UNIVERSITÀ DI PARMA

UNIVERSITÀ DEGLI STUDI PARMA

Dottorato di Ricerca in Tecnologie dell'Informazione

XXXVI Ciclo

ACTIVE CONTROL OF PERSONAL SOUND ZONES BY DIGITAL SIGNAL PROCESSING

Coordinatore:

Chiar.mo Prof. Marco Locatelli

Tutor:

Chiar.mo Prof. Riccardo Raheli

Dottorando: *Anatolij Borroni*

Anni 2020-2023

To my parents

Abstract

The advancement of audio technology in a world characterized by increasing connectivity and constant background noise has led to the development of personal sound zone (PSZ) systems. These systems are designed to provide tailored audio experiences to individual users in shared spaces. By utilizing several loudspeakers, PSZ systems can deliver distinct audio signals to various zones without the need for headphones, improving listener comfort and privacy, and reducing noise pollution.

PSZ systems find applications in various scenarios, including in-vehicle audio, public and home entertainment, and communication. They have the potential to offer benefits, such as individualized sound zones for passengers during car or public vehicle journeys, separate audio areas for multiple individuals at home and public spaces, and enhanced audio quality for phone calls, video conferencing, and online meetings. However, designing and implementing PSZ systems presents several challenges, including achieving high acoustic contrast (AC) among zones, minimizing signal distortion and interference, optimizing array configurations, and adapting to changing user preferences.

To control the sound field in PSZ systems, advanced digital signal processing techniques are required. These techniques have applications in immersive audio experiences, virtual and augmented reality, gaming, noise cancellation, and PSZ generation.

This thesis deals with various aspects of a PSZ systems with application in the automotive scenario.

Related to the capabilities of PSZ systems to control the soundfield in a particularly reverberant scenario, such as the cabin of a vehicle, this research introduces four methods for processing the measured impulse responses (IRs), reducing late reflections to enhance AC, robustness, and sound quality.

Various methods have been proposed for sound field control, such as pressure matching (PM) and acoustic contrast control (ACC) techniques. For a part of this work, the PM method is adopted, however, also the performance of the ACC method in terms of sound quality in a real-world system is investigated. Moreover, two techniques derived from the PM method are proposed to improve the AC, reproduction error and robustness in the considered scenario.

With the purpose of achieving a high fidelity of the reproduced audio, an acoustic pressure with flat spectrum and constant group delay is considered as target in most of the literature related to PM. However, this may not be the optimal target in order to improve AC maintaining high fidelity reproduction. For this reason, the first proposed technique involves the optimization of the target phase for PM with the aim of AC maximization.

One of the disadvantages of the original formulation of PM is that the performance is not robust against errors in the measurement positions with respect to the realistic positions of the listeners. With the purpose of solving this weakness, a statistical pressure matching (SPM) algorithm is developed. This technique allows to improve the reproduction fidelity and AC by using several measurements to average out the effect of the errors.

A significant aspect of PSZ systems involves designing filters to control the audio signals at the inputs of the loudspeakers based on the acoustic responses and assessing the resulting performance. Early evaluations were often performed under ideal conditions, thus, overestimating the system achievable performance in realistic conditions. To address these limitations, a stochastic model is proposed to generate mismatched frequency responses (FRs) for realistic performance prediction. This model considers complex coefficients in the frequency domain and perturbs the acoustic responses in the performance evaluation step.

Contents

List of acronyms

AC acoustic contrast ..... i, 2, 14, 30, 50, 79, 111, 123

ACC acoustic contrast control ..... ii, 2, 12, 49, 79, 124

BZ bright zone 7, 33, 80

DFT discrete Fourier transform 19, 30, 53

DH dummy head 15, 28, 56, 101

DTFT discrete-time Fourier transforms 48

DZ dark zone 7, 33, 86

FC frequency-correlated 5, 108, 112, 124

FDT frequency-dependent trimming 5, 41, 123

FEM finite element method 110

FFT fast Fourier transform 32, 52

FR frequency response ..... ii, 4, 8, 27, 52, 80, 109

IFFT inverse fast Fourier transform 52

IID independent and identically distributed 5, 95, 112, 124

IR impulse response ..... ii, 2, 12, 25, 41, 81, 110, 123

ISTFT inverse short-time Fourier transform 35, 60

MFR magnitude frequency response 43, 81

PM pressure matching ..... ii, 2, 10, 76, 79, 124

PSZ personal sound zone ..... i, 1, 8, 25, 41, 79, 109, 123

QMF quadrature mirror filter 42

RC raised cosine 43

RoGoED radius of gyration of the energy density 17, 70, 83, 123

SPM statistical pressure matching ..... ii, 3, 80, 124

STFT short-time Fourier transform 19, 32, 42

STOI short-time objective intelligibility 19, 25, 115

WPM weighted pressure matching 2, 94

Foreword

This research was performed in collaboration with ASK Industries S.p.A. (Reggio Emilia, Italy), which supported the activity within the Italian Ministry of Economic Development (MiSE)'s fund for the sustainable growth (F.C.S.) under grant agreement (CUP) B82C21000700005, project CGS (Connettività e Guida Sicura).

Introduction

In a world characterized by increasing interconnections and pervasive noise, the pursuit of personalized and immersive audio experiences has driven relentless innovation in the realm of audio technology, giving rise to research, experimentation, and implementation of personal sound zone (PSZ) systems.

The PSZ system is a novel technology that aims to create individualized listening experiences for multiple users in a shared environment. By using a system composed of several loudspeakers, these systems can, in principle, deliver different audio signals to different zones without the need for headphones. This can enhance the comfort and privacy of the listeners, as well as reduce noise pollution in the environment.

Applications of such systems span in-vehicle audio, home entertainment, and communication, providing individualized sound zones for passengers during long journeys, separate audio domains for multiple individuals in home settings, and clear audio quality during phone calls, video conferencing, and online meetings. However, designing and implementing these systems is not a trivial task. There are several challenges and trade-offs involved, such as achieving high acoustic contrast among the zones, minimizing the signal distortion and interference, optimizing the array configuration and effort, and adapting to the changing conditions and preferences of the users.

Besides the sound system, PSZ systems require advanced digital signal processing to control the sound field. Over the years, sound field control by digital signal processing has become a vibrant field of study within active acoustics,

finding applications in immersive audio experiences, virtual and augmented reality, gaming, noise cancellation, and personal sound zones generation.

Many sound field control methods have been proposed in the literature. These methods fall into two main categories: those based on analytic solutions derived from the Helmholtz equation, such as wave field synthesis [1] and ambisonics [2], and those based on optimizing specific criteria, such as pressure matching (PM) [3], weighted pressure matching (WPM) [4, 5], and acoustic contrast control (ACC) [6]. Other optimization techniques for establishing PSZ systems based on physical measures, such as the acoustic contrast and the reproduction error, are present in the literature. Some examples are the delay and sum beamforming [7], and the acoustic energy difference maximization methods [8]. Methods based on a perceptual model that accounts for the human hearing system were also proposed, see, e.g., [9].

More recently, audio signal processing by deep learning techniques [10] has also spread to sound field control for PSZ systems, e.g., [11–13], proving that they may outperform classical methods. However, these techniques require a large amount of training data that is difficult to acquire in the automotive application scenario considered in our research.

For this reason and due to its implementation simplicity, flexibility and capability of high fidelity of the reproduced audio [14], in this research, the PM algorithm is adopted as an effective and representative method in the automotive scenario. However, we also investigated the realistic performance of the ACC method in terms of the sound quality.

Related to the capability of the PSZ systems to control the reflected components of the acoustic channel, based on the idea of "trimming" introduced in [15], four methods for processing the channel impulse responses (IRs) by reducing the late reflections, with the aim of acoustic contrast (AC), robustness and sound quality enhancement are proposed and investigated.

In most of the literature related to PM, an ideal overall Dirac delta IR is set as the target for the bright acoustic region. However, in some recent works, e.g., [16] and [17], a performance improvement in terms of AC and

the so-called array effort is shown, by a proper design of the target sound field. In particular, in [17] a closed-form solution for the minimization of the array effort, under the constraint of a constant acoustic energy, is derived and analyzed by jointly optimizing the phase difference and amplitude ratio between the target pressure of two control points in the bright sound region.

Concerning this aspect, in this thesis, an algorithm for the maximization of the AC is proposed, by optimizing the phase difference between the target pressure of two control points in the bright sound region, so that the target amplitude is preserved. The idea is to maintain the sound pressure level in the bright region and perform cancellation in the opposite region by a proper design of the target signal phase, chosen in a limited discrete set by exhaustive search. In the psychoacoustic literature, phase or amplitude differences between the left and right ears of a person are referred to as the interaural phase difference (IPD) and interaural amplitude difference (IAD), respectively. They are used by the auditory system for sound source localization [18]. This means that if the control points are virtually located in proximity to the human ears, the proposed algorithm may introduce spatial effects in the reproduced sound. However, such effects are expected to have a minor impact on vocal sound.

One of the disadvantages of the PM method [3] is that the performance is not robust against perturbations, e.g., errors in the measurement positions with respect to the true positions of the listeners. In our realistic scenario, a regularized PM technique [19], aimed in theory to limit the energy emitted by the filters and increase the robustness of the performance against system perturbations [4, 20, 21], was also not effective. The most relevant works to solve this problem were carried out in [22, 23], that do not assure regularized performance in our setup, in [24] and subsequently in [25], that do not guarantee improvement at higher frequencies.

For this reason, a novel design method for PSZ filters, based on the PM algorithm improved to account for several measurements of the acoustic channel, is proposed. In particular, this technique, referred to as statistical pressure matching (SPM), uses several measurements to average out the effects of the

positioning errors, improve the fidelity of the reproduced audio and enhance the AC.

Essentially, the PSZ methods, such as PM, involve designing filters to control the audio signals at loudspeaker inputs based on the acoustic responses and assessing the resulting performance. Early evaluations and design of these systems were often performed under ideal conditions, which could involve idealized acoustic responses or anechoic environments, e.g., [1–3, 5, 14]. Indeed, in an ideal scenario, with no obstacles and boundaries, the propagation of the sound wave can be described by an attenuation and delay [26], whereas, in a realistic scenario, the reflected components become relevant in the loudspeaker-microphone transfer functions and these can be partially controlled. The narrower the environment considered, the more relevant the reflected components, so it becomes more difficult to control them. Therefore, such ideal conditions tend to overestimate the realistically achievable performance of these systems.

Previous research has delved into PSZ systems within automotive settings, as evident in works such as [16, 27–29]. These studies primarily relied on direct measurements for obtaining realistic performance assessments. However, throughout our research, access to an experimentally equipped vehicle was not consistently available, rendering this approach impractical. To overcome this limitation, we adopted an alternative strategy, involving one or few sets of intentionally mismatched measured frequency responses (FRs) in comparison to those utilized for filter design. This deliberate mismatch introduced subtle perturbations in the measured acoustic responses, enhancing the realism of our evaluations. However, a limited number of measurements may be available, whereas a larger number would be required for reliable assessment.

Various works dealing with the effects of configuration changes in the acoustic channel matrix and performance sensitivity concerning error in the measured transfer function can be found in the literature, e.g., [30–32]. Some other works in the literature are aimed at generating FRs for PSZ applications, e.g., [33–38]. However, none of these works is suitable for our purpose. Indeed, some works are too approximate, others are not realistic, too complex

or computationally heavy.

To address these limitations, in this thesis, a stochastic model is proposed to generate mismatched FRs for robust performance prediction. This model uses a Gaussian distribution to describe complex coefficients in the frequency domain, considering both independent and identically distributed (IID) and frequency-correlated (FC) samples. Unlike some prior works that introduce perturbations to acoustic responses before filter calculations, this approach adds perturbations in the performance evaluation step to avoid overestimation of the performance.

This dissertation deals with various aspects of a PSZ system. For this reason, it is organized into five main chapters. The first two chapters focus on the review and mathematical description of a PSZ system, the metrics adopted for performance evaluation, the experimental scenario and details regarding the simulation-based analysis. Our contributions are presented in the last three chapters, structured as follows.

- *Measurement processing for performance enhancement.* In Chapter 3, after providing an overview of subband filtering, frequency-dependent trimming (FDT) and windowing functions used for FDT, the proposed methods for processing the measured IRs, before the PSZ filter design stage, are described and analyzed by means of numerical simulations.

- *PSZ filter design methods.* In Chapter 4, some implementation aspects related to the ACC filter design method for applications in real-world systems are discussed. Specifically, the effects of the scaling method applied to these filters is investigated by means of numerical simulations and measurements performed in an experimentally set up vehicle. Thereafter, two novel PSZ filter design methods are presented. In particular, we develop the PM method with a focus on optimizing the target sound phase, and we also introduce a statistical approach to PM. We present and analyze the results through a combination of numerical simulations and real-world experiments carried on within the car cabin, comparing

the proposed methods with the original formulation of the PM.

- *Stochastic modeling of the measurement mismatch.* Chapter 5 delves into the development of stochastic models to generate mismatched FRs. The primary goal here is to enable robust, reliable, and realistic predictions regarding the performance of a PSZ system in the context of a vehicle scenario. These novel methods are subsequently subjected to a comprehensive evaluation, comparing their prediction capabilities with the performance estimations derived from in-situ measurements and the average results obtained using mismatched data.

Chapter 1

Review of personal sound zone systems

In this chapter, the problem of creating personalized and individual acoustic regions in a shared environment is described providing a numerical model. The algorithms used in this work, aimed at designing suitable filters to control the sound field, are reviewed. Furthermore, the main adopted evaluation metrics are described. Finally, an evaluation method for a realistic scenario adopted in this work is presented.

1.1 Problem definition

Consider the scenario illustrated in Fig. 1.1. We wish to be able to listen to audio content in a specific region, specified by a subset of M_B microphones of the microphones set $\{1, 2, \ldots, M\}$, namely the bright zone (BZ), and having silence, or listen to a different audio content, in a different region specified by a subset of M_D microphones, namely the dark zone (DZ).

The system includes L loudspeakers (sources) that generate the set of acoustic signals $\{u_1(f^{(k)}), \ldots, u_\ell(f^{(k)}), \ldots, u_L(f^{(k)})\}$, where $f^{(k)}$ is the k-th discrete frequency, and M microphones that measure the acoustic pressures

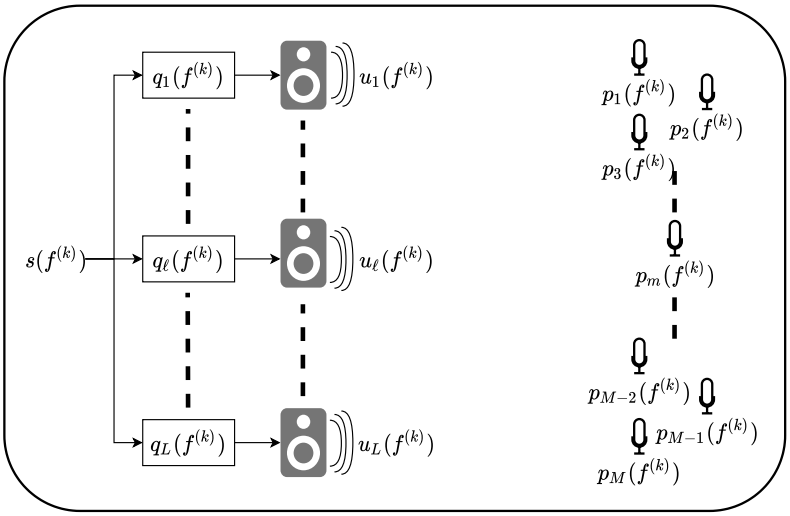


Figure 1.1: Considered scenario.

$\{p_1(f^{(k)}), \ldots, p_m(f^{(k)}), \ldots, p_M(f^{(k)})\}$, at specific points of interest (control points). The ℓ-th loudspeaker is driven by the signal $s(f^{(k)})$ filtered by the personal sound zone (PSZ) filter $q_\ell(f^{(k)})$. Moreover, assuming a linear relation between the sound produced by the loudspeakers and the acoustic pressures measured at the microphones, let $z_{m,\ell}(f^{(k)})$ be the frequency response (FR) between the m-th microphone and the ℓ-th loudspeaker. By the superposition principle, we can express the sound pressure at the k-th microphone ($k = 1, 2, \ldots, M$) as follows:

$$p_k(f^{(k)}) = \sum_{\ell=1}^{L} z_{k,\ell}(f^{(k)}) u_\ell(f^{(k)}). \tag{1.1}$$

At a fixed frequency, we can group the complex coefficients of all loudspeaker input signals into a column vector denoted as $\mathbf{u}(f^{(k)}) \in \mathbb{C}^{L\times 1}$:

$$\mathbf{u}(f^{(k)}) = \left[u_1(f^{(k)}) \ldots u_\ell(f^{(k)}) \ldots u_L(f^{(k)})\right]^T, \tag{1.2}$$

where $(\cdot)^T$ represents the transpose operation. Similarly, the sound pressures at

the microphones can be grouped into another column vector $\mathbf{p}(f^{(k)}) \in \mathbb{C}^{M\times 1}$:

$$\mathbf{p}(f^{(k)}) = \left[p_1(f^{(k)}) \dots p_k(f^{(k)}) \dots p_M(f^{(k)})\right]^T, \tag{1.3}$$

and the FRs between all loudspeaker-microphone pairs can be represented by a matrix $\mathbf{Z}(f^{(k)}) \in \mathbb{C}^{M\times L}$:

$$\mathbf{Z}(f^{(k)}) = \begin{bmatrix} z_{1,1}(f^{(k)}) & \cdots & z_{1,\ell}(f^{(k)}) & \cdots & z_{1,L}(f^{(k)}) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ z_{k,1}(f^{(k)}) & \cdots & z_{k,\ell}(f^{(k)}) & \cdots & z_{k,L}(f^{(k)}) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ z_{M,1}(f^{(k)}) & \cdots & z_{M,\ell}(f^{(k)}) & \cdots & z_{M,L}(f^{(k)}) \end{bmatrix}. \tag{1.4}$$

This allows us to describe the relationship between loudspeaker input signals and sound pressures at the microphones in matrix form:

$$\mathbf{p}(f^{(k)}) = \mathbf{Z}(f^{(k)})\mathbf{u}(f^{(k)}). \tag{1.5}$$

Note that, the matrix $\mathbf{Z}(f^{(k)})$ can be obtained by measuring the transfer functions between the electrical input of each loudspeaker and the electrical output of each microphone. In this way, each function $z_{m,\ell}(f^{(k)})$ incorporates also the FR of the ℓ-th loudspeaker and the k-th microphone.

To control the sound field, we can design a set of filters $\left\{q_1(f^{(k)}), \dots, q_L(f^{(k)})\right\}$ so that the input signal driving the ℓ-th loudspeaker in the frequency domain is given by:

$$u_\ell(f^{(k)}) = s(f^{(k)})q_\ell(f^{(k)}), \tag{1.6}$$

where $s(f^{(k)})$ is the audio signal to be reproduced. In this context, we consider the overall system as the linear system shown in Fig.1.2. Assuming $s(f^{(k)}) = 1$, we have $u_\ell(f^{(k)}) \equiv q_\ell(f^{(k)})$, and the matrix equation (1.5) becomes:

$$\mathbf{p}(f^{(k)}) = \mathbf{Z}(f^{(k)})\mathbf{q}(f^{(k)}). \tag{1.7}$$

To simplify the notation, henceforth, we will indicate the dependence on the discrete frequency $f^{(k)}$ with the superscript k, e.g., $\mathbf{Z}^{(k)} = \mathbf{Z}(f^{(k)})$, and it will be neglected where it is not needed.

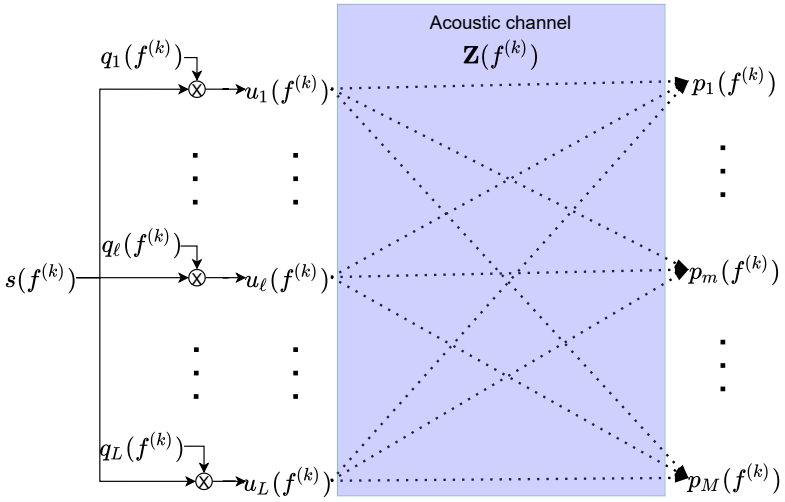


Figure 1.2: Scheme representing the linear system that relates the input audio signal, the filters used to control the sound field, the loudspeaker input signals and the sound pressures at the microphones.

1.2 PSZ filter design methods

1.2.1 Pressure matching method

The pressure matching (PM) algorithm attempts to achieve a desired target sound field $\hat{\mathbf{p}} \in \mathbb{C}^{M\times 1}$, hence, the idea is to find a vector $\mathbf{q}_{opt}$ that solves

$$\hat{\mathbf{p}} = \mathbf{Z}\mathbf{q}_{opt} \tag{1.8}$$

However, if the matrix $\mathbf{Z}$ is not square, there may not be an exact or unique solution. For this reason, [3] proposed to solve the system in the least square sense minimizing the overall error $\mathbf{e} = \mathbf{p} - \hat{\mathbf{p}}$ according to

$$J_{PM} = \|\mathbf{e}\|^2 = \mathbf{e}^H\mathbf{e} = (\mathbf{p} - \hat{\mathbf{p}})^H (\mathbf{p} - \hat{\mathbf{p}}) = (\mathbf{Z}\mathbf{q} - \hat{\mathbf{p}})^H (\mathbf{Z}\mathbf{q} - \hat{\mathbf{p}}) \tag{1.9}$$

where $(\cdot)^H$ denotes the complex conjugate and transpose operator. Considering $M > L$, and setting to zero the gradient of J with respect to $\mathbf{q}$, the optimal

solution is

$$\mathbf{q}_{opt} = \left(\mathbf{Z}^H\mathbf{Z}\right)^{-1}\mathbf{Z}^H\hat{\mathbf{p}} \tag{1.10}$$

where the matrix $\left(\mathbf{Z}^H\mathbf{Z}\right)^{-1}\mathbf{Z}^H$ is the Moore-Penrose pseudoinverse of the matrix $\mathbf{Z}$. If $M < L$, the matrix $\mathbf{Z}^H\mathbf{Z}$ has not an inverse because its size is $L \times L$ and its rank at most $M < L$, hence it is singular. In this case, the linear system (1.8) is indeterminate and admits infinite solutions. However, using the properties of the pseudoinverse matrix [39], the minimum norm solution can be expressed as

$$\mathbf{q}_{opt} = \mathbf{Z}^H\left(\mathbf{Z}\mathbf{Z}^H\right)^{-1}\hat{\mathbf{p}}. \tag{1.11}$$

However, 1.11 requires the inversion of a matrix that can still become close to singular for certain geometrical arrangements. This issue is known as ill-conditioning problem and may introduce significant errors in the computation.

In [19] a Tikhonov regularization parameter that solves the ill-conditioning problem and adds a constraint on the array effort was introduced. According to this approach, the cost function that must be minimized becomes

$$J_{PM} = \|\mathbf{e}\|^2 + \beta\mathbf{q}^H\mathbf{q} = \left(\mathbf{Z}\mathbf{q} - \hat{\mathbf{p}}\right)^H\left(\mathbf{Z}\mathbf{q} - \hat{\mathbf{p}}\right) + \beta\mathbf{q}^H\mathbf{q}. \tag{1.12}$$

Setting to zero the gradient of J with respect to $\mathbf{q}$, the optimal solution is

$$\mathbf{q}_{opt} = \begin{cases} \left(\mathbf{Z}^H\mathbf{Z} + \beta\mathbf{I}\right)^{-1}\mathbf{Z}^H\hat{\mathbf{p}} & \text{if } M > L \\ \left(\mathbf{Z} + \beta\mathbf{I}\right)^{-1}\hat{\mathbf{p}} & \text{if } M = L \\ \mathbf{Z}^H\left(\mathbf{Z}\mathbf{Z}^H + \beta\mathbf{I}\right)^{-1}\hat{\mathbf{p}} & \text{if } M < L. \end{cases} \tag{1.13}$$

An example of pseudo-code that describes the algorithm for the implementation of (1.13) can be found in [40].

In (1.13) the regularization parameter β is an additive term equal for all diagonal terms of the matrix $\mathbf{Z}^H\mathbf{Z}$ (or $\mathbf{Z}$ or $\mathbf{Z}\mathbf{Z}^H$) and ensures that the required matrix inversion is performed on a non singular matrix[1].

[1]For instance, if the matrix $\mathbf{Z}$ can be spectrally decomposed as $\mathbf{Z} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$, where $\mathbf{U}$ is a matrix of column-wise eigenvectors and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues, we have that $\mathbf{Z} + \beta\mathbf{I} = \mathbf{U}(\mathbf{\Lambda} + \beta)\mathbf{U}^{-1}$. Hence, β can also be added directly to all eigenvalues of the matrix $\mathbf{Z}$.

For implementation, when it is not specified, the considered target sound pressure is

$$\hat{p}_m(\omega) = \begin{cases} \dfrac{e^{-\jmath\omega M_0}}{\sqrt{N/2}} & m\text{-th control point in the BZ} \\ 0 & m\text{-th control point in the DZ} \end{cases} \tag{1.14}$$

where $M_0 = F_s/2$ is the delay expressed in samples such that the overall impulse response (IR) at the bright microphones is centered around half a second, where F_s is the sampling frequency in samples per second.

Since the working frequency range of the used loudspeakers does not cover the entire audible frequency range $[20, 20000]$ Hz, to avoid excessive power allocation out of the loudspeaker operational frequency range in the matrix inversion in (1.13), the regularization parameter is defined as

$$\beta = \frac{\beta_F}{\lambda^{max}} \tag{1.15}$$

where β_F is a proper scaling factor and λ^{max} is the maximum eigenvalue of the matrices $\mathbf{Z}^H\mathbf{Z}$, $\mathbf{Z}$ or $\mathbf{Z}\mathbf{Z}^H$ depending on the considered case in (1.13). In this way, for high λ^{max}, a small regularization parameter is used, while for small Λ_{max}, which means that the system has low capability to generate energy, a high value of β is used. In the numerical results described in the following chapters, the parameter β_F is set to 10^{-6} based on experimental considerations.

In this work, when it is not specified, the default design method for PSZ filters is PM as here described.

1.2.2 Acoustic contrast control

The acoustic contrast control (ACC) method, initially introduced with the brightness control method in [6], aims at maximizing the ratio between the sound pressure levels in two different acoustic regions. In the latest implementations, an indirect optimization [41] is usually adopted that consists in minimizing the average acoustic pressure in a dark region with a constraint

on the average acoustic pressure in a bright region and the introduction of a regularization parameter to improve the stability of the solution.

Consider to divide the $\mathbf{Z}$ matrix into two submatrices $\mathbf{Z}_B$ and $\mathbf{Z}_D$ encompassing the FRs between all the loudspeakers and the M_B microphones assumed to be in the BZ and the M_D microphones assumed to be in the DZ, respectively.

The spatial correlation matrix can be expressed as [6]

$$\mathbf{R} = \frac{1}{M}\mathbf{Z}^H\mathbf{Z} \tag{1.16}$$

so that, we can express the indirect ACC formulation by the cost function

$$J_{ACC} = \mathbf{q}^H\mathbf{R}_D\mathbf{q} + \lambda\mathbf{q}^H\mathbf{R}_B\mathbf{q} + \beta\mathbf{q}^H\mathbf{q}, \tag{1.17}$$

where λ is a Lagrange multiplier, β is a regularization parameter and $\mathbf{R}_B$ and $\mathbf{R}_D$ are the corresponding spatial correlation matrices for the bright and dark regions, respectively. The minimization of (1.17) is equivalent to the eigenvalue problem

$$\lambda\mathbf{q} = -\mathbf{R}_B^{-1}(\mathbf{R}_D + \beta\mathbf{I})\mathbf{q}, \tag{1.18}$$

where $\mathbf{I}$ is the identity matrix and $[\cdot]^{-1}$ indicates the matrix inversion operation. The optimal solution $\mathbf{q}_{opt}$ of (1.18) corresponds to the eigenvector associated with the minimum eigenvalue of the matrix $\mathbf{R}_B^{-1}(\mathbf{R}_D + \beta\mathbf{I})$, namely the eigenvector associated with the largest eigenvalue of the inverse matrix $(\mathbf{R}_D + \beta\mathbf{I})^{-1}\mathbf{R}_B$. Without any further processing of the solution, we cannot control the sound amplitude generated in the BZ, nor the sound phase.

To control the phase of the sound field in the BZ, in [4], the authors introduced a complex coefficient to correct the phase of the optimal vector $\mathbf{q}_{opt}$ given by the acoustic energy difference maximization algorithm [8]. The same solution can be applied to the ACC technique, therefore we multiply the vector $\mathbf{q}_{opt}$ found by minimizing (1.17) with

$$W = e^{\jmath\left(\hat{\phi} - \arg\left\{\mathbf{z}_{B,ref}^T\mathbf{q}_{opt}\right\}\right)} \tag{1.19}$$

where $\hat{\phi}$ is the desired phase and $\mathbf{z}_{B,ref}$ is a vector composed of the complex coefficients for a fixed frequency of the FRs between all the loudspeakers and a reference control point in the BZ, hence, $\arg\left\{\mathbf{z}_{B,ref}^{T}\mathbf{q}_{opt}\right\}$ is the phase generated by the solution of (1.17) for a fixed frequency evaluated in a reference bright control point.

A similar scaling procedure can be applied to have the desired amplitude at a reference microphone in the BZ, however, this will be further discussed in Chapter 4.

Following the principle of power allocation for PM filters, the regularization parameter is chosen as inversely proportional to the maximum eigenvalue $\lambda_{R_D}^{max}$ of the spatial correlation matrix $\mathbf{R}_D$ and the number of dark control points M_D. Accordingly, β in (1.17) is calculated as

$$\beta = \frac{1}{M_D}\frac{\beta_F}{\lambda_{R_D}^{max}} \tag{1.20}$$

where $\beta_F = 10^{-6}$ as for PM. Furthermore, no power is allocated for a given frequency if

$$\lambda_{R_B}^{max} < 10^{-2}\Lambda_{R_B}^{max} \tag{1.21}$$

where $\lambda_{R_B}^{max}$ is the maximum eigenvalue of the spatial correlation matrix $\mathbf{R}_B$ for a fixed frequency and $\Lambda_{R_B}^{max}$ is the maximum eigenvalue of the spatial correlation matrices $\mathbf{R}_B$ over all frequencies.

1.3 Performance metrics

1.3.1 Acoustic contrast

According to [42] and [16], acoustic contrast (AC) is defined as the ratio between the spatial averaged squared pressure in the BZ and that in the DZ, and is strictly related to the ACC method. Indeed, the ACC technique is based on the maximization of AC between two separated zones. In a dB scale, and for

a fixed $f^{(k)}$, we can express it as

$$C(f^{(k)}) = 10\log_{10}\left(\frac{M_D\|\mathbf{p}_B(f^{(k)})\|^2}{M_B\|\mathbf{p}_D(f^{(k)})\|^2}\right). \tag{1.22}$$

The previous formulation of AC specifies how good the contrast between the two sound zones at a given frequency is. The AC is typically represented as an average value over octave bands[2], widely used in acoustics. This approach also aims to simplify graphical visualization.

For some experiments, the AC is also measured directly in the vehicle. These measurements are performed as follows. A pink noise signal, whose sound is perceived in a more balanced and natural way by the human hearing, is filtered by filters obtained by the summation (according to the superposition principle) of the two PSZ filter sets designed for left and right ear regions of the driver, or the front passenger (codriver), and reproduced by the loudspeakers, while dummy heads (DHs) equipped with binaural microphones record the sound in the two zones. Then the energies of the recorded sounds are averaged over the left and right control points of the same acoustic region (driver and codriver). Finally, the AC is calculated from the ratio between the two energies.

1.3.2 Reproduction sound error

The sound error is suitable for evaluating the accuracy of the sound field reproduced in the sound zones. In agreement with [42] and [16], it can be expressed as the mean squared error between the target sound pressure and the reproduced sound pressure. In a logarithmic scale, for a fixed $f^{(k)}$, the mean square error normalized to the number of microphones can be written as

$$\varepsilon(f^{(k)}) = 10\log_{10}\left(\frac{1}{M_B}\|\mathbf{p}_B(f^{(k)}) - \hat{\mathbf{p}}_B(f^{(k)})\|^2\right). \tag{1.23}$$

[2] A band is said to be an octave in width when the upper band limit is twice the lower band limit. A one-third octave band is defined as a frequency band whose upper band limit is the lower band limit multiplied by the cube root of two [43].

1.3.3 Moments of PSZ filters

The moments of a continuous or discrete function provide quantitative measures related to the shape of the function. In probability theory and statistics, the moments are often used to compactly characterize the probability distribution of a random variable. Indeed, the first-order moment corresponds to the expected value, the central second-order one is the variance, the third-order one is associated with the skewness [44], and so on.

We are interested in values that quantify and describe the relevance of the so-called "pre-ringing", namely a leading component in the IRs of PSZ filters necessary to control the late reflections present in the measured IRs, and how the energy is distributed in the IRs of the designed filters. To this purpose, we consider the normalized energy temporal density, as in [15], that can be expressed as

$$U[k] = \frac{h^2[k]}{E} \tag{1.24}$$

where $h[k]$ is the amplitude of the k-th sample of a given filter IR and $E = \sum_\ell h^2[\ell]$ is its energy. Note that $\sum_k U[k] = 1$.

The central n-th order moment can be calculated by using its definition as [44]

$$\mu_n = \sum_{k=-\infty}^{+\infty} (k-c)^n \, x[k] \tag{1.25}$$

for a discrete function $x[k]$, where $c = 0$ for 1-st order moment, $c = \mu_1$ for $n > 1$, and $\sum_{k=-\infty}^{+\infty} x[k] = 1$. Since we deal with discrete time signals, the discrete case is here be considered.

Using (1.25), the first-order moment of (1.24) is

$$\mu_1 = \sum_k kU[k] = \frac{1}{E} \sum_k kh^2[k]. \tag{1.26}$$

It locates (in samples) the center of mass of the energy density (1.24). In acoustics, with reference to a loudspeaker-microphone IR, μ_1 is called center

time and affects speech intelligibility [45] since it is correlated with clarity[3].

Using (1.25), the second-order central moment of (1.24) is

$$\mu_2 = \sum_k (k-\mu_1)^2 U[k] = \frac{1}{E}\sum_k (k-\mu_1)^2 h^2[k] \tag{1.27}$$

where μ_1 is given by (1.26). It quantifies the dispersion of the evaluated function, i.e., in this case, the energy density, on the time axis. Indeed, in [15] the authors called it the central moment of the normalized energy temporal density and used it as an indicator of the filter compactness in the time domain.

The quantity μ_2 expresses the so-called variance in probability theory and statistics. However, from a graphical point of view, it makes more sense to consider the standard deviation, that is the square root of the variance [44]. Indeed, $\sqrt{\mu_2}$ gives a straightforward measure of the filter compactness in the time domain. By analogy with physical systems, $\sqrt{\mu_2}$ will be referred to as radius of gyration of the energy density (RoGoED) and measured in samples.

