Università Politecnica delle Marche
Scuola di Dottorato di Ricerca in Scienze dell'Ingegneria
Curriculum in Ingegneria Elettronica, Elettrotecnica e delle Telecomunicazioni

---

Efficient Algorithms for Immersive Audio Rendering Enhancement

Algoritmi efficienti per il Miglioramento della Riproduzione Sonora Immersiva

Ph.D. Dissertation of:
Valeria Bruschi

Advisor:
Prof. Stefania Cecchi

Curriculum Supervisor:
Prof. Franco Chiaraluce

XXXV edition - new series

Università Politecnica delle Marche
Scuola di Dottorato di Ricerca in Scienze dell'Ingegneria
Curriculum in Ingegneria Elettronica, Elettrotecnica e delle Telecomunicazioni

---

Efficient Algorithms for Immersive Audio Rendering Enhancement

Algoritmi efficienti per il Miglioramento della Riproduzione Sonora Immersiva

Ph.D. Dissertation of:
Valeria Bruschi

Advisor:
Prof. Stefania Cecchi

Curriculum Supervisor:
Prof. Franco Chiaraluce

XXXV edition - new series

Università Politecnica delle Marche
Scuola di Dottorato di Ricerca in Scienze dell'Ingegneria
Facoltà di Ingegneria
Via Brecce Bianche – 60131 Ancona (AN), Italy

"Stilicidi casus lapidem cavat"

(Lucrezio, *De rerum natura*, I 313)

Acknowledgments

Before starting with the discussion, I would like to thank the people who contributed to the realization of this thesis and who supported me during my journey.

I am extremely grateful to my advisor, professor Stefania Cecchi, who guided me during these three years and allowed me to grow professionally. I also thank her for the humanity she has always shown by establishing a precious friendship bond.

I thank my colleagues and friends Alessandro and Stefano, who have been essential during these three years both for their collaboration on several projects and for making the days at the office less heavy. I also thank all the people in the department with whom I have collaborated or shared experiences in the academic field.

I'd like to thank professor Vesa Välimäki for inviting me at the Department of Signal Processing and Acoustics of Aalto University, Finland, and for introducing me to his research group. This experience was fundamental for my career because it allowed me to acquire new knowledge and to meet many people who work in the same field as me.

I thank my parents because they have always been close to me and have always supported my choices, even the most difficult ones to accept. I also thank my brother Simone for emotionally supporting me, together with my parents, especially during the last very demanding and stressful period.

Finally, I thank my friends who have accompanied me over the years and all those people who, even if passing through, have been important to me and to my growth and whom I will remember forever.

Ancona, Febbraio 2023

Valeria Bruschi

Ringraziamenti

Prima di iniziare con la trattazione, vorrei ringraziare le persone che hanno contribuito alla realizzazione di questa tesi e che mi hanno supportato durante il mio percorso.

Un ringraziamento particolare va al mio tutor accademico, la professoressa Stefania Cecchi, che mi ha guidato durante questi tre anni e mi ha permesso di crescere professionalmente. La ringrazio inoltre per l'umanità che ha sempre dimostrato instaurando anche un prezioso legame di amicizia.

Ringrazio i miei colleghi e amici Alessandro e Stefano, che sono stati essenziali in questi tre anni sia per la collaborazione a vari progetti sia per aver reso meno pesanti le giornate in ufficio. Ringrazio anche tutte le persone del dipartimento con le quali ho collaborato o condiviso esperienze in ambito accademico.

Ringrazio inoltre il professore Vesa Välimäki per avermi ospitato presso il dipartimento di signal processing ed acustica dell'università Aalto, in Finlandia, e per avermi introdotto al suo gruppo di ricerca. Questa esperienza è stata fondamentale per il mio percorso perché mi ha permesso di acquisire nuove conoscenze e di incontrare molte persone che lavorano nel mio stesso ambito.

Ringrazio i miei genitori perché mi sono sempre stati vicino e hanno sempre sostenuto le mie scelte, anche quelle più difficili da accettare. Ringrazio anche mio fratello Simone per avermi supportato emotivamente, insieme ai miei genitori, specialmente nell'ultimo periodo particolarmente impegnativo e stressante.

Infine, ringrazio gli amici che mi hanno accompagnato in questi anni e tutte quelle persone che, anche se di passaggio, sono state importanti per me e per la mia crescita e che ricorderò per sempre.

Ancona, Febbraio 2023

Valeria Bruschi

Abstract

Immersive audio rendering is the process of creating an engaging and realistic sound experience in 3D space. In immersive audio systems, the head-related transfer functions (HRTFs) are used for binaural synthesis over headphones since they express how humans localize a sound source. HRTF interpolation algorithms can be introduced for reducing the number of measurement points and creating a reliable sound movement. Binaural reproduction can be also performed by loudspeakers. However, the involvement of two or more loudspeakers causes the problem of crosstalk. In this case, crosstalk cancellation (CTC) algorithms are needed to delete unwanted interference signals.

In this thesis, starting from a comparative analysis of HRTF measurement techniques, a binaural rendering system based on HRTF interpolation is proposed and evaluated for real-time applications. The proposed method shows good performance in comparison with a reference technique. The interpolation algorithm is also applied for immersive audio rendering over loudspeakers, by adding a fixed crosstalk cancellation algorithm, which assumes that the listener is in a fixed position. In addition, an adaptive crosstalk cancellation system, which includes the tracking of the listener's head, is analyzed and a real-time implementation is presented. The adaptive CTC implements a subband structure and experimental results prove that a higher number of bands improves the performance in terms of total error and convergence rate.

The reproduction system and the characteristics of the listening room may affect the performance due to their non-ideal frequency response. Audio equalization is used to adjust the balance of different audio frequencies in order to achieve desired sound characteristics. The equalization can be manual, such as in the case of graphic equalization, where the gain of each frequency band can be modified by the user, or automatic, where the equalization curve is automatically calculated after the room impulse response measurement. The room response equalization can be also applied to multichannel systems, which employ two or more loudspeakers, and the equalization zone can be enlarged by measuring the impulse responses in different points of the listening zone.

In this thesis, efficient graphic equalizers (GEQs), and an adaptive room response equalization system are presented. In particular, three low-complexity linear- and quasi-linear-phase graphic equalizers are proposed and deeply examined. Experiments confirm the effectiveness of the proposed GEQs in terms

of accuracy, computational complexity, and latency. Successively, a subband adaptive structure is introduced for the development of a multichannel and multiple positions room response equalizer. Experimental results verify the effectiveness of the subband approach in comparison with the single-band case. Finally, a linear-phase crossover network is presented for multichannel systems, showing great results in terms of magnitude flatness, cutoff rates, polar diagram, and phase response.

