. Control*, vol. 44, no. 7, pp. 1384–1406, 1999.

[168] S. Gunnarsson and L. Ljung, "Frequency domain tracking characteristics of adaptive algorithms," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 1072–1089, 1989.

[169] B. Ninness and H. Hjalmarsson, "Model structure and numerical properties of normal equations," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 48, no. 4, pp. 425–437, 2001.

[170] L. Ljung and T. Söderström, *Theory and practice of recursive identification*. MIT press, 1983.

[171] J. Cousseau, P. Diniz, G. Sentoni, and O. Agamennoni, "On orthogonal realizations for adaptive IIR filters," *Int. J. Circuit Theory Appls.*, vol. 28, no. 5, pp. 481–500, 2000.

[172] T. Söderström and P. Stoica, *System identification*. Prentice-Hall, 1989.

[173] T. Söderström and P. Stoica, "Instrumental variable methods for system identification," *Circuits Systems Signal Process.*, vol. 21, no. 1, pp. 1–9, 2002.

[174] L. Ljung, *System identification – Theory for the user*. Upper Saddle River, NJ: Prentice-Hall, 2nd ed., 1999.

[175] R. Pintelon and J. Schoukens, *System identification: a frequency domain approach*. John Wiley & Sons, 2nd ed., 2012.

[176] E. Reynders, "System identification methods for (operational) modal analysis: review and comparison," *Arch. Comput. Methods Eng.*, vol. 19, no. 1, pp. 51–124, 2012.

[177] P. Verboven, *Frequency-domain system identification for modal analysis*. PhD thesis, Vrije Universiteit Brussel, Brussels, 2002.

[178] S. Gannot and M. Moonen, "Subspace methods for multimicrophone speech dereverberation," *EURASIP J. Adv. Signal Process.*, vol. 2003, no. 11, p. 769285, 2003.

[179] P. Van Overschee and B. De Moor, *Subspace identification for linear systems: Theory–Implementation–Applications*. Springer Science & Business Media, 2012.

[180] M. Verhaegen and P. Dewilde, "Subspace model identification part 1. The output-error state-space model identification class of algorithms," *Int. J. Control*, vol. 56, no. 5, pp. 1187–1210, 1992.

[181] S. J. Qin, "An overview of subspace identification," *Comput. Chem. Eng.*, vol. 30, no. 10-12, pp. 1502–1513, 2006.

[182] M. Lovera, T. Gustafsson, and M. Verhaegen, "Recursive subspace identification of linear and non-linear Wiener state-space models," *Automatica*, vol. 36, no. 11, pp. 1639–1650, 2000.

[183] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, no. 3, pp. 657–682, 2014.

[184] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, vol. 375. Springer Science & Business Media, 1996.

[185] T. van Waterschoot, G. Rombouts, and M. Moonen, "Optimally regularized adaptive filtering algorithms for room acoustic signal enhancement," *Signal Process.*, vol. 88, no. 3, pp. 594–611, 2008.

[186] J. M. Gil-Cacho, M. Signoretto, T. van Waterschoot, M. Moonen, and S. H. Jensen, "Nonlinear acoustic echo cancellation based on a sliding-window leaky kernel affine projection algorithm," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 9, pp. 1867–1878, 2013.

[187] S. Van Vaerenbergh and L. A. Azpicueta-Ruiz, "Kernel-based identification of Hammerstein systems for nonlinear acoustic echo-cancellation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP '14)*, pp. 3739–3743, 2014.

[188] W. Liu, J. C. Principe, and S. Haykin, *Kernel adaptive filtering: a comprehensive introduction*, vol. 57. John Wiley & Sons, 2011.

[189] W. Liu, P. P. Pokharel, and J. C. Principe, "The kernel least-mean-square algorithm," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 543–554, 2008.

[190] J. Kivinen, A. J. Smola, and R. C. Williamson, "Online learning with kernels," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2165–2176, 2004.

[191] Y. Engel, S. Mannor, and R. Meir, "The kernel recursive least-squares algorithm," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2275–2285, 2004.

[192] Y. Lin and D. D. Lee, "Bayesian regularization and nonnegative deconvolution for room impulse response estimation," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 839–847, 2006.

[193] G. Enzner, "Bayesian inference model for applications of time-varying acoustic system identification," in *Proc. 18th Eur. Signal Process. Conf. (EUSIPCO '10)*, pp. 2126–2130, 2010.

[194] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

[195] R. Mignot, L. Daudet, and F. Ollivier, "Room reverberation reconstruction: Interpolation of the early part using compressed sensing," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 21, no. 11, pp. 2301–2312, 2013.