According to (1.25), the third-order central moment of (1.24)

$$\mu_3 = \sum_k (k-\mu_1)^3 U[k] = \frac{1}{E}\sum_k (k-\mu_1)^3 h^2[k] \tag{1.28}$$

gives a measure of the asymmetry of the sequence $U[k]$ about its center of mass μ_1. Indeed, let us decompose (1.28) as

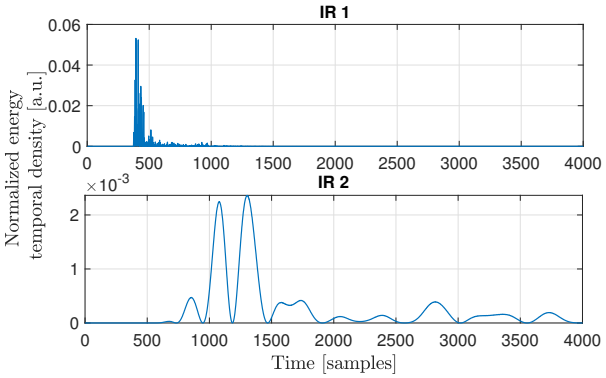
$$\mu_3 = \sum_{k=-\infty}^{k=\lfloor\mu_1\rfloor-1} (k-\mu_1)^3 U[k] + \sum_{k=\lfloor\mu_1\rfloor}^{k=+\infty} (k-\mu_1)^3 U[k] \tag{1.29}$$

where $\lfloor\cdot\rfloor$ is the floor function. Note that the first summation in (1.29) is negative ($k < \mu_1$), whereas the second summation is positive ($k > \mu_1$) due to the cubic exponent. If the sequence is symmetric with respect to its center of mass, the terms of the first summation cancel out with the corresponding terms in the second summation, so that μ_3 approaches zero. If the sequence

[3]The clarity factor is defined as the ratio between the early energy (direct component and first reflections) and late energy (reverberant component).

is skewed to the right, the second summation overcomes the first summation ($\mu_3 > 0$), and vice-versa if the sequence is skewed to the left ($\mu_3 < 0$). It is convenient to normalize (1.28) with respect to $\mu_2^{3/2}$ resulting in a dimensionless quantity [46].

As a practical example, let us consider the normalized temporal energy densities of two different IRs measured in a vehicle cabin. One IR (*IR 1*) is related to a loudspeaker operating in the low-frequency range and one (*IR 2*) to a loudspeaker working at medium-high frequencies. The two functions are shown in Fig. 1.3(a) and their parameters are summarized in the adjacent table. We can see in the figure that the temporal energy density of the loudspeaker operating at low-frequency range (*IR 2*) has a much higher temporal spreading than the other one (*IR 1*). Indeed, a high-frequency sound wave has a higher attenuation with the reflections with respect to a low-frequency sound wave giving a more compact IR measured between the source and the



(a)

Parameters	IR1	IR2
μ_1 [samples]	477	1855
$\sqrt{\mu_2}$ [samples]	176	1429
$\frac{\sqrt{\mu_3}}{\mu_2^{3/2}}$ (skewness)	42.69	4.78

(b)

Figure 1.3: (a) Normalized temporal energy densities of the IR for a loudspeaker operating at medium-high frequencies (top, *IR 1*) and a loudspeaker working at low frequencies (bottom, *IR 2*). The IRs are measured in the scenario described in Chapter 2. (b) Parameters of the normalized temporal energy densities plotted in (a).

measurement point. This fact is confirmed by the first- and the square root of the second-order moments that are higher for the second normalized temporal energy density. The normalized third-order moment confirms that both sequences are right-skewed with energies more concentrated on the left of the mass center[4].

1.3.4 Short-time objective intelligibility

The short-time objective intelligibility (STOI) measure gives a prediction of the percentage of correctly understood words averaged over a group of users [47] and, in general, it is defined between a clean signal and its degraded version. In Section 2.2, illustrative use cases of STOI for PSZ system evaluation are provided.

Let x and y be the clean and degraded signals, respectively. Performing the short-time Fourier transform (STFT) [48] of these signals, we denote with $\hat{x}(k,m)$ and $\hat{y}(k,m)$ the k-th discrete Fourier transform (DFT) element of the m-th frame of the signals x and y, respectively. Considering the clear signal x, we can define the norm of the j-th one-third octave band as

$$X_j(m) = \sqrt{\sum_{k=k_l(j)}^{k_u(j)-1} |\hat{x}(k,m)|^2} \tag{1.30}$$

where $k_l(j)$ and $k_u(j)$ are the indices of the lower and upper limits of the j-th one-third octave band, respectively. For a fixed one-third octave band j, (1.30) can be seen as an average temporal envelope of the signal x in that subband. Considering N consecutive time frames, we can form the vector

$$\mathbf{X}_{j,m} = [X_j(m-N+1), X_j(m-N+2), \ldots, X_j(m)]^T \tag{1.31}$$

where $[\cdot]^T$ denotes the vector transpose operator. The same can be defined for the degraded signal y as

$$\mathbf{Y}_{j,m} = [Y_j(m-N+1), Y_j(m-N+2), \ldots, Y_j(m)]^T . \tag{1.32}$$

[4] A right-skewed (or positive skewed) distribution has the mode on the left with respect to the mean and vice-versa for left-skewed [44].

Before the definition of the STOI, $\mathbf{Y}_{j,m}$ is normalized and clipped, hence, we define the new vector $\bar{\mathbf{Y}}_{j,m}$ populated by the terms

$$\bar{Y}_j(m-N+n) = \min\left(\frac{\|\mathbf{X}_{j,m}\|}{\|\mathbf{Y}_{j,m}\|}Y_j(m-N+n), \left(1+10^{-\gamma/20}\right)X_j(m-N+n)\right) \tag{1.33}$$

where γ is an arbitrary parameter and $\|\cdot\|$ indicates the Euclidean norm operator of the considered vector. The first argument of the $\min(\cdot)$ operator in (1.33) is an element of the scaled version of $\mathbf{Y}_{j,m}$ with norm equal to that of $\mathbf{X}_{j,m}$. The signal clipping operated by the $\min(\cdot)$ function reduces the underestimation of the intelligibility in case of severely degraded signal. To better understand, consider the following example. Consider a signal with energy gathered in a relatively small portion of the considered time frame and a degraded signal obtained by introducing noise outside that portion of the time frame. Without performing the clipping operation, the STOI would predict low intelligibility due to the noise, however, the considered noise does not perceptually degrade the intelligibility. This clipping operation limits the signal $\bar{\mathbf{Y}}_{j,m}$ to a level controlled by the parameter γ which is chosen empirically (in dB scale) [47].

For notation purposes, we define $\mathbf{M}(\mathbf{A}_{j,m})$ as the $N \times 1$ vector populated by N repetitions of the sample mean of the elements of the vector $\mathbf{A}_{j,m}$ (e.g., $\mathbf{X}_{j,m}$ and $\bar{\mathbf{Y}}_{j,m}$) which is calculated as

$$\mu_{j,m} = \frac{1}{N}\sum_{n=1}^{N} A_j(m-N+n). \tag{1.34}$$

Then, an intermediate intelligibility measure of the j-th one-third octave band and the $[(m-N+1),(m-N+2),\ldots,m]$ time frame can be expressed as [47]

$$d_{j,m} = \frac{\left(\mathbf{X}_{j,m}-\mathbf{M}(\mathbf{X}_{j,m})\right)^T\left(\bar{\mathbf{Y}}_{j,m}-\mathbf{M}\left(\bar{\mathbf{Y}}_{j,m}\right)\right)}{\|\mathbf{X}_{j,m}-\mathbf{M}(\mathbf{X}_{j,m})\| \ \left\|\bar{\mathbf{Y}}_{j,m}-\mathbf{M}\left(\bar{\mathbf{Y}}_{j,m}\right)\right\|}. \tag{1.35}$$

Finally, averaging over all the considered N_{obs} one-third octave bands and M

time frames, the STOI is given by

$$d = \frac{1}{MN_{obs}} \sum_{j,m} d_{j,m}. \tag{1.36}$$

The STOI (1.36) represents a physical measure, but does not fully account for the sound perception features of the human ears. For this reason, the authors in [47] carried out listening tests with the aim of mapping (1.36), i.e., a physical measure, into a perceptual measure, i.e., intelligibility, through the logistic function

$$f(d) = \frac{100}{1 + e^{ad+b}} \tag{1.37}$$

where d is given by (1.36), and the parameters a and b can be obtained by fitting the subjective data by a nonlinear least square procedure.

The intelligibility prediction model proposed in [47], obtained by mapping the STOI (1.36) by (1.37), showed overall better performance with respect to other objective intelligibility models for three different listening test experiments. However, it must be taken into account that the prediction model was validated considering the speech spectrum from 140 Hz to 4500 Hz, approximately, a sample rate of 10 kHz, and parameters γ in (1.33), a and b in (1.37) optimized for a specific type of noise introduced in the clear audio which differs from our case. Moreover, a possible underestimation of the intelligibility caused by the spectral differences of the clean and disturbed signals is claimed. Yet, STOI in [49, 50] used 16 kHz sampling, suggesting that the evaluation of audio signal with sampling frequency up to 16 kHz, greater than the 10 kHz considered in STOI [47], is not a concern.

Later, an extended version of the STOI [51] and more advanced methods for intelligibility prediction, such as [52], based on deep learning, were proposed. Moreover, non-intrusive intelligibility prediction algorithms, which do not require a reference clear signal, are present in the literature. A review of this last category can be found in [53]. While alternatives exist, we use the STOI from [47] for PSZ system performance evaluation due to its simplicity, established usage, and readily available software toolbox.

1.4 Performance evaluation in a realistic scenario

Early evaluations of these systems were often performed under ideal conditions, which could involve idealized acoustic responses or anechoic environments, e.g., [1–3, 5, 14]. However, such ideal conditions tend to overestimate the realistically achievable performance of these systems.

To obtain more realistic performance evaluations, researchers may conduct direct measurements in real-world settings or use numerical simulations. One approach involves using one or a few sets of measured FRs that are intentionally mismatched with respect to those used for filter design. This mismatch introduces slight perturbations in the measured acoustic responses, making the numerical evaluation more realistic.

In the following, $\mathbf{Z}_{des}$ refers to the matrix of FRs used for filter design and $\mathbf{Z}_{mm}$ as a (mismatched) matrix of FRs, i.e., different from $\mathbf{Z}_{des}$. Later on, we will explain better how and how much these matrices differ. Said that, we can model numerically the mismatch between $\mathbf{Z}_{des}$ and $\mathbf{Z}_{mm}$ with the error matrix

$$\mathbf{E} = \mathbf{Z}_{des} - \mathbf{Z}_{mm}. \tag{1.38}$$

Note that, having several acoustic channel matrices $\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_2, \ldots$ (each of them related to a different measurement of the acoustic channel), we can choose any one of them for filter design and one of the others for performance evaluation, as long as they are mismatched. For example, we can choose $\mathbf{Z}_1$ for filter design and $\mathbf{Z}_2$ for performance evaluation (assuming they are mismatched) but also vice-versa.

In this work, the PSZ filters are designed according to the considered cost function (e.g., PM or ACC) using the channel matrix $\mathbf{Z}_{des}$ and the evaluation is performed (e.g., AC and reproduction error) with a slightly different channel matrix $\mathbf{Z}_{mm}$, i.e., $\mathbf{Z}_{mm} \neq \mathbf{Z}_{des}$. This procedure enables numerical performance evaluation as if measured in the vehicle, providing a realistic analysis without additional measurements. Mismatched measurements involve repositioning microphones in the same locations. Fig. 1.4 compares AC curves

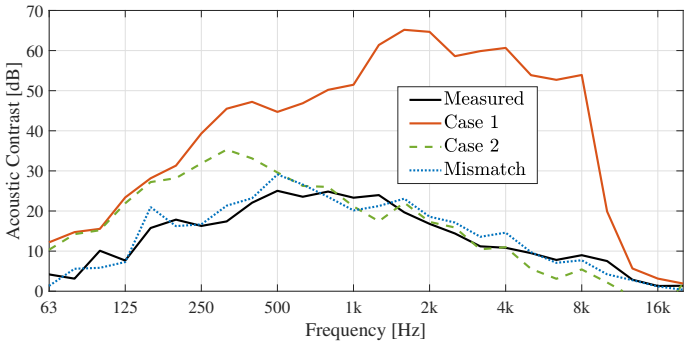


Figure 1.4: Comparison of AC curves obtained by numerical evaluation (*Case 1*, *Case 2* and *Mismatch*) and measured in a vehicle (*Measured*). The results are based on the configuration presented in Chapter 2.

obtained through numerical evaluation (obtained by filtering with measured FRs) with those directly measured in the vehicle. Indeed, we can observe that:

- the numerical evaluation based on the same set of FRs used for filter design, i.e., $\mathbf{Z}_{des}$, (labeled as *Case 1*) greatly overestimates the AC
- the numerical evaluation based on FRs measured after moving the microphones on its support, e.g., changing the height or the angle, (labeled as *Case 2*) still overestimates the AC, in particular at low frequencies
- the numerical evaluation with mismatched FRs (labeled as *Mismatch*), i.e., $\mathbf{Z}_{mm}$, predicts closely the AC measured in the vehicle (labeled as *Measured*).

Similar to the numerical evaluation of the AC (see Section 1.3.1), to measure the AC directly in the vehicle, a pink noise is filtered by the PSZ filters, designed for one of the regions, and reproduced, then, the signal is recorded by the microphones in the two sound regions. In the end, the ratio between the sound energies measured in the BZ and the energy in the DZ is calculated to obtain the AC.

Note that the AC curve labeled as *Case 2* is derived from moving the microphones on their support without removing and repositioning the sup-

port, which is crucial for mismatched measurements. Therefore, the removal and repositioning operation is necessary for a measurement to be considered properly mismatched. Measurements performed without this operation are not considered mismatched. Additionally, measurements before and after the remove-and-reposition operation are only considered mismatched if there is a slight height displacement, typically about 0.5 cm in our experimental setup.

Chapter 2

Experimental setup and simulation-based analysis

To perform a preliminary acoustic evaluation and comparison of the designed filters, a Simulink model was implemented to reproduce the audio perceived by the passengers in different acoustic regions in a vehicle cabin. However, because of the large number of filters that may be required for the simulation, the acoustic signals received at the ears of the passengers are calculated first. Hence, in the first step, the filtering of the input signals by the designed filters and the loudspeaker outputs with the vehicle impulse responses (IRs) are performed by a Matlab script. After all the required calculations, the audio signals are saved and they are ready to be loaded and reproduced by the Simulink model. This model is used as user interface for reproduction and comparison of the audio signals that should be perceived in the car with personal sound zone (PSZ) system.

In this chapter, the considered vehicle scenario and loudspeakers configuration, the Simulink model and the Matlab script will be described. Moreover, the implementation of short-time objective intelligibility (STOI) for PSZ system evaluation is explained.

2.1 Experimental scenario

During this work, two different loudspeaker configurations were primarily considered. However, both configurations feature 7 factory-installed loudspeakers, i.e., 2 front twitters, 2 front woofers, 2 pairs of rear twitter/woofer and a subwoofer placed in the trunk. Note that, each pair of rear twitter/woofer has twitter and woofer physically separated but driven by the same signal; hence, each pair is considered as a single loudspeaker by the filter design algorithms.

In addition to the factory-installed loudspeakers, the 2 configurations differ as follows.

A. The first considered configuration is schematized in Fig. 2.1. A few linear arrays of equally-spaced fullrange 1.5" loudspeakers, with an operational frequency range of about $[0.1, 10]$ kHz, spaced (center-to-center) by 4 cm (about 1.575"), were mounted in different positions in the cabin of the vehicle. Specifically, 2 arrays of 8 loudspeakers mounted on the dashboard (in front of the driver and codriver) with the loudspeaker membranes facing up, and 8 arrays of 4 loudspeakers mounted vertically on the 4 headrests (2 arrays for each headrest) directed towards the probable po-

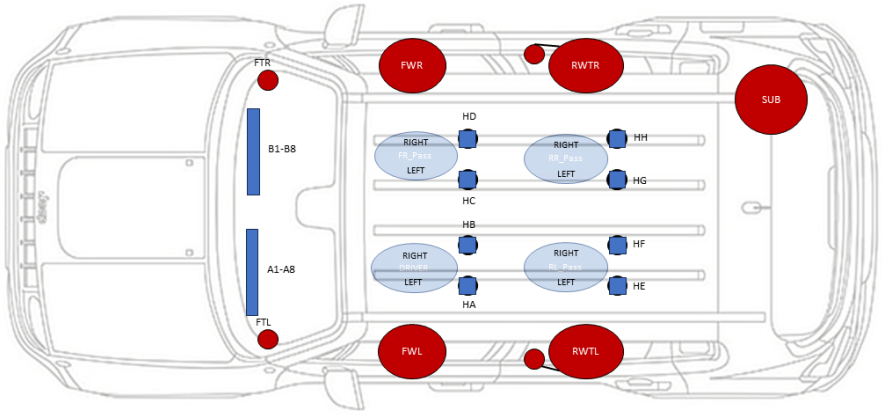


Figure 2.1: Scheme of the audio system in the vehicle for configuration A. Courtesy of ASK.

sition of the passenger heads

B. The second considered configuration is schematized in Fig. 2.2. 8 full-range loudspeakers, labeled with solid black circles, placed at the headrests (2 for each headrest of the vehicle) with the membrane directed towards the probable position of the ears of the passengers, and 5 full-range loudspeakers (bigger than the ones at the headrests), labeled with solid green circles, placed on the roof with the membrane directed towards the bottom of the vehicle.

The amplitude of the measured frequency responses (FRs) for the 2 types of the additional loudspeakers of the second configuration are plotted in Fig. 2.3. Note that, even if the 2 types of loudspeakers are both fullrange, those placed on the roof have a more limited operational frequency range with respect to the ones placed on the headrests and within the arrays.

In the numerical results, the first configuration will be referred to as *configuration A*, and the second one as *configuration B*.

In configuration A, the IRs were measured for all the 71 installed loudspeakers, however, only some of them were used for this work as a trade-off

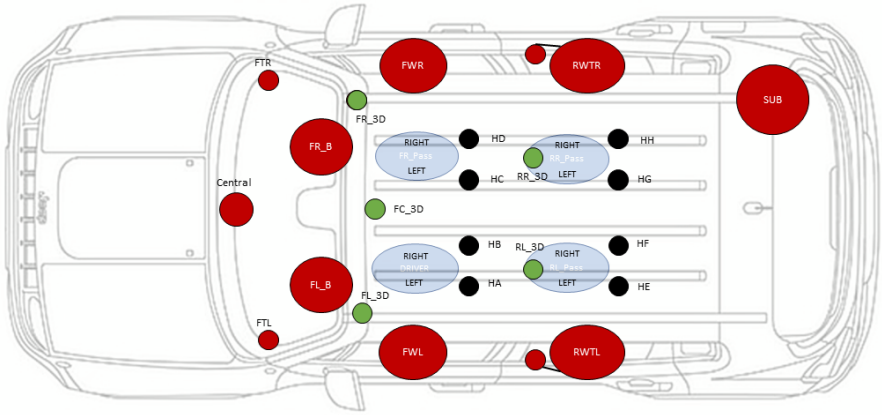


Figure 2.2: Scheme of the audio system in the vehicle for configuration B. Courtesy of ASK.

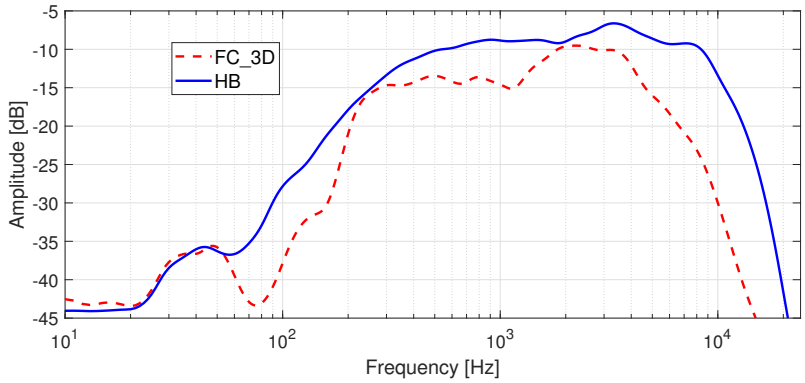


Figure 2.3: Amplitude of the measured FRs between the 2 loudspeakers, labeled as in Fig. 2.2, and the microphone placed in the probable position of the right ear of the driver passenger.

between performance and complexity. When it will be required, the loudspeakers used to control the sound regions will be listed.

Various measurements were taken for the considered configurations and they are reported below. All the IRs were measured considering an exponential sine sweep (ESS) [54] as a test signal sampled with sampling frequency $F_s =$ 48 kHz with a dummy head (DH) equipped with binaural microphones placed on top of the seats, as shown in Fig. 2.4.

A The chosen sound regions to be controlled were specified by the probable positions of the driver, codriver and rear passenger heads. For the driver and the codriver, different positions for each sound region were measured by moving back and forth the seat with steps of 3 cm, and up and down the DH along its support with steps of 2 cm. A total of 5 positions for each sound region were measured. We will refer to them as $(60, h60)$, $(30, h80)$, $(0, h100)$, $(-30, h120)$ and $(-60, h140)$, expressing the coordinates in mm. Note that the letter h stands for height and is used to make clear which coordinate is related to the horizontal axis and which one to vetical axis. For the rear passengers, the measurements were performed only by moving the DH up and down (again with steps

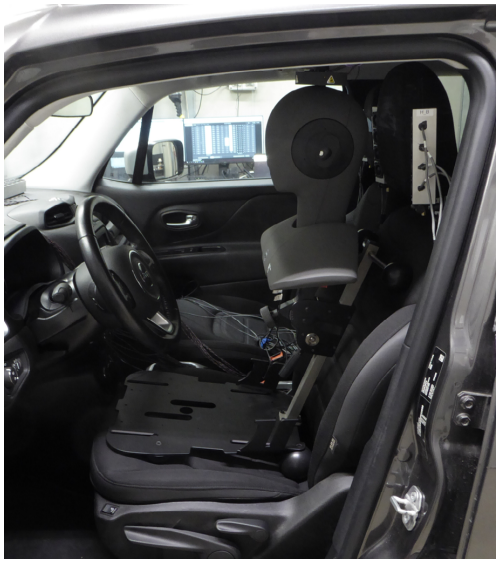


Figure 2.4: DH used for measurements positioned on the driver seat. Courtesy of ASK.

of 2 cm along its support) since the rear seats cannot be moved back and forth in the considered test vehicle. Moreover, a fixed head angle of 5 degrees relative to headrest position is considered, since the DH allows this setting. For the realistic performance evaluation (see Section1.4), a second measurement was performed for the same position at $(0, h100)$ by removing and repositioning the DH

B For the second configuration, the chosen sound regions to be controlled were specified by the probable positions of the driver, codriver. We performed 5 measurement sessions, the first 4 on the same day, varying head heights $(90, 100, 110$ mm relative to the base of the head support) and angles $(0, 5, 10$ degrees relative to headrest position), and a last one on a different day, focusing only on varying head heights $(80, 90, 95, 100, 105, 110, 115, 120$ mm). In each session, we measured the acoustic channel matrices for the various heights and possibly angles without moving the head supports. After each session, we removed and

repositioned the head supports to introduce a potential mismatch between sessions. The first 4 sessions led to overestimated acoustic contrast (AC) below 500 Hz (see comparison *Mismatch* and *Case2* in Fig. 1.4). This suggests that the acoustic channel responses in these 4 sessions remained partially correlated in this frequency range.

A discrete Fourier transform (DFT) with $N = 48000$ points was performed to get the FRs of the channel, which were then normalized to unit energy.

2.2 Simulation-based preliminary audio quality evaluation

2.2.1 Simulink model

The block diagram of the Simulink model root is shown in Fig. 2.5. As mentioned, the Simulink model only performs the routing to the audio device of the signals obtained with different designed filters and at different car positions.

In the blocks *ASK Filters*, *Filters 1* and *Filters 2* there are banks of *From Multimedia File* blocks that read the (stereo) audio signals calculated and saved by the Matlab script. All these signals are input to a *Multi-port Switch* block that outputs the input signal corresponding to the selected passenger position (driver, codriver, or rear passengers).

Once the audio signals for the position of interest are selected, the outputs of the different designed filter sets are input to the *Filter selector*. This block contains another *Multi-port Switch* block that outputs the audio signal filtered by the selected filter set.

The *Rotary Switch* blocks (one for the position and one for the filter set) are connected to *Constant* blocks, i.e., they control the value of the output constant, that drives the associated *Multi-port Switch* block. In this way, it is possible to change the reproduced audio during the simulation for direct comparison of different positions and filter sets.

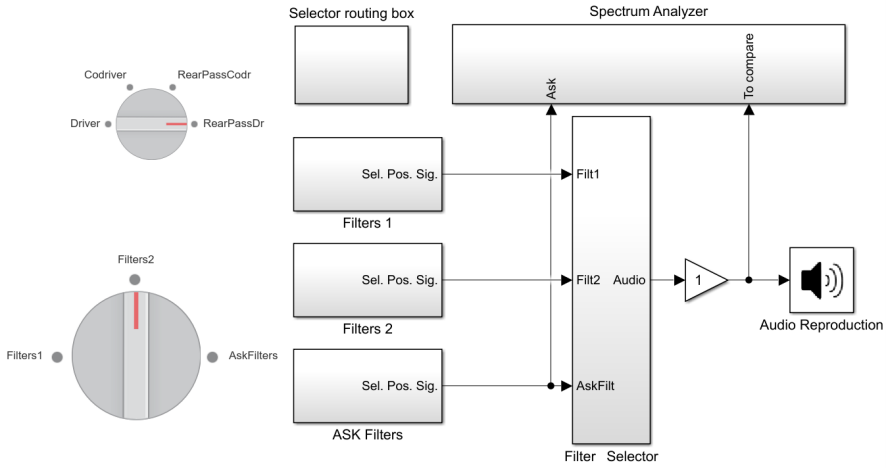


Figure 2.5: Block diagram of the Simulink model used to load, select and reproduce the audio perceived at the ears of the different passengers of the vehicle (driver, codriver, or rear passengers) with various PSZ filter sets.

A *Spectrum Analyzer* block is added to analyze the power spectrum of the reproduced audio and compare it with the power spectrum of the audio signal obtained with the filters designed by ASK. Indeed, the filters designed by ASK are chosen as a reference for performance evaluation.

2.2.2 Matlab script

The script mainly requires the IRs measured in the vehicle, the filters designed for the acoustic region control, two (stereo) audio traces and the position where they must be reproduced. For example, we can choose to beam the first audio trace to the driver and the second one to the other positions (or vice-versa).

Even if the processing does not have time limitations since it is performed offline before the reproduction, filtering by the convolution operator would require a considerable amount of computation time. For this reason, the filtering operation is implemented in the frequency domain. Furthermore, only the first 80-120 s of the audio traces are processed. This length is chosen to have enough time for a preliminary listening evaluation.

With the help of the flow chart in Fig. 2.6, we can describe the operations performed by the script as follows.

The measured IRs, the designed PSZ filters and the (stereo) audio traces are loaded. The main parameters at this step are the sampling frequency $F_s =$ 48 kHz, the number of samples after which the measured IRs are truncated $L_{IRs} = 24000$ samples, and the length of the filters $L_{filt} = 8192$ samples.

The short-time Fourier transform (STFT) of the audio traces are calculated separating the left and right channels. The length of the time frames (expressed in samples) is defined as

$$L_{frames} = N_{FFT} - L_{filt} - L_{IRs} \tag{2.1}$$

where $N_{FFT} = 48000$ points is the number of points of the fast Fourier transform (FFT) operation. The results are 4 matrices $\mathbf{A}_{1left}$, $\mathbf{A}_{1right}$, $\mathbf{A}_{2left}$ and

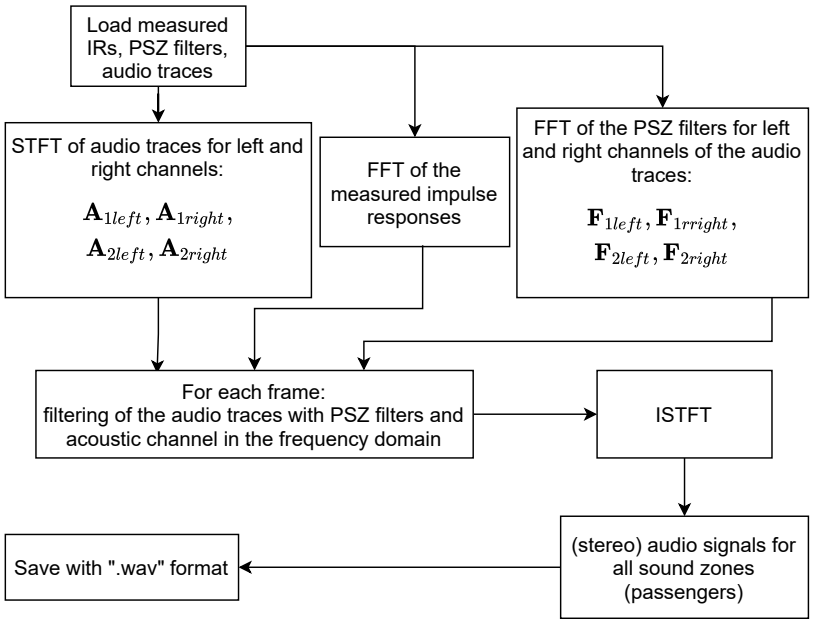


Figure 2.6: Flow chart of the operation performed by the Matlab script to calculate the audio signals to be reproduced by the Simulink model.

$\mathbf{A}_{2right}$, each of them with size $N_{frames} \times N_{FFT}$, where the subscripts indicate the number of the audio trace and if it is related to the left or the right channel, respectively.

Given the number of loudspeakers L, let $\mathbf{Z}_{i,left}$ and $\mathbf{Z}_{i,right}$ be the $L \times N_{FFT}$ matrices populated by the FFT of the measured IRs between the loudspeakers and the left and right control points, respectively, in the i-th acoustic region. Given the number of considered acoustic zones N_{zones}, for all $i = 1, 2, \ldots, N_{zones}$, these matrices are concatenated as follows

$$\mathbf{H} = \begin{bmatrix} \mathbf{Z}_{1,left} \\ \mathbf{Z}_{1,right} \\ \vdots \\ \mathbf{Z}_{N_{zones},left} \\ \mathbf{Z}_{N_{zones},right} \end{bmatrix} \tag{2.2}$$

forming an overall $2N_{zones}L \times N_{FFT}$ matrix. Note that here we consider the measurements performed only with the DH at a fixed position.

Let $\mathbf{Q}_{i,left}$ and $\mathbf{Q}_{i,right}$ be the $L \times N_{FFT}$ matrices populated by the FFT of PSZ filters designed assuming the left control point and the right control point as bright zone (BZ), respectively, and all other control points as dark zone (DZ). To reproduce the same audio trace in different acoustic regions, we have to apply the superposition principle to the different sets of filters designed for the different acoustic regions. Hence, considering two different audio traces, we can define two subsets of the acoustic regions I_1 and I_2 where we desire to listen to the first and the second audio traces, respectively. Therefore, we can define the overall $L \times N_{FFT}$ matrices

$$\begin{cases} \mathbf{F}_{1left} = \displaystyle\sum_{i \in I_1} \mathbf{Q}_{i,left} & \mathbf{F}_{1right} = \displaystyle\sum_{i \in I_1} \mathbf{Q}_{i,right} \\ \mathbf{F}_{2left} = \displaystyle\sum_{i \in I_2} \mathbf{Q}_{i,left} & \mathbf{F}_{2right} = \displaystyle\sum_{i \in I_2} \mathbf{Q}_{i,right}. \end{cases} \tag{2.3}$$

As it was done for the acoustic channel matrix, it is convenient to group

all the matrices (2.3) into a unique $4L \times N_{FFT}$ matrix as

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{1left} \\ \mathbf{F}_{1right} \\ \mathbf{F}_{2left} \\ \mathbf{F}_{2right} \end{bmatrix}. \tag{2.4}$$

At this point, we have to filter each frame of the audio traces (rows of the matrices $\mathbf{A}_{1left}$, $\mathbf{A}_{1right}$, $\mathbf{A}_{2left}$ and $\mathbf{A}_{2right}$) with the PSZ filters. To avoid the use of a *for* cycle to perform the filtering of the 4 input audio signals with the $4L$ filters, it is convenient to define the matrices $\mathbf{B}^g_{1left}$, $\mathbf{B}^g_{1right}$, $\mathbf{B}^g_{2left}$ and $\mathbf{B}^g_{2right}$ where all rows are copies of the g-th row of the matrices $\mathbf{A}_{1left}$, $\mathbf{A}_{1right}$, $\mathbf{A}_{2left}$ and $\mathbf{A}_{2right}$, respectively. Therefore, by concatenating them as

$$\mathbf{B}^g = \begin{bmatrix} \mathbf{B}^g_{1left} \\ \mathbf{B}^g_{1right} \\ \mathbf{B}^g_{2left} \\ \mathbf{B}^g_{2right} \end{bmatrix} \tag{2.5}$$

we can express the input of the loudspeakers as

$$\mathbf{L}_{in} = \sum_{r=0}^{3} \left[\mathbf{F} \odot \mathbf{B}^g\right]_{rL+1}^{(r+1)L} \tag{2.6}$$

where $\odot$ denotes the element-wise product, $[\cdot]_i^j$ indicates a matrix size reduction by selecting from the i-th to the j-th row of a matrix, and $\mathbf{L}_{in}$ is an $L \times N_{FFT}$ matrix where the ℓ-th row is the input signal to the ℓ-th loudspeaker.

Now, defining the $2N_{zones}L \times N_{FFT}$ matrix $\mathbf{L}^{2N_{zones}}$ as a vertical concatenation of $2N_{zones}$ copies of the matrix $\mathbf{L}$, we can express, for each frame, the acoustic signals for the left, $\mathbf{P}_{i,left}$, and the right, $\mathbf{P}_{i,right}$, control points of the i-th acoustic region as

$$\begin{cases} \mathbf{P}_{i,left} = \sum\limits_{rows} \left[\mathbf{L}^{2N_{zones}} \odot \mathbf{H}\right]_{(i-1)L+1}^{iL} \\ \mathbf{P}_{i,right} = \sum\limits_{rows} \left[\mathbf{L}^{2N_{zones}} \odot \mathbf{H}\right]_{iL+1}^{(i+1)L} \end{cases}. \tag{2.7}$$

Once we have collected $\mathbf{P}_{i,left}$ and $\mathbf{P}_{i,right}$ ($1 \times N_{FFT}$ matrices) for all frames of the input audio traces, an inverse short-time Fourier transform (ISTFT) [55] is performed to return to the time domain. The left and the right signals for each acoustic zone are then grouped to be saved in ".wav" format and ready to be reproduced by the Simulink model.

2.3 Implementation for STOI evaluation of a PSZ system

For the evaluation of the intelligibility, various audio contents were considered. As reference clean speeches, 15 tracks with 10 sentences each were used. Each sentence was taken from [56]. Note that, the same tracks were used in [47] for STOI evaluation, which motivated us to use them as well.

First of all, the selected speech track is filtered by the designed PSZ filters for the target sound region, without considering other audio contents. This means that we reproduce the speech in the target sound region in the absence of interfering signals reproduced for another sound region. Then, the obtained signal is filtered by the IRs of the considered channel. The signal at the output of this last filtering operation is considered as the clean speech signal, namely x in the previous section. The possible degradation introduced by these filtering operations is not taken into account in the intelligibility measure since these filtering operations are required to avoid underestimation of the STOI due to the delays introduced by the PSZ filters and the channel. This means that the degradation introduced to the clean signal x only consists of the signal leaking from the other sound regions, where different audio content is reproduced.