Active noise control (ANC) systems can be designed to reduce the effects of noise pollution and can be used simultaneously with an immersive audio system. The ANC works by creating a sound wave that has an opposite phase with respect to the sound wave of the unwanted noise. The additional sound wave creates destructive interference, which reduces the overall sound level.

Finally, this thesis presents an ANC system used for noise reduction. The proposed approach implements an online secondary path estimation and is based on cross-update adaptive filters applied to the primary path estimation that aim at improving the performance of the whole system. The proposed structure allows for a better convergence rate in comparison with a reference algorithm.

Sommario

Il rendering audio immersivo è il processo di creazione di un'esperienza sonora coinvolgente e realistica nello spazio 3D. Nei sistemi audio immersivi, le funzioni di trasferimento relative alla testa (head-related transfer functions, HRTFs) vengono utilizzate per la sintesi binaurale in cuffia poiché esprimono il modo in cui gli esseri umani localizzano una sorgente sonora. Possono essere introdotti algoritmi di interpolazione delle HRTF per ridurre il numero di punti di misura e per creare un movimento del suono affidabile. La riproduzione binaurale può essere eseguita anche dagli altoparlanti. Tuttavia, il coinvolgimento di due o più gli altoparlanti causa il problema del crosstalk. In questo caso, algoritmi di cancellazione del crosstalk (CTC) sono necessari per eliminare i segnali di interferenza indesiderati.

In questa tesi, partendo da un'analisi comparativa di metodi di misura delle HRTF, viene proposto un sistema di rendering binaurale basato sull'interpolazione delle HRTF per applicazioni in tempo reale. Il metodo proposto mostra buone prestazioni rispetto a una tecnica di riferimento. L'algoritmo di interpolazione è anche applicato al rendering audio immersivo tramite altoparlanti, aggiungendo un algoritmo di cancellazione del crosstalk fisso, che considera l'ascoltatore in una posizione fissa. Inoltre, un sistema di cancellazione crosstalk adattivo, che include il tracciamento della testa dell'ascoltatore, è analizzato e implementato in tempo reale. Il CTC adattivo implementa una struttura in sottobande e risultati sperimentali dimostrano che un maggiore numero di bande migliora le prestazioni in termini di errore totale e tasso di convergenza.

Il sistema di riproduzione e le caratteristiche dell'ambiente di ascolto possono influenzare le prestazioni a causa della loro risposta in frequenza non ideale. L'equalizzazione viene utilizzata per livellare le varie parti dello spettro di frequenze che compongono un segnale audio al fine di ottenere le caratteristiche sonore desiderate. L'equalizzazione può essere manuale, come nel caso dell'equalizzazione grafica, dove il guadagno di ogni banda di frequenza può essere modificato dall'utente, o automatica, la curva di equalizzazione è calcolata automaticamente dopo la misurazione della risposta impulsiva della stanza. L'equalizzazione della risposta ambientale può essere applicata anche ai sistemi multicanale, che utilizzano due o più altoparlanti e la zona di equalizzazione può essere ampliata misurando le risposte impulsive in diversi punti della zona di ascolto.

In questa tesi, equalizzatori grafici efficienti e un sistema adattativo di equalizzazione d'ambiente. In particolare, sono proposti e approfonditi tre equalizzatori grafici a basso costo computazionale e a fase lineare e quasi lineare. Gli esperimenti confermano l'efficacia degli equalizzatori proposti in termini di accuratezza, complessità computazionale e latenza. Successivamente, una struttura adattativa in sottobande è introdotta per lo sviluppo di un sistema di equalizzazione d'ambiente multicanale. I risultati sperimentali verificano l'efficienza dell'approccio in sottobande rispetto al caso a banda singola. Infine, viene presentata una rete crossover a fase lineare per sistemi multicanale, mostrando ottimi risultati in termini di risposta in ampiezza, bande di transizione, risposta polare e risposta in fase.

I sistemi di controllo attivo del rumore (ANC) possono essere progettati per ridurre gli effetti dell'inquinamento acustico e possono essere utilizzati contemporaneamente a un sistema audio immersivo. L'ANC funziona creando un'onda sonora in opposizione di fase rispetto all'onda sonora in arrivo. Il livello sonoro complessivo viene così ridotto grazie all'interferenza distruttiva.

Infine, questa tesi presenta un sistema ANC utilizzato per la riduzione del rumore. L'approccio proposto implementa una stima online del percorso secondario e si basa su filtri adattativi in sottobande applicati alla stima del percorso primario che mirano a migliorare le prestazioni dell'intero sistema. La struttura proposta garantisce un tasso di convergenza migliore rispetto all'algoritmo di riferimento.

Foreword

During my period at Dipartimento di Ingegneria dell'Informazione of Università Politecnica delle Marche as Ph.D. student, I had the pleasure to work with the research group led by Prof. Stefania Cecchi. The contents of this thesis have been partially included in the following publications.

- **V. Bruschi**, S. Nobili, S. Cecchi, and F. Piazza, "An Innovative Method for Binaural Room Impulse Responses Interpolation," in 148th Convention of the Audio Engineering Society (AES), May 2020.
- **V. Bruschi**, S. Nobili, and S. Cecchi, "A Real-Time Implementation of a 3D Binaural System based on HRIRs Interpolation," in 12th International Symposium on Image and Signal Processing and Analysis (ISPA), Sep. 2021.
- **V. Bruschi**, S. Nobili, A. Terenzi, and S. Cecchi, "An Improved Approach for Binaural Room Impulse Responses Interpolation in Real Environments," in 152nd Convention of the Audio Engineering Society (AES), May 2022.
- **V. Bruschi**, S. Nobili, and S. Cecchi, "Real Time Binaural Synthesis of Moving Sound Sources over Loudspeakers," in IEEE Immersive and 3D Audio: from Architecture to Automotive (I3DA), Sep. 2021.
- **V. Bruschi**, S. Nobili, F. Bettarelli, and S. Cecchi, "Listener-position Sub-band Adaptive Crosstalk Canceller using HRTFs Interpolation for Immersive Audio Systems," in 150th Convention of the Audio Engineering Society (AES), May 2021.
- **V. Bruschi**, S. Nobili, A. Terenzi, and S. Cecchi, "A Low-Complexity Linear-Phase Graphic Audio Equalizer Based on IFIR Filters," IEEE Signal Process. Lett., vol. 28, pp. 429–433, Feb. 2021.
- **V. Bruschi**, V. Välimäki, J. Liski, and S. Cecchi, "Linear-Phase Octave Graphic Equalizer," J. Audio Eng. Soc., Special Issue on Audio Filter Design, Jun. 2022.
- **V. Bruschi**, V. Välimäki, J. Liski, S. Cecchi et al., "A Low-Latency Quasi- Linear-Phase Octave Graphic Equalizer," in International Conference on Digital Audio Effects, Sep. 2022, pp. 94–100.