[196] C. Papayiannis, C. Evers, and P. A. Naylor, "Sparse parametric modeling of the early part of acoustic impulse responses," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO '17)*, pp. 678–682, 2017.

[197] R. Mignot, G. Chardon, and L. Daudet, "Low frequency interpolation of room impulse responses using compressed sensing," *IEEE Trans. Audio Speech Lang. Process.*, vol. 22, no. 1, pp. 205–216, 2014.

[198] W. Jin and W. B. Kleijn, "Theory and design of multizone soundfield reproduction using sparse methods," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 12, pp. 2343–2355, 2015.

[199] E. Fernandez-Grande and A. Xenaki, "Compressive sensing with a spherical microphone array," *J. Acoust. Soc. Am.*, vol. 139, no. 2, pp. EL45–EL49, 2016.

[200] N. Antonello, E. De Sena, M. Moonen, P. A. Naylor, and T. van Waterschoot, "Room impulse response interpolation using a sparse spatio-temporal representation of the sound field," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 10, pp. 1929–1941, 2017.

[201] G. Vairetti, T. van Waterschoot, M. Moonen, M. Catrysse, and S. H. Jensen, "Sparse linear parametric modeling of room acoustics with orthonormal basis functions," *Proc. 22nd Eur. Signal Process. Conf. (EUSIPCO '14)*, 2014.

[202] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.

[203] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, 2010.

[204] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[205] B. Bank, "Loudspeaker and room response equalization using parallel filters: Comparison of pole positioning strategies," in *Proc. AES 51st Int. Conf.: Loudspeakers and Headphones*, Aug 2013.

[206] M. Schoenle, N. Fliege, and U. Zölzer, "Parametric approximation of room impulse responses by multirate systems," in *Proc. 1993 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '93)*, vol. 1, pp. 153–156, IEEE, 1993.

[207] M. Karjalainen, P. A. Esquef, P. Antsalo, A. Mäkivirta, and V. Välimäki, "Frequency-zooming ARMA modeling of resonant and reverberant systems," *J. Audio Eng. Soc.*, vol. 50, no. 12, pp. 1012–1029, 2002.

[208] S. J. Elliott and P. A. Nelson, "Multiple-point equalization in a room using adaptive digital filters," *J. Audio Eng. Soc.*, vol. 37, no. 11, pp. 899–907, 1989.

[209] S. Bharitkar, P. Hilmes, and C. Kyriakakis, "Robustness of spatial average equalization: A statistical reverberation model approach," *J. Acoust. Soc. Am.*, vol. 116, no. 6, pp. 3491–3497, 2004.

[210] S. Bharitkar and C. Kyriakakis, "A cluster centroid method for room response equalization at multiple locations," in *Proc. 2001 IEEE Workshop Appls. Signal Process. Audio Acoust. (WASPAA '01)*, pp. 55–58, IEEE, 2001.

[211] K. Lakhdhar, M. Jaidane, H. Shaiek, and J. Boucher, "Iterative equalization of room transfer function using biquadratic filters," in *Proc. 2009 IEEE Instrum. Meas. Technol. Conf. (I2MTC '09)*, pp. 1463–1466, IEEE, 2009.

[212] R. J. Oliver and J.-M. Jot, "Efficient multi-band digital audio graphic equalizer with accurate frequency response control," in *Preprints Audio Eng. Soc. Conv. 139*, Oct 2015.

[213] T. van Waterschoot and M. Moonen, "A pole-zero placement technique for designing second-order IIR parametric equalizer filters," *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 8, pp. 2561–2565, 2007.

[214] U. Zölzer, *Digital audio signal processing.* John Wiley & Sons, 2008.

[215] G. Ramos and J. J. Lopez, "Filter design method for loudspeaker equalization based on IIR parametric filters," *J. Audio Eng. Soc*, vol. 54, no. 12, pp. 1162–1178, 2006.

[216] H. Behrends, A. von dem Knesebeck, W. Bradinal, P. Neumann, and U. Z olzer, "Automatic equalization using parametric IIR filters," *J. Audio Eng. Soc*, vol. 59, no. 3, pp. 102–109, 2011.

[217] B. D. Radlovic and R. A. Kennedy, "Nonminimum-phase equalization and its subjective importance in room acoustics," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 6, pp. 728–737, 2000.

[218] B. D. Radlovic, R. C. Williamson, and R. A. Kennedy, "Equalization in an acoustic reverberant environment: Robustness results," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 311–319, 2000.

[219] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 36, no. 2, pp. 145–152, 1988.

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

[221] F. Lim, W. Zhang, E. A. Habets, and P. A. Naylor, "Robust multichannel dereverberation using relaxed multichannel least squares," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 9, pp. 1379–1390, 2014.