Note that, before filtering by the PSZ filters, the signals are normalized to have the same peak level.

Four possible audio signals, mainly with different spectral behavior with respect to the speech tracks, were considered as interference:

1. a rock/pop music, i.e., Draft Punk - "Get lucky"
2. a classic music, i.e., Vivaldi - "The winter"

3. house music, i.e., DJ Funk - "Scrub the ground"
4. a generic male speech.

One of the 15 tracks and one of the interfering audio are taken and processed according to the method described previously in this section, generating the audio signals for the driver and the codriver. Hence, we have one of the speech tracks degraded by the other audio in one sound region, named y_T, and one of the audio signal tracks degraded by the speech track in the other sound region, named y_A. Note that, the subscripts T in y_T and A in y_A stand for *track*, to indicate that the main contribution of the signal is given by the considered speech track, and *audio*, to indicate that the main contribution of the signal is given by one of the listed audio tracks, respectively.

Note that, the processing to generate the audio files described in the previous section is performed with a sample rate of 48 kHz, despite the algorithm developed in [47] was evaluated considering a sample rate of 10 kHz. However, the Matlab script made available by the same authors already implements a down sampling to overcome this problem. Hence, our signals are downsampled to 10 kHz in the following analysis.

Then, the procedure is repeated for each of the 15 speech tracks and, finally, an arithmetic mean of the obtained 15 intelligibility values is performed.

The evaluation of (1.35) between x and y_T leads to the prediction of the intelligibility of the speech track with respect to the interfering audio content - the higher the intelligibility value, the better. However, for the performance analysis of the PSZ system, the evaluation of (1.35) performed between x and y_A is also interesting. Indeed, this second evaluation still leads to the prediction of an intelligibility value, but is related to a quantity that represents the privacy level achieved by the system, which means that a lower value is better. With sound zone privacy, we mean the state of being not able to understand completely or in part the words contained in an audio content reproduced for a sound zone in which we are not located. A related work on the reduction of speech intelligibility, i.e., improving privacy, can be found in [57]. For this reason, in the next section, the results of both evaluations will

be presented and discussed. Moreover, the intelligibility will be analyzed as a function of a gain applied to the speech track after the normalization to the peak value.

The intelligibility and privacy, expressed in percentage of words understood, versus a suitable power gain, applied only to the speech track, achieved in the considered sound regions and for various interfering audio contents according to the procedure described above, is presented in Fig. 2.7.

In each of the subfigures in Fig. 2.7, we have 4 curves:

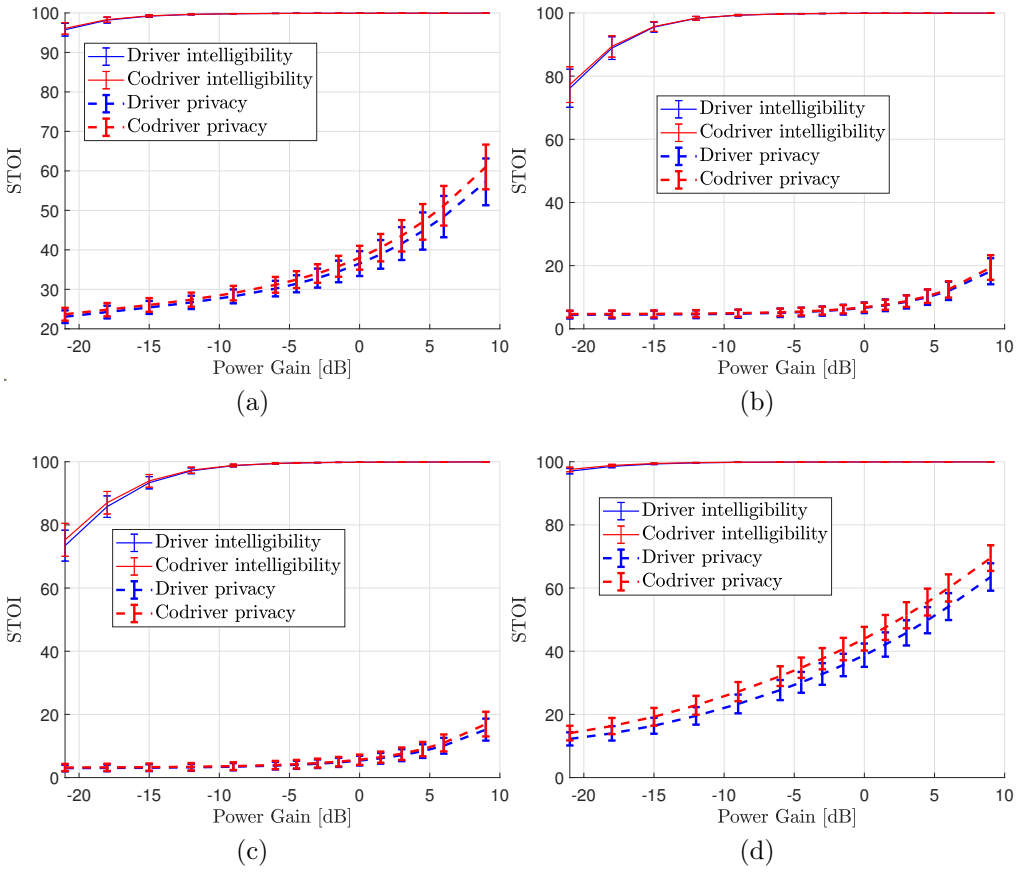


Figure 2.7: Prediction of intelligibility and privacy versus power gain applied to the speech track with second audio content: (a) generic male speech, (b) rock/pop music, (c) house music, (d) classic music.

- *Driver intelligibility* - speech track reproduced at the driver position which leaks towards the codriver, audio content at the codriver position which leaks towards the driver position, and STOI evaluated in the driver position
- *Codriver intelligibility* - speech track reproduced at the codriver position which leaks towards the driver position, audio content at the driver position which leaks towards the codriver position, and STOI evaluated in the codriver position
- *Driver privacy* - speech track reproduced at the driver position which leaks towards the codriver position, audio content at the codriver position which leaks towards the driver position, and STOI evaluated in the codriver position
- *Codriver privacy* - speech track reproduced at the codriver position which leaks towards the driver position, audio content at the driver position which leaks towards the codriver position, and STOI evaluated in the driver position.

The upper curves are predictions of intelligibility, i.e., the higher the better, and the bottom curves are predictions of privacy, i.e., the lower the better.

A first observation is that both intelligibility and privacy do not increase linearly, despite the gain applied to the speech track increases linearly. This may be justified by the fact that the STOI in (1.36) is mapped into intelligibility by the non linear function (1.37).

Another observation is the lower intelligibility achieved with interfering rock/pop and house music with respect to generic male speech or classic music. This behavior seems reasonable since interfering rock/pop and house music have much more energy contributions in the same frequency range of the speech track than interfering generic male speech or classic music, and this decreases the value of STOI. About the interfering generic male speech, we also have to consider that in this type of audio content, there are silent parts of the signal contributing to increase the intelligibility of the speech track of interest. The same reason justifies the lower values of privacy when rock/pop

and house music is of interest with respect to the other two audio genres.

We can also observe that the perceptual levels, both intelligibility and privacy, of the codriver are consistently slightly better than those of the driver. This is due to the fact that the audio channels for the 2 positions in the car are not symmetrical, hence, the relative performance is not exactly the same.

An interesting case use of the presented curves is for the straightforward choice of the power gain tradeoff value to be applied to the speech signal. Indeed, if we want a minimum intelligibility value, we can see directly the corresponding value of the achieved privacy, and vice-versa.

For comparison, the curves of the predicted intelligibility and privacy with the PSZ system not active are plotted in Fig. 2.8. We can see that intelligibility and privacy curves overlap, which means that the speech is perceived in the same way in both the sound regions and there is no separation between them.

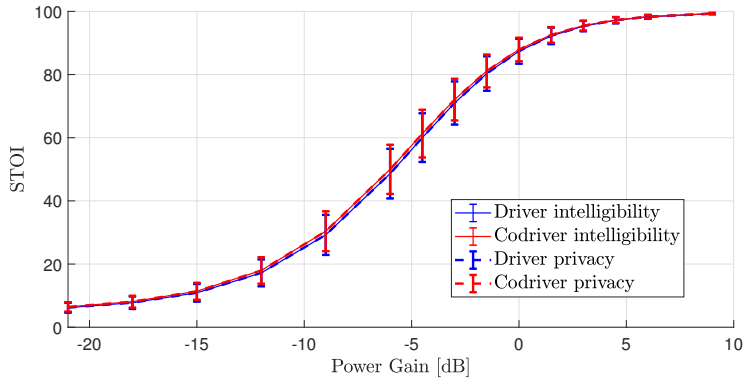


Figure 2.8: Prediction of intelligibility and privacy versus power gain applied to the speech track with rock/pop music with the PSZ system not active.

Chapter 3

Pre-processing of measurements for the performance enhancement of PSZ filters

In this chapter, following the terminology adopted in [15], we refer to the operation of properly truncating an impulse response (IR) as "trimming". The designed filters may exhibit "pre-ringing", namely a leading component in their IRs necessary to control the late reflections present in the measured IRs. To mitigate this pre-ringing, a frequency-dependent trimming (FDT) was proposed in [15]. With the aim of robustness improvement, an algorithm based on an exhaustive search of the trimming lengths in comparison with the empirical solution adopted in [15] is proposed and evaluated by means of numerical simulations and measurements performed in a vehicle equipped for personal sound zone (PSZ) system tests.

Furthermore, three additional methods for the selection of the trimming lengths are investigated, with a direct focus on the mitigation of the early and late energy in the measured IRs of the loudspeaker-microphone pairs. In

the literature, this operation is also called room equalization, dereverberation, compensation, correction or other. In [58], digital signal processing techniques that aim at improving sound reproduction, e.g., least-squares optimization, frequency domain deconvolution, etc., were reviewed. Our target is similar but not the same because we wish to improve the performance of the PSZ system, and we try to achieve this goal in a simpler manner.

The first investigated method is based on the assumption that the first peak of the magnitude of the short-time Fourier transform (STFT) of a measured IR evaluated at a fixed frequency is associated, for that frequency, with the arrival of the direct sound component. Hence, by windowing around that point and extracting the first lobe or a portion of it, we should be able to remove early and late reflections.

For the second method, we assume that we can extract the direct sound component by comparing two IRs obtained by repeated measurements performed at about the same position, i.e., by evaluating their crosscorrelation, in which mainly the early and late components differ.

In the last method, assuming that the energy vanishes earlier as the considered frequency increases [59], we perform windowing with lengths inversely proportional to the considered frequency and scaled by a proper scalar, to be set by trial and error.

This chapter is organized as follows. In Section 3.1, a literature review of the quadrature mirror filters (QMFs) [60] is accomplished and a filterbank for fractional-octave bands filtering is defined. The effect of trimming in the time and frequency domain, and possible alternative solutions are discussed in Section 3.2. In Section 3.3, some windowing functions are reviewed. In Section 3.4, the four proposed algorithms for trimming length design are presented. Finally, in Section 3.5, the results obtained by the numerical simulations and the measurements are reported and discussed. Conclusions are drawn in Section 3.6.

3.1 Octave-band filterbank design

3.1.1 Quadrature mirror filters

QMFs have been extensively used for splitting a signal into two or more subbands in the frequency domain so that each subband signal can be processed independently [60]. These filters take this name from the symmetry of their modulus around the normalized angular frequency $\pi/2$.

Considering the scheme in Fig. 3.1(a), where we wish to divide a discrete signal $x[n]$ into 2 subbands with equal bandwidth, the prototypes of the magnitude frequency responses (MFRs) of the QMFs are shown in Fig. 3.1(b). The modulus of the two filters are complementary to one, i.e., they are related as

$$|H_0(\omega)| + |H_1(\omega)| = 1 \tag{3.1}$$

where ω is the normalized angular frequency.

If we assume that $H_0(\omega)$ and $H_1(\omega)$ are both finite impulse response filters with linear phase $e^{-\jmath\omega T_0}$, where T_0 is the delay introduced by the filter, the output of the system in Fig. 3.1(a) can be written as

$$X'(\omega) = X(\omega)\left[H_0(\omega) + H_1(\omega)\right] = X(\omega)e^{-\jmath\omega T_0} \tag{3.2}$$

which corresponds to a delayed version of the input $X(\omega)$.

An example of a QMF prototype is the raised cosine (RC) function with modulus defined as

$$|H_0(\omega)| = \begin{cases} 1 & |\omega| \leqslant \dfrac{\pi}{2}(1-\alpha) \\ 0.5 + 0.5\cos\left[\dfrac{1}{\alpha}\left(|\omega| - \dfrac{\pi}{2}(1-\alpha)\right)\right] & \dfrac{\pi}{2}(1-\alpha) < |\omega| \leqslant \dfrac{\pi}{2}(1+\alpha) \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

where α is the roll-off factor. Examples of $|H_0(\omega)|$ and its complementary $|H_1(\omega)|$ for some values of the roll-off parameter are shown in Fig. 3.2.

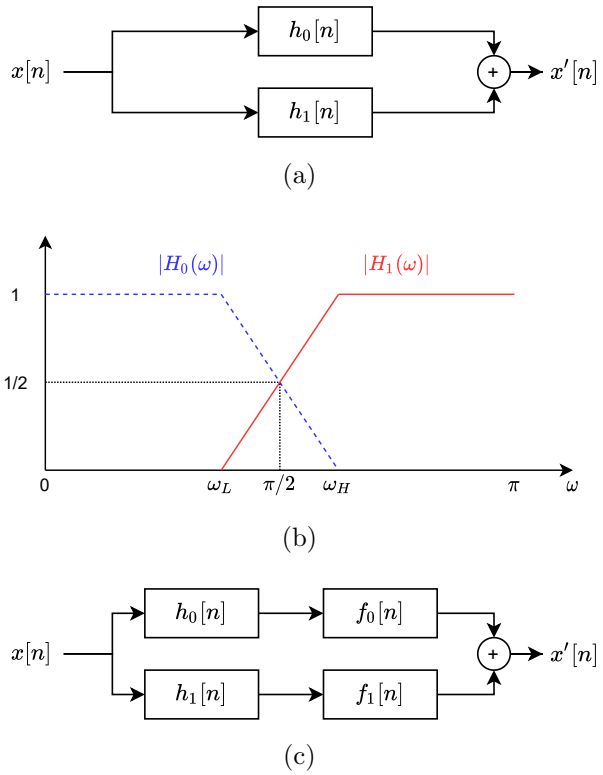


Figure 3.1: (a) Scheme of 2-subband division and reconstruction. (b) Prototype of MFRs of QMFs. (c) Scheme for subband signal processing. Note that, $\omega_L = \pi/2 - \Delta\omega$ and $\omega_H = \pi/2 + \Delta\omega$ since the filters are symmetric with respect to $\pi/2$.

Considering the subband signal processing system in Fig. 3.1(c), the output can be written as

$$\begin{aligned} X'(\omega) &= X(\omega)H_0(\omega)F_0(\omega) + X(\omega)H_1(\omega)F_1(\omega) \\ &= \begin{cases} X(\omega)H_0(\omega)F_0(\omega) & |\omega| \leqslant \omega_L \\ X(\omega)H_1(\omega)F_1(\omega) & |\omega| \geqslant \omega_H \\ X(\omega)H_0(\omega)F_0(\omega) + X(\omega)H_1(\omega)F_1(\omega) & \omega_L < |\omega| < \omega_H \end{cases} \end{aligned} \tag{3.4}$$

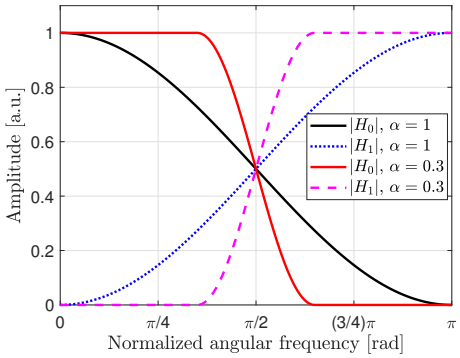


Figure 3.2: Amplitude of the QMFs obtained with the RC function for some values of the roll-off factor α and 2-channel splitting.

where ω_L and ω_H are the lower and upper normalized angular frequency of the transition band of the prototype filter $H_0(\omega)$, $H_1(\omega)$, $F_0(\omega)$ and $F_1(\omega)$.

The most critical part of (3.4) is in the range (ω_L, ω_H) where there is interference between the two subband signals. This interference can be reduced by shortening the transition bandwidth of the prototype filter $H_0(\omega)$. Considering the RC function (3.3), the transition bandwidth is controlled by the roll-off factor α.

3.1.2 Variable filter bandwidth

For audio applications, we are interested in performing signal processing over a fraction of the octave bands. For this purpose, it may be useful to define the following function

$$T_{tr}(\omega, \alpha, \omega_L, \omega_{tr}) = \begin{cases} 1 & \text{if } \omega \in \Omega_{low} \\ 0.5 + 0.5\cos\left[\dfrac{1}{\alpha}\left(|\omega - \omega_L| - \dfrac{\omega_{tr}}{2}(1-\alpha)\right)\right] & \text{if } \omega \in \Omega_{tr} \\ 0 & \text{if } \omega \in \Omega_{high} \end{cases} \tag{3.5}$$

where

$$\begin{cases} \Omega_{low} = \left\{\omega : \omega_L \leqslant |\omega| \leqslant \omega_L + \dfrac{\omega_{tr}}{2}(1-\alpha)\right\} \\ \Omega_{tr} = \left\{\omega : \omega_L + \dfrac{\omega_{tr}}{2}(1-\alpha) < |\omega| \leqslant \omega_L + \dfrac{\omega_{tr}}{2}(1+\alpha)\right\} \\ \Omega_{high} = \left\{\omega : \omega_L + \dfrac{\omega_{tr}}{2}(1+\alpha) < |\omega| \leqslant \omega_L + \omega_{tr}\right\} \end{cases} \tag{3.6}$$

the subscript "tr" stands for *transition* and $\omega_{tr} = \omega_H - \omega_L$ is the transition bandwidth given the upper and lower normalized angular frequencies of the transition band. The function (3.5) describes an amplitude transition from 1 to 0 (or vice-versa) in a general interval $[\omega_L, \omega_H]$. We can note that (3.5) becomes (3.3) with $\omega_L = 0$ and $\omega_H = \pi$.

Now we have to define the filter bank $\{H_0(\omega), \ldots, H_i(\omega), \ldots, H_{N_{obs}-1}(\omega)\}$ where $H_i(\omega)$ is the filter for the i-th fractional octave band and N_{obs} is the number of the considered fractional octave bands. The modulus of the filter corresponding to the first octave band can be expressed as

$$|H_0(\omega)| = \begin{cases} 1 & \text{if } |\omega| < \omega_L^0 \\ T_{tr}(\omega, \alpha, \omega_L^0, \omega_{tr}^0) & \text{if } \omega_L^0 \leqslant |\omega| \leqslant \omega_H^0 \\ 0 & \text{otherwise} \end{cases} \tag{3.7}$$

where ω_L^0 and ω_H^0 are the lower and upper normalized angular frequencies of the transition band for the first octave band. Likewise, we can define the filter corresponding to the i-th octave band as

$$|H_i(\omega)| = \begin{cases} 1 - |H_{i-1}(\omega)|, & \text{if } \omega_L^{i-1} \leqslant |\omega| \leqslant \omega_H^{i-1} \\ 1 & \text{if } \omega_H^{i-1} < |\omega| < \omega_L^i \\ T_{tr}(\omega, \alpha, \omega_L^i, \omega_{tr}^i) & \text{if } \omega_L^i \leqslant |\omega| \leqslant \omega_H^i \\ 0 & \text{otherwise} \end{cases} \tag{3.8}$$

where ω_L^i and ω_H^i are the lower and upper normalized angular frequencies of the transition band for the i-th octave band, and ω_L^{i-1} and ω_H^{i-1} are the

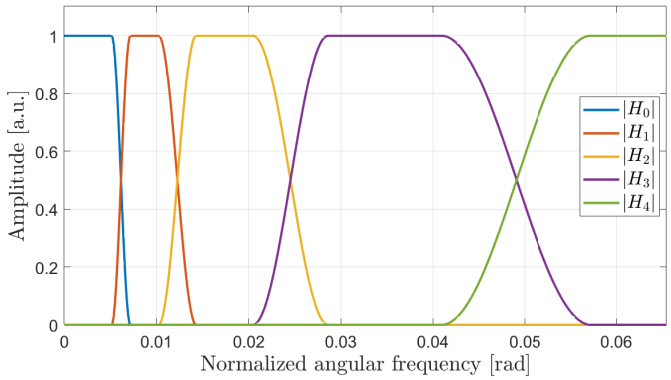


Figure 3.3: MFRs of the first 5 filters of the filterbank designed with (3.7) and (3.8) for 1-octave band division. The parameter α is set to 0.5.

lower and upper normalized angular frequencies of the transition band for the $(i-1)$-th octave band. Note that (3.7) is extended from the first considered octave band to include the frequency zero for simplicity, i.e., $\omega_L^0 = 0$, and it is a low pass filter, whereas (3.8) corresponds to a band pass filter. Furthermore, we consider $\omega_H^0 = 2\pi(F_c^0/F_s)$, $\omega_L^{i-1} = 2\pi(F_c^{i-1}/F_s)$, $\omega_H^{i-1} = \omega_L^i = 2\pi(F_c^i/F_s)$, $\omega_H^i = 2\pi(F_c^{i+1}/F_s)$, where F_c^i is the central frequency of the i-th octave band (with $i = 0, 1, \ldots, N-1$) and F_s is the sampling frequency, both expressed in hertz.

The MFRs of the first five filters of a 1-octave band filterbank defined by (3.7) and (3.8) are shown in Fig. 3.3.

Similarly to (3.2), assuming that the filters of the filterbank $\{H_0(\omega), \ldots, H_{N_{obs}-1}(\omega)\}$ have the same linear phase $e^{-\jmath\omega T_0}$, for construction, the sum of an input signal, band-limited to the frequency $\omega_L^{N_{obs}-1}$, filtered by the filterbank is a delayed version of itself.

3.2 Frequency-dependent trimming

In enclosed spaces, such as the cabin of a vehicle, the measured IRs may be very long due to late reflections. The computation of the filters aimed at

sound region control may produce filters with significant pre-ringing, which is required to control the late reflections of the measured IRs. This component in the IR of the filters may be perceived as unnatural and disturbing in the produced sound [61]. Furthermore, even if these filters allow to control very well the exact position where the loudspeaker-microphone IRs are measured, the robustness to system variations, such as the position, may be poor.

In the time domain, the trimming operation proposed in [15] is a multiplication of an input signal $x[n]$ by a window function $w[n]$, defined as

$$x_w[n] = x[n]w[n] \tag{3.9}$$

where $x_w[n]$ is the windowed version of the input signal. Given a time interval $[n_l, n_h]$, with $n_h > n_l$, a window function can be any function equal to zero outside and greater than zero inside the interval. An example of a windowing function is the rectangular function, that is equal to a constant inside the windowing interval and zero outside. Other examples of windowing functions are Hanning, Hamming or Blackman functions [62].

It is well-known that the multiplication in the time domain corresponds to the convolution in the frequency domain (and vice-versa) [63]. Indeed, we have that

$$x_w[n] = x[n]w[n] \iff X_w(\omega) = \frac{1}{2\pi} X(\omega) \otimes W(\omega) \tag{3.10}$$

where $X(\omega)$ and $W(\omega)$ are the continuous-frequency discrete-time Fourier transforms (DTFT) of the signals $x[n]$ and $w[n]$, respectively, ω denotes the normalized angular frequency and $\otimes$ indicates the circular convolution operator.

Let us consider as input signal a discrete-time cosine sequence at frequency F_c defined as

$$x[n] = \cos[\omega_c n] \tag{3.11}$$

where $\omega_c = 2\pi(F_c/F_s)$ is the normalized angular frequency of the signal and

F_s is the sampling frequency. The DTFT of (3.11) is [63]

$$X(\omega) = \sum_{k=-\infty}^{\infty} \pi \left[\delta(\omega - \omega_c + 2\pi k) + \delta(\omega + \omega_c + 2\pi k)\right]. \tag{3.12}$$

Therefore, for any window function $W(\omega)$, assuming negligible the possible spectral aliasing introduced in case of $W(\omega)$ non band-limited, the right side of (3.10) in the interval $-\pi \leq \omega < \pi$ can be written as

$$X_w(\omega) = \frac{1}{2}\left[W(\omega - \omega_c) + W(\omega + \omega_c)\right]. \tag{3.13}$$

This means that starting from a signal that has the energy concentrated at the frequencies $\pm\omega_c$, the windowing causes an energy leakage [64] in the frequency domain (or spectral leakage), i.e., the energy is spread onto the neighboring frequencies. Furthermore, as the window length decreases, much more energy spreads out and the energy centered at $\pm\omega_c$ decreases. This phenomenon can be observed in Fig. 3.4 where the discrete-time cosine sequence is windowed with a rectangular function characterized by three different window lengths.

In [15], the authors proposed FDT as a solution to improve the performance of the acoustic contrast control (ACC) algorithm. The critical aspect of this solution is the choice of the window lengths for each subband and the subbands

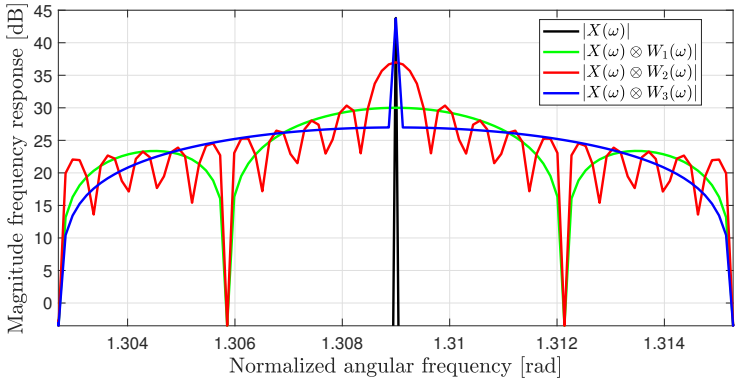


Figure 3.4: MFRs of a cosine sequence with $F_c = 10$ kHz, $F_s = 48$ kHz and its windowed version with a rectangular function of lengths $L_1 = 2000$ (green), $L_2 = 10000$ (red) and $L_3 = 47000$ samples (blue).

themselves. They opted for empirical window lengths applied on 1/3-octave bands, which is a common practice for acoustic, electro-acoustic and audio systems [65]. With these choices, they were able to achieve a higher acoustic contrast (AC), in particular at the higher frequencies, reduce the pre-ringing in the designed filters and increase the robustness of the system with respect to the case without FDT. However, this empirical solution may not be optimal or suitable for any scenario and configuration.

According to the previous considerations, the trimming operation causes an energy leakage in the frequency domain. In case of FDT, the energy leakage is different if we perform a subband filtering before or after the windowing operation, as shown in the block diagrams in Fig. 3.5. However, since the goal of the FDT is the control of information (reduction or cancellation of the late reflections), the two methods produce almost the same effect on the processed IR, as shown in Fig. 3.6 where a loudspeaker-microphone pair IR measured in a car cabin is frequency-dependent trimmed according to the two methods in Fig. 3.5. Furthermore, the energy leakage in the frequency domain can be minimized by using a proper windowing function such as Hanning, Hamming or Blackman functions [62].

If we consider the effect of the FDT in the frequency domain, we can note, both in Figs. 3.4 and 3.7, that the windowing performs a smoothing of the MFR. In Fig. 3.7, a FDT is applied according to the block diagram in Fig. 3.5(a) (windowing followed by filtering).

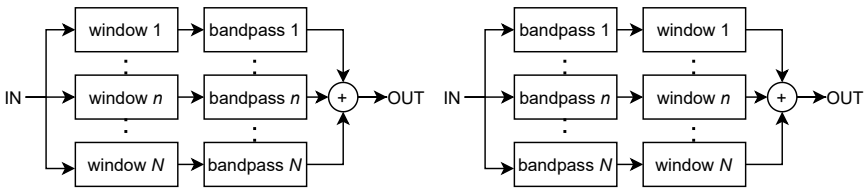


Figure 3.5: Block diagrams of the FDT for subbands filtering after (a) and before (b) the windowing operation.

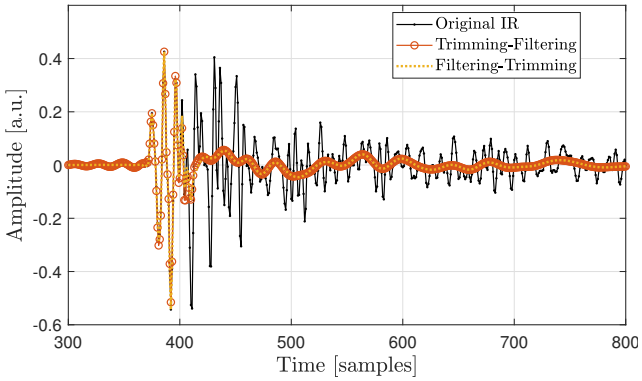


Figure 3.6: Microphone-loudspeaker pair IR measured in a vehicle cabin, and its frequency-dependent trimmed versions with windowing followed by filtering (*Trimming-Filtering*) and filtering followed by windowing (*Filtering-Trimming*).

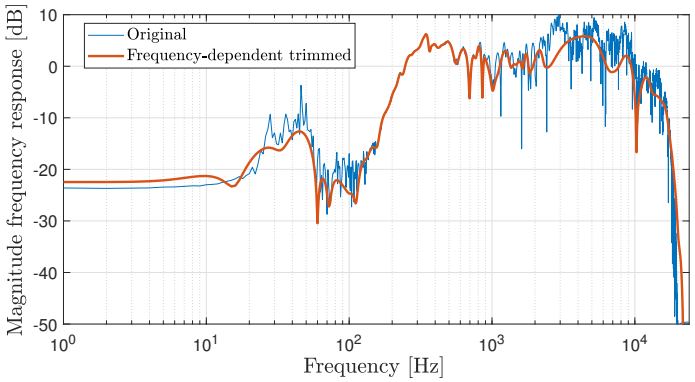


Figure 3.7: MFR measured in a vehicle cabin and its frequency-dependent trimmed version performed according to the scheme in Fig. 3.5(a) (windowing followed by filtering).

The smoothing (or averaging) of the MFR is in agreement with the purpose of the FDT. Firstly, the approximation of the MFR shape should increase the robustness of the solution by allowing the calculation of filters that do not control perfectly the measurement points but perform better (on average) in the surrounding regions. Second, the smoothing of the peaks and notches of the higher frequencies reduces the presence of peaks and notches in the MFR (at the higher frequencies) of the designed filters. Indeed, the peaks at

high frequencies may be perceived as disturbing in the listening and a high variability of peaks and notches in the frequency response (FR) of the designed filters may corresponds to a long pre-ringing in their IRs.

Based on the observed effect of the FDT on the MFRs, an alternative solution by signal processing performed directly in the frequency domain may be the smoothing (or averaging) of the FR, or FR interpolation using variable frequency resolution. Furthermore, a variable frequency resolution to perform fast Fourier transform (FFT) and inverse fast Fourier transform (IFFT) should be investigated.

The mentioned alternatives can be operated on the measured acoustic FRs or directly on FRs of the designed filters. In any case, as for FDT, the optimal choice of some parameters, such as the windowing lengths, the windowing function, smoothing (or averaging) or interpolation, the considered subbands, etc., must be derived.

3.3 Windowing functions

Since a windowing function is required for the trimming operation, some windowing functions and their time-frequency characteristics are discussed in the following.

Assuming L is the length in samples of the window and denoting by n the sample index, we consider the Blackman window, defined as

$$w[n] = 0.42 - 0.5\cos\left[\frac{2\pi n}{L-1}\right] + 0.08\cos\left[\frac{4\pi n}{L-1}\right] \tag{3.14}$$

the Gaussian window, described by

$$w[n] = e^{-n^2/(2\sigma^2)} \tag{3.15}$$

where $\sigma = (L-1)/(2r)$ and r is a width factor and the Tukey window, defined

as

$$w[n] = \begin{cases} 0.5\left\{1 + \cos\left[\frac{2\pi(n-\alpha(L-1)/2)}{\alpha(L-1)}\right]\right\} & \text{if } 0 \leq n \leq \frac{\alpha L}{2} \\ 1 & \text{if } \frac{\alpha L}{2} < n < L-1-\frac{\alpha L}{2} \\ 0.5\left\{1 + \cos\left[\frac{2\pi(n-\alpha(L-1)/2-(L-1))}{\alpha(L-1)}\right]\right\} & \text{if } L-1-\frac{\alpha L}{2} \leq n \leq L-1 \end{cases} \tag{3.16}$$

where $0 \leq \alpha \leq 1$. We also consider the Chebyshev window as described in [66, 67].

In Figs. 3.8 and 3.9, the IRs and MFRs of some symmetric windowing functions with length $L = 127$ samples are plotted considering a sampling frequency of $F_s = 48$ kHz and using 1000 points for the discrete Fourier transform (DFT).

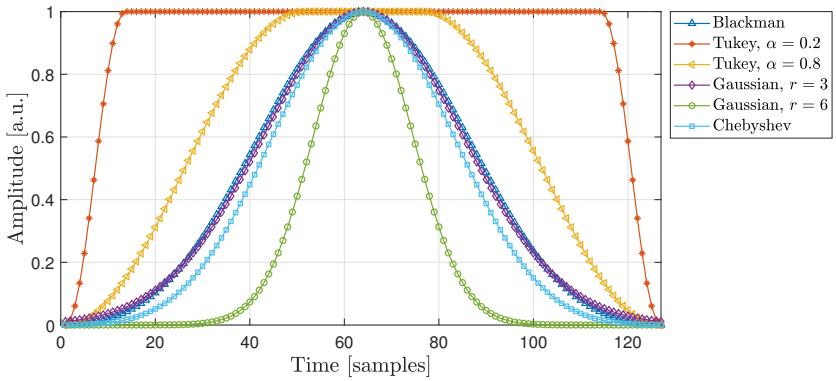


Figure 3.8: IRs of some symmetric windowing functions of length $L = 127$ samples.

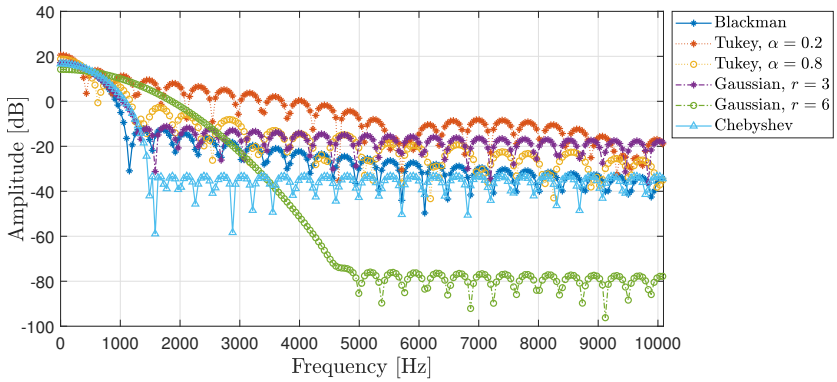


Figure 3.9: MFRs of the symmetric windowing functions presented in Fig. 3.8 obtained by DFT with a frequency resolution of 48 Hz.

If we wish to apply a window function to an IR of a loudspeaker-microphone pair $x[n]$, we have to multiply the IR by the windowing function centered around a fixed time instant (in samples). Possible time instants are the locations of the IR energy peak value N_{max} or the beginning of the IR defined as N_{in} such that $x^2[N_{in}] \geq \gamma x^2[N_{max}]$, where $0 < \gamma < 1$ is to be selected empirically. To clarify the difference between N_{in} and N_{max}, these points are marked in an example IR in Fig. 3.10.