- S. Cecchi, A. Terenzi, **V. Bruschi**, A. Carini, and S. Orcioni, "A Subband Implementation of a Multichannel and Multiple Position Adaptive Room Response Equalizer," Applied Acoustics, vol. 173, p. 107702, 2021.
- **V. Bruschi**, S. Nobili, A. Terenzi, and S. Cecchi, "Using Interpolated FIR Technique for Digital Crossover Filters Design," in Proceedings of the 30th European Signal Processing Conference (EUSIPCO), 2022, pp. 214–218.
- S. Nobili, **V. Bruschi**, F. Bettarelli, and S. Cecchi, "An Efficient Active Noise Control System with Online Secondary Path Estimation for Snoring Reduction," in IEEE 29th European Signal Processing Conference (EUSIPCO), 2021, p. 7.
- S. Nobili, **V. Bruschi**, F. Bettarelli, and S. Cecchi, "A Real Time Subband Implementation of an Active Noise Control System for Snoring Reduction," in IEEE 12th International Symposium on Image and Signal Processing and Analysis (ISPA), 2021, pp. 109–114.
- **V. Bruschi**, A. Terenzi, N. Dourou, S. Spinsante, and S. Cecchi, "Comparative Analysis of HRTFs Measurement using In-ear Microphones," in Sensors, Collection "Advanced Techniques for Acquisition and Sensing", MDPI, 2023.
- **V. Bruschi**, N. Dourou, S. Nobili, and S. Cecchi, "A New Method for HRTFs Interpolation based on Inverse Distance Weighting," in 154th Convention of Audio Engineering Society, 2023.
- **V. Bruschi**, N. Dourou, S. Nobili, and S. Cecchi, "Perceptual Study of a Low-Computational Linear-Phase Digital Crossover Network," in 154th Convention of Audio Engineering Society, 2023.
- S. Cecchi, **V. Bruschi**, S. Nobili, A. Terenzi, and V. Välimäki, "Crossover Networks: A Review," J. Audio Eng. Soc., Special Issue on New Trends in Audio Effects, Part 2, 2023.

Some contents of the thesis are also reported in the following paper, currently under review.

- S. Nobili, **V. Bruschi**, A. Terenzi and S. Cecchi, "A Non-Uniform Delayless Approach to Active Noise Control," IEEE Trans. on Control Systems Technology, 2023.

Contents

List of Figures

List of Tables

List of Abbreviations

ANC	Active Noise Control
BRIR	Binaural Room Impulse Response
BRTF	Binaural Room Transfer Function
CTC	Crosstalk Cancellation
DI	Distortion Index
DSP	Digital Signal Processing
DTW	Dynamic Time Warping
EQ	Equalizer
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FxLMS	Filtered-x Least Mean Square
GEQ	Graphic Equalizer
GP	Gravity Point
HATS	Head and Torso Simulator
HP	Highpass
HRIR	Head-Related Impulse Response
HRTF	Head-Related Transfer Function
IFIR	Interpolated Finite Impulse Response
IIR	Infinite Impulse Response
ILD	Interaural Level Difference
IR	Impulse Response
IRS	Inverse Repeated Sequence
ITD	Interaural Time Difference
LMS	Least Mean Square
LP	Lowpass

LSD	Log-Spectral Distance
LSFI	Least Square Frequency Invariant
LTI	Linear Time-Invariant
MIMO	Multi-Input Multi-Output
MISO	Multi-Input Single-Output
MLS	Maximum Length Sequence
MSE	Mean Squared Error
MT	Mixing Time
NLMS	Normalized Least Mean Square
NUTS	NU-Tech Satellite
OPS	Orthogonal Periodic Sequence
PM	Pressure Matching
PPS	Perfect Periodic Sequence
RACE	Recursive Ambiophonic Crosstalk Elimination
RIR	Room Impulse Response
RRE	Room Response Equalization
RT	Real-Time
RTF	Room Transfer Function
SAF	Subband Adaptive Filtering
SD	Spectral Deviation
SIMO	Single-Input Multi-Output
SISO	Single-Input Single-Output
VBAP	Vector Base Amplitude Panning

Chapter 1

Introduction

Immersive audio rendering uses technologies that increase the realism of sound in a virtual environment. The spatial audio creates a three-dimensional soundscape in which sounds appear to come from different directions. Digital signal processing (DSP) algorithms are employed to enhance the audio immersive scenario and to advance the listening experience [7]. Two different approaches can be employed for the development of a 3D audio system [8]. The first consists of multichannel systems, in which a large number of loudspeakers is engaged to reproduce the desired effect in the listening zone. The second is binaural synthesis, in which the acoustic signals, enriched with directional cues, are directly reproduced at the ears of the listener through headphones or loudspeakers. In particular, binaural reproduction is obtained by filtering the input signal with the appropriate pair of head-related transfer functions (HRTFs) [9], which describe the acoustic path between the sound source and the listener's ears and contain the localization cues, including the sound diffraction of the torso, head and ear's pinna of the listener. HRTFs can be measured with standard binaural mannequins [10] or with in-ear microphones installed on real subjects [11]. The involvement of real subjects allows the estimation of personalized HRTFs, but the measurements can be affected by head movements and microphone position [12]. Binaural synthesis can be easily reproduced using headphones, which provide a perfect channel separation. However, the long-term use of headphones may be troublesome and annoying. An alternative is using stereo loudspeakers in front of the listener. In this case, the channels are not separated due to the interference between the two loudspeaker signals. This phenomenon is called crosstalk and it can be reduced with crosstalk cancellation (CTC) algorithms. CTC usually consists of the inversion of the matrix containing the four HRTFs that represent the four acoustic paths between the loudspeakers and the listener's ears [13]. CTC can be achieved through fixed or adaptive solutions. Fixed CTC techniques assume that the listener does not move from the sweet spot, i.e., the zone where the cancellation has the maximum effect. Contrarily, adaptive CTC algorithms can adaptively update the CTC filters according to the listener position, detected by head tracking. The audio rendering can