[222] I. Kodrasi, S. Goetze, and S. Doclo, "Regularization for partial multichannel equalization for speech dereverberation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 9, pp. 1879–1890, 2013.

[223] P. D. Hatziantoniou and J. N. Mourjopoulos, "Generalized fractional-octave smoothing of audio and acoustic responses," *J. Audio Eng. Soc.*, vol. 48, no. 4, pp. 259–280, 2000.

[224] L.-J. Brännmark and A. Ahlén, "Spatially robust audio compensation based on SIMO feedforward control," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1689–1702, 2009.

[225] L.-J. Brännmark, A. Bahne, and A. Ahlén, "Compensation of loudspeaker–room responses in a robust MIMO control framework," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 6, pp. 1201–1216, 2013.

[226] M. Kolundžija, C. Faller, and M. Vetterli, "Multi-channel low-frequency room equalization using perceptually motivated constrained optimization," in *Proc. 2012 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '12)*, pp. 533–536, IEEE, 2012.

[227] N. Kaplanis, S. Bech, S. Tervo, J. Pätynen, T. Lokki, T. van Waterschoot, and S. H. Jensen, "Perceptual aspects of reproduced sound in car cabin acoustics," *J. Acoust. Soc. Am.*, vol. 141, no. 3, pp. 1459–1469, 2017.

[228] D. Botteldooren, "Finite-difference time-domain simulation of low-frequency room acoustic problems," *J. Acoust. Soc. Am.*, vol. 98, no. 6, pp. 3302–3308, 1995.

[229] L. Savioja and U. P. Svensson, "Overview of geometrical room acoustic modeling techniques," *J. Acoust. Soc. Am.*, vol. 138, no. 2, pp. 708–730, 2015.

[230] E. De Sena, N. Antonello, M. Moonen, and T. Van Waterschoot, "On the modeling of rectangular geometries in room acoustic simulations," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 4, pp. 774–786, 2015.

[231] E. De Sena, H. Hacıhabiboğlu, Z. Cvetković, and J. O. Smith, "Efficient synthesis of room acoustics via scattering delay networks," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 9, pp. 1478–1492, 2015.

[232] F. Wefers, *Partitioned convolution algorithms for real-time auralization*, vol. 20. Logos Verlag Berlin GmbH, 2015.

[233] J. Atkins, A. Strauss, and C. Zhang, "Approximate convolution using partitioned truncated singular value decomposition filtering," in *Proc. 2013 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '13)*, pp. 176–180, IEEE, 2013.

[234] J. S. Abel and K. J. Werner, "Distortion and pitch processing using a modal reverberator architecture," in *Proc. 18th Int. Conf. Digital Audio Effects (DAFx '15)*, 2015.

[235] T. van Waterschoot, G. Rombouts, P. Verhoeve, and M. Moonen, "Double-talk-robust prediction error identification algorithms for acoustic echo cancellation," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 846–858, 2007.

[236] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic acoustic echo cancellation-an overview of the fundamental problem," *IEEE Signal Process. Lett.*, vol. 2, no. 8, pp. 148–151, 1995.

[237] M. Ali, "Stereophonic acoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation," in *Proc. 1998 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '98)*, vol. 6, pp. 3689–3692, IEEE, 1998.

[238] M. Mboup and M. Bonnet, "On the adequateness of IIR adaptive filtering for acoustic echo cancellation," in *Proc. 6th Eur. Signal Process. Conf. (EUSIPCO '92)*, pp. 111–114, 1992.

[239] A. P. Liavas and P. A. Regalia, "Acoustic echo cancellation: Do IIR models offer better modeling capabilities than their FIR counterparts?," *IEEE Trans. Signal Process.*, vol. 46, no. 9, pp. 2499–2504, 1998.

[240] S. Gudvangen and S. Flockton, "Modelling of acoustic transfer functions for echo cancellers," *IEE Proc.-Vis Image Signal Process.*, vol. 142, no. 1, pp. 47–51, 1995.

[241] L. Salama and J. Cousseau, "Efficient echo cancellation based on an orthogonal adaptive IIR realization," in *Proc. SBT/IEEE Int. Telecommun. Symp. (ITS'98)*, vol. 2, pp. 434–437, IEEE, 1998.

[242] J. Y. Wen, N. D. Gaubitch, E. A. Habets, T. Myatt, and P. A. Naylor, "Evaluation of speech dereverberation algorithms using the MARDY database," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC 2006)*, (Paris, France), 2006.

[243] E. Hadad, F. Heese, P. Vary, and S. Gannot, "Multichannel audio database in various acoustic environments," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC 2014)*, (Antibes-Juan Les Pins), pp. 313–317, 2014.