Applying the windowing around N_{max}, to avoid mitigation of the loudspeaker-microphone pair IR component before N_{max}, it makes sense to use an asymmetric windowing function. In Figs. 3.11(a) and 3.11(b), the IRs and MFRs of asymmetric windowing functions are depicted. They are obtained by setting the first $\lceil L/2 \rceil - 1$ samples of the first, third, fourth and last windowing functions in Fig. 3.8 to 1.

Considering their time behavior, the symmetric windowing functions, in particular the narrowest ones such as the Blackman, Gaussian, Chebyshev, are more critical to be used. Indeed, these windows suppress most of the IR to which they are applied and there is the risk of losing a significant portion of this IR if the window is not centered in the correct position. In the frequency

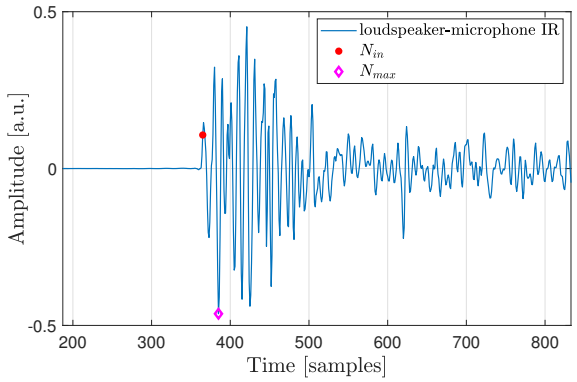
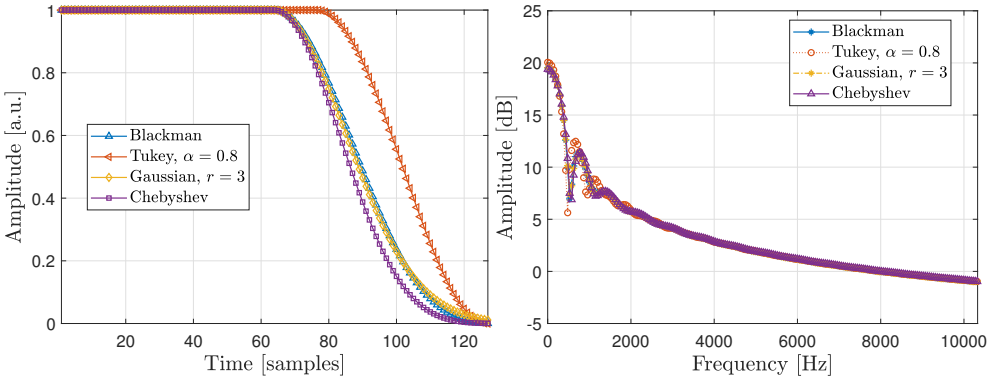


Figure 3.10: Location of the time instants N_{in} and N_{max} of a measured IR of a loudspeaker-microphone pair on which windowing is applied. To identify the beginning of the IR, $\gamma = 10^{-2}$ is considered.



(a) (b)

Figure 3.11: (a) IRs of some asymmetric windowing functions of length $L = 127$ samples and (b) their MFRs obtained by DFT with a frequency resolution of 48 Hz.

domain, these windowing functions perform a better averaging of the MFR of the input signal, reducing the peaks and notches. This capability is given by the width of the main lobe of the windowing function in the frequency domain, as well as the secondary lobes.

On the other hand, with the asymmetric windowing functions, the risk of losing a significant IR portion of the loudspeaker-microphone pair is reduced. However, the reduction capability of peaks and notches in the MFR is less than with the symmetric windowing functions.

According to the previous observations, the Tukey window with a low value of the parameter α (close to 0) is a good compromise between the reduction capability of the peaks and notches and the low risk of losing a significant portion of the loudspeaker-microphone IR. However, all windowing functions will be considered in the following evaluation of the AC.

3.4 Proposed trimming methods

For the results in this chapter, as sources, within configuration A, the 16 loudspeakers of the front arrays and the 7 factory-installed loudspeakers are

used to design the filters. Furthermore, the filters are designed considering the set of IRs with 5 control points for each ear, whereas the numerical evaluations are performed with another set of IRs measured only in the position $(0, h100)$ (see configuration A, Section 2.1) and the direct performance measurements are made by placing the dummy heads (DHs) to have the ear in the same position as $(0, h100)$. In the following sections of this chapter, a mismatched IR set, with respect to the reference IR set used for PSZ filter design, will be referred to as secondary IR set.

For simplicity, here, we label as *A1-A8* and *B1-B8* the loudspeakers of the array in front of the driver and the array in front of the codriver, respectively, as FWL and FWR the left and right front woofers, as $RWTL$ and $RWTR$ the left and right rear woofers-twitters, as FTL and FTR the left and right front twitters, and as SUB the subwoofer.

3.4.1 Optimal trimming lengths based on exhaustive search

As mentioned early, the main issue of the FDT [15] is the design of the windowing lengths for each subband and the width of the subbands themselves. Hereafter, an algorithm aimed at the design of the windowing lengths will be proposed with the help of the flow chart in Fig. 3.12. The algorithm is based on an exhaustive search over all possible windowing lengths to maximize the average AC evaluated in the neighboring control points with respect to a reference one.

Considering configuration A and assuming to design the PSZ filters for the driver and codriver, we start by partitioning all the measured IRs into sets: a first set composed of the IRs measured with the DH in the position $(60, h60)$, a second set composed by the IRs measured with the DH in the position $(30, h80)$, and so on. Among these IR sets, we select a reference one that will be used to calculate the PSZ filters, and one or more secondary sets that will be used to design the trimming lengths.

Once the reference IR set is chosen, we calculate the PSZ filters by applying a FDT with various lengths to the IRs of the reference set. Hence, we start from

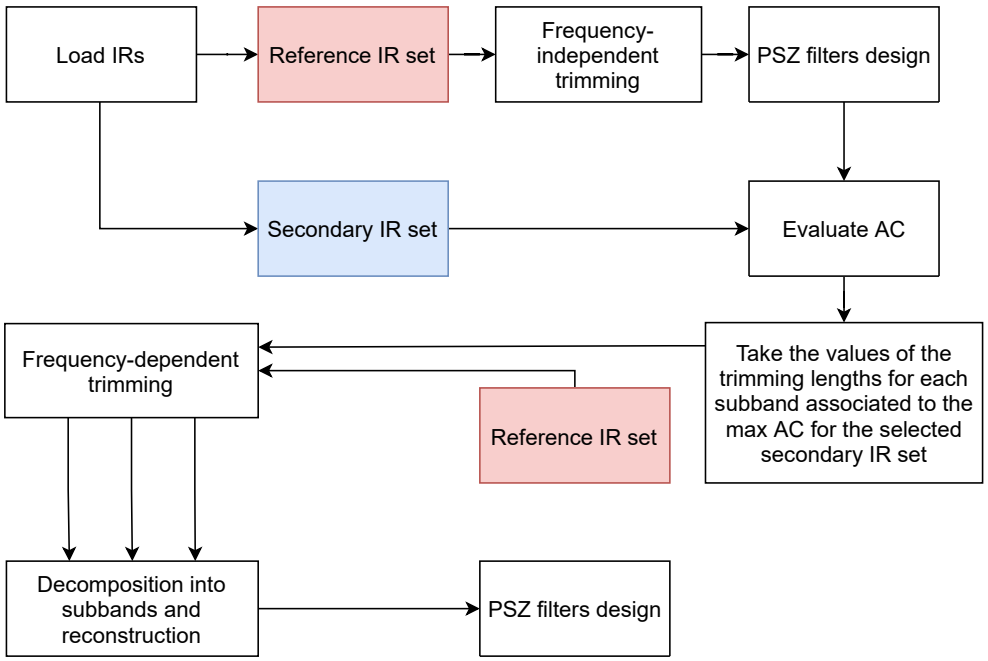


Figure 3.12: Flow chart of the proposed algorithm for the design of FDT lengths.

a windowing length and we calculate the PSZ filters applying that windowing length to all the reference IRs, we increase the windowing length and we calculate another set of PSZ filters using the updated windowing length, and so on. At the end, we collect various sets of PSZ filters obtained by applying FDT with different windowing lengths.

At this point, for each PSZ filter set we evaluate the (spatially averaged) AC that we achieve with that PSZ filter set by using the secondary IR set. In this way, we collect AC curves as a function of the frequency ω and the trimming length L_{trim} that can be denoted as $C(\omega, L_{trim})$. Since it is reasonable to consider 1/3-octave bands [65], the AC curves are averaged over the 1/3-octave bands and the result can be denoted as $\bar{C}(i, L_{trim})$ where i refers to the subband index.

At this point, for each subband, we choose the trimming length that maximizes the average AC in the subband. More formally, considering N_{obs} sub-

bands, for all $i = 0, ..., N-1$ we select

$$L_{trim}^{opt}(i) = \arg\max_{L_{trim}} \bar{C}(i, L_{trim}). \tag{3.17}$$

Once we have found $L_{trim}^{opt}(i)$ for all subbands, we consider again the IRs of the reference set, we perform a FDT and subband decomposition on each of the IR and we reconstruct the full-band IRs (according to the scheme in Fig. 3.5(a)). Finally, we calculate the PSZ filters using the frequency-dependent trimmed IRs.

3.4.2 Short-time Fourier transform-based trimming

This method is based on the time-frequency analysis of the signals obtained by the STFT of the measured loudspeaker-microphone IRs and the assumption that the first peak of the STFT magnitude for a fixed frequency corresponds to the arrival time of the direct sound component.

The flow chart in Fig. 3.13 describes this algorithm. Let $x[n]$ be a measured IR of a loudspeaker-microphone pair. The 1/3-octave band filtering is performed by the filterbank defined previously in this chapter, so that, given

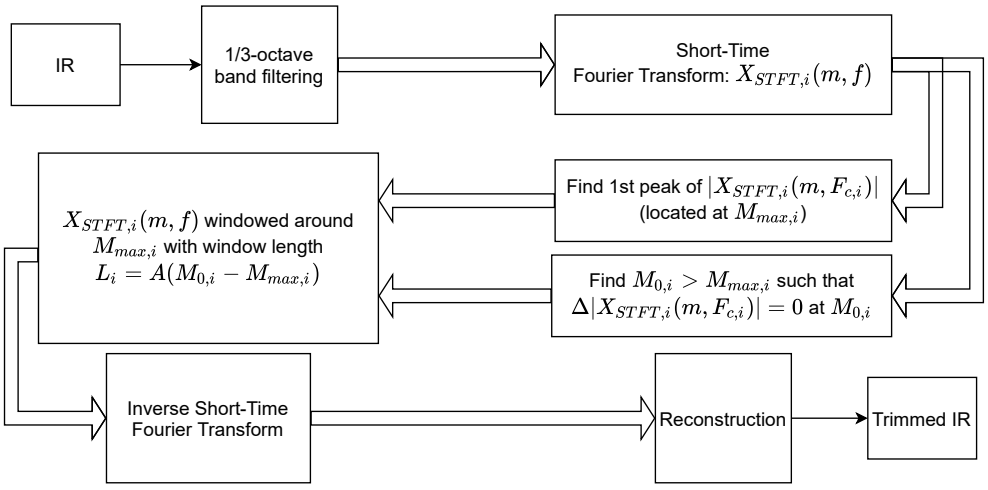


Figure 3.13: Flow chart of the STFT-based trimming algorithm.

the input $x[n]$, the outputs are $\{x_0[n], \ldots, x_i[n], \ldots, x_{N_{obs}-1}[n]\}$ where $x_i[n]$ is i-th subband component of $x[n]$, i.e., the signal filtered by the i-th 1/3-octave band filter, and N_{obs} is the number of considered octave bands. The first considered band has central frequency 79 Hz ($i = 0$) and is extended to the range from 0 to 85 Hz. The last one has central frequency of 16 kHz and is extended to the range from about 15 kHz to the maximum representable frequency of 24 kHz. Hence, the initial and terminal bands are not 1/3-octave ones. Instead, the internal 1/3-octave bands have center frequencies

$$F_{c,i} = \left\lfloor 10^3 2^{(i-11)/3} \right\rceil \text{ Hz} \tag{3.18}$$

where and $\lfloor \cdot \rceil$ is the rounding operator, and bandwidth

$$B_i = \left\lceil F_{c,i} 2^{1/6} - F_{c,i}/2^{1/6} \right\rceil \tag{3.19}$$

such that $F_{c,11} = 1$ kHz, where $\lceil \cdot \rceil$ indicates the ceiling operator and i ranges from 1 to 22.

Considering the i-th subband component, a STFT is performed giving to the output $X_{STFT_i}(m, f)$ where m is the index of the frame. The length of the frames is chosen to be proportional to the bandwidth B_i of the 1/3-octave band considered, with central frequency $F_{c,i}$, according to

$$L_{STFTframe,i} = 2^{\left\lceil \log_2 \frac{F_s}{B_i/5} \right\rceil} \tag{3.20}$$

where F_s is the sampling frequency. The frame lengths (3.20) allow to have at least 5 points in the bandwidth of interest, e.g., considering the 1/3-octave band with a central frequency of 1 kHz and bandwidth of about 232 Hz, the frame length is 2048 samples and allows a frequency resolution of about 23.4 Hz, giving 9 points in the bandwidth of interest. Furthermore, the frames are taken with an overlap of 95% of the frame samples and they are windowed with a Hamming window.

Given $X_{STFT,i}(m, f)$, we search for the location of the first peak of its magnitude at the fixed frequency $F_{c,i}$, i.e., $M_{max,i}$. Once the first peak is

found, we search for the position of the first local minimum $M_{0,i}$ after $M_{max,i}$ according to

$$M_{0,i} = \min \left\{ \underset{m>M_{max,i}}{\arg\min} \; | \, \Delta \, |X_{STFT,i}(m,f)| \, | \right\} \tag{3.21}$$

where $\Delta y[m] = y[m+1] - y[m]$ denotes the forward difference operator applied to a sequence $y[m]$. Given $M_{max}^{(k)}$ and $M_0^{(k)}$, we can finally define the length of the window as

$$L_i = 2 \left\lceil A \left(M_{0,i} - M_{max,i} \right) \right\rceil \tag{3.22}$$

where $A > 0$ is a suitable constant to be found by trial and error. In Figs. 3.14 and 3.15, two examples of the time relation between the STFT magnitude and the window applied to the STFT are reported.

The STFT magnitude in Fig. 3.15 can justify the choice of windowing around the first peak instead of the global maximum location. In fact, the first peak must correspond to the arrival of the direct sound component while the global maximum can occur later due to reflections that combine constructively.

Finally, each component $X_{STFT,i}(m,f)$ is windowed, and after taking the inverse short-time Fourier transform (ISTFT), all N_{obs} components are used to reconstruct the trimmed version of $x[n]$. Note that the same windowing function is applied for all frequency indices of $X_{STFT,i}(m,f)$, since it is a 2-dimensional function.

3.4.3 Crosscorrelation-based trimming

This method is based on the evaluation of the maximum crosscorrelation coefficient between the IRs used to calculate PSZ filters and the IRs used to evaluate their performance.

In principle, we may be able to replicate some measurements, i.e., by placing the same loudspeakers and microphones in exactly the same positions. However, in a realistic scenario, e.g., in a vehicle, the early and late reflections in the measured IRs may change due to slight errors in the placement of the

microphones. This behavior can be observed in Fig. 3.16, highlighted in the inset, where a comparison between reference and secondary IRs is presented. We can note that the two IRs differ after the first almost identical part.

A mismatch between reference and evaluation IRs allows us to approach a more realistic system performance evaluation. We refer to performance measures, in particular the AC, that almost do not change in a neighborhood of the measurement point. This means that we may be able to reduce the pre-ringing and improve performance robustness with respect to the measurement point by designing PSZ filters such that they do not control the IR component that differs in the secondary IR. So, the idea is to mitigate early and late reflections of the IRs by applying crosscorrelation-based trimming.

The algorithm is described in the flow chart in Fig. 3.17. The reference and evaluation IRs, named $x_{ref}[n]$ and $x_{eval}[n]$, respectively, are filtered by the filterbank giving as output their 1/3-octave band components $\{x_{ref,1}[n], \ldots, x_{ref,i}[n], \ldots, x_{ref,N_{obs}})[n]\}$ and $\{x_{eval,1}[n], \ldots, x_{eval,i}[n], \ldots, x_{eval,N_{obs}})[n]\}$. Considering the same 1/3-octave

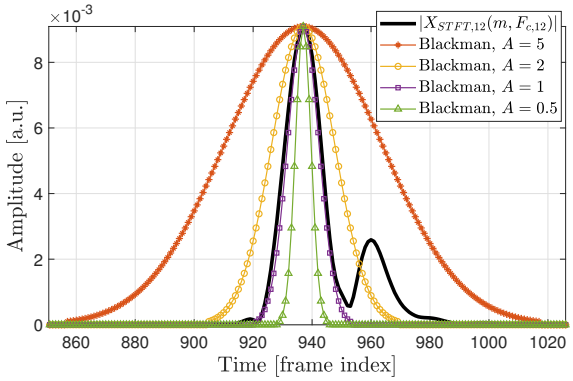


Figure 3.14: Example of time relation between the STFT magnitude at the fixed frequency of $F_c = 1260$ Hz and Blackman windows obtained according to (3.22), with various values of the parameter A. The STFT is calculated from the 12-th component of the measured IR between loudspeaker *A1* and the left driver microphone at $(0, h100)$.

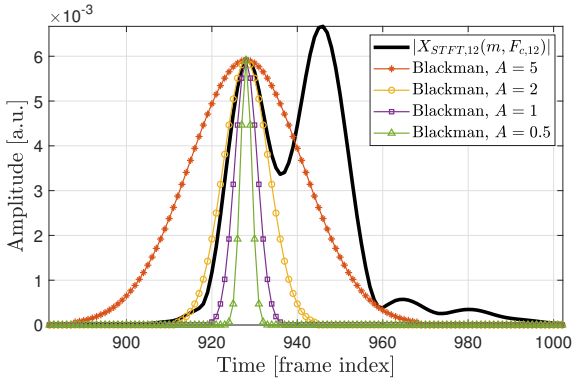


Figure 3.15: Example of time relation between the STFT magnitude at the fixed frequency of $F_c = 1260$ Hz and Blackman windows obtained according to (3.22), with various values of the parameter A. The STFT is calculated from the 12-th component of the measured IR between loudspeaker $A6$ and the left driver microphone at $(0, h100)$.

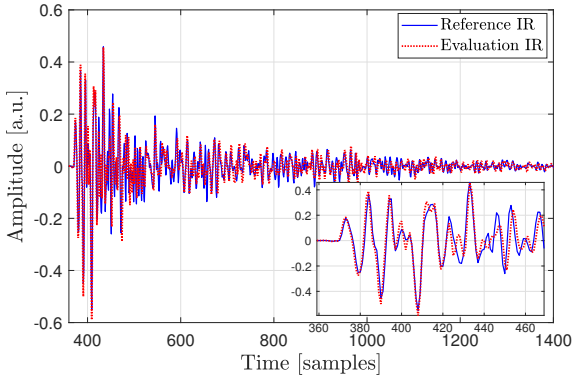


Figure 3.16: Comparison between reference and evaluation IRs. The measurements are taken between $A1$ and the left driver microphone at $(0, h100)$. A zoom of the x axis allows to better see the differences of the IR belonging to reference and secondary sets.

band components for both IRs, we search for the location of their maximum values as

$$\begin{cases} N_{max,ref,i} = \arg\max_n \ \upsilon\left(x^2_{ref,i}[n]\right) \\ N_{max,eval,i} = \arg\max_n \ \upsilon\left(x^2_{eval,i}[n]\right) \end{cases} \tag{3.23}$$

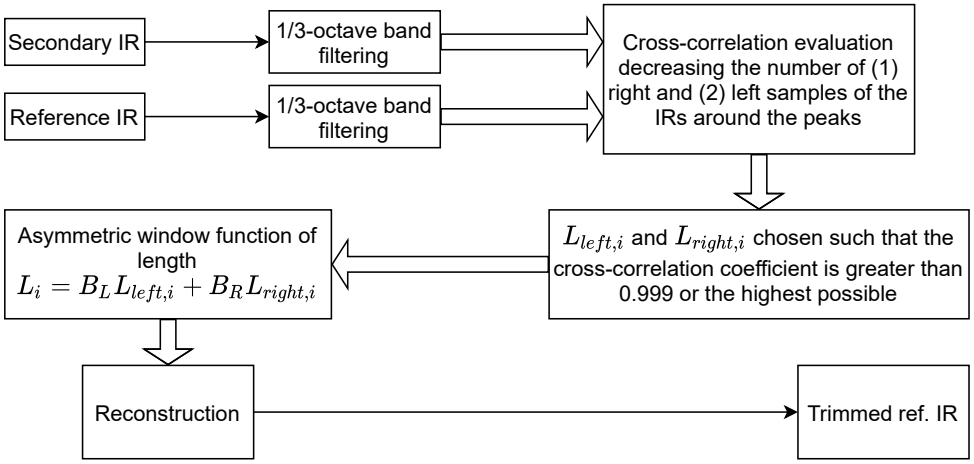


Figure 3.17: Flow chart of the crosscorrelation-based trimming algorithm.

where $\upsilon(\cdot)$ is the envelop function, and we set a maximum number of samples on the left as $L_{left,max,i}$ and on the right as $L_{right,max,i}$ of the $N_{max,ref,i}$ and $N_{max,eval,i}$ points that are considered for the first evaluation of the crosscorrelation. $L_{left,max,i}$ is the number of samples between $N_{max,ref,i}$ and $N_{in,left,i}$ where $N_{max,ref,i}$ is the first sample such that

$$\upsilon\left(x^2_{ref,i}[N_{in,left,i}]\right) \geq 10^{-2}\upsilon\left(x^2_{ref,i}[N_{max,ref,i}]\right). \tag{3.24}$$

The threshold of 10^{-2} with respect to the instantaneous energy of the signal is chosen empirically considering it as a reference value to assume the signal started or ended.

Assume now that the IRs have length equal to F_s samples, i.e., 1 second, and define the reflected version of $x_{ref,i}$ around $F_s/2$ samples as

$$\check{x}_{ref,i}[n] = x_{ref,i}[-n + F_s] \tag{3.25}$$

for index $n = 1, 2, \ldots, F_s$. By using (3.25) in (3.23), we can calculate the corresponding point $\check{N}_{max,ref,i}$. After that, we can set $N_{in,right,i}$ such that

$$\upsilon\left(\check{x}^2_{ref,i}[N_{in,right,i}]\right) \geq 10^{-2}\upsilon\left(\check{x}^2_{ref,i}[\check{N}_{max,ref,i}]\right). \tag{3.26}$$

At this point, $L_{right,max,i}$ is the number of samples between $\check{N}_{max,ref,i}$ and $N_{in,right,i}$. In Fig. 3.18(a), the relation between $x_{ref,i}$, $N_{max,ref,i}$ and $N_{in,left,i}$ is illustrated. In Fig. 3.18(b), the relation between $\check{x}_{ref,i}$, $\check{N}_{max,ref,i}$ and $N_{in,right,i}$ is illustrated. Note that, the mirroring operation is motivated by the implementation in Matlab to find the position in the tail of the considered signal where the magnitude of the signal is above a given threshold.

We begin by evaluating the crosscorrelation coefficient between the right parts of the IRs components, i.e.

$$\begin{cases} x_{ref,i}[n] & \text{for } N_{max,ref,i} \leq n \leq N_{max,ref,i} + L_{right,max,i} - \ell \\ x_{eval,i}[n] & \text{for } N_{max,eval,i} \leq n \leq N_{max,eval,i} + L_{right,max,i} - \ell \end{cases} \tag{3.27}$$

where $\ell = 0, 1, \ldots, L_{right,max,i} - 1$. We increase ℓ, starting from zero, until a crosscorrelation coefficient of 0.999 is achieved with ℓ^*_{right} or until $\ell = L_{right,max,i} - 1$ is reached. In case of the second condition, we take ℓ^*_{right} such that the crosscorrelation coefficient is maximum. We choose a value of 0.999 as a practical sufficiently-large value. Given ℓ^*_{right}, the right length of the window (in samples) is set as

$$L^{(k)}_{right} = L_{right,max,i} - \ell^*_{right}. \tag{3.28}$$

The same procedure is performed on the left parts of the IRs components, i.e., between

$$\begin{cases} x_{ref,i}[n] & \text{if } N_{max,ref,i} - L_{left,max,i} + \ell \leq n \leq N_{max,ref,i} \\ x_{eval,i}[n] & \text{if } N_{max,eval,i} - L_{left,max,i} + \ell \leq n \leq N_{max,eval,i} \end{cases} \tag{3.29}$$

where $\ell = 0, 1, \ldots, L_{left,max,i} - 1$. Proceeding as in the right case, we can determine ℓ^*_{left} which assumes the same meaning of ℓ^*_{right} on the left side of $N_{max,ref,i}$ and $N_{max,eval,i}$. Then, the left length of the window (in samples) is set as

$$L_{left,i} = L_{left,max,i} - \ell^*_{left}. \tag{3.30}$$

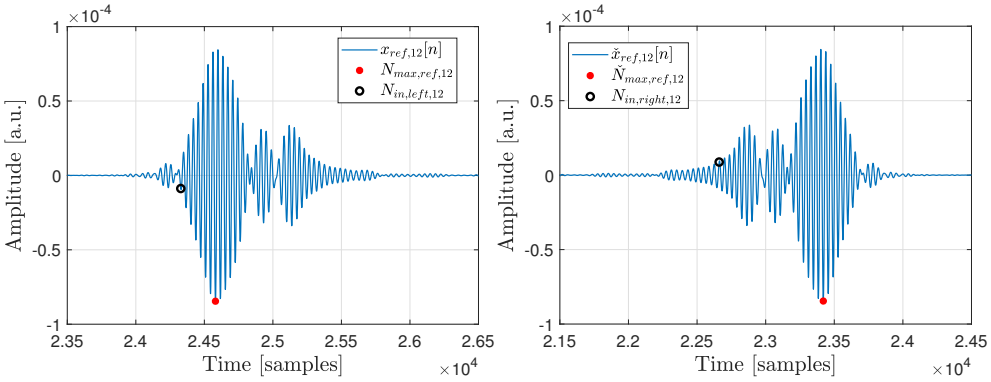


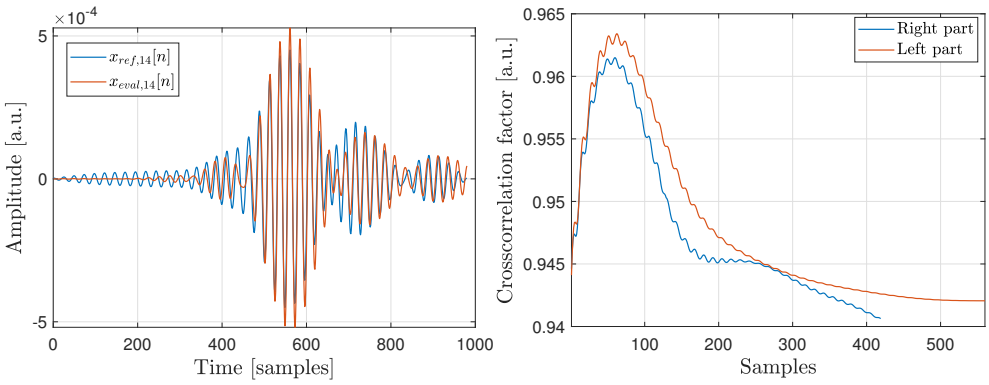
Figure 3.18: (a) Relation between $x_{ref,i}$, $N_{max,ref,i}$ and $N_{in,left,i}$. (b) Relation between $\check{x}_{ref,i}$, $\check{N}_{max,ref,i}$ and $N_{in,right,i}$. The 12-th component of the measured IR between loudspeaker $A6$ and the left driver microphone at $(0, h100)$ is considered.

An example of this procedure is shown in Fig. 3.19, where, in Part (a), the signals are aligned with respect the location of their peak values ($N_{max,ref,i}$ and $N_{max,eval,i}$) and they are visualized from $N_{max,ref,i} - L_{left,max,i}$ to $N_{max,ref,i} + L_{right,max,i}$. In Part (b), the crosscorrelation coefficient is plotted as a function of ℓ. The algorithm takes ℓ^*_{left} and ℓ^*_{right} that, in this case, achieve the maximum crosscorrelation coefficient since they do not reach the value of 0.999.

The length of the windowing function is then

$$L_i = B_L L_{left,i} + B_R L_{right,i} \tag{3.31}$$

where $B_L > 0$ and $B_R > 0$ are suitable constants to be determined by trial and error, with the aim of increasing or maintaining the AC achieved without trimming and removing as much as possible the late reflections in the IRs. The resulting windowing function will be asymmetric since, in general, $B_L L_{left,i} \neq B_R L_{right,i}$, and it will be formed by the first $B_L L_{left,i}$ samples of a symmetric window of length $2B_L L_{left,i}$ and the last $B_R L_{right,i}$ samples of a symmetric window of length $2B_R L_{right,i}$.



(a) (b)

Figure 3.19: (a) 14-th component (F_c = 2 kHz) of reference and evaluation IRs measured between loudspeaker $A8$ and the left driver microphone at $(0, h100)$. (b) Crosscorrelation coefficient versus the number of samples ℓ considered for the left and right parts of the signals.

Finally, after windowing, the N_{obs} components are used to reconstruct the trimmed version of the reference IR.

3.4.4 Frequency-proportional trimming

The last method is based on the assumption that the sound energy of the high-frequency components expires faster than the lower-frequency components [59]. For this reason, it may be reasonable to set a window length as a function of the considered frequency range. In particular, we consider a function inversely proportional to the central frequencies of the 1/3-octave bands. The flow chart of the algorithm is shown in Fig. 3.20.

At a glance, by comparing the flow charts of the proposed methods (see Figs. 3.12, 3.13, 3.12 and 3.20), we can note that this last method is the easiest one to implement and the one that perform the FDT faster due to limited number of operations required.

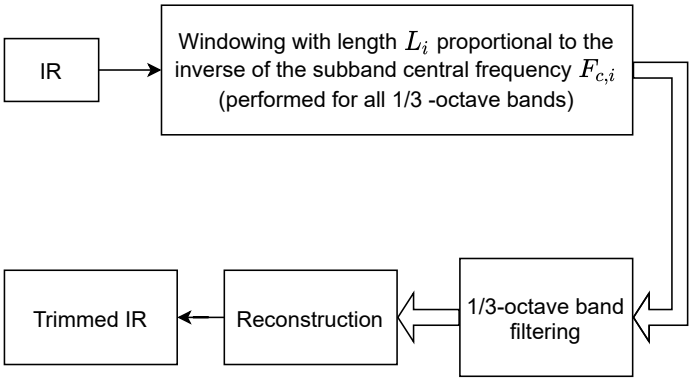


Figure 3.20: Flow chart of the frequency-proportional trimming algorithm.

For each N_{obs}, the input IR is windowed with length

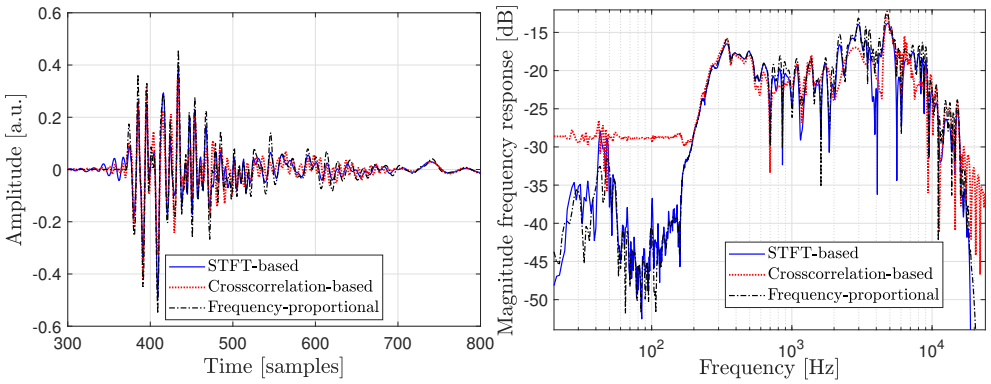
$$L_i = 2 \left\lceil \frac{DF_s}{F_{c,i}} \right\rceil \tag{3.32}$$

where $D > 0$ is a suitable scalar to be found by trial and error. The output signals are filtered by the filterbank for the 1/3-octave band analysis and reconstructed to get the trimmed version of the IR. By far, this is the simplest and most straightforward method for trimming.

Note that, $F_s/F_{c,i}$ corresponds to a period of a sinusoidal signal oscillating at the frequency $F_{c,i}$ expressed in samples. This means that the i-th window with length (3.32) extracts about $2D$ periods of the i-th subcomponent of the IR.

For comparison, the trimmed versions of the same IR obtained with the three different algorithms are plotted in Fig. 3.21(a). In Fig. 3.21(b), the corresponding MFRs are depicted.

Note that, the oscillation at the beginning of the IRs obtained by STFT- and crosscorrelation-based algorithms are undesired and caused by the 1/3-octave band filtering performed before trimming. In the frequency domain, we can note that the method based on the crosscorrelation shapes the MFR, allocating an excess of energy in the low frequency (between 0 and 200 Hz) and the high frequency (between 16 and 24 kHz) ranges. This is another unwanted effect.



(a) (b)

Figure 3.21: (a) Trimmed IRs obtained by the proposed three methods. The original IR is measured between loudspeaker *A1* and the left driver microphone at $(0, h100)$. For the STFT-based method, A is set to 0.5. For crosscorrelation-based trimming, A and B are set to 5. For the frequency-proportional algorithm, A is set to 20. In all methods, the Tukey window is used with $\alpha = 0.2$ and centered at the peak values of the windowed signals. (b) MFRs of the trimmed IRs shown in (a) and obtained with the proposed three methods.

3.5 Numerical Results

3.5.1 Optimal trimming lengths based on exhaustive search

As reference IRs, the set of IRs measured in the position $(0, h100)$ is used to design the filters, and the IR sets measured in the positions $(30, h80)$ and $(-30, h120)$ are used for AC evaluation for trimming length selection.

The trimming length is applied starting from the beginning of the non-zero IR region, defined as the sample n_R such that

$$|h_{n_R}| \geq 0.3\,(\max\ |h_i|) \tag{3.33}$$

where h_i is the i-th sample of the IR. The implemented trimming function can

be expressed as

$$w[i] = \begin{cases} 1 & \text{if } i < n_R + \lfloor \frac{L_{trim}}{2} \rfloor \\ 0.5 + 0.5 \cos\left[\left(i - n_R - \lceil \frac{L_{trim}}{2} \rceil - 1\right) \frac{\pi}{\lceil L_{trim}/2 \rceil}\right] & \\ & \text{if } n_R + \lfloor \frac{L_{trim}}{2} \rfloor \leq i \leq n + L_{trim} \\ 0 & \text{otherwise} \end{cases} \tag{3.34}$$

where L_{trim} is the trimming length, $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ are the floor and ceil functions, respectively. The trimming lengths considered for the exhaustive search are in the range $[2, 22000]$ samples.