be enhanced also by audio equalization techniques, which aim at compensating for the non-ideal transfer function of the reproduction system and of the environment. Many equalization approaches exist in the literature [14], and they can be classified into manual or automatic equalizers (EQs). Manual EQs allow the user to adjust several parameters according to his/her equalization preference. Graphic equalizers (GEQs) are an example of manual equalization, where only the gains of the frequency bands can be selected. Automatic EQs concern room equalization and involve the use of one or more microphones to estimate the room impulse responses (RIRs) that must be equalized. The equalization zone can be enlarged by considering multiple microphone positions for the RIR measurements. In addition, multichannel equalization can also be taken into account when more loudspeakers are involved. In the case of multichannel systems, loudspeakers can reproduce different frequency bands and crossover networks are applied to divide the signal spectrum [15]. Finally, the audio rendering can be affected also by environmental noise that may disturb the listening experience. For this reason, active noise control (ANC) algorithms can be implemented to reduce unwanted noise. The basic idea of ANC techniques is to generate an additional sound signal, called antinoise, that is out of phase with the signal that must be deleted. In general, a microphone detects the undesired signal and a loudspeaker reproduces the antinoise. The primary path is defined by the path between the noise source and the reference microphone, while the secondary path is between the loudspeaker and the microphone. Different ANC implementations that aim at estimating the primary and the secondary path can be found in the literature [16].

In this thesis, innovative systems for immersive audio rendering enhancement are presented. In particular, different HRTF measurement procedures are presented and analyzed through experimental comparisons. Successively, a binaural synthesis technique based on HRTF interpolation is proposed. The HRTF interpolation is applied to reduce the measurement points maintaining a convincing 3D audio experience. The binaural system is used for both headphones and loudspeakers reproduction. In the case of loudspeakers playback, a fixed recursive CTC algorithm is applied. Moreover, an adaptive subband crosstalk canceller is also presented using the HRTF interpolation for the CTC filters update. Regarding audio equalization, three linear- and quasi-linear-phase graphic equalizers are implemented employing interpolated finite impulse response (IFIR) filters. Whereas room equalization is presented by a multichannel and multiple positions adaptive equalizer based on a subband RIRs identification approach. Concerning multichannel systems, a linear-phase crossover network based on IFIR filters is presented. Finally, an active noise control system, which uses a subband structure in the primary path estimation and involves an online secondary path estimation, is discussed.

1.1 Thesis outline

This thesis is organized as follows.

Chapter 2 presents immersive audio rendering algorithms. Starting with a comparative analysis of HRTF measurement methods, a binaural synthesis system based on HRTFs is proposed for the reproduction over headphones and loudspeakers. A fixed and an adaptive subband crosstalk cancellation algorithm are involved in loudspeaker reproduction.

Chapter 3 proposes audio equalization methods for audio rendering enhancement. Three linear- and quasi-linear-phase graphic equalizer implementations are discussed and an adaptive multichannel equalization procedure is presented. Moreover, a linear-phase crossover network is explored for multichannel systems.

Chapter 4 analyzes an innovative subband active noise control algorithm with online secondary path estimation, based on subband adaptive filtering applied to the primary path estimation.

Chapter 5 reports the conclusions of the thesis.

1.2 Thesis contributions

The main contributions of this thesis are listed in the following.

Chapter 2

- The binaural synthesis of moving sound sources implements a HRTF interpolation algorithm that improves a previous method. The system is also applied to the case of a reverberant environment by adding the automatic calculation of the mixing time, i.e., the transition point between the early reflections and the reverberant tail of the impulse responses. This contribution is presented in [17–19].

- The recursive ambiophonic crosstalk elimination (RACE) is added to the binaural system for loudspeakers reproduction, considering the listener in a fixed position in front of two loudspeakers. This contribution is presented in [20].

- An existing subband crosstalk cancellation algorithm is improved to obtain an adaptive solution with the addition of a head tracker and HRTF interpolation to update the CTC filters. This contribution is presented in [21].

Chapter 3

- Three graphic equalizers based on interpolated FIR filters are developed. A linear-phase uniform GEQ is obtained by the implementation of a parallel structure. A linear-phase octave GEQ is obtained by a tree structure derived from a prototype filter. Finally, a low-latency quasi-linear-phase octave GEQ is derived from the linear-phase one by designing the first band with an infinite impulse response (IIR) filter. This contribution is presented in [22–24].
- A multichannel and multiple positions adaptive room response equalizer is developed, by improving a previous system with a subband structure in the identification procedure. This contribution is presented in [25].
- A low-complexity linear-phase digital crossover network based on interpolated FIR filters is proposed. This contribution is presented in [26].

Chapter 4

- An active noise control system with online secondary path estimation is improved by adding a subband adaptive filtering (SAF) structure in the primary path estimation. This contribution is presented in [27, 28].

Chapter 2

Immersive Audio Rendering

Immersive audio rendering is a technique used to create a realistic 3D soundscape, achieved by simulating virtual sound sources. For this reason, it is important to know how humans localize a sound source. The localization features are contained in head-related transfer functions, widely used to produce immersive scenarios. In this context, this chapter first presents a comparative analysis of HRTF measurements and then proposes effective algorithms of digital signal processing for binaural synthesis reproduction over headphones and loudspeakers. In addition, crosstalk cancellation algorithms are presented for loudspeaker reproduction. Figure 2.1 shows an overview of immersive audio rendering systems. This chapter is organized as follows. Section 2.1 describes HRTF measurement methodologies. Section 2.2 presents a system for binaural synthesis through headphones using HRTFs interpolation. Section 2.3 reports systems for the immersive reproduction over loudspeakers, introducing crosstalk cancellation algorithms. Finally, Section 2.4 concludes the chapter.

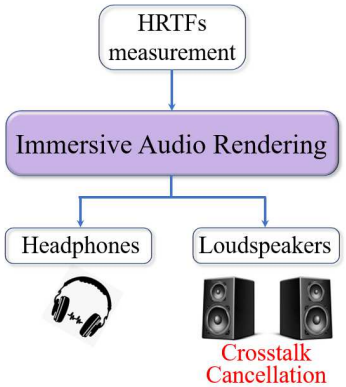


Figure 2.1: Overview of immersive audio rendering systems.