[244] J. Eaton, N. D. Gaubitch, A. H. Moore, and P. A. Naylor, "The ACE challenge - corpus description and performance evaluation," in *Proc. 2015 IEEE Workshop Applicat. Signal Process. Audio Acoust. (WASPAA 2015)*, pp. 1–5, IEEE, 2015.

[245] R. Stewart and M. B. Sandler, "Database of omnidirectional and B-format room impulse responses," in *Proc. 2010 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP '10)*, (Dallas, USA), pp. 165–168, 2010.

[246] M. Jeub, M. Schäfer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *Proc. Int. 16th Conf. Digital Signal Process.*, (Santorini, Greece), pp. 1–5, 2009.

[247] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP J. Adv. Signal Process.*, vol. 2009, p. 6, 2009.

[248] S. E. Olive, P. L. Schuck, J. G. Ryan, S. L. Sally, and M. E. Bonneville, "The detection thresholds of resonances at low frequencies," *J. Audio Eng. Soc.*, vol. 45, no. 3, pp. 116–128, 1997.

[249] W. Klippel and R. Werner, "Loudspeaker distortion - measurement and perception, part 1: Regular distortion defined by design," in *26th Tonmeistertagung*, (Leipzig, Germany), 2010.

[250] W. Klippel and R. Werner, "Loudspeaker distortion - measurement and perception, part 2: Irregular distortion caused by defects," in *26th Tonmeistertagung*, (Leipzig, Germany), 2010.

[251] W. Klippel and U. Seidel, "Measurement of impulsive distortion, rub and buzz and other disturbances," in *Preprints Audio Eng. Soc. Conv. 114*, (Amsterdam, The Netherlands), 2003.

[252] W. Klippel, "Rub and buzz and other irregular loudspeaker distortion (tutorial)," in *Preprints Audio Eng. Soc. Conv. 134*, (Rome, Italy), 2013.

[253] R. C. Heyser, "Acoustical measurements by time delay spectrometry," *J. Audio Eng. Soc.*, vol. 15, no. 4, pp. 370–382, 1967.

[254] P. M. Clarkson, J. Mourjopoulos, and J. Hammond, "Spectral, phase, and transient equalization for audio systems," *J. Audio Eng. Soc.*, vol. 33, no. 3, pp. 127–132, 1985.

[255] M. Holters, T. Corbach, and U. Zölzer, "Impulse response measurement techniques and their applicability in the real world," in *Proc. 12th Int. Conf. Digital Audio Effects (DAFx 2009)*, 2009.

[256] ISO 3382-2:2008, "Acoustics - measurements of room acoustic parameters - part 2: Reverberation time in ordinary rooms," 2008.

[257] F. Jacobsen, "A note on acoustic decay measurements," *J. Sound Vib.*, vol. 115, no. 1, pp. 163–170, 1987.

[258] M. Brookes, "VOICEBOX: A speech processing toolbox for MATLAB," Imperial College London, Software Library, 2011.

[259] IEC 60268-13:1998, "Sound system equipment - part 13: Listening tests on loudspeakers," 1998.

[260] Genelec Oy, *Genelec 7050B Active Subwoofer - Operating manual*, 2005. D0061R001.

[261] D. Keele Jr., "Low-frequency loudspeaker assessment by nearfield sound-pressure measurement," *J. Audio Eng. Soc.*, vol. 22, no. 3, pp. 154–162, 1974.

[262] IEC 61260-1:1995, "Electroacoustics – octave-band and fractional-octave-band filters – part 1: Specifications," Geneva, Switzerland: International Electrotechnical Commission, 1995.

[263] Y.-P. Lin and P. Vaidyanathan, "A Kaiser window approach for the design of prototype filters of cosine modulated filterbanks," *IEEE Signal Process. Lett.*, vol. 5, no. 6, pp. 132–134, 1998.

[264] G. Vairetti, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. van Waterschoot, "Room acoustic system identification using orthonormal basis function models," in *Proc. AES 60th Int. Conf.: DREAMS (Dereverb. Reverb. Audio Music Speech)*, (Leuven, Belgium), AES, 2016.

[265] B. Bank and J. O. Smith, III, "A delayed parallel filter structure with an FIR part having improved numerical properties," in *Preprints Audio Eng. Soc. Conv. 136*, 2014.

[266] G. Oliveira, A. Da Rosa, R. Campello, J. Machado, and W. Amaral, "An introduction to models based on Laguerre, Kautz and other related orthonormal functions - part I: linear and uncertain models," *Int. J. Model. Ident. Control*, vol. 14, pp. 121–132, 2011.

[267] G. Chardon and L. Daudet, "Optimal subsampling of multichannel damped sinusoids," in *Proc. 2010 IEEE Sensor Array Multichannel Signal Process. Workshop (SAM 2010)*, (Jerusalem, Israel), pp. 25–28, IEEE, 2010.