In Figs. 3.22(a) and 3.22(b), the AC contour plots are shown, evaluated at the reference control points (used for filter design) and at the secondary control points. Clearly, the maximum of AC at the reference control point is achieved without applying trimming. However, we are more interested in the second contour plot, where we can see that we can slightly improve the performance

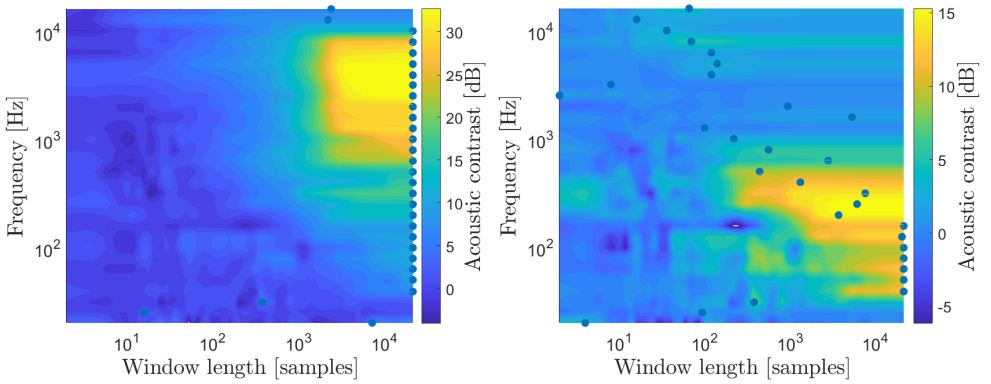


Figure 3.22: (a) Contour plots of the AC as a function of frequency and trimming lengths evaluated at the reference control points. (b) Contour plots of the AC as a function of frequency and trimming lengths evaluated at the secondary control points. The marks indicate the coordinates of the maximum value for each 1/3-octave band.

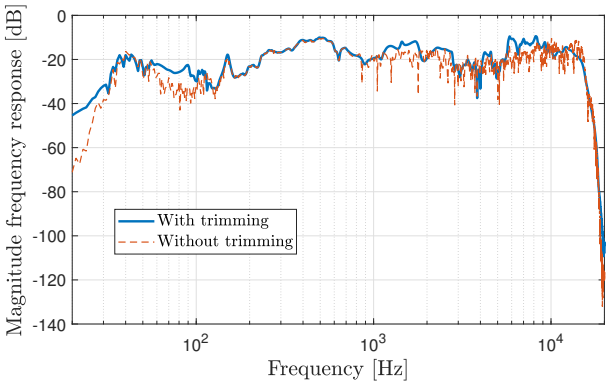


Figure 3.23: MFRs of the filters designed for *A1* to control the left ear region of the driver with and without FDT.

at the secondary control points by applying a FDT to the reference IRs.

The subband filtering is performed with the filters defined in Section 3.1 according to (3.7) and (3.8) with $\alpha = 0.5$. Note that, due to the operational frequency range of the sound system, the first 1/3-octave bands until the one centred at 39 Hz are integrated into the low-pass filter H_0 and the last considered 1/3-octave band is the one centred at 16 kHz.

As described in the section about FDT, we can see in Fig. 3.23 the smoothing effect of the trimming operation on the MFR of a filter designed for a loudspeaker array. Furthermore, as expected, we can clearly observe the reduction of the pre-ringing in the IRs of the filters and the overall IRs at the control points[1] shown in Figs. 3.24 and 3.25, respectively. This is also confirmed by the radius of gyration of the energy density (RoGoED) evaluated for the IRs of the filters and the overall IRs at the control points shown in Fig. 3.26.

Finally, we can consider the AC achieved with and without FDT shown in Fig. 3.27, where we can observe only a small gain of the AC over the 8 kHz subband with the frequency-dependent trimmed solution.

[1]With the overall response at a specific control point (or microphone) we mean that we are considering the superposition effect of all the acoustic responses, between all the loudspeakers and the specific microphone, in cascade with the associated PSZ filters.

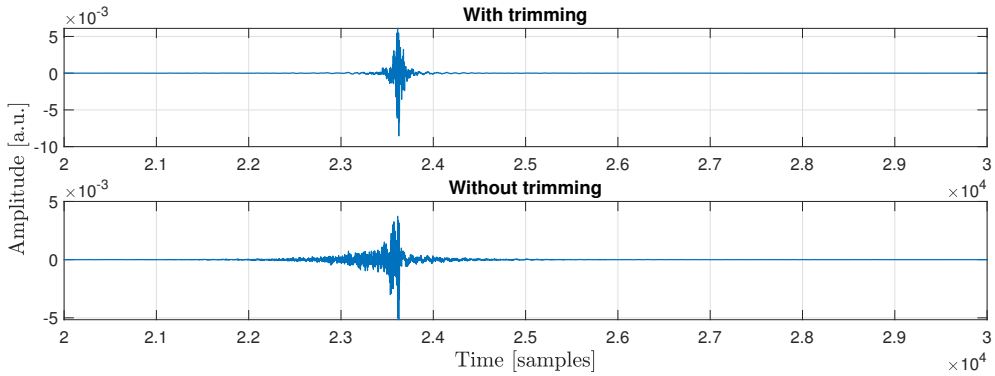


Figure 3.24: IRs of the filters designed for *A1* to control the left ear region of the driver with and without FDT.

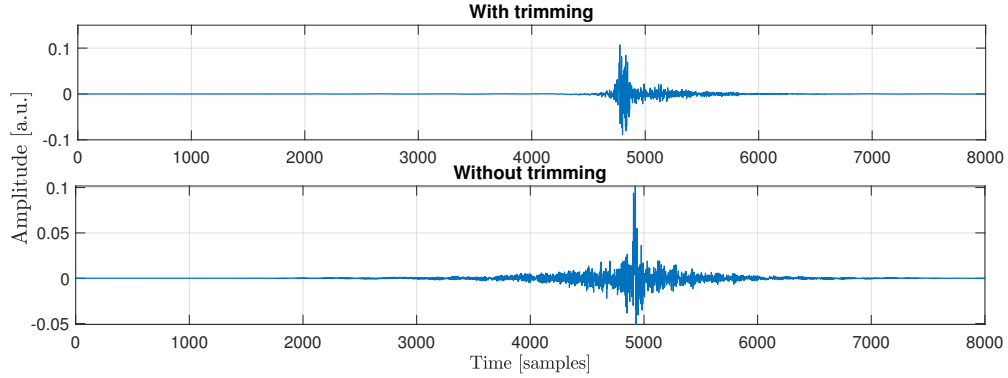


Figure 3.25: Overall IRs evaluated at the left control point of the driver. The PSZ filters are designed to achieve the bright region at this point with and without FDT.

3.5.2 STFT-based, crosscorrelation-based and frequency-proportional trimming methods

First of all, consider the AC achieved using the STFT-based trimming algorithm, and various windowing functions and values of the parameter A. The curves are shown in Figs. 3.28 and 3.29.

We can note that there is almost no performance difference considering the various windowing functions. Furthermore, the reduction of the window

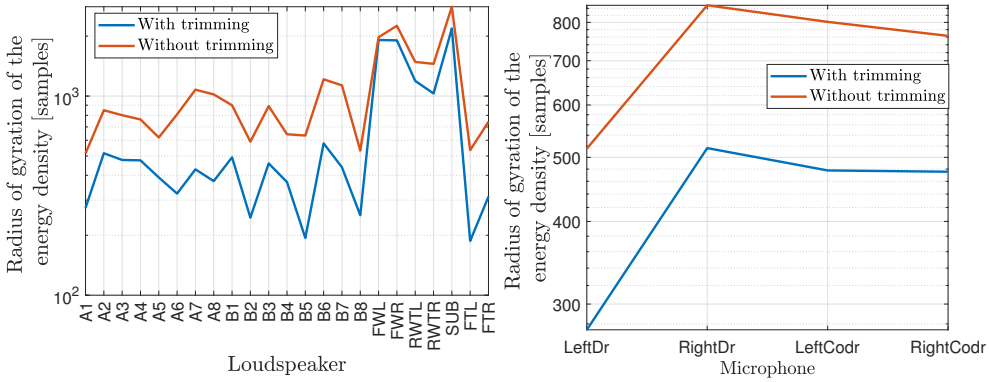


Figure 3.26: RoGoED evaluated for (a) the IRs of the PSZ filters designed with and without FDT and (b) the overall IRs obtained with these filters in the considered acoustic regions.

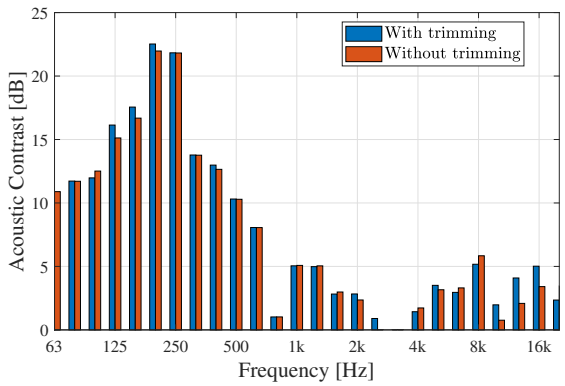


Figure 3.27: AC, averaged over 1/3-octave bands, evaluated between the driver (bright region) and codriver (dark region) obtained with PSZ filters designed with and without FDT.

lengths decreases, on average, the AC.

The AC obtained with PSZ filters, designed using the IRs trimmed with the crosscorrelation-based method and various windowing functions, are plotted in Fig. 3.30. In Fig. 3.31, the RoGoED (compactness factor, see Section 1.3.3) is shown for the filters considered for the AC curves in Fig. 3.30(a). Even for

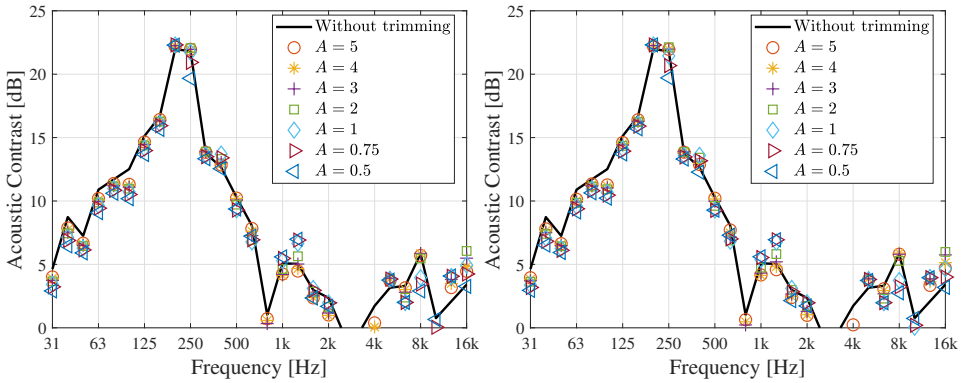


Figure 3.28: AC, averaged over 1/3-octave bands, evaluated between the driver (bright) and codriver (dark) and obtained with PSZ filters designed using the IRs trimmed with the STFT-based method, (a) Blackman and (b) Chebyshev windowing function with various values of the parameter A.

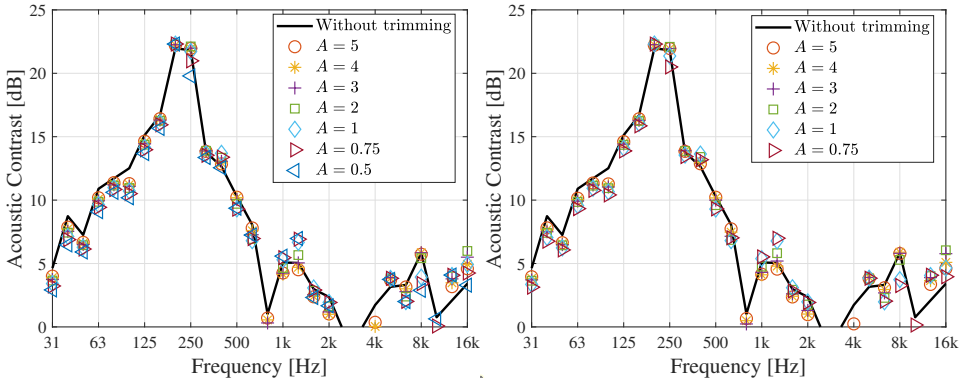


Figure 3.29: AC, averaged over 1/3-octave bands, evaluated between the driver (bright) and codriver (dark) and obtained with PSZ filters designed using the IRs trimmed with the STFT-based method, (a) Gaussian ($r = 3$) and (b) Tukey windowing function with various values of the parameter A.

the cases with the largest values of the parameters B_L and B_R (Fig. 3.30(a)), for which there is a small reduction of the RoGoED, there is, on average, a loss of the AC. We can note that the use of the Tukey function with a small value of the parameter α achieves a lower reduction of the AC compared with

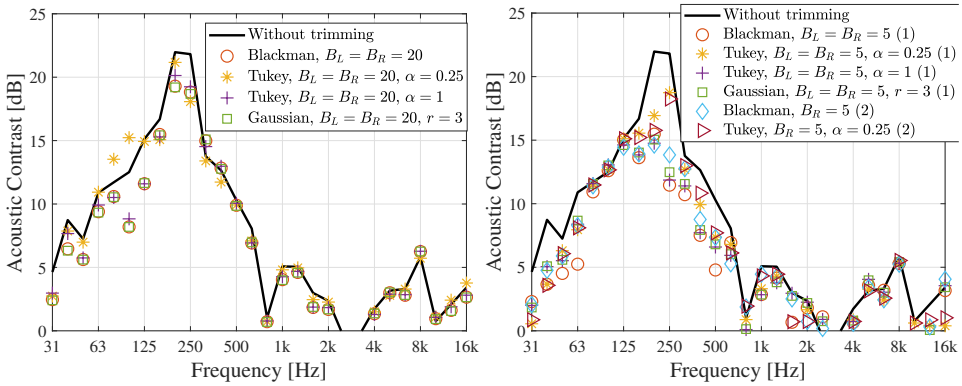


Figure 3.30: AC, averaged over 1/3-octave bands, evaluated between the driver (bright) and codriver (dark) and obtained with PSZ filters designed using the IRs trimmed with the crosscorrelation-based method and various windowing functions. (a) The complete window (left and right parts) is applied centering it at the peak values of the windowed signals. In (b), (1) indicates that the complete window (left and right parts) is applied centering it at the peak values of the windowed signals and (2) indicates that only the right part of the window is applied.

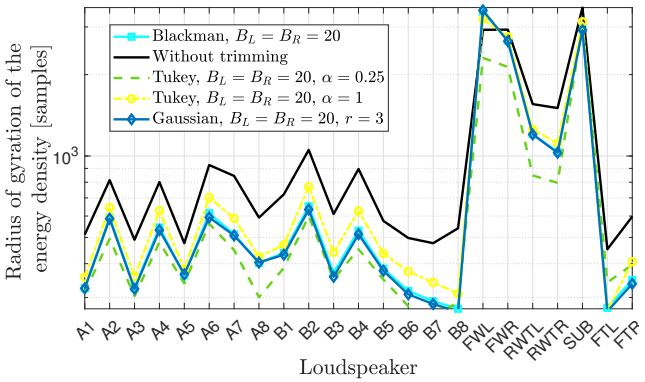


Figure 3.31: RoGoED of the PSZ filters (used to derive Fig. 3.30(a)) designed using the IRs trimmed with the crosscorrelation-based method and various windowing functions. The complete window (left and right parts) is applied centering it at the peak values of the windowed signals.

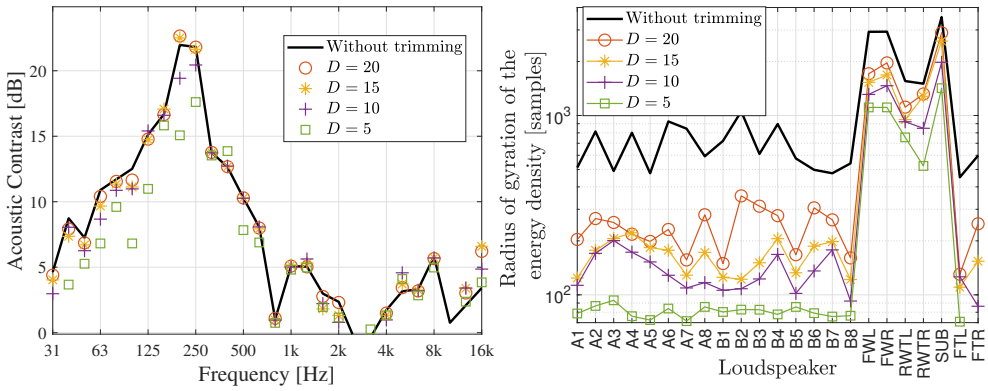
the other windowing functions.

For the last method, i.e., the one based on the frequency proportionality

with windowing performed around the peak value of the windowed signals, the AC curves and the RoGoED values of the associated filters are shown in Figs. 3.32(a) and 3.32(b), respectively. For the case of windowing performed around the beginning of the windowed signals, refer to Figs. 3.33(a) and 3.33(b). For this method, only the Tukey windowing function with $\alpha = 0.25$ is considered because it is the windowing function that allows to achieve the lowest values of RoGoED with the lowest loss of the AC (on average). For this method, $D = 15$ seems to be a good trade-off between the filter IR compactness and the achieved AC. About the different position of the windowing function (peak value or beginning), comparing Fig. 3.32 with Fig. 3.33 we can note almost no difference between the two alternatives.

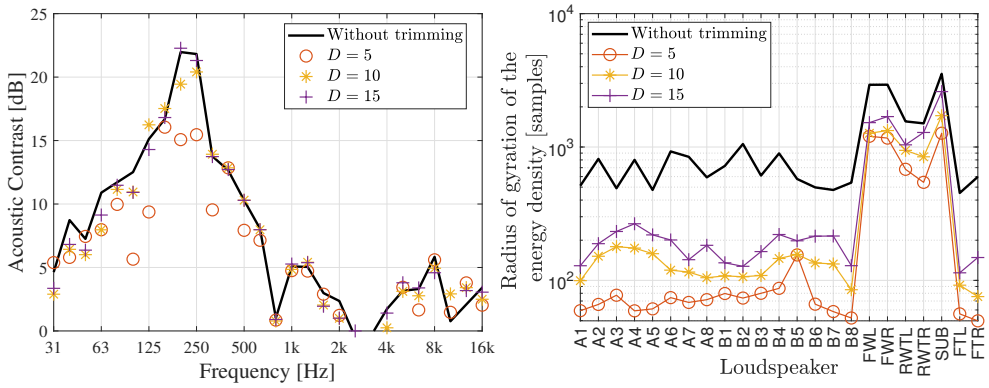
3.6 Conclusions

Building upon the concept of FDT, originally introduced in [15], we have developed an algorithm based on an exhaustive search for optimal trimming



(a) (b)

Figure 3.32: (a) AC, averaged over 1/3-octave bands, evaluated between the driver (bright) and codriver (dark) and obtained with PSZ filters designed using the IRs trimmed with the frequency-proportional window lengths with various values of the parameter D and Tukey function with $\alpha = 0.25$. In this case, the windows are centered at the peak value of the windowed signal. (b) RoGoED of the PSZ filters.



(a) (b)

Figure 3.33: (a) AC, averaged over 1/3-octave bands, evaluated between the driver (bright) and codriver (dark) and obtained with PSZ filters designed using the IRs trimmed with the frequency-proportional window lengths with various values of the parameter D and Tukey function with $\alpha = 0.25$. In this case, the windows are centered at the beginning of the windowed signal. (b) RoGoED of the PSZ filters

lengths. Additionally, we have presented a comprehensive review of QMFs to establish a filterbank tailored for fractional-octave band filtering. Furthermore, our exploration encompassed an analysis of the impact of FDT in both time and frequency domains, along with an examination of the moments of IR energy density.

The core focus of our algorithm is to exhaustively search for the optimal trimming lengths. This search is geared toward maximizing the AC, as evaluated numerically at two control points situated in close proximity to those used for filter design with the pressure matching (PM) method. As a result, we have effectively managed to mitigate a significant portion of the pre-ringing observed in the IRs of the designed filters, including the overall IRs evaluated at the control points. However, it is important to note that despite our efforts, there is no discernible performance improvement concerning the AC. Additionally, the sound timbre achieved with the proposed trimming lengths does not achieve the desired result when compared to the empirical solution

employed in [15].

Moreover, we have introduced three new algorithms designed to assist in selecting appropriate windowing lengths. These algorithms are grounded in simple assumptions and offer valuable insights. Our investigation has also delved into a discussion of various windowing functions and their temporal and spectral characteristics.

In the context of the exhaustive search method, our reported results indicate minimal improvement in the AC when employing the proposed trimming techniques. However, our endeavors have successfully reduced the pre-ringing in the IRs of the designed filters, minimizing the loss of AC.

Upon comparing the performance achieved with various windowing functions outlined in this report, it becomes evident that the Tukey function with a small parameter α (close to zero) emerges as the most suitable choice. This particular windowing function stands out for its ability to reduce the RoGoED factor with minimal loss of AC.

Among the four algorithms, the frequency-proportional trimming method stands as the most efficient choice in terms of RoGoED factor reduction, AC preservation, implementation simplicity, and execution time efficiency. In contrast, both the crosscorrelation-based and AC maximization-based methods demand significantly longer execution times due to their exhaustive search for optimal window lengths aimed at maximizing crosscorrelation coefficients and AC, respectively. The STFT-based method runs somewhat faster but still lags behind the efficiency of the frequency-proportional approach.

Chapter 4

PSZ filter design methods

The acoustic contrast control (ACC) method designs filters to maximize the acoustic contrast (AC) between two sound regions. However, the acoustic characteristics of sound produced by the application of these filters and the potential challenges associated with implementing them in a real-world system have remained topics lacking sufficient clarification within the related literature. For this reason, these topics are investigated by means of numerical simulations and "on the field" measurements performed in a vehicle equipped for personal sound zone (PSZ) system tests. Specifically, a rescaling method of the ACC solution is treated.

Recent research, exemplified in works such as [16] and [17], has highlighted the potential for enhanced performance in terms of AC through the strategic design of the target sound field employed by the pressure matching (PM) algorithm. However, it is important to note that altering the target sound field can have adverse effects on perceived audio quality. In essence, optimizing the target sound field to achieve maximum AC leads the PM algorithm solution to converge to the solution provided by the ACC algorithm [17]. To address the objective of AC maximization, we introduce a design method for the target sound field in the context of the PM algorithm.

Lastly, a general solution based on the PM method, aimed to minimize

the average reproduction error and improve the robustness, namely statistical pressure matching (SPM), is proposed and compared with a few methods available in the literature. A performance comparison with the original PM method is carried out by means of numerical simulation and measurements performed directly in the vehicle.

This chapter is organized as follows. In Section 4.1, the rescale method of the ACC filters is discussed providing the numerical results. In Section 4.2, the proposed algorithm for the design of the target sound field is outlined and the numerical results are shown. In Section 4.3, the SPM method is compared theoretically with similar methods found in the literature, and a numerical performance comparison with the original formulation of the PM is presented. Finally, in Section 4.4, overall conclusions are drawn.

4.1 Remarks on scaling methods of ACC filters

In most of the reviewed literature, the possible scaling of the solution given by the ACC algorithm was not considered because it does not affect the performance in terms of AC. In the evaluation of the reproduction error in the bright zone (BZ), it may seem obvious to rescale the solution for each considered frequency values to obtain the desired sound level over the entire sound spectrum. Initially, we assumed that the scaling method was not a relevant aspect and we implemented a gain calculated for each frequency (frequency-dependent gain), however, the listening test with filters designed with the ACC algorithm showed that the produced sound was very disturbing with highly perceivable distortions at high frequencies. At first, we attributed this problem to the algorithm itself. However, after an analysis in the time and frequency domains of the designed filters and some trials, we concluded that the problem was related to the method used to calculate the gains. Indeed, with the implementation of a gain equal for all frequencies (frequency-independent gain), the highly perceivable distortions at high frequencies disappeared. So, we calculate $\mathbf{q}_{opt}$ for each frequency and the entire frequency responses (FRs) at the

bright control points obtained with these solutions, we calculate the squared pressure averaged in the frequency domain for a reference bright point (but we can also take the average between multiple bright control points), and we calculate the unique gain for all designed filters.

The need of a frequency-independent gain can be justified as follows. The solutions obtained by the ACC algorithm do not control the pressure level in the BZ at all and the reached pressure level for close frequencies can be very different, creating a spectrum (evaluated numerically) with many notches and peaks. A frequency-dependent gain compensates for all these notches and peaks creating complementary peaks and notches in the FRs of the designed filters that pre-distort the produced sound. On the other hand, a frequency-independent gain only scales the spectrum shape. From another point of view, a frequency-dependent gain changes the amplitude of each frequency component modifying completely the impulse responses (IRs) of the filters designed in the frequency domain.

Another problem found in the practical implementation of the ACC algorithm is the power allocation at frequencies outside the operational frequency range of the system. Indeed, even if the loudspeakers are not able to generate power at the considered frequency, the solution of (1.17) is a unitary norm vector that cannot have all null values. To avoid this problem, a control of the eigenvalues of the spatial correlation matrix $\mathbf{R}_B$ must be performed. Indeed, the eigenvalues of $\mathbf{R}_B$ can be related to the capacity of the audio system to generate power at a specific frequency.

In accordance with the above, the filters designed to control the phase and using a frequency-dependent or frequency-independent gain will be referred to by the acronyms YFD and YFI, respectively. We will refer to the designed filters with the acronyms NFD and NFI, if no phase control and the frequency-dependent or the frequency-independent scaling, respectively, is adopted[1].

The frequency-dependent gain is calculated to achieve a flat magnitude frequency response (MFR) of the desired amplitude level at the central control

[1] Y/N refer to yes/no phase control, respectively.

point of the bright region, whereas the frequency-independent gain is calculated to achieve on average a flat MFR of the desired amplitude level.

4.1.1 Numerical results

Configuration A is used in this section for the numerical results, specifically, the same configuration used in Section 3.5, based on 7 factory-installed loudspeakers, 2 linear arrays on the dashboard and 8 headrest arrays.

We can begin by considering the MFRs of the filters designed for the first loudspeaker of the array in front of the driver to control the left ear region of the driver and considering the control of the phase and the frequency-dependent and independent scaling approaches. These curves are shown in Fig. 4.1. The curves of the filters calculated without considering the phase control are missing since a change of the filter phase will not change the MFR of the filter. As mentioned in Chapter 1, the high variation between notches and peaks at high frequencies can be observed in the solution obtained with the frequency-dependent scaling gain.

In the time domain, the phase control also influences the designed filters. Indeed, as can be observed in Fig. 4.2, there is an average difference of more than 1000 samples of the RoGoED between the filters designed with or without

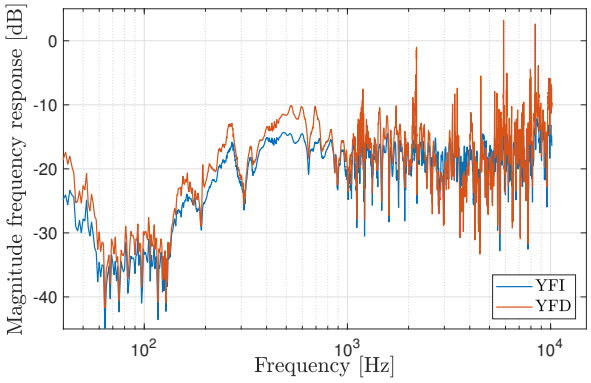


Figure 4.1: MFRs of the filters designed for *A1* to control the left ear region of the driver according to the frequency-dependent and independent gain approaches with phase control.

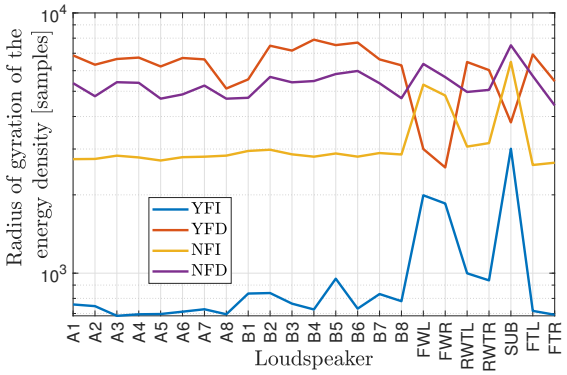


Figure 4.2: Radius of gyration of the energy density (RoGoED) of the PSZ filters designed according to all combinations of the phase control and frequency-dependent and independent scaling methods.

phase control. This difference is larger considering different scaling methods. Note that for each loudspeaker a filter for each ear of the driver and codriver (4 filters) is designed and the values in Fig. 4.2 are given by averaging 4 different values of $\sqrt{\mu_2}$ (1.27).

In Fig. 4.3, the IRs of the filters designed for loudspeaker $A1$ according to all combinations of the phase control and frequency-dependent and independent scaling methods are reported. Here, we can see that the YFI solution is much more compact with respect to the others, whereas the differences between the YFD, NFI and NFD solutions are more difficult to observe.

Now, we can analyze what we obtain at the evaluation control points with the filters designed according to the various methods.

We can start by considering the overall MFRs. These are shown in Figs. 4.4(a) and 4.4(b), where the first one is obtained with control of the phase and the second one without it. A first observation regards the reached amplitude level. Since these evaluations are performed using a different set of channel IRs, i.e., a mismatch between the IRs used for filter design and the IRs used for the numerical evaluations is present, none of the designed solutions achieves the flat target (0 dB). However, even if in the limited frequency range of about $[50, 2000]$ Hz only, the YFD solution achieves the closest level

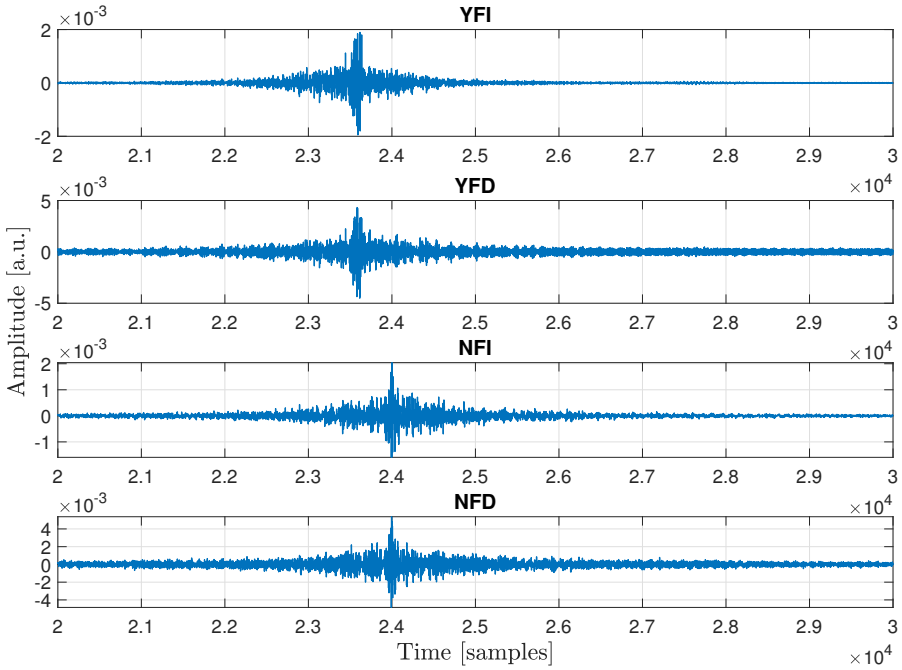
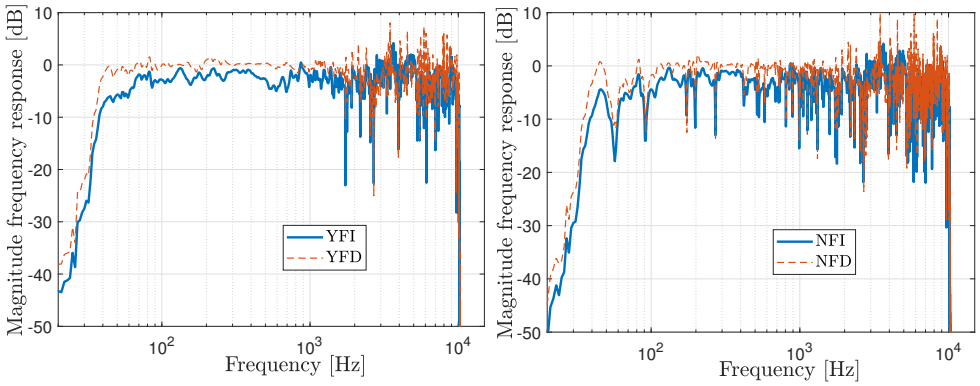


Figure 4.3: IRs of the filters designed for *A1* to control the left ear region of the driver according to all combinations of the phase control and frequency-dependent and independent scaling methods.



(a) (b)

Figure 4.4: Overall MFRs evaluated at the left control point of the driver. The PSZ filters are designed to achieve the bright region at this point (a) with phase control and (b) without, for both the frequency-dependent and independent scaling approaches.

to the target. In general, we can see that the solutions that adjust the phase of the filter coefficients perform better with respect to the solutions that do not control the phase, i.e., the NFD and NFI solutions. In particular, we can see in Fig. 4.4(b) that there are some notches at low frequencies where the ear is sensible. Anyhow, the peaks at high frequencies caused by the peaks in the filter responses and the mismatched channel responses are the main problem, producing an audible annoying sound both in the vehicle and simulation.

Observing the overall IRs evaluated at the left control point of the driver, shown in Fig. 4.5, there is no straightforward way to conclude with the same interpretations made about the overall MFRs. It is clearly visible that the YFI solution achieves a lower pre-ringing with respect to the other solutions

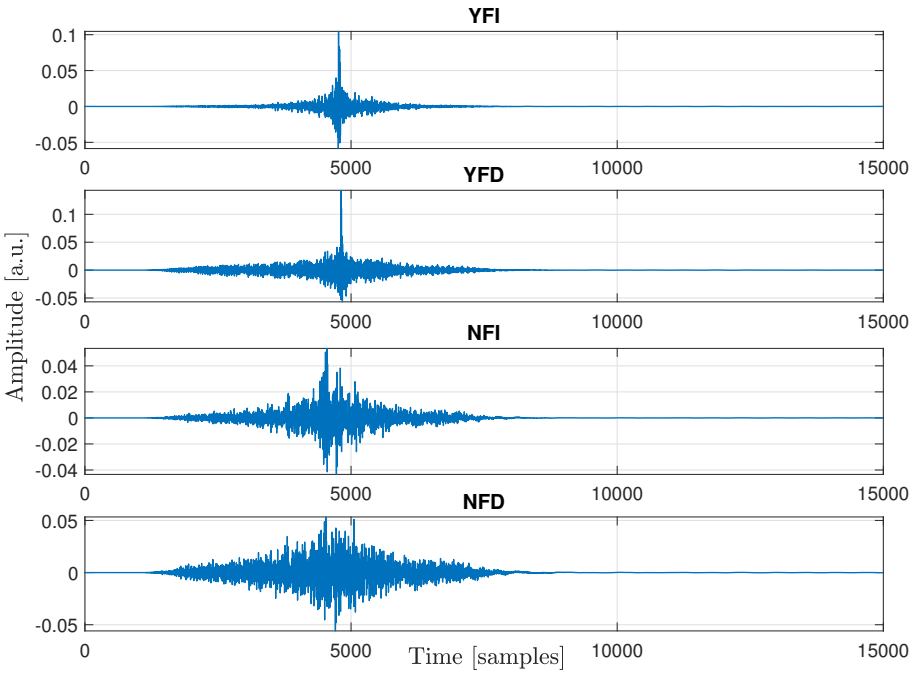


Figure 4.5: Overall IRs evaluated at the left control point of the driver. The PSZ filters are designed to achieve the bright region at this point according to all combinations of the phase control and frequency-dependent and independent scaling methods.