2.1 Head-related transfer functions

The head-related transfer functions (HRTFs), in the frequency domain, or the head-related impulse responses (HRIRs), in the time domain, are special functions that contain the human localization cues, such as interaural level difference (ILD) and interaural time difference (ITD), as shown in Figure 2.2. The ILD is the sound level difference between the signal that reaches the left ear and the one that reaches the right ear (cf. Figure 2.2(a)). The ITD is the time delay of sound arrival between the left and the right ear (cf. Figure 2.2(b)). Moreover, HRTFs also contain the spectral cues which characterize the sound localization over the vertical plane. HRTFs are used in immersive sound to simulate the way sound waves interact with a person's head and ears. They can be measured using binaural mannequins or in-ear microphones. In this section, the definition of HRTFs is reported and a comparative analysis of HRTFs measurement techniques is presented.

2.1.1 Background on HRTFs measurement

The HRTFs are individual since they depend on the shape of the head, pinnae, and torso which are different for each human being. However, HRTFs can be measured using head and torso simulators that standardize the body dimensions considering the median values among many adults [10]. In the literature, standard head and torso simulators are generally employed for the creation of HRTFs databases [1] and the study of measurement limits, such as the directional resolution [29, 30] and the distance between the sound source and the head [31, 32]. Personalized HRTFs can be measured using in-ear microphones to catch the individual differences lost by standardization [11, 12, 33–38]. Perceptual studies demonstrated that the use of personalized HRTFs improves the source localization and the fidelity of the binaural rendering [39–41], but measurements conducted on real subjects could be affected by head movements [12] and the in-ear microphone position [37, 42]. The problem of head movements can be avoided by the use of neck supports [34], or by detecting the listener's head position by a tracking system for each HRTF [35, 36]. The position of the microphone is an important aspect since the sound pressure distribution along the ear canal is non-uniform [43–47]. The best microphone position changes depending on the type of study and application and it could be at the entrance of the ear canal [10, 48, 49], inside the ear canal [43, 50] or close to the eardrum [44, 51, 52]. Previous research proved that most of the localization cues can be captured close to the eardrum [53–55], even if it is an unpleasant position for the involved subjects.

Focusing on the HRTF measurement algorithms employed in the literature

[42], the most popular ones are the pseudo-random sequences (maximum length sequence (MLS), inverse repeated sequence (IRS), and Golay code) [56,57] and sweep signals (linear and exponential sweeps) [58]. However, HRTF measurements could be affected by several problems, such as nonlinear distortions of the electro-acoustic systems, environmental noises, reflections from the environments, sound source characteristics, and temperature variations [42, 59]. The measurement inside a controlled environment (e.g., anechoic chamber) can solve the problems derived by the environment, while nonlinear distortions can be avoided by choosing the appropriate procedure and stimuli [60–62]. In [63], perfect periodic sequences (PPSs) and orthogonal periodic sequences (OPSs) are applied for HRTFs measurement in a real car environment, proving robustness towards nonlinearities.

2.1.2 Head-related transfer functions definition

The sound waves perceived by a listener are affected by the head, pinnae, and torso of the listener [9]. Head-related transfer functions $H_{\mathrm{L,R}}$ represent the effect of these characteristics in the frequency domain, while head-related impulse responses $h_{\mathrm{L,R}}$ are the respective functions in the time domain. For each source position, two HRTFs are defined, i.e., one for the left ear (L) and another for the right ear (R). HRTFs may depend on the frequency ω and on the position of the sound source in terms of distance ρ, azimuth ϑ and elevation

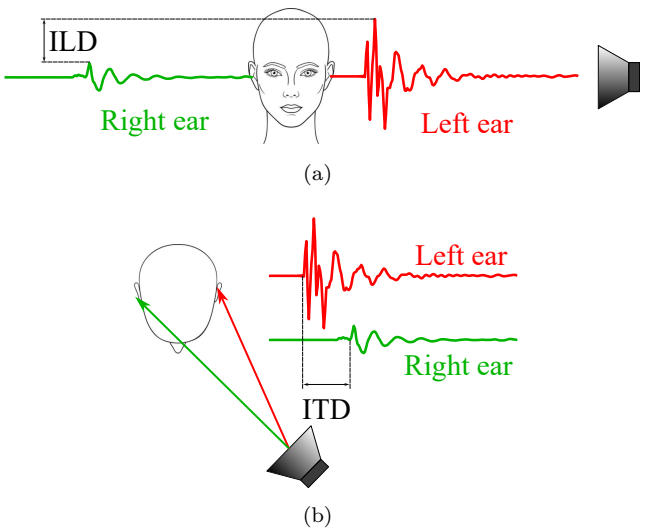


Figure 2.2: Illustration of (a) the interaural level difference (ILD) and (b) the interaural time difference (ITD).

φ and can be defined as follows [9]:

$$H_{\mathrm{L,R}}(\omega, \rho, \vartheta, \varphi) = \frac{G_{\mathrm{L,R}}(\omega, \rho, \vartheta, \varphi)}{T(\omega, \rho, \vartheta, \varphi)}, \tag{2.1}$$

where $G_{\mathrm{L,R}}(\omega, \rho, \vartheta, \varphi)$ is the transfer function of the path between the sound source and the ear canal entrance, while $T(\omega, \rho, \vartheta, \varphi)$ is the room transfer function of the acoustic path between the sound source and the listener position (considering the point at the center of the head) without the presence of the listener. For example, Figure 2.3 shows the left HRTF in both the time and frequency domain for the frontal position, i.e., $\vartheta = 0°$ and $\varphi = 0°$, taken from the MIT HRTF database of [1]. The peaks and notches of the magnitude frequency response, shown in Figure 2.3(b), are caused by the head, pinna, and torso of the listener and they change with the direction of the sound source. HRTFs are usually measured in anechoic environments in order to avoid reflections and reverberation of real rooms. Moreover, HRTFs are not affected by the distance ρ when it is bigger than 1 m [9]. For this reason, far-field HRTFs databases are created considering the same distance ρ. Contrarily, when the measurement is carried out in reverberant environments, the sound source distance influences the resulting impulse response that, in this case, is called binaural room impulse response (BRIR), composed by the respective HRIR and the room impulse response (RIR) [5]. In the frequency domain, the binaural room transfer function (BRTF) is defined by $G_{\mathrm{L,R}}(\omega, \rho, \vartheta, \varphi)$ and can be obtained by inverting Equation (2.1). More in general, an impulse response (IR) represents the acoustic path between two points in the space and completely characterizes a causal linear time-invariant (LTI) system. The impulse response can be truncated and expressed by a finite number of coefficients as long as the length is large enough to contain all the characteristics of the environment. For this reason,

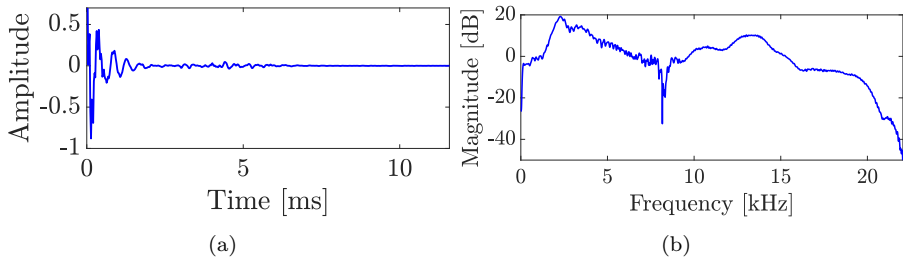


Figure 2.3: Head-related (a) impulse response (in the time domain) and (b) transfer function (in the frequency domain) of the left ear for the frontal position, i.e., $\vartheta = 0°$ and $\varphi = 0°$. The HRTF is taken from the MIT HRTF database of [1].