[268] T. van Waterschoot, "KU Leuven ESAT Speech Lab room impulse response database," Tech. Rep. ESAT-STADIUS TR 15-74, KU Leuven, Belgium, Aug. 2015.

[269] M. Karjalainen, P. A. A. Esquef, P. Antsalo, A. Mäkivirta, and V. Välimäki, "AR/ARMA analysis and modeling of modes in resonant and reverberant systems," in *Preprints Audio Eng. Soc. Conv. 112*, 2002.

[270] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Stat. Soc. B. Met.*, pp. 267–288, 1996.

[271] S. Mallat and Z. Zhang, "Adaptive time-frequency decomposition with matching pursuits," in *Proc. IEEE-SP Int. Symp. Time-Freq. and Time-Scale Anal.*, pp. 7–10, IEEE, 1992.

[272] Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Conf. Rec. 27^{th} Asilomar Conf. on Signals, Systems, Comput.*, pp. 40–44, 1993.

[273] G. H. Golub and C. F. Van Loan, *Matrix computations*. Baltimore and London: The Johns Hopkins University Press, 3rd ed., 1996.

[274] G. Vairetti, N. Kaplanis, S. Bech, E. De Sena, M. Moonen, and T. van Waterschoot, "The subwoofer room impulse response (SUBRIR) database," *J. Audio Eng. Soc.*, vol. 65, May 2017.

[275] G. Vairetti, T. van Waterschoot, M. Moonen, M. Catrysse, and S. H. Jensen, "An automatic model-building algorithm for sparse approximation of room impulse responses with orthonormal basis functions," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC 2014)*, (Antibes-Juan Les Pins), 2014.

[276] A. Kaelin, A. Lindgren, and G. Moschytz, "Simplified adaptive IIR filters based on optimized orthogonal prefiltering," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 42, no. 5, pp. 326–333, 1995.

[277] M. Campi, R. Leonardi, and L. A. Rossi, "Generalized super-exponential method for blind equalization using Kautz filters," in *Proc. IEEE Signal Process. Workshop Higher-Order Stat. (SPW-HOS '99)*, pp. 107–111, 1999.

[278] S. Narayan, A. Peterson, and M. Narasimha, "Transform domain LMS algorithm," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 31, no. 3, pp. 609–615, 1983.

[279] J. Lee and C. Un, "Performance of transform-domain LMS adaptive digital filters," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 34, no. 3, pp. 499–510, 1986.

[280] B. Farhang-Boroujeny and S. Gazor, "Selection of orthonormal transforms for improving the performance of the transform domain normalised LMS algorithm," in *IEE Proc. F (Radar Signal Process.)*, vol. 139, pp. 327–335, 1992.

[281] S. Gazor and B. Farhang-Boroujeny, "Quantization effects in transform-domain normalized LMS algorithm," *IEEE Trans. Circuits Syst. II: Analog Digital Signal Process.*, vol. 39, no. 1, pp. 1–7, 1992.

[282] F. Beaufays, "Transform-domain adaptive filters: An analytical approach," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 422–431, 1995.

[283] D. R. Morgan and S. G. Kratzer, "On a class of computationally efficient, rapidly converging, generalized NLMS algorithms," *IEEE Signal Process. Lett.*, vol. 3, no. 8, pp. 245–247, 1996.

[284] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electron. Commun. Jpn. (Part I: Commun.)*, vol. 67, no. 5, pp. 19–27, 1984.

[285] A. Mader, H. Puder, and G. U. Schmidt, "Step-size control for acoustic echo cancellation filters–an overview," *Signal Process.*, vol. 80, no. 9, pp. 1697–1719, 2000.

[286] J.-K. Hwang and Y.-P. Li, "Variable step-size LMS algorithm with a gradient-based weighted average," *IEEE Signal Process. Lett.*, vol. 16, no. 12, pp. 1043–1046, 2009.

[287] Y. Zhang, N. Li, J. A. Chambers, and Y. Hao, "New gradient-based variable step size LMS algorithms," *EURASIP J. Adv. Signal Process.*, vol. 2008, p. 105, 2008.

[288] J. Nocedal and S. J. Wright, *Numerical Optimization*, ch. 3 Line Search Methods, pp. 30–65. Springer, 2nd ed., 2006.

[289] I. R. Titze, *Principles of voice production.* Englewood Cliffs, Prentice Hall, 1994.

[290] J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," *IEEE Signal Process. Lett.*, vol. 6, no. 1, pp. 1–3, 1999.