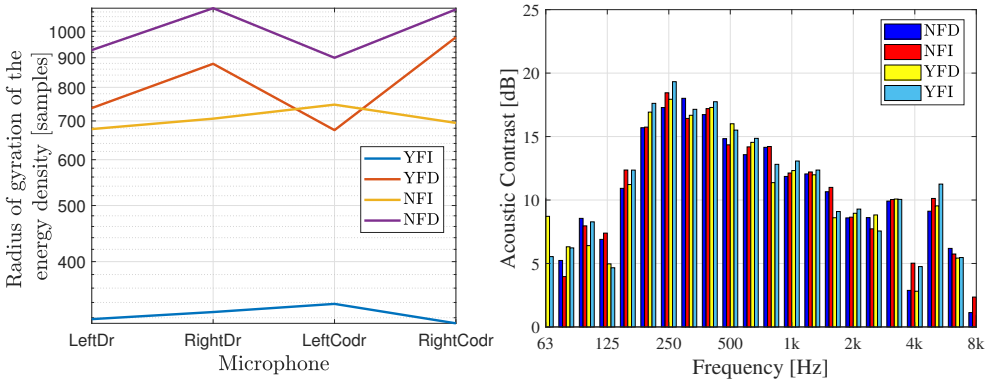
while the NFD is the worst one. Comparing the YFD and NFI solutions, the first seems better, however, the listening tests performed in the vehicle and numerical simulation show that this is not the case. The same behaviour can be observed in the values of the RoGoED plotted in Fig. 4.6(a) where the curves for the NFI and YFD exhibit very close average values over the evaluation control points. This fact may be justified by the misalignment at the control points (no phase control) of the low-frequency contributions allocating low-frequency energy components around the central time μ_1, increasing the pre-ringing and the RoGoED. However, unlike the high-frequency pre-ringing caused by the peaks at high frequencies in the MFR of the YFD solution, the low-frequency pre-ringing is not perceived as an annoying sound. Note that, the values of the RoGoED in Fig. 4.6(a) are referred to the overall IRs obtained with the filters designed to have the bright region at the same zone where the IRs are evaluated. The acronyms LeftDr, RightDr, LeftCodr and RightCodr stand for left driver, right driver, left codriver and right codriver, respectively.

Finally, we can consider the AC achieved by the different combinations of the phase control and the gain methods. The AC measured in the vehicle between the driver (BZ) and the codriver (dark zone (DZ)) are shown in Fig. 4.6(b). Unexpectedly, all the methods achieve almost the same performance in terms of AC. Taking into account also the listening tests, a clear improvement is achieved in terms of sound quality going from the NFD solution to the YFI solution; hence, this leads to the possibility of sound quality and timbre improvement by proper processing of the ACC solution without loss of performance in terms of the AC.

4.2 PM with target phase optimization

In almost all the literature, the target sound pressure vector is chosen as

$$\hat{\mathbf{p}} = \begin{bmatrix} \hat{\mathbf{p}}_B \\ \hat{\mathbf{p}}_D \end{bmatrix} \tag{4.1}$$



(a) (b)

Figure 4.6: (a) RoGoED of the overall IRs evaluated in the sound regions and (b) AC, averaged over 1/3-octave bands and measured in the vehicle between the driver (bright region) and codriver (dark region), considering PSZ filters designed according to all combinations of the phase control and frequency-dependent and independent scaling methods.

where[2] $\hat{\mathbf{p}}_B = \left[\hat{p}_{B,1} \cdots \hat{p}_{B,M_B}\right]^T$ is the vector composed of the sound pressures at the M_B control points assumed to be in the BZ with

$$\hat{p}_{B,i} = 1 \qquad i = 1, \ldots, M_B \tag{4.2}$$

$\hat{\mathbf{p}}_D = \left[\hat{p}_{D,1} \cdots \hat{p}_{D,M_D}\right]^T$ is the vector composed of the sound pressures at the M_D control points assumed to be in the DZ with

$$\hat{p}_{D,i} = 0 \qquad i = 1, \ldots, M_D. \tag{4.3}$$

This means that we are imposing flat targets in the frequency domain (or Dirac pulses in the time domain) for all the bright control points and null targets for all the dark control points. However, except for almost perfectly controlled environments such as anechoic chambers, in realistic scenarios it is not possible to have an IR without reflections that contribute to the reverberation, i.e., an ideal Dirac pulse. Furthermore, for music listening, some reverberation is

[2] $[\cdot]^T$ is the vector or matrix transpose operator

required because it adds qualities of fullness, warmth and cohesion to the musical piece [68]. Lastly, recent work, e.g., [16] and [17], proved that a proper target sound field may improve performance in terms of AC and array effort.

A proper design of the target sound field may also be a possible solution to reduce the pre-ringing in the IRs of the designed filters. Indeed, if we consider setting a target with time characteristics analogous to the natural behaviors of a sound propagating in the considered environment, i.e., direct sound component plus early and late reflections, the algorithm should reduce the effort for reflection control. In other words, it is not advisable to remove early and late reflections completely, thus reducing the pre-ringing in the designed filter. According to this idea, the target sound field may be one of the loudspeaker-microphone pair measured IRs, as considered in [17], a measured overall IR[3] with an equalized sound system, or an overall IR designed by using the delay and sum beamforming algorithm [7] to align all the loudspeaker contributions for one of the bright control points.

In our experiment, based on the superposition principle, we consider the design of different PSZ filter sets, for each ear of the vehicle passengers, to create multiple differentiated sound regions and achieve the stereo effect. Hence, two PSZ filter sets are designed for each passenger with the same target sound field at each ear.

In the following, the index $i = 1, \ldots, N_{sz}$ denotes the sound zone, where N_{sz} is the number of individual and differentiated considered sound regions (e.g., the vehicle passengers), and the index $j = 1, 2$ specifies the control point associated with the left and right ear for each passenger, respectively.

In order to improve the AC, optimization of the phase difference between the targets for the left and right ears is proposed. To this purpose, consider the optimal filter coefficients given by the PM method (1.13) for the j-th control

[3]IRs measured with all loudspeakers operating at the same time with the same input signal.

point of the i-th passenger

$$\mathbf{q}_{(i,j)}^{opt} = \begin{cases} \left(\mathbf{Z}_{(i,j)}^H \mathbf{Z}_{(i,j)} + \beta \mathbf{I}\right)^{-1} \mathbf{Z}_{(i,j)}^H \hat{\mathbf{p}}_{(i,j)} & \text{if } M-1 > L \\ \left(\mathbf{Z}_{(i,j)} + \beta \mathbf{I}\right)^{-1} \hat{\mathbf{p}}_{(i,j)} & \text{if } M-1 = L \\ \mathbf{Z}_{(i,j)}^H \left(\mathbf{Z}_{(i,j)} \mathbf{Z}_{(i,j)}^H + \beta \mathbf{I}\right)^{-1} \hat{\mathbf{p}}_{(i,j)} & \text{if } M-1 < L \end{cases} \tag{4.4}$$

where $\mathbf{Z}_{(i,j)} \in \mathbb{C}^{M-1 \times L}$ and $\hat{\mathbf{p}}_{(i,j)} \in \mathbb{C}^{M-1 \times 1}$ are the channel matrix and the target vector, respectively, for the j-th control point of the i-th passenger, M is the number of microphones and L is the number of loudspeakers. In particular, the matrix $\mathbf{Z}_{(i,j)}$ results by removing from the overall channel matrix $\mathbf{Z} \in \mathbb{C}^{M \times L}$ the row vector populated by the coefficients related to the control point $\bar{j}$ of the i-th passenger specified by the single element of the set $\{\{1,2\} \setminus \{j\}\}$, i.e., $z_{(i,\bar{j}),\ell}$ for $\ell = 1, ..., L$.

To simplify the expression, we can group all terms out of the target vector as $\mathbf{A}_{(i,j)}$, so that (4.4) can be written as

$$\mathbf{q}_{(i,j)}^{opt} = \mathbf{A}_{(i,j)} \hat{\mathbf{p}}_{(i,j)} \tag{4.5}$$

where

$$\mathbf{A}_{(i,j)} = \begin{cases} \left(\mathbf{Z}_{(i,j)}^H \mathbf{Z}_{(i,j)} + \beta \mathbf{I}\right)^{-1} \mathbf{Z}_{(i,j)}^H & \text{if } M-1 > L \\ \left(\mathbf{Z}_{(i,j)} + \beta \mathbf{I}\right)^{-1} & \text{if } M-1 = L \\ \mathbf{Z}_{(i,j)}^H \left(\mathbf{Z}_{(i,j)} \mathbf{Z}_{(i,j)}^H + \beta \mathbf{I}\right)^{-1} & \text{if } M-1 < L. \end{cases} \tag{4.6}$$

To better understand how the matrix $\mathbf{Z}_{(i,j)}$ is formed, consider a system with $N_{sz} = 2$ sound zones, 2 microphones for each zone, and $L = 5$ loudspeakers, so that $\mathbf{Z}$ is a 4×5 matrix which can be written according to the notation introduced above as

$$\mathbf{Z} = \begin{bmatrix} z_{(1,1),1} & z_{(1,1),2} & z_{(1,1),3} & z_{(1,1),4} & z_{(1,1),5} \\ z_{(1,2),1} & z_{(1,2),2} & z_{(1,2),3} & z_{(1,2),4} & z_{(1,2),5} \\ z_{(2,1),1} & z_{(2,1),2} & z_{(2,1),3} & z_{(2,1),4} & z_{(2,1),5} \\ z_{(2,2),1} & z_{(2,2),2} & z_{(2,2),3} & z_{(2,2),4} & z_{(2,2),5} \end{bmatrix}. \tag{4.7}$$

Assume we want to design the filters for control point 1 of the sound zone 1, hence we need to form the matrix $\mathbf{Z}_{(1,1)}$. Then, we need to remove the coefficients related to the control point 2 of the sound zone 1 from the overall channel matrix (4.7), i.e., the row vector

$$\begin{bmatrix} z_{(1,2),1} & z_{(1,2),2} & z_{(1,2),3} & z_{(1,2),4} & z_{(1,2),5} \end{bmatrix}. \tag{4.8}$$

The filter coefficients for the i-th passenger, assumed to be in the bright region, are defined as the superposition of the filters in (4.5), for $j = 1, 2$, as

$$\mathbf{q}_i = \mathbf{q}_{(i,1)}^{opt} + \mathbf{q}_{(i,2)}^{opt}. \tag{4.9}$$

In order to enable a stereo effect, the design of $\mathbf{q}_{(i,1)}^{opt}$ and $\mathbf{q}_{(i,2)}^{opt}$, relative to the control points at the two ears, should be performed separately. In the design of each one, the control point related to the other ear could be taken into account or not. If we take it into account in the design to achieve the stereo effect, we should differentiate the target acoustic pressure at the two control points. However, this can lead to a poor control of the sound field at low frequencies since the two control points in the BZ are close to each other. For this reason, the control point related to the other ear is not considered in the filter design.

Using (4.9), we can express the AC (1.22) (in linear scale) between the bright and dark sound regions as

$$C = \frac{M-2}{2} \frac{\left\| \mathbf{Z}_i \left(\mathbf{q}_{(i,1)}^{opt} + \mathbf{q}_{(i,2)}^{opt} \right) \right\|^2}{\left\| \mathbf{Z}_D \left(\mathbf{q}_{(i,1)}^{opt} + \mathbf{q}_{(i,2)}^{opt} \right) \right\|^2} \tag{4.10}$$

where $M-2$ is the number of all control points outside the BZ, $\mathbf{Z}_i$ is the matrix formed by the coefficients related to the two control points of the i-th passenger, i.e., $z_{(i,1),\ell}$ and $z_{(i,2),\ell}$ for $\ell = 1, ..., L$, and $\mathbf{Z}_D$ is formed by removing the coefficients of the matrix $\mathbf{Z}_i$ from the overall channel matrix $\mathbf{Z}$.

An arbitrary phase rotation of the elements of the target vector $\hat{\mathbf{p}}$ can be described as $\hat{\mathbf{p}} = \hat{\mathbf{P}}\hat{\mathbf{\Phi}}$, in which $\hat{\mathbf{P}} = \mathrm{diag}\left(\hat{\mathbf{p}}\right)$ is an equivalent diagonal matrix

of the target vector and $\hat{\mathbf{\Phi}} = \left[e^{\jmath\hat{\phi}_1}, \ldots, e^{\jmath\hat{\phi}_{M-1}}\right]^T$. Note that, the absence of rotation can be described by $\hat{\mathbf{\Phi}} = \mathbf{1}_{M-1}$ where $\mathbf{1}_{M-1}$ is a column vector populated by $M-1$ ones.

Therefore, applying a phase rotation to the control points of the i-th sound zone, (4.9) can be written as

$$\mathbf{q}_i = \mathbf{A}_{(i,1)}\hat{\mathbf{P}}_{(i,1)}\hat{\mathbf{\Phi}}_{(i,1)} + \mathbf{A}_{(i,2)}\hat{\mathbf{P}}_{(i,2)}\hat{\mathbf{\Phi}}_{(i,2)}. \tag{4.11}$$

In order to simplify the expression, we assume that $\hat{\mathbf{\Phi}}_{(i,1)} = \mathbf{1}_{M-1}$ or $\hat{\mathbf{\Phi}}_{(i,2)} = \mathbf{1}_{M-1}$ and $\hat{\mathbf{P}}_{(i,1)} = \hat{\mathbf{P}}_{(i,2)}$. Then, we can maximize the AC given by (4.10) with respect to a phase difference between the left and right ear targets. Note that if the matrices $\mathbf{A}_{(i,1)}$ and $\mathbf{A}_{(i,2)}$ are different, a phase rotation applied to left or right targets will will result in different performance.

Finally, considering a specific frequency and assuming to set to zero all targets for the control points in the dark zone, it is reasonable to concentrate on phase shift $\hat{\phi}_m \in (-\pi, \pi]$ where the index m indicates the bright control point.

As described above, $\hat{\phi}_m$ is optimized for each frequency of the spectrum. Accordingly, the optimized pressure spectrum could be characterized by a non-constant group delay that may introduce distortions in the reproduced sound. For this reason, in order to reduce the distortions, 1/3-octave band optimization is also considered. In this case, a constant phase shift is set across the 1/3-octave bands.

4.2.1 Numerical results

For the optimization of the target phase, we follow a brute force approach by discretizing the possible phases and searching for for the best value. To this purpose, 100 equally spaced phase values in the range $(-\pi, \pi]$ are used. Moreover, in order to maximize the AC between the driver and codriver positions, the phase for the left control point of the driver is optimized. For the results presented in this section, the trimming operation is not performed. Configu-

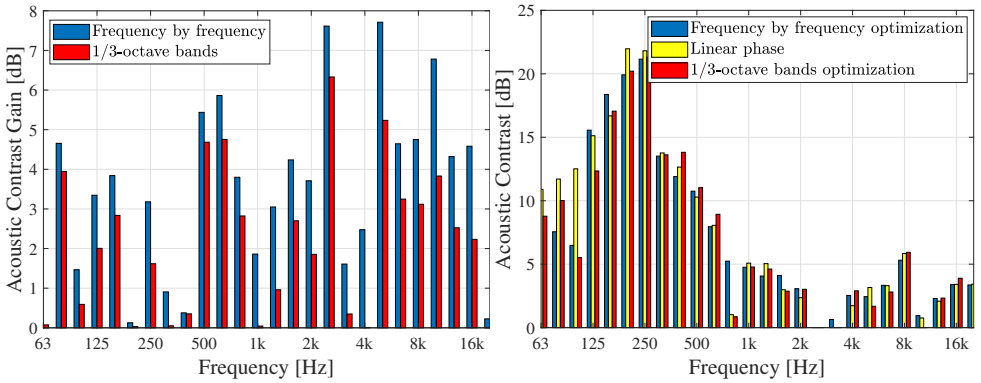


Figure 4.7: (a) AC gain achieved in the ideal case with phase optimized for each frequency of the spectrum and across the 1/3-octave bands. (b) AC achieved in the realistic case with phase optimized for each frequency of the spectrum and across the 1/3-octave bands. For reference, the AC achieved without any phase optimization is reported in (b).

ration A is used in this section for the numerical results, specifically, the same used in Section 3.5.

In this analysis, ideal and realistic cases are considered. Hence, for the ideal case, the performance is evaluated with the same set of channel IRs used for filter design, whereas, for the realistic case, the evaluation is performed with a mismatched set.

The AC gain in dB achieved in the ideal case by the phase target optimization is shown in Fig. 4.7(a). This is obtained as the difference between the AC in dB achieved with and without phase target optimization. The AC evaluated in the realistic case, with phase target optimization and without it, is shown in Fig. 4.7(b). In the ideal case, the phase target optimization performed for each frequency of the spectrum achieves a higher potential AC with respect to the 1/3-octave band optimization. However, in the realistic case, the change of the target has less impact on the reproduced sound. In both cases, spatial effects may influence the reproduced sound, due to the phase difference between the left and right ear targets.

In a realistic condition, the performance suffers a reduction of the AC in some of the 1/3-octave bands, in particular at low frequencies. Some improvement of the AC is achieved between 350 Hz and 560 Hz with the 1/3-octave band optimization. A more relevant gain of the AC is achieved in the 1/3-octave band centered at 794 Hz, with the optimization performed for each point of the spectrum.

Considering the realistic performance, in order to maintain a certain level of AC and reduce the distortion caused by the possibly non-constant group delay, it is possible to limit the target phase optimization to the frequencies which exhibit AC gain.

4.3 Statistical PM

One of the disadvantages of the first formulation of the PM method [3] is that the performance is not robust against perturbations, e.g., errors in the measurement positions with respect to the realistic positions of the listeners. Later in [19], a regularization parameter was introduced to solve the ill-conditioning problem of the matrix to be inverted by the algorithm, limit the energy emitted by the filters and increase the robustness of the performance against system perturbations [4, 20, 21]. With additional pre- and post-processing, we have so far adopted and implemented the latter formulation.

However, with our system configuration of microphones and loudspeakers, even the original formulation of the PM [19] with the regularization parameter does not achieve robust performance. We can observe this fact in Fig. 4.8 where the AC curves evaluated with the same set of measured IRs used to design the filters $\mathbf{Z}_{des}$ and with the measured mismatched IRs $\mathbf{Z}_{mm}$ are plotted for various values of the regularization parameter, which is linearly proportional to β_F used during the filters design stage (see Section 1.2.1). Indeed, with larger regularization parameter, e.g., with $\beta_F = 10^{-4}$, which should achieve more robust performance, the AC decreases in some frequency range of interest, such as around 500 Hz. On the other hand, even a very small regularization

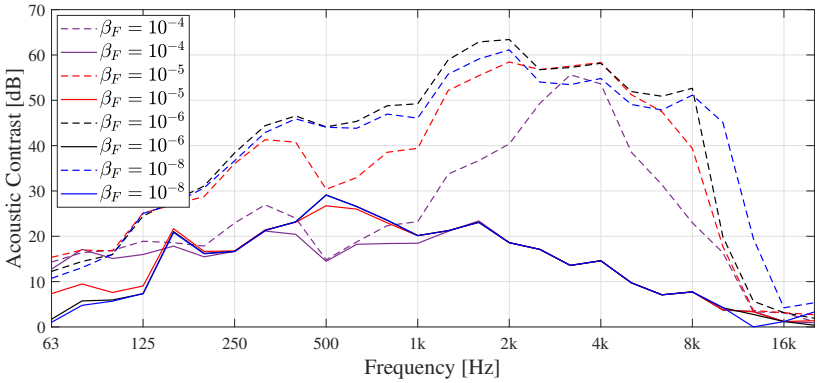


Figure 4.8: AC evaluated with the same set of measured IRs used to design the filters $\mathbf{Z}_{des}$ (dashed lines) and with the measured mismatched IRs $\mathbf{Z}_{mm}$ (solid lines), for various values of the parameter β_F.

parameter does not assure improvement of the realistic AC, in particular at higher frequencies.

In [4] and then in [5], an alternative formulation of the PM was proposed with the aim of controlling the trade-off between directivity performance and reproduction accuracy of the desired sound field, namely the so-called weighted pressure matching (WPM). Indeed, the denomination refers to the weights assigned to the minimization of the different terms in the PM cost function, in particular, the reproduction errors of the different control points located in various positions. We shall discuss this solution in Section 4.3.3.

Later, statistical models of the system were proposed to improve the robustness of the PSZ algorithms against perturbations of the system. A first related work was carried out in [41]. In this work, given the statistics of a diffuse field for an almost ideal scenario accounting for system perturbations, a closed-form solution for the ACC method was derived. However, the considered scenario is too simple compared to ours and, practically, the authors derived an alternative numerical solution for the regularization.

In [69], a variation of the ACC method was optimized taking into account a statistical model, i.e., uniform and normal distributions of the uncertainties

in the FRs of the loudspeakers. Another robust solution of the ACC against uncertainties was proposed in [70], hence, as the previous solutions in [41,69], it is not suitable for our purpose since we are investigating a robust PM solution.

In [22], a framework to improve the robustness of the PSZ methods, by taking into account the statistics of additive and multiplicative measurement errors, was proposed. Similarly to [41], the framework proposed in [22] allows the derivation of a regularization parameter to improve the robustness of the PSZ filters against system perturbations.

Later in [23], a PSZ solution incorporating coarse error estimation was proposed to improve robustness. Here, the regularization is performed with the addition of an error matrix in the optimal solution of the PM instead of the usual regularization parameter β. However, as we saw in Fig. 4.8, the optimization of the regularization parameter alone, proposed in this new solution and the previous one [22], does not assure regularized performance in our setup.

In [24], a PM solution that minimizes the reproduction error averaged with respect to an estimated distribution of the channel matrix was proposed. The distribution of the channel matrix is estimated by the Bayesian inference method from the measured channel matrices. However, the proposed solution is evaluated up to 300 Hz and it seems not able to improve the robustness in case of noise with independent and identically distributed (IID) samples. Hence, it does not guarantee improvement at higher frequencies, where it is reasonable to assume uncorrelated errors between the various loudspeaker-microphone pairs.

In [25], a regularization method based on the solution of [24], where the statistics are extracted by repeated measurements, was proposed. The improved robustness is also achieved by the mismatching method, i.e., by repeating the measurements by removing and then attempting to position again, in the same previous position, the measurement instrument. Note that, we just adopted this measurement method starting from the beginning of this work with the aim of reliable performance evaluation.

4.3.1 Optimization of the average reproduction error

Considering a stochastic model of the acoustic channel matrix $\mathbf{Z}$ with a given distribution, for the fixed discrete frequency $f^{(k)}$, we wish to minimize the average reproduction error

$$J(\mathbf{q}) = \mathbb{E}\left\{\|\mathbf{Z}\mathbf{q} - \hat{\mathbf{p}}\|^2\right\} + \beta\|\mathbf{q}\|^2 \tag{4.12}$$

where $\mathbf{q}$ is the vector of the optimal filter coefficients, $\hat{\mathbf{p}}$ is the vector of the target sound field coefficients and β is the regularization parameter. Note that β is still needed to avoid power allocation at lower and higher frequencies (see Section 1.2.1).

To find the optimal solution of (4.12), it is convenient to expand the expression as

$$J(\mathbf{q}) = \mathbf{q}^H\mathbb{E}\left\{\mathbf{Z}^H\mathbf{Z}\right\}\mathbf{q} - \hat{\mathbf{p}}^H\mathbb{E}\left\{\mathbf{Z}\right\}\mathbf{q} - \mathbf{q}^H\mathbb{E}\left\{\mathbf{Z}^H\right\}\hat{\mathbf{p}} + \|\hat{\mathbf{p}}\|^2 + \beta\|\mathbf{q}\|^2. \tag{4.13}$$

Then, setting to zero the gradient of $J(\mathbf{q})$ with respect to $\mathbf{q}$, the optimal solution is

$$\mathbf{q}_{SPM} = \left(\mathbb{E}\left\{\mathbf{Z}^H\mathbf{Z}\right\} + \beta\mathbf{I}_L\right)^{-1}\mathbb{E}\left\{\mathbf{Z}^H\right\}\hat{\mathbf{p}} \tag{4.14}$$

where $\mathbf{I}_L$ is a $L\times L$ identity matrix and the subscript SPM stands for statistical pressure matching.

By substituting (4.14) into (4.13) the achieved minimum average reproduction error can be expressed as

$$J(\mathbf{q}_{SPM}) = \|\hat{\mathbf{p}}\|^2 - \hat{\mathbf{p}}^H\mathbb{E}\left\{\mathbf{Z}\right\}\left(\mathbb{E}\left\{\mathbf{Z}^H\mathbf{Z}\right\} + \beta\mathbf{I}_L\right)^{-1}\mathbb{E}\left\{\mathbf{Z}^H\right\}\hat{\mathbf{p}}. \tag{4.15}$$

Note that, considering the IID Gaussian model, we can obtain similar results of [24] from (4.14). Indeed, the optimal solution for this distribution becomes

$$\mathbf{q}_{SPM,IID} = \left(\mathbb{E}\left\{\mathbf{Z}^H\right\}\mathbb{E}\left\{\mathbf{Z}\right\} + (\sigma^2 + \beta)\mathbf{I}_L\right)^{-1}\mathbb{E}\left\{\mathbf{Z}^H\right\}\hat{\mathbf{p}} \tag{4.16}$$

where β could be negligible, assuming we properly design or estimate σ, and $\mathbb{E}\left\{\mathbf{Z}\right\} = \mathbf{Z}_{des}$.

Unlike the original formulation of the PM, which exhibits 3 different solutions as a function of the number of microphones and loudspeakers [19], i.e.,

$$\mathbf{q}_{PM} = \begin{cases} \left(\mathbf{Z}^H\mathbf{Z} + \beta\mathbf{I}_L\right)^{-1}\mathbf{Z}^H\hat{\mathbf{p}} & M > L \\ \left(\mathbf{Z} + \beta\mathbf{I}_L\right)^{-1}\hat{\mathbf{p}} & M = L \\ \mathbf{Z}^H\left(\mathbf{Z}\mathbf{Z}^H + \beta\mathbf{I}_M\right)^{-1}\hat{\mathbf{p}} & M < L \end{cases} \tag{4.17}$$

the solution (4.14) does not require a similar diversification since the matrix to be inverted, even without the regularization term, is well-conditioned provided a well-behaved distribution of the channel matrix is considered.[4]

An empirical suboptimal solution can also be obtained by analogy with (4.17), for the corresponding case with $M < L$, and expressed as

$$\mathbf{q}^*_{SPM} = \mathbb{E}\left\{\mathbf{Z}^H\right\}\left(\mathbb{E}\left\{\mathbf{Z}\mathbf{Z}^H\right\} + \beta\mathbf{I}_M\right)^{-1}\hat{\mathbf{p}}. \tag{4.18}$$

This solution is clearly a suboptimal one with respect to (4.14). We denote this solution by the superscript $*$. The fact that (4.18) corresponds to a suboptimal solution can be seen in Fig. 4.9, which will be discussed in more detail in the section about the numerical results, where the cost function (4.12) is compared for $\mathbf{q}_{SPM}$ and $\mathbf{q}^*_{SPM}$ in the ideal case, i.e., the solution and the cost are calculated considering the same 9 sets of the acoustic channels, without regularization parameter, and assuming $\mathbf{Z}$ is uniformly distributed over these sets.

Note that, in (4.12), we consider designing PSZ filters to achieve the same target sound field for all the possible measurement points. Indeed, in [24], the measurements were carried out at approximately the same position, i.e., except for a small mismatch. However, if the displacement of the various measurement positions increases too much, the proposed method still maximizes the average reproduction error but it may increases the reproduction error in the reference position.

[4] As discussed in Section 4.3.3, this condition is verified if enough realizations of the acoustic channel are taken into account.

4.3.2 Empirical distribution

The optimal solution (4.14) requires the knowledge of the distribution of the channel matrix $\mathbf{Z}$ or, at least, its first and second-order moments. The particular solution (4.16), obtained by assuming the IID Gaussian distribution of the channel matrix elements, is one of the simplest cases. However, the assumption of IID coefficients does not fully account for the characteristics of the acoustic channel.

In a realistic scenario, several acoustic channel matrices can be measured and made available. A reasonable probability can be associated with each measured channel matrix, resulting in the definition of a discrete distribution for $\mathbf{Z}$. Let R be the number of available measurements and $\pi_r = \mathrm{P}\{\mathbf{Z} = \mathbf{Z}_r\}$ be the probability of the r-th realization, or measurement, of the audio channel matrix. Then, (4.12) can be written in general as

$$J(\mathbf{q}) = \sum_{r=1}^{R} \|\mathbf{Z}_r\mathbf{q} - \hat{\mathbf{p}}\|^2 \pi_r + \beta\|\mathbf{q}\|^2. \tag{4.19}$$

Assuming that repeated measurements are performed aiming at the same position, i.e., at very close positions, we can assume that the channel matrix has uniform distribution, i.e., $\pi_r = 1/R$ for $r = 1, 2, \ldots, R$. Hence, (4.19) simplifies to

$$J(\mathbf{q}) = \frac{1}{R}\sum_{r=1}^{R} \|\mathbf{Z}_r\mathbf{q} - \hat{\mathbf{p}}\|^2 + \beta\|\mathbf{q}\|^2. \tag{4.20}$$

Assuming again a general distribution specified by the probability π_r assigned to the r-th measured channel matrix $\mathbf{Z}_r$, we can express (4.19) in an alternative compact and elegant form. To this purpose, let

$$\mathfrak{Z} = \mathrm{diag}\,(\mathbf{Z}_1, \ldots, \mathbf{Z}_r, \ldots, \mathbf{Z}_R) \in \mathbb{C}^{MR \times LR} \tag{4.21}$$

be a block diagonal matrix that collects the R measurements,

$$\mathrm{rep}_U(\mathbf{A}) = [\mathbf{A}^T, \ldots, \mathbf{A}^T, \ldots, \mathbf{A}^T]^T \in \mathbb{C}^{D_1 U \times D_2} \tag{4.22}$$

be a vector or matrix populated by U repetitions of the vector or matrix $\mathbf{A} \in \mathbb{C}^{D_1 \times D_2}$ and

$$\mathfrak{P} = \text{diag}\left(\sqrt{\pi_1}\mathbf{I}_M, \ldots, \sqrt{\pi_r}\mathbf{I}_M, \ldots, \sqrt{\pi_R}\mathbf{I}_M\right) \in \mathbb{R}^{MR \times MR}. \tag{4.23}$$

We can express the cost function in the matrix form

$$J(\mathbf{q}) = \|\mathfrak{P}\left[\mathfrak{Z}\ \text{rep}_R(\mathbf{q}) - \text{rep}_R(\hat{\mathbf{p}})\right]\|^2 + \beta\|\mathbf{q}\|^2 \tag{4.24}$$

and the optimal solution can be written as

$$\mathbf{q}_{SPM} = \left[\text{rep}_R(\mathbf{I}_L)^H \mathfrak{Z}^H \mathfrak{P}^2 \mathfrak{Z} \text{rep}_R(\mathbf{I}_L) + \beta\mathbf{I}_L\right]^{-1} \left[\text{rep}_R(\mathbf{I}_L)^H \mathfrak{Z}^H \mathfrak{P}^2 \text{rep}_R(\mathbf{I}_M)\right] \hat{\mathbf{p}}. \tag{4.25}$$

4.3.3 Relation with other solutions in the literature

For a comparison with the proposed SPM method, we report for convenience a general expression of the cost function of the WPM formulated in [4]

$$J_{WPM}(\mathbf{q}) = \lambda_B\|\mathbf{Z}_B\mathbf{q} - \hat{\mathbf{p}}_B\|^2 + \lambda_G\|\mathbf{Z}_G\mathbf{q} - \hat{\mathbf{p}}_G\|^2 + \lambda_D\|\mathbf{Z}_D\mathbf{q} - \hat{\mathbf{p}}_D\|^2 + \beta\|\mathbf{q}\|^2 \tag{4.26}$$

where the subscripts B, G and D denote the bright, grey and dark regions, respectively, and λ_* is the weight attributed to the various terms. We can see some similarity in the expression (4.26) with (4.19). However, (4.26) was formulated with the aim of weighting the minimization of the reproduction errors in the different sound regions, whereas (4.19) weights the reproduction error of the possible realizations of the acoustic channel attributing the same weight to the different sound regions, and with the purpose of minimizing the average error.

On the other hand, the proposed formulation in [24], and later in [25], has more similarities with the PM formulation proposed by us. For convenience, we report this solution, which can be expressed as

$$\mathbf{q}_{prob} = \left(\mathbb{E}\left\{\mathbf{Z}^H\right\}\mathbb{E}\left\{\mathbf{Z}\right\} + \mathbf{\Sigma}\right)^{-1}\mathbb{E}\left\{\mathbf{Z}^H\right\}\hat{\mathbf{p}} \tag{4.27}$$

where $\mathbf{\Sigma} = \mathrm{diag}\,(T_1, \ldots, T_\ell, \ldots, T_L)$ with $T_\ell = \sum_{m=1}^{M} \mathrm{Var}\,\{z_{m,\ell}\}$. Comparing (4.27) with (4.16), we can see that the main difference is in the term to be inverted since in [25] the channel coefficients were considered independent. However, this independence is not true in general. Indeed, at low frequencies the channel coefficients can be correlated over the different microphones. Hence, (4.27) does not fully account for the statistics of the measured acoustic channel matrices. By comparing with (4.16), (4.27) is a particular and simplified solution of (4.14).

Moreover, in (4.27), the ill-conditioning problem is present. Indeed, the matrix $\mathbb{E}\left\{\mathbf{Z}^H\right\}\mathbb{E}\left\{\mathbf{Z}\right\} \in \mathbb{C}^{L\times L}$ is still rank deficient if $M < L$ and, since $\mathrm{Var}\,\{z_{m,\ell}\}$ is proportional to the power transmitted from the ℓ-th loudspeaker to the m-th microphone, the regularization term may not be able to regularize the inversion sufficiently at the frequencies where the system is unable to produce energy, e.g., outside the operational frequency range of the loudspeakers.

On the contrary, assuming enough realizations of the acoustic channel are available, $\mathbb{E}\left\{\mathbf{Z}^H\mathbf{Z}\right\} \in \mathbb{C}^{L\times L}$ becomes a full-rank matrix even in the case of $M < L$ [5]. However, to enable proper adjustment outside the operational frequencies of the loudspeakers, the regularization term in (4.27) is still useful since it performs a different regularization taking into account the properties of each loudspeaker separately. A similar approach was considered in [72], where, however, the regularization parameter was user-defined for each of the diagonal elements of $\mathbf{\Sigma}$.

4.3.4 Implementation of the statistical PM

The solution (4.25) might seem the fastest way to calculate the PSZ filter since, for each discrete frequency $f^{(k)}$, it requires loading the audio channel matrices and concatenating them to build the various matrices required in the solution. However, it is not so straightforward, considering the design of filters

[5] If we have enough realizations such that we can write $\mathbb{E}\left\{\mathbf{Z}^H\mathbf{Z}\right\} = \sum_{r=1}^{L} \mathbf{u}_r\mathbf{v}_r$ with $\mathbf{u}_1, \ldots, \mathbf{u}_L$ and $\mathbf{v}_1, \ldots, \mathbf{v}_L$ being linearly independent, Sylvester's rank inequality [71] assure us that $\mathrm{rank}\left(\mathbb{E}\left\{\mathbf{Z}^H\mathbf{Z}\right\}\right) = L$.

aimed at stereo reproduction. Indeed, as we have done so far, in the design of the PSZ filters for one of the ears, the FRs related to the other ear are removed (see Section 4.2) and this operation may be tricky to be performed on the overall matrix $\mathfrak{Z}$. For this reason, the solution (4.14) was implemented as described hereafter.