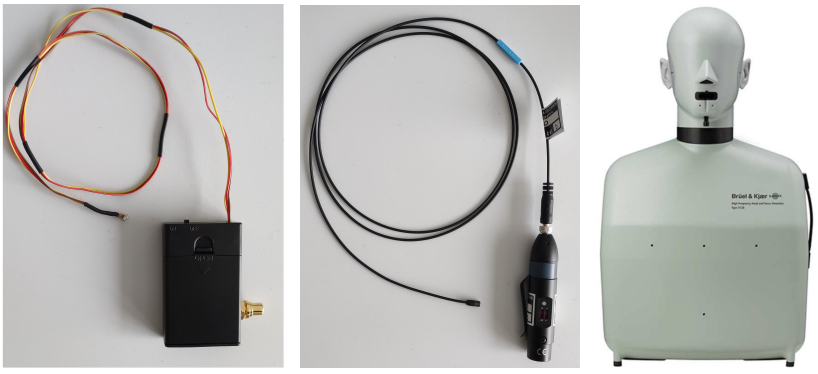
it can be seen as a finite impulse response (FIR) filter of order N and length $N + 1$. In the literature, several techniques for HRTFs measurement can be found, varying the type and the position of microphones, and a comparative analysis of these techniques is reported in the following.

2.1.3 Comparative analysis of HRTFs measurements

HRTFs are strongly individual and they can be measured using standard head simulators or in-ear microphones worn by real subjects. In the second case, the position of the in-ear microphone may affect the HRTF measurement. In this context, this section presents HRTF measurements, comparing the results obtained by a standard binaural mannequin with the ones obtained by two different in-ear microphones. Moreover, the influence of the microphone position is investigated using the ear of the mannequin. Finally, HRTF measurements on two real subjects have been carried out to examine the HRTF differences.

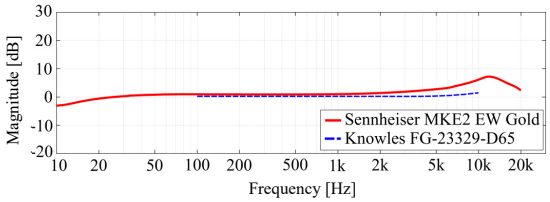
Hardware setup

HRTFs can be measured by using standard simulators, equipped with fixed microphones, or in-ear microphones. Focusing on microphones, the sensors used must be as small as possible, indeed this is important for two main reasons: first, since the sensor is placed inside the ear canal, a small device could be installed without being too annoying for the subject, then, a small form factor is also important in order to minimize any modification in the ear form which

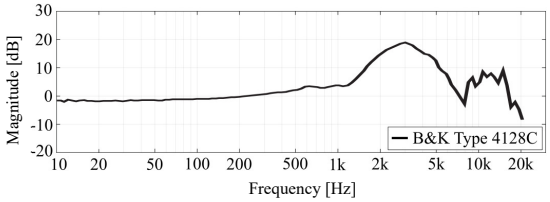


(a) (b) (c)

Figure 2.4: (a) Knowles FG-23329-D65 with its battery power supply, (b) Sennheiser MKE2-EW Gold microphone with its power supply and signal conditioner MZA900P, and (c) Brüel & Kjær head and torso simulator (HATS) Type 4128C.



(a)



(b)

Figure 2.5: Frequency responses of (a) Knowles FG-23329-D65 [2], Sennheiser MKE2 EW Gold [3], and (b) Brüel & Kjær head and torso simulator (HATS) Type 4128C [4], provided by the manufacturers.

can alter the measured responses. Another important aspect of the microphone is its frequency response, which should be as flat as possible.

For this analysis, the HRTF measurements have been carried out with two different microphones, i.e., the Knowles FG-23329-D65 and the Sennheiser MKE2-EW Gold. Figure 2.4(a)-(b) shows photos of the microphones and Figure 2.5(a) shows their frequency responses. The Knowles FG-23329-D65 [2] is an electret condenser omnidirectional microphone. Its dimensions are about a few millimeters in diameter, as shown in Figure 2.4(a), allowing an easy placement inside the ear canal. The Knowles microphone has a very flat frequency response in a reasonable band between 100 Hz and 10 kHz, as visible in Figure 2.5(a). It has also a very limited power consumption (50 μA), so just two AA batteries are needed to power the microphone avoiding noise problems. The Sennheiser MKE2-EW [3] Gold is a condenser Lavalier omnidirectional microphone. It features a wide frequency range from 20 Hz to 20 kHz, as reported in Figure 2.5(a), and an almost flat frequency response below 5 kHz. Figure 2.4(b) shows the microphone with its power supply/signal conditioner Sennheiser MZA-900P. The microphone is powered by a 48V phantom line and generates a low-impedance balanced output. In comparison with the Knowles, the Sennheiser features a slightly bigger capsule with a thicker wire which makes the placement more difficult, on the other side the Sennheiser has a wider frequency response and more robust construction and it can be easily powered by any modern soundcard.

The HRTFs measured with the two microphones are compared with the ones measured with the Brüel & Kjær head and torso simulator (HATS) Type 4128C, i.e., a binaural mannequin used as a reference and shown in Figure 2.4(c). The frequency response of the mannequin is reported in Figure 2.5(b). In this case, the magnitude response is not flat due to the effect of the ear of the dummy head. For the measurements with the B&K simulator, the mannequin is connected to its power supply B&K PS 2829. Moreover, the microphones and a Genelec 8020 A are connected to the Scarlett Focusrite 2i2 soundcard, managed by a computer that uses the NU-Tech software [64] for the acquisitions. The measurements have been carried out inside a semianechoic chamber and taking into account only the left ear. The experimental setup is shown in Figure 2.6.