[291] G. Waters, "Sound quality assessment material – recordings for subjective tests: User's handbook for the EBU–SQAM compact disk," tech. rep., European Broadcasting Union (EBU), 1988.

[292] J. Chen, J. Benesty, and Y. A. Huang, "Time delay estimation in room acoustic environments: an overview," *EURASIP J. Adv. Signal Process.*, vol. 2006, no. 1, p. 026503, 2006.

[293] T. Gänsler and J. Benesty, "Stereophonic acoustic echo cancellation and two-channel adaptive filtering: an overview," *Int. J. Adapt. Control Signal Process*, vol. 14, no. 6, pp. 565–586, 2000.

[294] T. Hoya, J. A. Chambers, and P. A. Naylor, "Low complexity ϵ-NLMS algorithms and subband structures for stereophonic acoustic echo cancellation," in *Proc. 1999 Int. Workshop Acoust. Echo Noise Control (IWAENC '99)*, pp. 36–39, 1999.

[295] V. Välimäki and J. D. Reiss, "All about audio equalization: solutions and frontiers," *Appl. Sciences*, vol. 6, no. 5, p. 129, 2016.

[296] M. Karjalainen, E. Piirilä, A. Järvinen, and J. Huopaniemi, "Comparison of loudspeaker equalization methods based on DSP techniques," *J. Audio Eng. Soc*, vol. 47, no. 1/2, pp. 14–31, 1999.

[297] A. Gabrielsson, B. Lindström, and O. Till, "Loudspeaker frequency response and perceived sound quality," *J. Acoust. Soc. Am.*, vol. 90, no. 2, pp. 707–719, 1991.

[298] B. C. Moore and C.-T. Tan, "Perceived naturalness of spectrally distorted speech and music," *J. Acoust. Soc. Am.*, vol. 114, no. 1, pp. 408–419, 2003.

[299] R. Plomp and H. Steeneken, "Effect of phase on the timbre of complex tones," *J. Acoust. Soc. Am.*, vol. 46, no. 2B, pp. 409–421, 1969.

[300] R. Bristow-Johnson, "The equivalence of various methods of computing biquad coefficients for audio parametric equalizers," in *Preprints Audio Eng. Soc. Conv. 97*, 1994.

[301] P. Regalia and S. Mitra, "Tunable digital frequency response equalization filters," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 35, no. 1, pp. 118–120, 1987.

[302] U. Zölzer and T. Boltze, "Parametric digital filter structures," in *Preprints Audio Eng. Soc. Conv. 99*, 1995.

[303] F. E. Toole and S. E. Olive, "The modification of timbre by resonances: Perception and measurement," *J. Audio Eng. Soc*, vol. 36, no. 3, pp. 122–142, 1988.

[304] A. Gray and J. Markel, "Distance measures for speech processing," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 24, no. 5, pp. 380–391, 1976.

[305] B. C. Moore and C.-T. Tan, "Development and validation of a method for predicting the perceived naturalness of sounds subjected to spectral distortion," *J. Audio Eng. Soc*, vol. 52, no. 9, pp. 900–914, 2004.

[306] G. Ramos and P. Tomas, "Improvements on automatic parametric equalization and cross-over alignment of audio systems," in *Preprints Audio Eng. Soc. Conv. 126*, 2009.

[307] H. Rosenbrock, "An automatic method for finding the greatest or least value of a function," *Comput. J*, vol. 3, no. 3, pp. 175–184, 1960.

[308] T. Corbach, A. von dem Knesebeck, K. Dempwolf, M. Holters, P. Sorowka, and U. Zölzer, "Automated equalization for room resonance suppression," in *Proc. 12th Int. Conf. Digital Audio Effects (DAFx09)*, 2009.

[309] J. S. Abel and D. P. Berners, "Filter design using second-order peaking and shelving sections.," in *Proc. 30th Int. Comput. Music Conf. (ICMC)*, (Miami, USA), 2004.

[310] Z. Chen, Y. Liu, G. Geng, and F. Yin, "Optimal design of digital audio parametric equalizer," *J. Inform. Comput. Sci.*, vol. 11, no. 1, pp. 57–66, 2014.

[311] M. Hayes, J. Lim, and A. Oppenheim, "Signal reconstruction from phase or magnitude," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 28, no. 6, pp. 672–680, 1980.

[312] J. O. Smith, *Spectral pre-processing for audio digital filter design.* Ann Arbor, MI: Michigan Publishing, University of Michigan Library, 1983.

[313] P. A. Regalia, S. K. Mitra, and P. Vaidyanathan, "The digital all-pass filter: A versatile signal processing building block," *Proc. IEEE*, vol. 76, no. 1, pp. 19–37, 1988.