Once the target ear for the PSZ filters is selected, the acoustic channel matrices, namely $\mathbf{Z}_{r,full}$, are loaded sequentially. After an acoustic channel matrix is loaded, we derive the matrix $\mathbf{Z}_r$ by removing the FRs related to the opposite ear. Then, for each discrete frequency, the calculation of $\mathbb{E}\left\{\mathbf{Z}^H\mathbf{Z}\right\}$ and $\mathbb{E}\left\{\mathbf{Z}^H\right\}$ is performed by a sample mean, i.e., assuming that $\mathbf{Z}$ is uniformly distributed. At each loaded acoustic channel matrix, the values of $\mathbb{E}\left\{\mathbf{Z}^H\mathbf{Z}\right\}$ and $\mathbb{E}\left\{\mathbf{Z}^H\right\}$ are computed, i.e.,

$$\begin{cases} \mathbb{E}\left\{\mathbf{Z}^H\right\} = \dfrac{1}{R}\displaystyle\sum_{r=1}^{R}\mathbf{Z}_r^H \\ \mathbb{E}\left\{\mathbf{Z}^H\mathbf{Z}\right\} = \dfrac{1}{R}\displaystyle\sum_{r=1}^{R}\mathbf{Z}_r^H\mathbf{Z}_r \end{cases} \tag{4.28}$$

Once this procedure is performed for all the considered acoustic channel matrices, we calculate $\mathbf{q}_{SPM}$ according to (4.14) for each discrete frequency. The procedure is repeated for each sound zone considered, and for each ear in case of stereo audio signal (see Section 4.2).

4.3.5 Numerical results

For the results in this section, Configuration B, described in Section 2.1, is adopted. In particular, within the measurements performed for Configuration B, the ones with the dummy head (DH) at 5 degrees only, excluding the one at 115 mm for symmetry, for a total of 19 measurements, are used.

About the filters used for the analysis, $\mathbf{q}_{SPM}$ and $\mathbf{q}_{SPM}^*$ are designed using 9 acoustic channel matrices (first 3 measurement sessions at the heights 90, 100 and 110 mm), whereas $\mathbf{q}_{PM}$ is designed using the channel matrix measured at the central position (at 100 mm) of the first measurement session. Note

that, for the mismatch evaluation we used the acoustic channel measured in the 5-th measurement session.

The first interesting result is shown in Fig. 4.9. Here, a comparison of the reproduction error computed numerically, using (4.28) by averaging over 9 channel matrices to evaluate (4.12), achieved with $\mathbf{q}_{SPM}$, $\mathbf{q}^*_{SPM}$ and $\mathbf{q}_{PM}$, is depicted. For the proposed method there is no mismatch, whereas, there is a partial mismatch for the original PM. Moreover, we did not use the regularization parameter, i.e., $\beta = 0$, and we assumed that the channel matrices have uniform distribution. As we can see, the proposed method outperforms the original PM formulation (4.17), as well (4.18), in terms of the reproduction error, in the ideal condition.

In Fig. 4.10, a comparison of the reproduction error, evaluated numerically in the mismatch condition, for various DH heights, achieved with $\mathbf{q}_{SPM}$ and $\mathbf{q}_{PM}$ is depicted. Here, the curves labeled with *average* are obtained using (4.12) by averaging over 5 different heights of the DH, i.e., among the acoustic channel matrices measured at 80, 90, 100, 110 and 120 mm. Comparing the curves obtained for the central height, the methods achieve almost the same performance with no relevant differences. However, on average, $\mathbf{q}_{SPM}$

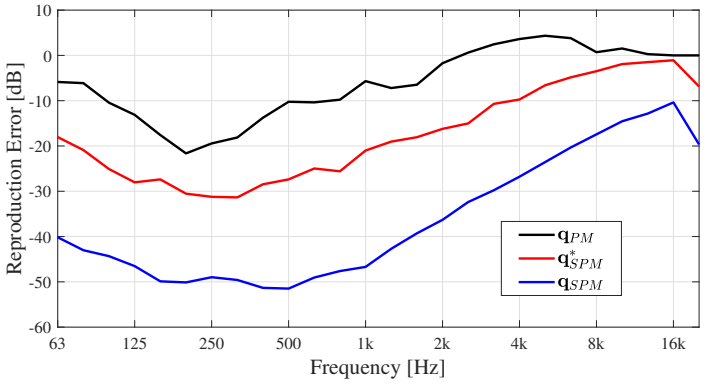


Figure 4.9: Reproduction error, evaluated numerically according to (4.12), by averaging over 9 measured channel matrices, achieved with $\mathbf{q}_{SPM}$ and $\mathbf{q}^*_{SPM}$, calculated with the same 9 channel matrices, and $\mathbf{q}_{PM}$, also calculated using one of the 9 channel matrices. Note that, for error evaluation and filter design, the parameter β is set to zero and it is assumed that $\mathbf{Z}$ has uniform distribution.

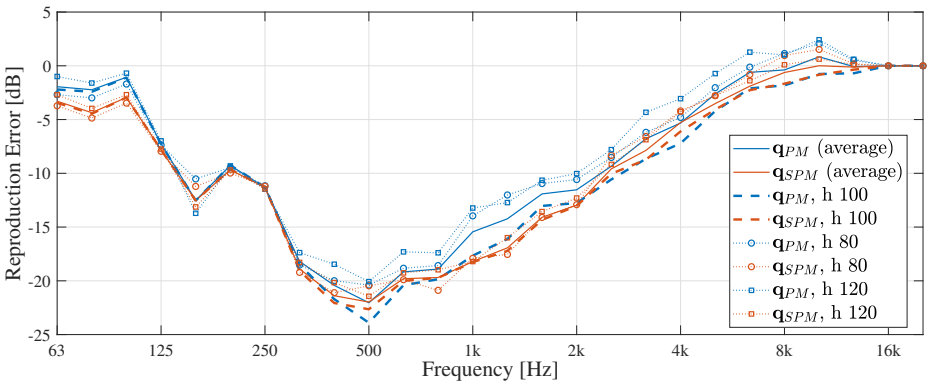
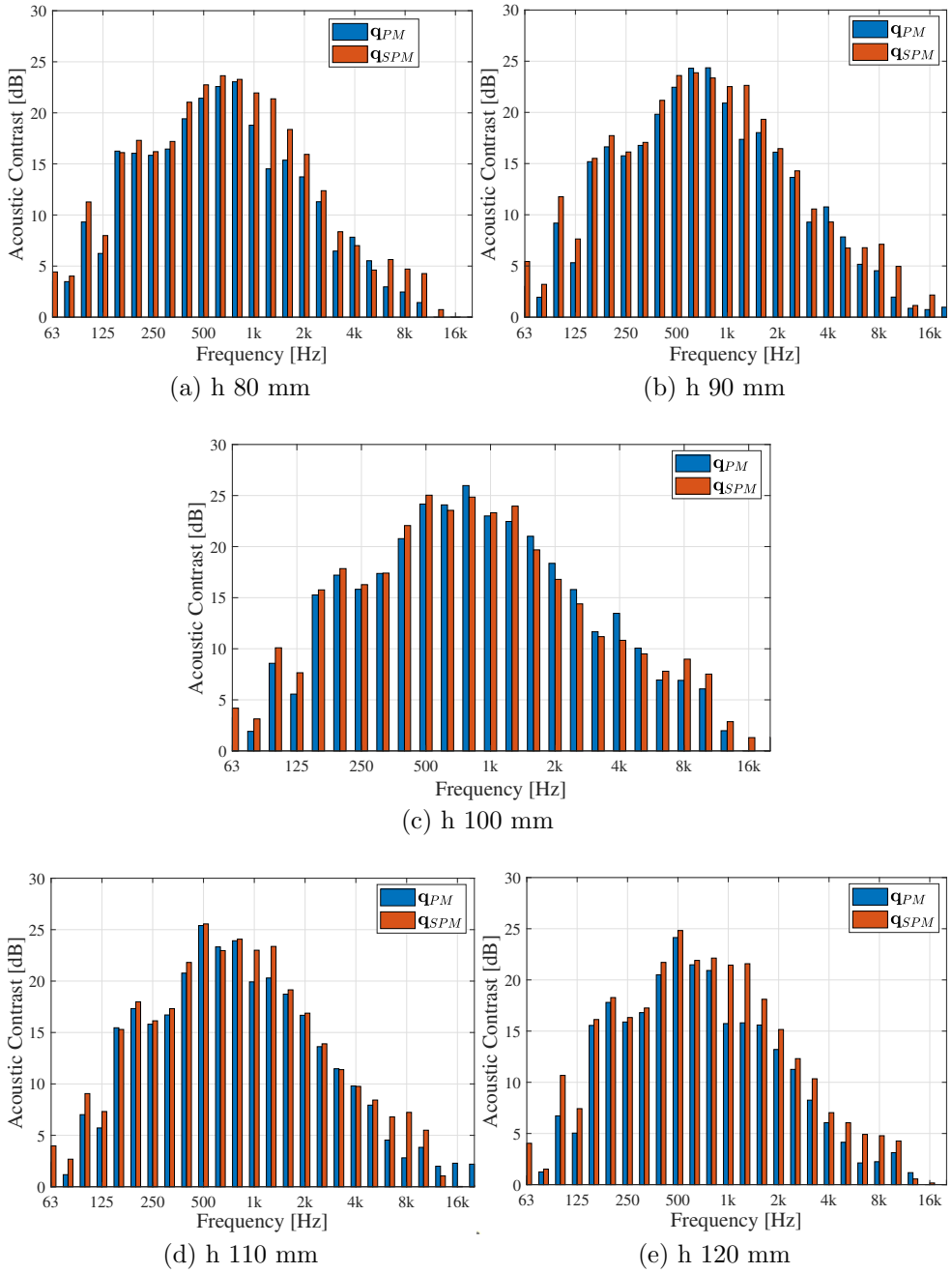


Figure 4.10: Reproduction error, evaluated numerically in the mismatch condition for various heights of the DH and an average value, achieved with $\mathbf{q}_{SPM}$ and $\mathbf{q}_{PM}$. For the average value, the heights 80, 90, 100, 110 and 120 mm are considered.

achieves a lower reproduction error over almost all the frequencies of interest, with gains up to 2.5 dB with respect to $\mathbf{q}_{PM}$. Moreover, considering the evaluations for the other heights, we can see that $\mathbf{q}_{SPM}$ is able to achieve almost the same performance, with small variations, for all the various heights. On the other hand, the reproduction error achieved by $\mathbf{q}_{PM}$ increases as the distance from the central height increases. Overall, the proposed PM method is capable of achieving up to 5 dB lower reproduction error than the original PM at 2 cm away from the central position.

In Fig. 4.11, the AC measured in the vehicle between the driver, assumed to be in the bright zone, and the codriver, assumed to be in the dark zone, for various heights of the DHs are depicted. Note that, in the positioning of the DHs, the same heights in the driver and codriver position were considered. At the central position, the two methods achieve about the same performance in terms of AC, as observed for the reproduction error in Fig. 4.10. The improvement achieved by the proposed method becomes clearly visible as we move the DH away from the central height. Indeed, between 800 Hz and 2 kHz, at heights 80 and 120 mm, the proposed PM is capable of achieving up to 5 dB higher AC than the original PM.



(a) h 80 mm (b) h 90 mm

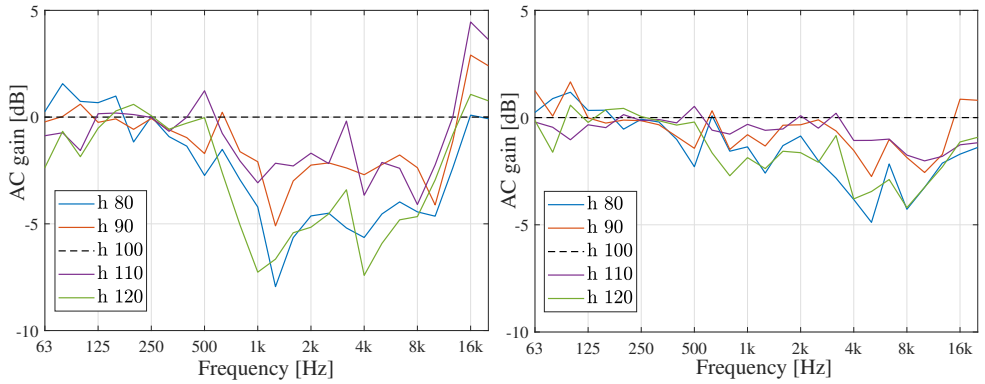
(c) h 100 mm

(d) h 110 mm (e) h 120 mm

Figure 4.11: AC between the driver (bright) and codriver (dark) measured in the vehicle for the proposed and original PM methods, positioning the DHs at heights of (a) 80, (b) 90, (c) 100, (d) 110 and (e) 120 mm.

A reveling way to observe the robustness of the proposed method against the displacement of the head with respect to the central position is presented in Fig. 4.12. Here, the AC gain, defined as the difference between the AC measured at a given height and the AC measured at the central height, i.e., at 100 mm, is depicted for various heights. Indeed, since it expresses a variation of the AC from a reference value, we can clearly see that the proposed PM achieves a lower variation of the AC than the original PM as the distance increases from the central height, in particular between 500 Hz and 4 kHz. Note that, these curves are obtained by the results presented in Fig. 4.11.

For completeness, a comparison of intelligibility and privacy, evaluated numerically in the mismatched condition, between the proposed PM and the original PM is depicted in Fig. 4.13. Here, we can observe that the two methods have almost the same performance. This means that the proposed method is able to reduce the reproduction error without degrading intelligibility or privacy.



(a) Original PM (b) Proposed PM

Figure 4.12: AC gains evaluated at the various heights with respect to the central position at 100 mm of (a) the original and (b) proposed PM methods. The curves are obtained from the results presented in Fig. 4.11.

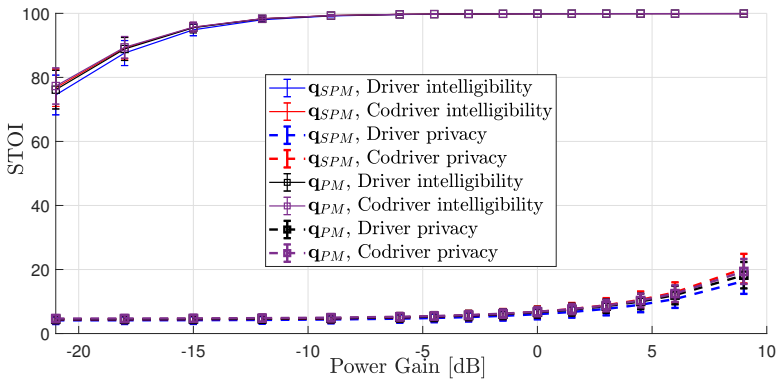


Figure 4.13: Comparison of intelligibility and privacy achieved with the proposed and original PM methods. The evaluation is performed with the acoustic channel matrix measured at the height of 100 mm in the 5-th measurement session. The interfering audio content is rock/pop music.

4.4 Conclusions

We have undertaken a comprehensive examination of the challenges associated with implementing filters designed using the ACC method. Our investigation involves a combination of numerical evaluations and in-vehicle measurements. Specifically, we focus on two key aspects: the gain method required to scale the ACC solutions and the phase equalization needed to attain the desired signal level in the brighter region.

The acoustic assessment of the ACC filters, carried out by simulations and real-world vehicle measurements, sheds light on critical aspects. It becomes apparent that a frequency-dependent gain method is unsuitable for practical implementation, as it introduces high-frequency distortions.

In contrast, our exploration leads to a viable solution: a frequency-independent gain method, complemented by phase equalization, effectively eliminates distortions without compromising AC performance. However, it is worth noting that the resulting sound may not achieve the desired timbre. Nevertheless, the results indicate the possibility of further signal processing to address this issue without sacrificing acoustic contrast.

Stricly related to the PSZ filter design, a phase target design method for AC maximization with the PM algorithm is proposed, focusing on optimizing the phase target for the left control point at the driver position to maximize the AC between driver and codriver zones.

Performance evaluation in a realistic automotive environment using measured IRs is carried out. The analysis includes both an ideal case where passenger ear positions perfectly match the control points used for optimization and a more realistic case accounting for IR mismatches.

The results demonstrate a significant improvement in AC in the ideal case, but the improvement is more limited when IR mismatches are considered, i.e., in a realistic scenario. Furthermore, given the limited number of measurements, the robustness of the proposed algorithm requires further investigation.

Listening tests based on the filters designed with the proposed algorithm, show that the potentially non-constant group delay and the target phase difference between left and right control points may introduce distortions and spatial effects in the reproduced sound. However, these effects have a limited impact on voice messages, such as those used for driver assistance or navigation.

Lastly, based on the idea of using various measurements in the PSZ filter design stage, we propose a novel method aimed at minimizing the average reproduction error on the basis of various acoustic channel matrices obtained from a set of measurements or a known distribution. Unlike previous methods in the literature, our approach is more versatile, as it does not rely on specific assumptions about the acoustic channel matrices. In cases where an IID channel model is assumed, our solution can be related to the latest methods found in the literature.

In our study, we had access to diverse measurement sets and carried out a comprehensive performance analysis and comparison. We used numerical simulations and experimental measurements to evaluate our proposed method against the original formulation of the PM method. Our proposed PM method achieves an average improvement of up to 2.5 dB and up to 5 dB in scenarios

with the largest considered mismatches. Additionally, it yields lower reproduction errors compared to the original PM method.

Furthermore, the proposed PM method demonstrates the potential for up to a 5 dB gain in AC, directly measured in the vehicle, within the frequency range of interest. This gain is observed when the listener position was displaced by up to 2 cm from the original PM-optimized position. Remarkably, our method achieves nearly identical performance in terms of AC, intelligibility, and privacy when compared to the original method at the same optimized position, with no perceived audio distortion.

These results underscore the robustness of the proposed PM method in handling mismatches compared to the original PM formulation. However, further research is needed to carry on experimental performance comparisons with the statistical methods found in the literature. An intriguing avenue for future work could involve analyzing the PSZ filters obtained through our method, where the cost function is optimized using synthesized channel matrices generated by the frequency-correlated (FC) model introduced in the next Chapter 5.

Chapter 5

Stochastic modeling of the measurement mismatch

Although we are able to obtain realistic performance curves by numerical simulations (as described in Chapter 1), both these curves and those measured in the vehicle are just based on a limited number of possible measurements. To achieve a more reliable performance evaluation, we would need to take many more measurements, which may be impractical. For this reason, we propose a stochastic model of mismatched frequency responses (FRs) starting from a measured set of FRs, with the goal of more reliable performance assessment.

5.1 Review of the literature

In [30], an interesting work dealing with the effects of configuration changes in the acoustic channel matrix was presented, in particular for a vehicle interior. However, no numerical model was provided.

A related work appears in [31], where the performance sensitivity of the brightness control method [6] with respect to an error in the measured transfer function caused by various reasons, such as microphone position mismatch, is investigated. The extension to the general personal sound zone (PSZ) method

is presented in [32]. However, since an anechoic environment is considered by the authors, these results are not suitable for our case.

As we discuss in the following, some other works in the literature are aimed at generating acoustic channel responses for PSZ system applications. Specifically, some works are very simple for our purpose, whereas others are not realistic or too complex and very computationally heavy. According to [36], the methods for impulse response (IR) generation can be mainly classified as wave-based, ray-based and statistical. More recently, neural network-based methods, such as [33], were also proposed.

In the wave-based category, we can find the use of the free-space sound propagation model, used in the older references about PSZ systems. This model is easy to implement and not computationally heavy, but it is valid only in anechoic or free-field environments. A more advanced wave-based approach is the finite element method (FEM) [73], which can achieve accurate results. However, the need for the boundary condition definition is not suitable and efficient for a very complex environment such as a vehicle. Moreover, even if we are able to define correctly the boundary conditions for our purpose, FEM is extremely computationally demanding.

Two common ray-based methods are the ray-tracing [34] and image methods [35]. Several works attempted to improve the accuracy and the computational efficiency of the image source method. Indeed, in one of the latest work [74], an open-source Python library, able to exploit the computational capabilities of a Graphic Processing Unit (GPU) and model accurately all the characteristics of a realistic IR, is proposed. However, the image method is limited to regular geometries that are formed by plane surfaces [36], hence, it is not suitable for the vehicle environment.

The ray-tracing method may be applied to irregular geometries [36]. Similarly to its implementation in the video game industry [37] for light propagation, optimized algorithms for acoustic applications were developed, e.g., see [38]. However, an accurate 3-dimensional model of the vehicle interior is still necessary. Moreover, since ray-tracing methods can be approximated by

energy propagation [36], this approximate method may not be recommended for application to the PSZ system, for which knowledge of the signal phases cannot be neglected.

The statistical modeling methods, such as the statistical energy analysis [75], are not suitable for the design and simulation of the soundfield in a virtualized space [76]. However, they are suitable for the prediction of the average behaviors of a system.

Based on this idea, we propose a Gaussian model, characterized by empirical parameters, for the generation of mismatched FRs from measured FRs with the aim of robust performance prediction.

The idea is to have a model that synthesizes acoustic channel matrices, from a few initial ones, and enables a robust numerical evaluation of the performance, in particular the acoustic contrast (AC).

5.2 Methodology

Let $\mathbf{Z}_{D,gen}$ and $\mathbf{Z}_{B,gen}$ be the channel matrices generated by the model associated to the dark and bright sound regions, respectively, and let $\mathbf{Z}_{D,mm_r}$ and $\mathbf{Z}_{B,mm_r}$ be the channel matrices of the r-th mismatched measurement, with respect to the one used to design the vectors of PSZ filter coefficients $\mathbf{q}^{(k)}$, associated with the dark and bright sound regions, respectively. With the goal of predicting the realistic performance, we wish that the AC, averaged over the possible realizations, approximates the average measured AC

$$\mathbb{E}\left\{\frac{\left\|\mathbf{Z}_{B,gen}^{(k)}\mathbf{q}^{(k)}\right\|^2}{\left\|\mathbf{Z}_{D,gen}^{(k)}\mathbf{q}^{(k)}\right\|^2}\right\} \simeq \frac{1}{R}\sum_{r=1}^{R}\frac{\left\|\mathbf{Z}_{B,mm_r}^{(k)}\mathbf{q}^{(k)}\right\|^2}{\left\|\mathbf{Z}_{D,mm_r}^{(k)}\mathbf{q}^{(k)}\right\|^2} \tag{5.1}$$

where $\mathbb{E}\{\cdot\}$ represents the expectation with respect to the possible realizations of the audio channel matrix and the right term corresponds to the arithmetic mean of the AC evaluated with R measured mismatched channel matrices.

5.3 IID model

In the first attempt, an independent and identically distributed (IID) Gaussian model was considered. However, even if the IID model was able to achieve good prediction of the AC, it was not able to reproduce the acoustical effect of a measured mismatched acoustic response, therefore, a frequency-correlated (FC) Gaussian model was studied for our purpose.

Let $\mathbf{Z}_{des}^{(k)} \in \mathbb{C}^{M\times L}$ be the matrix of the measured transfer functions used for PSZ filter design for the k-th discrete frequency $f^{(k)}$, where M and L are the number of the measurement points (microphones) and the number of the loudspeakers of the system, respectively. Let $\mathbf{N}^{(k)} \in \mathbb{C}^{M\times L}$ be a random matrix populated by

$$n_{m,\ell}^{(k)} = \sigma\left(f^{(k)}\right)\left(a_{m,\ell}^{(k)} + \jmath\, b_{m,\ell}^{(k)}\right) \tag{5.2}$$

where m and ℓ are the microphone and loudspeaker indices, respectively, $a_{m,\ell}^{(k)}, b_{m,\ell}^{(k)}$ are normal random variables, i.e., $a_{m,\ell}^{(k)}, b_{m,\ell}^{(k)} \sim \mathcal{N}(0, 1/2)$, $\jmath = \sqrt{-1}$ is the imaginary unit and $\sigma\left(f^{(k)}\right)$ is a function used to assign the variance to $n_{m,\ell}^{(k)}$; indeed, by construction, we can note that $\mathrm{Var}\left\{n_{m,\ell}^{(k)}\right\} = \sigma^2\left(f^{(k)}\right)$. We can now define a stochastic model of a mismatched acoustic channel matrix as

$$\mathbf{Z}_N^{(k)} = \mathbf{Z}_{des}^{(k)} + \mathbf{N}^{(k)}. \tag{5.3}$$

Then, for the k-th discrete frequency $f^{(k)}$, the average AC can be expressed as

$$C^{(k)} = \mathbb{E}\left\{\frac{M_D\left\|\left(\mathbf{Z}_{B,des}^{(k)} + \mathbf{N}_B^{(k)}\right)\mathbf{q}^{(k)}\right\|^2}{M_B\left\|\left(\mathbf{Z}_{D,des}^{(k)} + \mathbf{N}_D^{(k)}\right)\mathbf{q}^{(k)}\right\|^2}\right\} \tag{5.4}$$

where M_D and M_B are the number of microphones in the dark and bright sound zones, respectively.

We can note that as $\sigma\left(f^{(k)}\right)$ goes to zero, the AC (5.4) converges to the ideal value (e.g., see curve *Case 1* in Fig. 1.4); as $\sigma\left(f^{(k)}\right)$ becomes larger, the matrices $\mathbf{N}_B^{(k)}$ and $\mathbf{N}_D^{(k)}$ dominates the numerator and denominator, respectively, and the AC (5.4) approaches one (assuming that $\mathbf{N}_B^{(k)}$ and $\mathbf{N}_D^{(k)}$ have

the same statistics). Basically, the aim of (5.3) is to properly set the statistics of $\mathbf{N}^{(k)}$ to approximate the matrix $\mathbf{E}$ in (1.38).

We wish to properly define the function $\sigma^2\left(f^{(k)}\right)$ in order to achieve (5.1). In the following, we propose a heuristic function based on numerical results. For the k-th discrete frequency $f^{(k)}$, we express the variance of $n_{m,\ell}^{(k)}$ as

$$\sigma^2\left(f^{(k)}\right) = \begin{cases} K\lambda\left(f^{(k)}\right)\left[f^{(k)}/F_h\right]^{\alpha_1} & f^{(k)} < F_{ce} \\ K\lambda\left(f^{(k)}\right)\left[f^{(k)}/F_h\right]^{\alpha_2} & f^{(k)} \geq F_{ce} \end{cases} \tag{5.5}$$

where K, α_1, α_2 are arbitrary parameters to be determined, $\lambda\left(f^{(k)}\right)$ is a function that accounts for the characteristics of the channel matrix $\mathbf{Z}_{des}^{(k)}$, and F_h and F_{ce} are arbitrary frequencies.

The above model can be motivated by the following argument. In the presence of a constant delay caused by the mismatched position, the resulting phase error is proportional to the frequency $f^{(k)}$. Therefore, we can infer that the standard deviation of the modeling error increases linearly with the frequency; then, the variance grows linearly with $[f^{(k)}]^2$. Hence, as a general model that accounts for a possible mismatched position, we assume that the term $[f^{(k)}]^\alpha$, with the proper setting of the exponent α, accounts for the increasing variance of the modeling error as the frequency increases. This model can account for the special case of a simple phase error by setting $\alpha_1 = \alpha_2 = 2$. In order to have more flexibility, a piecewise model with two regions, which are characterized by two different exponents α_1 and α_2, is adopted.

The factor $\lambda\left(f^{(k)}\right)$ accounts for the characteristics of the audio channel. Indeed, there is no need to introduce noise where the system does not operate. More generally, the idea is to allow the model to be accurate for the various loudspeakers, microphones and environment configurations. As we discuss in the following, the case of $\lambda\left(f^{(k)}\right)$ set to a proper constant, i.e., frequency independent, is also considered. In particular, two cases are investigated, the first with

$$\lambda\left(f^{(k)}\right) = \lambda^{max}\left(f^{(k)}\right) \tag{5.6}$$

where $\lambda^{max}\left(f^{(k)}\right)$ is the maximum eigenvalue of the matrices $\mathbf{Z}_{des}^{(k)^H}\mathbf{Z}_{des}^{(k)}$ or $\mathbf{Z}_{des}^{(k)}\mathbf{Z}_{des}^{(k)^H}$[1] (see Section 1.2.1), and the second with

$$\lambda\left(f^{(k)}\right) = \bar{\lambda}^{max} = \frac{1}{9900}\sum_{f^{(k)}=10^2}^{10^4}\lambda^{max}\left(f^{(k)}\right) \tag{5.7}$$

which corresponds to the mean of $\lambda^{max}\left(f^{(k)}\right)$ between 100 Hz and 10 kHz with steps of 1 Hz. Note that, considering the mean value is an arbitrary choice; indeed, the maximum value of $\lambda^{max}\left(f^{(k)}\right)$ in the same frequency range $(100, 10000)$ Hz, namely Λ^{max}, could have been chosen.

The value of F_{ce} in (5.5) sets the frequency at which the exponent changes from α_1 to α_2. The factor $1/F_h$ fixes the value of variance at the frequency F_h for any α_1 and α_2. Note that, to avoid discontinuity in $\sigma^2\left(f^{(k)}\right)$ at the point of change of the exponent from α_1 to α_2, we must set $F_h = F_{ce}$. The parameter K helps to adjust the overall level of the variance.

5.4 Frequency-correlated model

The main problem of the IID Gaussian model (5.2) is the possibly abrupt change of $\mathbf{Z}_N^{(k)}$ between two consecutive frequencies. Indeed, if this happens, a clearly perceivable distortion in the audio, verified by listening tests, is introduced. In particular, a noise similar to white noise is perceivable in the audio generated by using the synthesized mismatched channel matrix (5.3). This means that the model (5.2) is not suitable for evaluating perceptual measures, since it does not correspond to the effect introduced by a physical mismatch of the microphone position. However, as we show in the result section, this model is still suitable for evaluating physical measures, such as AC and reproduction error. This can be justified by the fact that even if the positioning error is assumed to be fixed and equal for all the frequencies, a random position error is assumed in the model (5.2). This effect is neglected by averaging, as it is

[1] Recall that the matrix $\mathbf{Z}_{des}^{(k)}\mathbf{Z}_{des}^{(k)^H}$ has the same nonzero eigenvalues of $\mathbf{Z}_{des}^{(k)^H}\mathbf{Z}_{des}^{(k)}$.

usually done in the measurement of AC and reproduction error, for which their fractional octave band values are analyzed.

We now propose a FC model for which samples of consecutive frequencies are correlated. For the m-th and ℓ-th pair, the correlated sample for the k-th frequency can be expressed as

$$x_{m,l}^{(k)} = \rho_{m,\ell}^{(k)} x_{m,\ell}^{(k-1)} + \sqrt{1 - \left|\rho_{m,\ell}^{(k)}\right|^2} \left(a_{m,\ell}^{(k)} + \jmath\, b_{m,\ell}^{(k)}\right) \tag{5.8}$$

where $a_{m,\ell}^{(k)}, b_{m,\ell}^{(k)} \sim \mathcal{N}(0, 1/2)$ and $\rho_{m,l}^{(k)}$ is the complex correlation coefficient between the k-th and $(k-1)$-th samples. Note that, by construction, $x_{m,\ell}^{(k)}$ obtained by (5.8) is still zero mean and unit variance, hence, the synthesized mismatched acoustic channel matrix can be expressed as

$$\mathbf{Z}_C^{(k)} = \mathbf{Z}_{des}^{(k)} + \mathbf{C}^{(k)} \tag{5.9}$$

where $\mathbf{C}^{(k)}$ is populated by

$$c_{m,\ell}^{(k)} = \sigma\left(f^{(k)}\right) x_{m,\ell}^{(k)} \tag{5.10}$$

so that the variance of the samples $c_{m,\ell}^{(k)}$ is still $\sigma^2\left(f^{(k)}\right)$ as for the IID case.

We propose this model to achieve a realistic sound effect using the synthesized mismatched channel matrix. Listening tests confirm our success, as the additional noise introduced by the IID model (5.3) is eliminated. With proper adjustment of $\sigma\left(f^{(k)}\right)$, the sound generated with $\mathbf{Z}_C^{(k)}$ closely resembles that of a measured mismatched channel matrix.

5.5 Numerical results

Matrices synthesized from (5.3) and (5.9) represent possible realizations of the generated stochastic mismatched channel. The proposed models aim to provide a statistical measure of the PSZ system performance. The AC results are presented as the mean of AC curves from various realizations. For the short-time objective intelligibility (STOI), the average over the 15 speech tracks

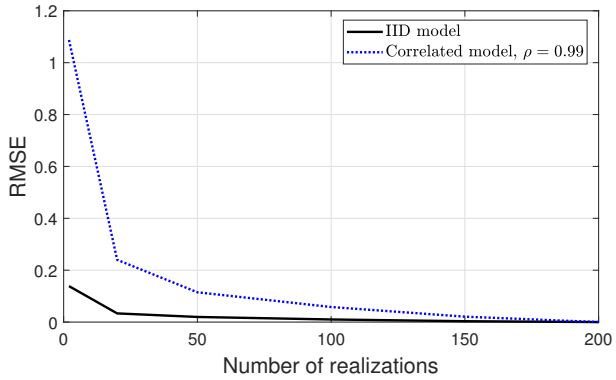


Figure 5.1: Convergence of the AC performance with the number of realizations considered in the average.

(see Section 2.3) is assumed sufficient for performance convergence. Different speech tracks use distinct realizations of the synthesized channel matrix for signal propagation.

In Fig. 5.1, the root mean square error between the AC value, obtained with a given number of channel realizations, and the AC curve obtained with 200 channel realizations, assumed sufficient for convergence, is shown. The IID model (5.3) shows rapid AC convergence due to its frequency domain variability, whereas the FC model (5.9) requires more realizations for AC convergence. Nevertheless, as illustrated in Fig. 5.2, when both models have the same parameters and an adequate number of realizations, their AC converges to the same value. This agreement allows us to draw similar conclusions about the AC prediction for both IID and FC models.

As a trade-off between the reliability of the results and execution time[2], we consider the evaluation with 50 IID and FC realizations, whereas, as a realistic reference curve for AC, we considered the mean between the AC curves obtained numerically with all the measured mismatched channel matrices.

Note that the curve evaluated with $\mathbf{Z}_{des}^{(k)}$ is shown only in Fig. 5.3. Compared with this curve, we can clearly observe the reduction of the AC in the

[2]Time required to generate all the realizations of the channel matrix and perform the evaluation of the AC for each realization.

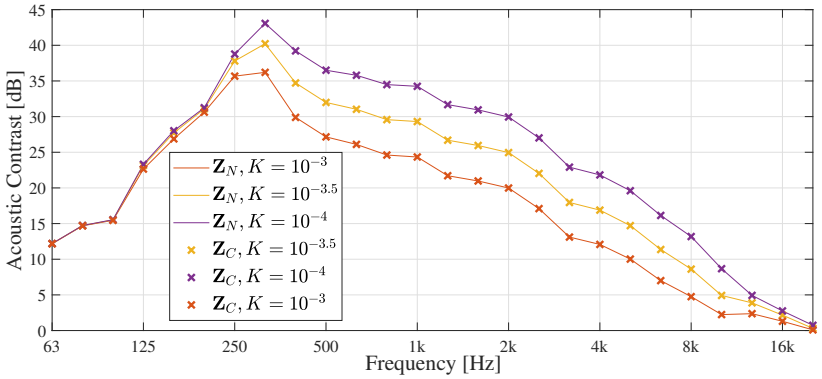


Figure 5.2: AC predicted by the IID model (labeled as $\mathbf{Z}_N$) and the FC model (labeled as $\mathbf{Z}_C$) with $\rho_{m,\ell}^{(k)} = 0.99$, both models with $\lambda\left(f^{(k)}\right) = \lambda^{max}\left(f^{(k)}\right)$, for various values of the parameter K and 200 realizations. Furthermore, $\alpha_1 = 2$, $F_h = 855$ Hz and $F_{ce} = F_s/2$.

other curves caused by the introduced noise, in particular at high frequencies.

In general, we can observe that the AC prediction curves get closer to the ideal one, as K becomes smaller, and 0, otherwise. Indeed, as K goes to zero, the matrix $\mathbf{Z}_N^{(k)}$, approaches $\mathbf{Z}_{des}^{(k)}$. As K increases, the introduced term $\mathbf{N}^{(k)}$ dominates in the numerator and denominator of (5.4) and, since they have elements with the same variance, the mean value of $C^{(k)}$ approaches 1 (i.e., 0 dB).