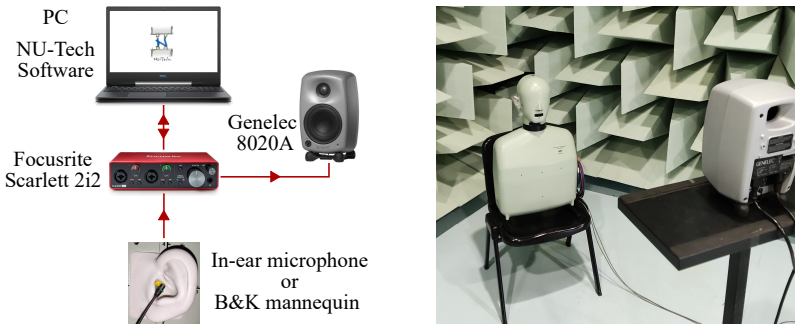
Experimental results

The experiments have been carried out to analyze two main comparisons listed below:

1. a comparison between the HRTFs measured with the in-ear miniature microphones placed in different points on the B&K mannequin ear canal and the HRTFs measured by the internal microphone of mannequin considering different positions of the sound source (cf. Figure 2.7);

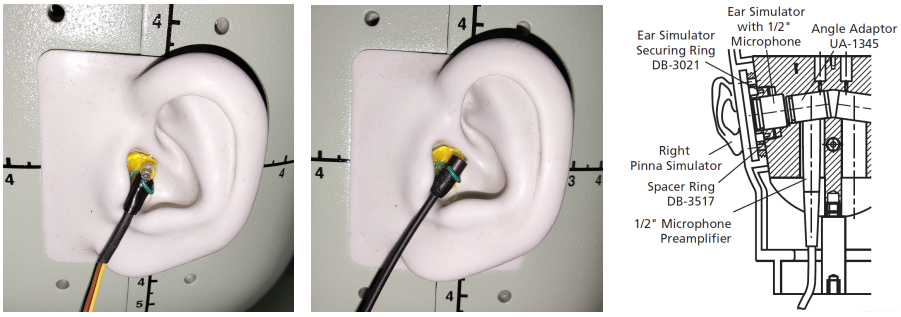
2. a comparison of individual HRTFs measured on five real subjects with the two in-ear microphones for different positions of the sound source (cf. Figure 2.8).

The two microphones have been settled on the left ear of the mannequin and of the subjects by means of a hook fixed on earplugs, as shown in Figure 2.7(a)-(b). The position of the microphone inside the head of the B&K HATS



(a) (b)

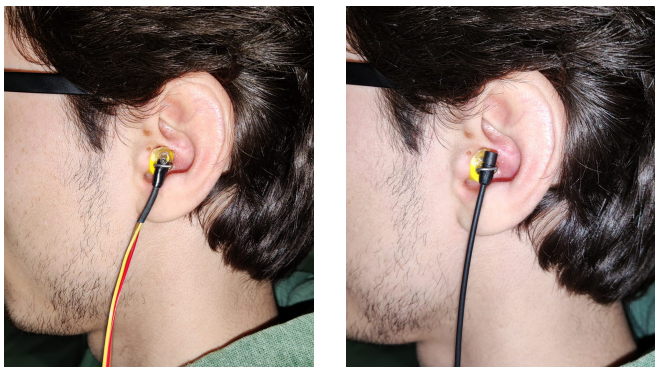
Figure 2.6: (a) scheme and (b) photo of the experimental setup used for HRTF measurements.



(a) (b) (c)

Figure 2.7: Photos of the (a) Knowles and (b) Sennheiser microphones on the mannequin ear, and (c) section of the B&K head [4]

.



(a) (b)

Figure 2.8: Photos of the (a) Knowles and (b) Sennheiser microphones on the ear of a real subject.

is shown in Figure 2.7(c). The measured HRTFs are compared in terms of frequency magnitude response and log-spectral distance (LSD) [63]. The LSD is calculated between the reference HRTF of the dummy ear $H_{\mathrm{HATS}}(k)$ and the one measured with the in-ear microphone $H_{\mathrm{MIC}}(k)$ as

$$\mathrm{LSD} = \sqrt{\frac{1}{k_2 - k_1 + 1} \sum_{k=k_1}^{k_2} \left[10 \log_{10} \frac{\left|H_{\mathrm{HATS}}(k)\right|^2}{\left|H_{\mathrm{MIC}}(k)\right|^2} \right]^2}, \tag{2.2}$$

where k_1 and k_2 delimit the frequency range within which the LSD is estimated, defined as $B = [k_1 \frac{F_\mathrm{s}}{K}, k_2 \frac{F_\mathrm{s}}{K}] = [100\,\mathrm{Hz}, 10\,\mathrm{kHz}]$, with $K = 4096$ the number of frequency bins for the FFT computation, and $F_\mathrm{s} = 48\,\mathrm{kHz}$ the sampling

frequency.

The first experiment aims at analyzing the differences between the HRTFs measured with the mannequin and the HRTFs measured with the in-ear microphones settled on the dummy ear. The in-ear microphones have been placed in the mannequin ear considering four different positions as shown in Figure 2.9(c). Figure 2.7 shows the real microphone placement for position P1. Also four different positions of the sound source have been taken into account varying the azimuth ϑ and the elevation φ, as shown in Figure 2.9(a)-(b), i.e., $\vartheta = 0°, 45°$, and $\varphi = 0°, 15°$.

The aim of this experiment is to investigate how much the microphone position influences the HRTF measurement and the results are shown in Figure 2.10 for both the Knowles (in the first column) and the Sennheiser (in the second column) microphones. It is evident that, apart from a different gain, the three microphones (i.e., the dummy ear, the Knowles, and the Sennheiser microphones) show similar HRTFs at the low frequencies. In fact, the trend of the frequency responses is almost the same for all four considered source positions. More in detail, in all the cases the HRTFs measured with the Sennheiser microphone are more similar to the ones measured with the mannequin up to 2 kHz, especially when the sound source is in front of the listener (cf. Figures 2.10(a)-(b)). Moreover, the position of the microphone affects the frequency response only at frequencies higher than 4 kHz. To better analyze these behaviors, LSD values have been calculated and reported in Table 2.1. These results confirm the performance of the Sennheiser microphone which reaches the lowest value in comparison with Knowles microphone for different microphone positions and different source positions. Furthermore, focusing on LSD value variations for the four microphone positions at the same source position, Sennheiser microphone shows the best performance in comparison with Knowles one demonstrating a better robustness through the microphone posi-

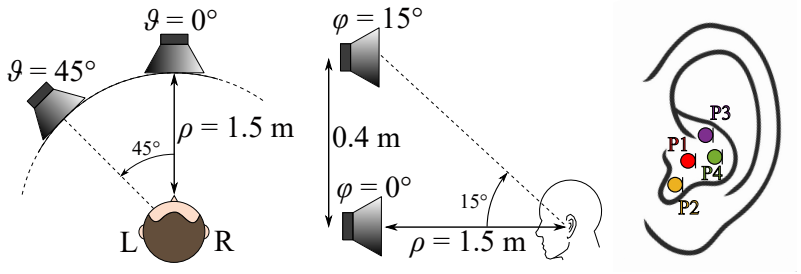


Figure 2.9: Positions of the sound source in terms of (a) azimuth and (b) and (c) positions of the in-ear microphone considered in experiment 1.