[314] J.-M. Jot, "Proportional parametric equalizers - application to digital reverberation and environmental audio processing," in *Preprints Audio Eng. Soc. Conv. 139*, 2015.

[315] F. Toole, "The measurement and calibration of sound reproducing systems," *J. Audio Eng. Soc*, vol. 63, no. 7/8, pp. 512–541, 2015.

[316] J. Abildgaard Pedersen and K. Thomsen, "Fully automatic loudspeaker-room adaptation – the RoomPerfect system," in *Proc. AES 32nd Int. Conf.: DSP For Loudspeakers*, 2007.

[317] Radiohead, "Burn the witch," in *A Moon Shaped Pool*, XL Records, 2016. (12 s-20 s).

[318] S. Cecchi, L. Palestini, P. Peretti, L. Romoli, F. Piazza, and A. Carini, "Evaluation of a multipoint equalization system based on impulse response prototype extraction," *J. Audio Eng. Soc*, vol. 59, no. 3, pp. 110–123, 2011.

[319] A. Bahne, L.-J. Brännmark, and A. Ahlén, "Symmetric loudspeaker-room equalization utilizing a pairwise channel similarity criterion," *IEEE Trans. Signal Process.*, vol. 61, no. 24, pp. 6276–6290, 2013.

[320] A. Bahne and A. Ahlén, "Optimizing the similarity of loudspeaker–room responses in multiple listening positions," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 24, no. 2, pp. 340–353, 2016.

[321] M. Johansson, L.-J. Brännmark, A. Bahne, and M. Sternad, "Sound field control using a limited number of loudspeakers," in *Proc. AES 36th Int. Conf.: Automotive Audio*, 2009.

[322] S. Berthilsson, A. Barkefors, L.-J. Brännmark, and M. Sternad, "Acoustical zone reproduction for car interiors using a MIMO MSE framework," in *Proc. AES 48th Int. Conf.: Automotive Audio*, 2012.

[323] A. Barkefors, M. Sternad, and L.-J. Brännmark, "Design and analysis of linear quadratic gaussian feedforward controllers for active noise control," *IEEE/ACM Trans. Audio Speech Lang. Proc.*, vol. 22, no. 12, pp. 1777–1791, 2014.

[324] L.-J. Brännmark and M. Sternad, "Controlling the impulse responses and the spatial variability in digital loudspeaker-room correction.," in *Proc. 2015 Int. Symp. ElectroAcoust. Tech. (ISEAT '15)*, 2015.

[325] L.-J. Brännmark, M. Sternad, and A. Ahlén, "Spatially robust audio precompensation," Mar. 20 2008. US Patent App. 12/052,031.

[326] L.-J. Brännmark, M. Sternad, and M. Johansson, "Sound field control in multiple listening regions," May 28 2009. US Patent App. 12/453,958.

[327] L.-J. Brännmark, A. Ahlén, and A. Bahne, "Audio precompensation controller design using a variable set of support loudspeakers," Mar. 22 2012. US Patent App. 14/009,215.

[328] A. Bahne, L.-J. Brännmark, and A. Ahlén, "Audio precompensation controller design with pairwise loudspeaker channel similarity," Aug. 23 2016. US Patent 9,426,600.

[329] L.-J. Brännmark, "Robust audio precompensation with probabilistic modeling of transfer function variability," in *Proc. 2009 IEEE Workshop Appls. Signal Process. Audio Acoust. (WASPAA '09)*, pp. 193–196, IEEE, 2009.

[330] P. Jarske, S. Mitra, and Y. Neuvo, "Variable linear phase FIR filters," in *Proc. 1988 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '88)*, pp. 1463–1466, IEEE, 1988.

[331] P. Jarske, Y. Neuvo, and S. K. Mitra, "A simple approach to the design of linear phase FIR digital filters with variable characteristics," *Signal Process.*, vol. 14, no. 4, pp. 313–326, 1988.

[332] P. Stoica, R. L. Moses, *et al.*, *Spectral analysis of signals*, vol. 452. Upper Saddle River, NJ: Pearson Prentice Hall, 2005.

[333] A. G. Constantinides and W. Li, "Digital filter design using root moments for sum-of-all-pass structures from complete and partial specifications," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 315–324, 2006.

[334] L.-J. Brännmark and A. Ahlén, "Robust loudspeaker equalization based on position-independent excess phase modeling," in *Proc. 2008 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '08)*, pp. 385–388, IEEE, 2008.

[335] L.-J. Brännmark and A. Ahlén, "Variable control of the pre-response error in mixed phase audio precompensation," in *Proc. 2009 IEEE Workshop Appls. Signal Process. Audio Acoust. (WASPAA '09)*, pp. 197–200, IEEE, 2009.