In Fig. 5.3, we can see that this choice of parameters for $\sigma\left(f^{(k)}\right)$ does not allow to fit the realistic curve of the AC. In particular, considering a fixed value of the parameter K and $\alpha_1 = 0$, between 250 Hz and 1 kHz, the model underestimates the AC, whereas, it overestimates the AC between 2 kHz and 4 kHz. For $\alpha_1 = 2$, the model for $K = 10^{-3}$ seems to better fit the realistic performance in the range $[1, 6]$ kHz, but it underestimates the AC between 500 Hz and 1 kHz and greatly overestimates it for the frequencies below 500 Hz.

In Fig. 5.4 the AC with the same settings of Fig. 5.3 is shown, except for $\lambda\left(f^{(k)}\right) = \lambda^{max}\left(f^{(k)}\right)$. The curves for $\alpha_1 = 0$ (Part a), still do not fit

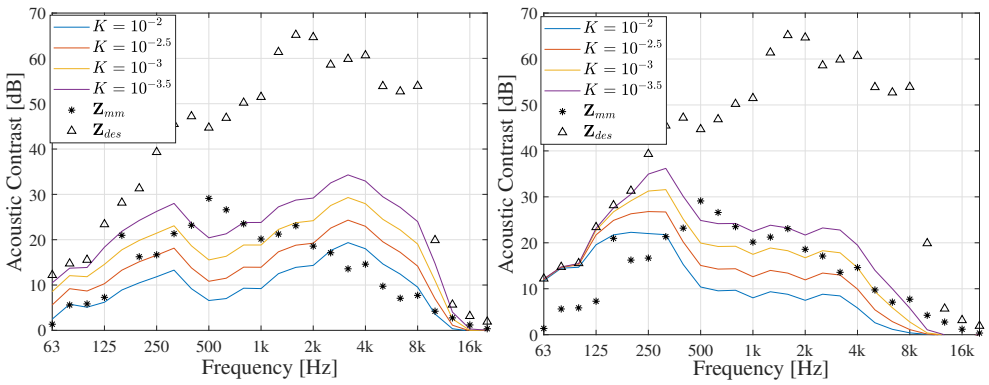


Figure 5.3: AC predicted with the Gaussian model, averaged over 50 realizations, (a) $\alpha_1 = 0$ and (b) $\alpha_1 = 2$, $\lambda\left(f^{(k)}\right) = \bar{\lambda}^{max}$, for various values of the parameter K, $F_h = 855$ Hz and $F_{ce} = F_s/2$. For comparison, the AC evaluated with the measured mismatched channel matrix, i.e., $\mathbf{Z}_{mm}^{(k)}$, and with the same channel matrix used for filters design, i.e., $\mathbf{Z}_{des}^{(k)}$, are shown.

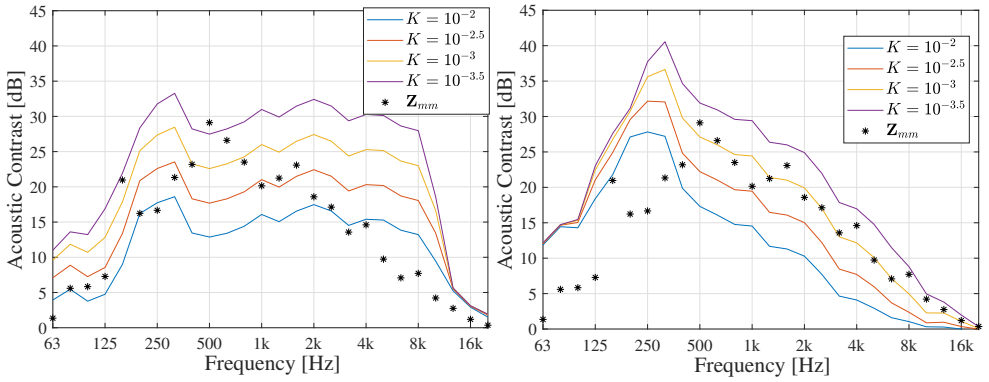


Figure 5.4: AC predicted with the Gaussian model, averaged over 50 realizations, (a) $\alpha_1 = 0$ and (b) $\alpha_1 = 2$, $\lambda\left(f^{(k)}\right) = \lambda^{max}\left(f^{(k)}\right)$, for various values of the parameter K, $F_h = 855$ Hz and $F_{ce} = F_s/2$. For comparison, the AC evaluated with the measured mismatched channel matrix, i.e., $\mathbf{Z}_{mm}^{(k)}$, is shown.

the measured AC, however, for $\alpha_1 = 2$ (Part b), the curve for $K = 10^{-3}$ predicts quite well the behavior of the realistic AC curve, obtained with the mismatched measured responses, above about 500 Hz.

We can now empirically optimize the variance parameters in (5.5) to match the realistic AC curve. The empirically optimized variance is

$$\sigma^2\left(f^{(k)}\right) = \begin{cases} \dfrac{\bar{\lambda}^{max}}{10^{2.25}}\left[\dfrac{f^{(k)}}{428}\right]^0 & f^{(k)} < 428 \text{ Hz} \\ \dfrac{\lambda^{max}\left(f^{(k)}\right)}{10^{3.75}}\left[\dfrac{f^{(k)}}{428}\right]^2 & f^{(k)} \geq 428 \text{ Hz} \end{cases} \tag{5.11}$$

It is also essential to consider perceptual evaluations, which will lead us to shift our focus to the FC model. Listening tests, as explained in Chapter 2, are performed using synthesized acoustic channel matrices. With the IID model, background noise, similar to white noise, is evident in the background but does not interfere perceptually with intelligibility, as seen in Fig. 5.5. However, at the extreme power gain values, intelligibility and privacy are overestimated.

In contrast, the FC model (5.9), with or without the optimized standard deviation, produces audio without background noise or distortions. Perceptually, it closely resembles audio generated with measured mismatched channel

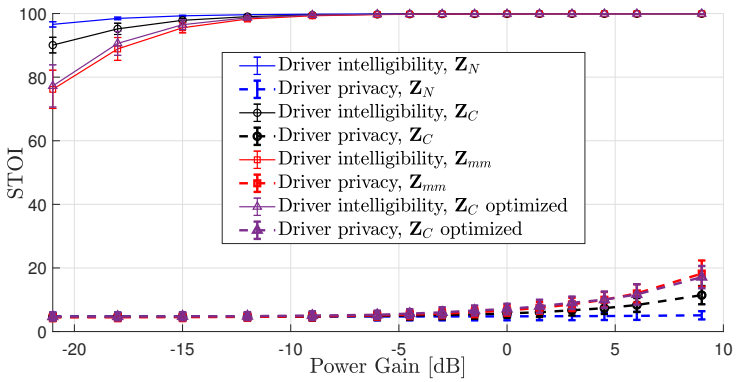


Figure 5.5: Intelligibility and privacy evaluated with the IID model and the FC model with $\rho_{m,\ell}^{(k)} = 0.99$, both models with $\lambda\left(f^{(k)}\right) = \lambda^{max}\left(f^{(k)}\right)$, $K = 10^{-3}$, $\alpha_1 = 2$, $F_h = 855$ Hz and $F_{ce} = F_s/2$. The considered interfering audio content is rock/pop music. For comparison, the intelligibility and privacy evaluated with the measured mismatched channel matrix, and the ones predicted by the FC model with the optimized parameters (5.11), are depicted.

matrices in the vehicle, as shown in the intelligibility and privacy curves in Fig. 5.5.

As mentioned earlier, an accurate AC prediction is achieved by utilizing the FC model with the same empirically optimized variance (5.11) but considering more realizations in the averaging process, as shown in Fig. 5.6. In this figure, a comparison of the AC curves obtained numerically from the measurements, with the proposed model with empirically optimized variance (5.11), with statistics estimated from the measurements and measured directly in the vehicle, is reported.

Fig. 5.7 presents a comparison of the empirical standard deviation (5.11) in the proposed method, used for every microphone-loudspeaker pair, and the standard deviations estimated for few significant pairs of microphone-loudspeaker acoustic response mismatches (1.38). In particular, those between the left driver microphone and two loudspeakers, i.e., the front left woofer and

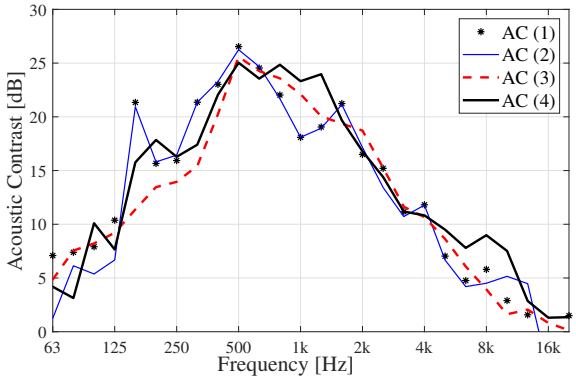


Figure 5.6: Comparison of AC predicted numerically by using the measured mismatched channel matrices (1), the FC model with mean and variance extracted from measurements (2) and empirically optimized (3). For instance, the FC model adopted $\rho_{m,\ell}^{(k)} = 0.99$, zero mean, empirically optimized variance (5.11) and 200 realization for averaging operation. For reference, the AC measured directly in the vehicle (4) is reported.

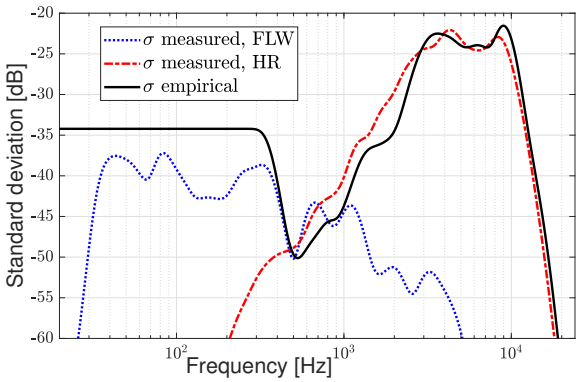


Figure 5.7: Empirically defined standard deviation (5.11) and standard deviation estimated from the mismatched measurements, i.e., the statistics of (1.38) for the front left woofer (FLW) and the headrest speaker (HR). Note that the empirical curve is valid for every microphone-loudspeaker pair and the estimated ones are referred to the microphone placed at the left ear of the driver.

the closest headrest speaker are considered. These loudspeakers were chosen for analysis because they contribute most to the generation and control of the sound field at the left microphone of the driver position. It can be seen that the empirical curve (5.11) approximates the envelope of the two estimated ones quite well.

5.6 Conclusions

A stochastic model is proposed for generating acoustic channel matrices from initially measured data, with the aim of predicting the realistic performance of a PSZ system. Specifically, this model utilizes IID and FC Gaussian models with zero mean and an empirically defined variance function to populate a mismatch matrix added to the measured data.

Performance comparisons, considering AC, STOI, and listening tests, were carried out for the proposed model. These comparisons were based on direct

measurements in the vehicle and numerical simulations using additional channel matrices measured in the cabin of the vehicle, assumed to be realistic. In the realistic automotive scenario under consideration, with the refinement of model parameters, the results indicate that the IID model may be suitable for a quick and accurate prediction of AC. However, only the FC model, which also demonstrates this capability, enables an accurate prediction of the perceptual metric and a realistic reproduction of the acoustic channel, whereas, the IID model generates an unwanted background noise.

The drawback of the FC model is that it requires a larger number of realizations for performance convergence. In evaluating the IID and FC models, STOI has played a crucial role in highlighting the advantages of the correlated model. Without considering this metric, the sole evaluation of AC would not demonstrate the improvements offered by the FC model, and the listening experience would be less indicative of its superiority over the IID model.

Ultimately, the FC model allows for accurate numerical predictions of realistic performance in terms of both physical and perceptual metrics without the need for numerous repeated measurements.

General conclusions

Various aspects of a personal sound zone (PSZ) system are studied in this thesis: from the processing of the acoustic channel measurements to the performance evaluation stage.

In the pursuit of enhancing PSZ systems, several advancements and investigations have been undertaken. The research begins with the development of an algorithm based on the concept of frequency-dependent trimming (FDT) and an exhaustive search for optimal trimming lengths. This algorithm seeks to maximize the acoustic contrast (AC) through the numerical evaluation of two control points, resulting in a reduction of pre-ringing in the impulse response (IR) of the designed filters. However, it is important to note that there is no significant improvement in AC, and the sound timbre falls short of expectations when compared to an empirical solution.

Additionally, three algorithms are introduced to assist in selecting appropriate windowing lengths for FDT operation. Among various windowing functions, the Tukey function with a small parameter is found to be the most suitable choice due to its ability to reduce the radius of gyration of the energy density (RoGoED) with minimal loss of AC.

Among the various proposed algorithms, the frequency-proportional trimming method stands out as the most efficient choice for RoGoED factor reduction, AC preservation, simplicity of implementation, and execution time efficiency, whereas, other methods, such as cross-correlation-based and AC maximization-based approaches, demand significantly longer execution times.

The study further explores challenges associated with implementing filters designed using the acoustic contrast control (ACC) method. It is discovered that a frequency-dependent gain method introduces high-frequency distortions, prompting the introduction of a frequency-independent gain method complemented by phase equalization, which effectively eliminates distortions without sacrificing AC.

Furthermore, a phase target design method for AC maximization is proposed, focusing on optimizing the phase target for control points to maximize AC in a realistic automotive environment. While significant improvements in AC are observed in ideal cases, the performance is more limited in scenarios accounting for IR mismatches. Listening tests reveal potential non-constant group delay and spatial effects in the reproduced sound but with a limited impact on voice messages.

The research also introduces a novel method, namely statistical pressure matching (SPM), for minimizing the average reproduction error based on various acoustic channel matrices obtained from measurements or a given distribution, with results indicating improved AC and intelligibility outperforming the original pressure matching (PM).

Additionally, a stochastic model is proposed for generating acoustic channel matrices, highlighting the superiority of frequency-correlated (FC) model over independent and identically distributed (IID) model, especially in realistic automotive environments, and the potential in predicting and achieving realistic and robust performance prediction without the need of several measurements.

List of publications

The list of publications based on this thesis work at time of preparation of this final version is here reported.

(c1) A. Borroni, M. Martalò, A. Costalunga, C. Tripodi, and R. Raheli, "A Stochastic Model of the Acoustic Response inside the Cabin of an Automobile", in *2023 AEIT Intern. Conf. on Electrical and Electronic Technologies for Automotive* (AEIT AUTOMOTIVE), Modena, Italy, July 2023, pp. 1-6, doi:10.23919/AEITAUTOMOTIVE58986.2023.10217256

(c2) A. Borroni, M. Martalò, A. Costalunga, C. Tripodi, and R. Raheli, "Pressure Matching with Optimized Target Phase for Personal Sound Zone System", in *Proc. 2022 Intern. Conf. on Electrical, Computer, Communications and Mechatronics Engineering* (ICECCME), Maldives, November 2022, pp. 1-6, doi:10.1109/ICECCME55909.2022.9988138

(c3) A. Borroni, M. Martalò, A. Costalunga, C. Tripodi, and R. Raheli, "Experimental Analysis of Individual Listening Zone Algorithms in Controlled Environments", in *2021 Immersive and 3D Audio: from Architecture to Automotive* (I3DA), Bologna, Italy, September 2021, pp. 1-7, doi:10.1109/I3DA48870.2021.9610882

(j1) A. Borroni, M. Martalò, A. Costalunga, C. Tripodi, and R. Raheli, "Stochastic Models of the Acoustic Response in a Car Cabin", submitted

for journal publication, 2024

(j2) A. Borroni, M. Martalò, A. Costalunga, C. Tripodi, and R. Raheli, "Statistical Pressure Matching for Robustness Enhancement of Personal Sound Zone Systems", in preparation for journal publication, 2024

Oral Presentations

(o1) A. Borroni, M. Martalò, A. Costalunga, C. Tripodi, and R. Raheli, "Stochastic Models of the Acoustic Response in a Car Cabin", Annual Meeting of GTTI (oral presentation), Roma, September 2023.

In (c2) we presented the method and results shown in Section 4.2 of this thesis; the works presented partially in (c1), during the event (o1) and in article (j1), to be submitted, are based on the contents of Chapter 5; the article (j2), in preparation, presents the method and results of Section 4.3.

Bibliography

[1] J. Ahrens, R. Rabenstein, S. Spors, The theory of wave field synthesis revisited, in: Proc. Audio Eng. Soc. Conv. 124, Amsterdam, The Netherlands, 2008, pp. 1–19, url:aes.org/e-lib/browse.cfm?elib=14488.

[2] M. A. Poletti, Three-dimensional surround sound systems based on spherical harmonics, J. Audio Eng. Soc. 53 (11) (2005) 1004–1025, url:semanticscholar.org/CorpusID:15875606.

[3] O. Kirkeby, P. A. Nelson, Reproduction of plane wave sound fields, J. Acoust. Soc. Amer. 94 (5) (1993) 2992–3000. doi:10.1121/1.407330.

[4] M. Shin, F. M. Fazi, P. A. Nelson, F. C. Hirono, Controlled sound field with a dual layer loudspeaker array, J. Sound and Vibration 333 (16) (2014) 3794–3817. doi:10.1016/j.jsv.2014.03.025.

[5] F. Olivieri, F. M. Fazi, S. Fontana, D. Menzies, P. A. Nelson, Generation of private sound with a circular loudspeaker array and the weighted pressure matching method, IEEE/ACM Trans. Audio, Speech, and Language Processing 25 (8) (2017) 1579–1591. doi:10.1109/TASLP.2017.2700945.

[6] J.-W. Choi, Y.-H. Kim, Generation of an acoustically bright zone with an illuminated region using multiple sources, J. Acoust. Soc. Amer. 111 (4) (2002) 1695–1700. doi:10.1121/1.1456926.

[7] B. D. Van Veen, K. M. Buckley, Beamforming: A versatile approach to spatial filtering, IEEE Acoustic Speach Signal Proc. Mag. 5 (2) (1988)

4–24. doi:10.1109/53.665.

[8] M. Shin, S. Q. Lee, F. M. Fazi, P. A. Nelson, D. Kim, S. Wang, K. H. Park, J. Seo, Maximization of acoustic energy difference between two spaces, J. Acoust. Soc. America 128 (1) (2010) 121–131. doi:10.1121/1.3438479.

[9] N. de Koeijer, Sound zones with a cost function based on human hearing, Master's thesis, Delft University of Technology, url:repository.tudelft.nl (September 2021).

[10] H. Purwins, B. Li, T. Virtanen, J. Schlüter, S.-Y. Chang, T. Sainath, Deep learning for audio signal processing, IEEE J. Selected Topics in Signal Processing 13 (2) (2019) 206–219. doi:10.1109/JSTSP.2019.2908700.

[11] L. Comanducci, F. Antonacci, A. Sarti, A deep learning-based pressure matching approach to soundfield synthesis, in: Proc. 2022 Intern. Workshop Acoustic Signal Enhancement (IWAENC), Bamberg, Germany, 2022, pp. 1–5. doi:10.1109/IWAENC53105.2022.9914712.

[12] X. Hong, B. Du, S. Yang, M. Lei, X. Zeng, End-to-end sound field reproduction based on deep learning, J. Acoust. Soc. America 153 (5) (2023) 3055–3064. doi:10.1121/10.0019575.

[13] H. Sallandt, P. Krah, M. Lemke, Supervised learning for multi zone sound field reproduction under harsh environmental conditions, arXiv preprint (2021). doi:10.48550/arXiv.2112.07349.

[14] T. Betlehem, W. Zhang, M. A. Poletti, T. D. Abhayapala, Personal sound zones: Delivering interface-free audio to multiple listeners, IEEE Signal Process. Mag. 32 (2) (2015) 81–91. doi:10.1109/MSP.2014.2360707.

[15] M. Ebri, N. Strozzi, F. M. Fazi, A. Farina, L. Cattani, Individual listening zone with frequency-dependent trim of measured impulse responses, in: Audio Eng. Soc. Conv. 149, New York, USA, 2020, url:aes.org/e-lib/browse.cfm?elib=20946.

[16] X. Liao, J. Cheer, S. Elliott, S. Zheng, Design of a loudspeaker array for personal audio in a car cabin, Audio Eng. Soc. 65 (3) (2017) 226–238. doi:10.17743/jaes.2016.0065.

[17] W. Gallian, F. M. Fazi, C. Tripodi, N. Strozzi, A. Costalunga, Optimisation of the target sound fields for the generation of independent listening zones in a reverberant environment, in: 2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA), Bologna, Italy, 2021, pp. 1–10. doi:10.1109/I3DA48870.2021.9610888.

[18] J. Blauert, Spatial hearing: The psychophysics of human sound localization, MIT press, 1997. doi:10.7551/mitpress/6391.001.0001.

[19] O. Kirkeby, P. A. Nelson, F. Orduna-Bustamante, H. Hamada, Local sound field reproduction using digital signal processing, J. Acoust. Soc. Amer. 100 (3) (1996) 1584–1593. doi:10.1121/1.416060.

[20] P. Coleman, P. Jackson, M. Olik, M. Olsen, M. Møller, J. Pedersen, The influence of regularization on anechoic performance and robustness of sound zone methods, J. Acoust. Soc. America 133 (5) (2013) 3344–3344. doi:10.1121/1.4805658.

[21] P. Coleman, P. Jackson, M. Olik, M. Møller, M. Olsen, J. Pedersen, Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array, J. Acoust. Soc. America 135 (4) (2014) 1929–1940. doi:10.1121/1.4866442.

[22] Q. Zhu, P. Coleman, M. Wu, J. Yang, Robust acoustic contrast control with reduced in-situ measurement by acoustic modeling, J. audio Eng. Soc. 65 (6) (2017) 460–473, url:aes.org/e-lib/browse.cfm?elib=18779.

[23] Q. Zhu, P. Coleman, M. Wu, J. Yang, Robust reproduction of sound zones with local sound orientation, J. Acoust. Soc. America 142 (1) (2017) EL118–EL122. doi:10.1121/2F1.4994685.

[24] M. B. Møller, J. K. Nielsen, E. Fernandez-Grande, S. K. Olesen, On the influence of transfer function noise on sound zone control in a room,

IEEE/ACM Trans. Audio, Speech, and Language Process. 27 (9) (2019) 1405–1418. doi:10.1109/TASLP.2019.2921151.

[25] Y. Qiao, E. Choueiri, The performance of a personal sound zone system with generic and individualized binaural room transfer functions, in: Audio Eng. Soc. Conv. 152, The Hague, Netherlands, 2022, url: aes.org/e-lib/browse.cfm?elib=21692.

[26] F. Olivieri, F. Fazi, M. Shin, P. Nelson, Pressure-matching beamforming method for loudspeaker arrays with frequency dependent selection of control points, in: Audio Eng. Soc. Conv. 138, Vol. 2, Warsaw, Poland, 2015, pp. 890–899, url:aes.org/e-lib/browse.cfm?elib=17746.

[27] N. Yanagidate, J. Cheer, S. Elliott, T. Toi, Car cabin personal audio: Acoustic contrast with limited sound differences, in: Audio Eng. Soc. Conference: 55th Intern. Conference: Spatial Audio, Audio Engineering Society, Helsinki, Finland, 2014, url:aes.org/e-lib/browse.cfm?elib=17354.

[28] J.-W. Choi, J. S. Hong, Y. H. Kim, Generation of personal sound zones in the interior and exterior of automotive vehicles, in: Proc. Inter-Noise 2015, International Institute of Noise Control Engineering, San Francisco, California, USA, 2015.

[29] L. Vindrola, M. Melon, J.-C. Chamard, B. Gazengel, Use of the filtered-x least-mean-squares algorithm to adapt personal sound zones in a car cabin, J. Acoust. Soc. America 150 (3) (2021) 1779–1793. doi:10.1121/10.0005875.

[30] W.-H. Cho, J. Chang, S.-H. Shin, Effect of configuration changes on the acoustic transfer function of a vehicle interior, Appl. Acoust. 193 (2022) 108759.

[31] J.-Y. Park, M.-H. Song, J. Chang, Y.-H. Kim, Performance degradation due to transfer function errors in acoustic brightness and contrast control: sensitivity analysis, in: Proc. 20th Intern. Congr. Acoustics, (ICA) 2010, Sydney, Australia, 2010, pp. 1–7.

[32] J.-Y. Park, J.-W. Choi, Y.-H. Kim, Acoustic contrast sensitivity to trans-

fer function errors in the design of a personal audio system, J. Acoust. Soc. America 134 (2013) EL112–8. doi:10.1121/1.4809778.

[33] A. Ratnarajah, S.-X. Zhang, M. Yu, Z. Tang, D. Manocha, D. Yu, Fast-rir: Fast neural diffuse room impulse response generator, in: Proc. IEEE Intern. Conf. Acoustics, Speech and Signal Process. (ICASSP), Singapore, 2022, pp. 571–575. doi:10.48550/arXiv.2110.04057.

[34] A. Kulowski, Algorithmic representation of the ray tracing technique, Appl. Acoust. 18 (1985) 449–469. doi:10.1016/0003-682X(85)90024-6.

[35] J. B. Allen, D. A. Berkley, Image method for efficiently simulating small-room acoustics, J. Acoust. Soc. Amer. 65 (4) (1979) 943–950. doi:10.1121/1.382599.

[36] E. A. Habets, *Room impulse response generator*, Tech. rep. 2.4 (Technische Universiteit Eindhoven, Tech. Rep. 2.4, January 2006).

[37] T. Oakden, M. Kavakli, Graphics processing in virtual production, in: Proc. 2022 14th Intern. Conf. Comput. and Automation Eng. (ICCAE), Brisbane, Australia, 2022, pp. 61–64. doi:10.1109/ICCAE55086.2022.9762415.

[38] L. Pisha, S. Yadegari, Specular path generation and near-reflective diffraction in interactive acoustical simulations, IEEE Trans. Visualization and Comput. Graph. (2023) 1–13. doi:10.1109/TVCG.2023.3238662.

[39] J. C. A. Barata, M. Hussein, The Moore–Penrose pseudoinverse: A tutorial review of the theory, Brazilian J. of Physics 42 (1-2) (2012) 146–165. doi:10.1007/s13538-011-0052-z.

[40] F. Olivieri, F. M. Fazi, M. Shin, P. A. Nelson, F. Simone, L. Yue, Numerical comparison of sound field control strategies under free-field conditions for given performance constraints, in: Inst. of Acoust. - Reprod. Sound 2014, Vol. 1, Birmingham, UK, 2014, url:eprints.soton.ac.uk/370525.

[41] S. J. Elliott, J. Cheer, J.-W. Choi, Y. Kim, Robustness and regularization of personal audio systems, IEEE Trans. Audio, Speech, and Lan-

guage Process. 20 (7) (2012) 2123–2133. doi:10.1109/TASL.2012.2197613.

[42] M. B. Møller, M. Olsen, F. Jacobsen, A hybrid method combining synthesis of a sound field and control of acoustic contrast, in: Proc. Audio Eng. Soc. Conv. 132, Budapest, Hungary, 2012, pp. 565–572, url:aes.org/e-lib/browse.cfm?elib=16265.

[43] Wikipedia, Octave band — Wikipedia, the free encyclopedia, https://en.wikipedia.org/wiki/Octave_band, [Online; controllata il 17-October-2023] (2023).

[44] J. Tabak, Probability and statistics: The science of uncertainty, Infobase Publishing, 2014, url:utstat.toronto.edu/mikevans/jeffrosenthal/book.

[45] S. Duangpummet, J. Karnjana, W. Kongprawechnon, M. Unoki, Blind estimation of speech transmission index and room acoustic parameters based on the extended model of room impulse response, Appl. Acoust. 185 (2022). doi:10.48550/arXiv.2212.13009.

[46] M. G. Bulmer, Principles of statistics, Courier Corporation, 1979.

[47] C. H. Taal, R. C. Hendriks, R. Heusdens, J. Jensen, An algorithm for intelligibility prediction of time–frequency weighted noisy speech, IEEE Trans. Audio, Speech, and Language Process. 19 (7) (2011) 2125–2136. doi:10.1109/TASL.2011.2114881.

[48] Short-time fourier transform, https://it.mathworks.com/help/signal/ref/stft.html, accessed: 18/10/2021.

[49] J. Donley, C. Ritz, W. B. Kleijn, Multizone soundfield reproduction with privacy- and quality-based speech masking filters, IEEE/ACM Trans. Audio, Speech, and Language Process. 26 (6) (2018) 1041–1055. doi:10.1109/TASLP.2018.2798804.

[50] T. Lee, J. K. Nielsen, M. G. Christensen, Signal-adaptive and perceptually optimized sound zones with variable span trade-off filters, IEEE/ACM Trans. Audio, Speech, and Language Process. 28 (2020) 2412–2426. doi:10.1109/TASLP.2020.3013397.

[51] J. Jensen, C. H. Taal, An algorithm for predicting the intelligibility of speech masked by modulated noise maskers, IEEE/ACM Trans. Audio, Speech, and Language Process. 24 (11) (2016) 2009–2022. doi:10.1109/TASLP.2016.2585878.

[52] R. E. Zezario, S.-W. Fu, F. Chen, C.-S. Fuh, H.-M. Wang, Y. Tsao, Deep learning-based non-intrusive multi-objective speech assessment model with cross-domain features, IEEE/ACM Trans. Audio, Speech, and Language Process. 31 (2022) 54–70. doi:10.1109/TASLP.2022.3205757.

[53] Y. Feng, F. Chen, Nonintrusive objective measurement of speech intelligibility: A review of methodology, Biomedical Signal Process. and Control 71 (2022) 103204. doi:10.1016/j.bspc.2021.103204.

[54] A. Farina, Advancements in impulse response measurements by sine sweeps, in: Proc. Audio Eng. Soc. Conv. 122, Vienna, Austria, 2007, url:aes.org/e-lib/browse.cfm?elib=14106.

[55] Inverse short-time fourier transform, https://it.mathworks.com/help/signal/ref/istft.html, accessed: 18/10/2021.

[56] K. Wagener, Factors influencing sentence intelligibility in noise, BIS Verlag, 2004, url:oops.uni-oldenburg.de/460/1/wagfac04.

[57] F. Chen, J. Adcock, S. Krishnagiri, Audio privacy: Reducing speech intelligibility while preserving environmental sounds, in: Proc. 16th ACM Intern. Conf. Multimedia, Assoc. for Computing Machinery, New York, NY, USA, 2008, p. 733–736. doi:10.1145/1459359.1459472.

[58] S. Cecchi, A. Carini, S. Spors, Room response equalization—a review, Appl. Sciences 8 (1) (2018) 1–47. doi:10.3390/app8010016.

[59] S. W. Rienstra, A. Hirschberg, An introduction to acoustics, Eindhoven Univ. of Technol., 2004, https://www.win.tue.nl/~sjoerdr/papers/boek.pdf, Accessed: 15/03/2022.

[60] S. K. Agrawal, O. Sahu, Two-channel quadrature mirror filter bank: An overview, ISRN Signal Process. 2013 (2013) 1–10. doi:10.1155/2013/815619.

[61] H. R. van Maanen, Temporal decay: a useful tool for the characterization of resolution of audio systems?, in: Audio Eng. Soc. Conv. 94, Berlin, Germany, 1993, url:aes.org/e-lib/browse.cfm?elib=6663.

[62] J. Tan, B. Wen, Y. Tian, M. Tian, Frequency convolution for implementing window functions in spectral analysis 36 (5) (2017) 2198–2208. doi:10.1007/s00034-016-0403-7.

[63] B. P. Lathi, Signal processing and linear systems, Oxford Univ. Press New York, 1998, url:signal-processing-and-linear-systems-9780190299040.

[64] F. J. Harris, On the use of windows for harmonic analysis with the discrete Fourier transform, Proc. IEEE 66 (1) (1978) 51–83. doi:10.1109/PROC.1978.10837.

[65] P. Hatziantoniou, J. Mourjopoulos, Generalized fractional-octave smoothing of audio and acoustic responses, J. Audio Eng. Soc. 48 (4) (2000) 259–280, url:aes.org/e-lib/browse.cfm?elib=12070.

[66] Chebyshev window, https://it.mathworks.com/help/signal/ref/chebwin.html, accessed: 15/02/2022.

[67] D. S. P. Committee, Programs for Digital Signal Processing, IEEE Press, New York, 1979.

[68] P. N. Reinhart, P. E. Souza, Effects of varying reverberation on music perception for young normal-hearing and old hearing-impaired listeners, Trends in hearing, SAGE 22 (January 2018). doi:10.1177/2331216517750706.

[69] Y. Cai, L. Liu, M. Wu, J. Yang, Robust time-domain acoustic contrast control design under uncertainties in the frequency response of the loudspeakers, in: INTER-NOISE and NOISE-CON Congr. and Conf. Proc., Vol. 249, Institute of Noise Control Engineering, 2014, pp. 5775–5780.

[70] J. Zhang, L. Shi, M. G. Christensen, W. Zhang, L. Zhang, J. Chen, Robust acoustic contrast control with positive semidefinite constraint using iterative POTDC algorithm, in: Proc. 2022 Intern. Workshop Acous-

tic Signal Enhancement (IWAENC), 2022, pp. 1–5. doi:10.1109/IWAENC53105.2022.9914730.

[71] G. Matsaglia, G. P. H. Styan, Equalities and inequalities for ranks of matrices †, Linear and Multilinear Algebra 2 (3) (1974) 269–292. doi:10.1080/03081087408817070.

[72] T. Afghah, E. Patros, M. Puckette, A pseudoinverse technique for the pressure-matching beamforming method, in: Audio Eng. Soc. Conv. 145, New York, NY, USA, 2018, url:aes.org/e-lib/browse.cfm?elib=19778.

[73] A. G. Prinn, A review of finite element methods for room acoustics, Acoustics 5 (2) (2023) 367–395. doi:10.3390/acoustics5020022.

[74] D. Diaz-Guerra, A. Miguel, J. R. Beltran, gpuRIR: A python library for room impulse response simulation with GPU acceleration, Multimedia Tools and Appl. 80 (4) (2021) 5653–5671. doi:10.1007/s11042-020-09905-3.

[75] F. J. Fahy, Statistical energy analysis: a critical overview, Philosoph. Trans.: Physical Sciences and Eng. 346 (1681) (1994) 431–447, url:jstor.org/stable/54287.

[76] L. Savioja, J. Huopaniemi, T. Lokki, R. Väänänen, Creating interactive virtual acoustic environments, J. Audio Eng. Soc. 47 (9) (1999) 675–705, url:aes.org/e-lib/browse.cfm?elib=12095.

Acknowledgments

I would like to express my sincere gratitude to my supervisors, Professor Riccardo Raheli from University of Parma and Marco Martalò from University of Cagliari, for their guidance, support, and encouragement throughout my PhD journey. They have been very generous with their time and expertise, and have helped me overcome the challenges and difficulties I faced during my research. They have also inspired me with their passion and enthusiasm for this research.

I would also like to thank Alessandro Costalunga, Carlo Tripodi, and Nicolò Strozzi from ASK Industries S.p.A. for their collaboration and assistance in conducting the in-situ measurements. They have provided me with valuable data and feedback that have improved the quality and validity of my work.

I am grateful to the University of Parma and ASK Industries S.p.A. for providing me with the opportunity and the resources to pursue my PhD degree. I also appreciate the financial support from the Italian Ministry of Economic Development (MiSE)'s under the project CGS (Connettività e Guida Sicura).

I would like to dedicate this thesis to my family and friends, who have always been there for me with their love, care, and support. They have given me the motivation and the strength to pursue my dreams and goals.

Finally, I would like to thank all the people who have contributed directly or indirectly to this thesis. I apologize if I have inadvertently omitted anyone. I hope that this thesis will be a useful contribution to the field of personal sound zone systems.