[336] H. Akçay and P. Heuberger, "A frequency-domain iterative identification algorithm using general orthonormal basis functions," *Automatica*, vol. 37, no. 5, pp. 663–674, 2001.

[337] J. J. Shynk *et al.*, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, no. 1, pp. 14–37, 1992.

[338] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3790–3804, 2006.

[339] A. Carini, S. Cecchi, L. Romoli, and G. L. Sicuranza, "Legendre nonlinear filters," *Signal Process.*, vol. 109, pp. 84–94, 2015.

[340] J. Gil-Cacho, *Adaptive Filtering Algorithms for Acoustic Echo Cancellation and Acoustic Feedback Control in Speech Communication Applications*. PhD thesis, KU Leuven, 2013.

[341] J. G. Stoddard and J. S. Welsh, "Volterra kernel identification using regularized orthonormal basis functions." *ArXiv e-prints*, https://arxiv.org/abs/1804.07429, Apr. 2018.

Curriculum Vitae

Giacomo Vairetti was born in Lecco (Italy) on December 22, 1984. He received the B.Sc. in 2010 and the M.Sc. (cum laude) in 2012, both in Computer Engineering at Politecnico di Milano (Italy). Since 2013, he has been a Ph.D. student in Electrical Engineering at KU Leuven (Belgium), where he was also a Marie Curie Fellow in the EU-funded ITN "Dereverberation and Reverberation of Audio, Music and Speech" (DREAMS). He was a visiting student at the Signal Processing and Acoustics Department of Aalto University (Finland) in 2012 and at the Electronic Systems Department of Aalborg University (Denmark) in 2014. His research interests are in signal processing and system identification applied to room acoustic modeling, sound synthesis, and audio reproduction.

Publication List

International Journal Papers

1. **G. Vairetti**, N. Kaplanis, E. De Sena, S. H. Jensen, S. Bech, M. Moonen, and T. van Waterschoot, "The Subwoofer Room Impulse Response (SUBRIR) database," *J. Audio Eng. Soc.*, vol. 65, no. 5, pp. 389–401, May 2017.

2. **G. Vairetti**, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. van Waterschoot, "A scalable algorithm for physically motivated and sparse approximation of room impulse responses with orthonormal basis functions," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 7, pp. 1547—1561, Jul. 2017.

3. **G. Vairetti**, E. De Sena, S. H. Jensen, M. Moonen, and T. van Waterschoot, "Orthonormal basis functions adaptive filters for room acoustic signal enhancement," Submitted for publication to *Signal Process.*, Elsevier, Apr. 2018.

4. **G. Vairetti**, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. van Waterschoot, "An automatic design procedure for low-order IIR parametric equalizers," Submitted for publication to *J. Audio Eng. Soc.*, Apr. 2018.

International Conference Papers

1. **G. Vairetti**, T. van Waterschoot, M. Moonen, M. Catrysse, and S. H. Jensen, "Sparse linear parametric modeling of room acoustics with orthonormal basis functions", in *Proc. 22nd Eur. Signal Process. Conf. (EUSIPCO '14)*, Lisbon, Portugal, pp. 1–5, Sep. 2014.

2. **G. Vairetti**, T. van Waterschoot, M. Moonen, M. Catrysse, and S. H. Jensen, "An automatic model-building algorithm for sparse approximation of room impulse responses with orthonormal basis functions", in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC 2014)*, Antibes, France, pp. 248–252, Sep. 2014.

3. **G. Vairetti**, E. De Sena, T. van Waterschoot, M. Moonen, M. Catrysse, N. Kaplanis, and S. H. Jensen, "A physically-motivated parametric model for compact representation of room impulse responses based on orthonormal basis functions", in *Proc. 10th Eur. Congr. Expo. Noise Control Eng. (EuroNoise 2015)*, Maastricht, The Netherlands, pp. 149–154, Jun. 2015.

4. **G. Vairetti**, E. De Sena, M. Catrysse, S.H. Jensen, M. Moonen and T. van Waterschoot, "Room acoustic system identification using orthonormal basis function models", in *Proc. AES 60th Int. Conf.: DREAMS (Dereverb. Reverb. Audio Music Speech)*, Leuven, Belgium, pp. 7(3), Feb. 2016.

5. **G. Vairetti**, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen and T. van Waterschoot, "Multichannel identification of room acoustic systems with adaptive IIR filters based on orthonormal basis functions", in *Proc. 41st IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP 2016)*, Shangai, China, pp. 16–20, Mar. 2016.

FACULTY OF ENGINEERING TECHNOLOGY
DEPARTMENT OF ELECTRICAL ENGINEERING
Stadius Center for Dynamical Systems,
Signal Processing and Data Analytics
Kasteelpark Arenberg 10, B-3001 Leuven