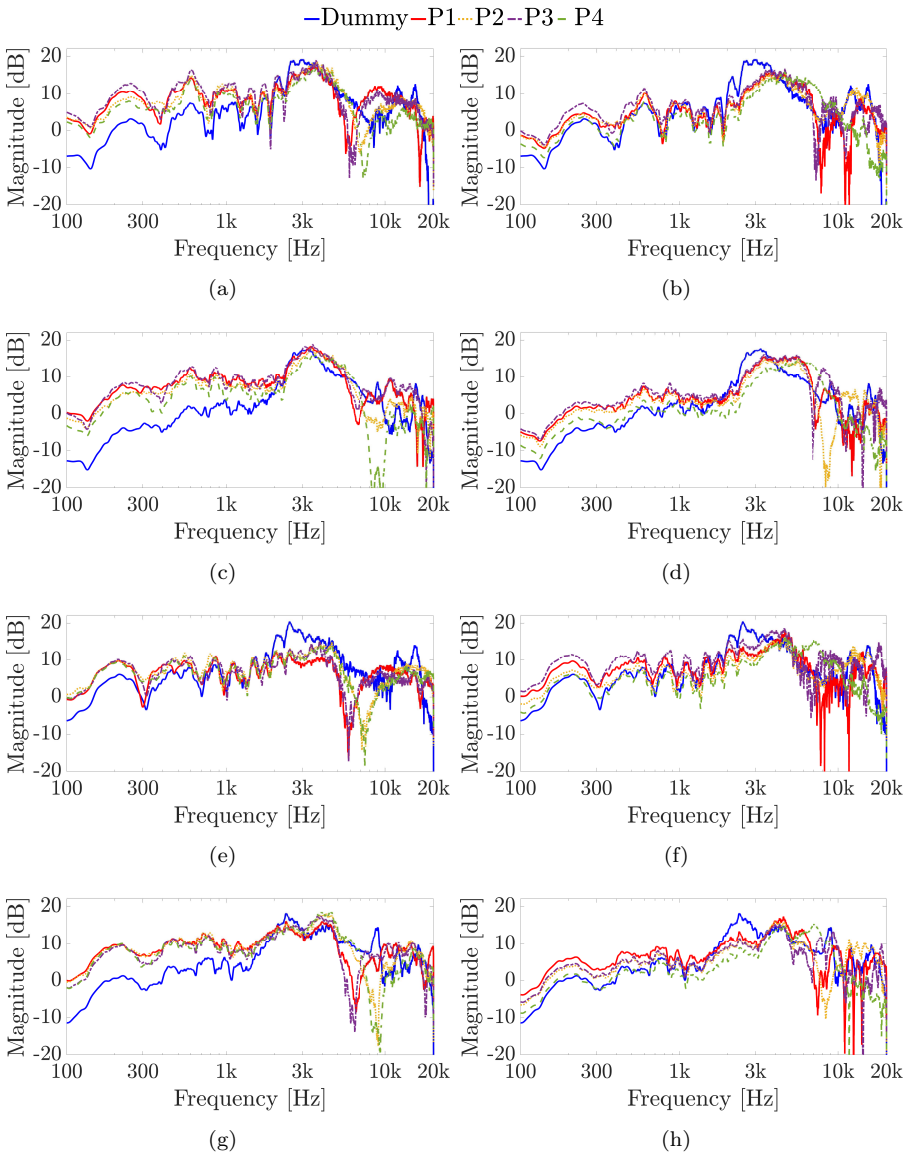


Figure 2.10: Experiment 1: HRTFs comparison considering the different in-ear microphone positions of Fig. 2.9(c) for sound source positions (a)-(b) with $\vartheta = 0°$ and $\varphi = 0°$, (c)-(d) with $\vartheta = 45°$ and $\varphi = 0°$, (e)-(f) with $\vartheta = 0°$ and $\varphi = 15°$, and (g)-(h) with $\vartheta = 45°$ and $\varphi = 15°$, using (a)-(c)-(e)-(g) the Knowles microphone and (b)-(d)-(f)-(h) the Sennheiser microphone.

tion.

The second experiment involves five real subjects wearing alternatively the

Table 2.1: LSD values (in dB) obtained in experiment 1. For each source position and microphone position, the lowest value of the LSD is highlighted.

Source pos.	Mic. pos.	Know.	Senn.
$\vartheta = 0°, \varphi = 0°$	P1	1.9	**1.4**
	P2	1.4	**1.1**
	P3	2.0	**1.4**
	P4	**1.5**	1.7
$\vartheta = 45°, \varphi = 0°$	P1	1.5	**1.1**
	P2	**1.6**	2.3
	P3	**1.3**	1.4
	P4	3.4	**1.5**

Source pos.	Mic. pos.	Know.	Senn.
$\vartheta = 0°, \varphi = 15°$	P1	1.9	**1.3**
	P2	1.5	**0.9**
	P3	1.5	**1.3**
	P4	**1.6**	1.7
$\vartheta = 45°, \varphi = 15°$	P1	1.9	**1.7**
	P2	2.6	**1.9**
	P3	2.1	**1.3**
	P4	3.2	**1.4**

two in-ear microphones. Results of experiment 2 are reported in Figure 2.11, considering two azimuth angles of the sound source, i.e., $\vartheta = 0°$ and $\vartheta = 45°$ and an elevation of $\varphi = 0°$, as shown in Figure 2.9(a). The measurements on subjects are compared with the HRTFs measured with the same microphone fixed on the dummy ear. In this case, the central position P1 of Figure 2.9(c) has been chosen for the acquisitions. As expected, each subject has a different frequency response due to the ear's shape but a comparison between

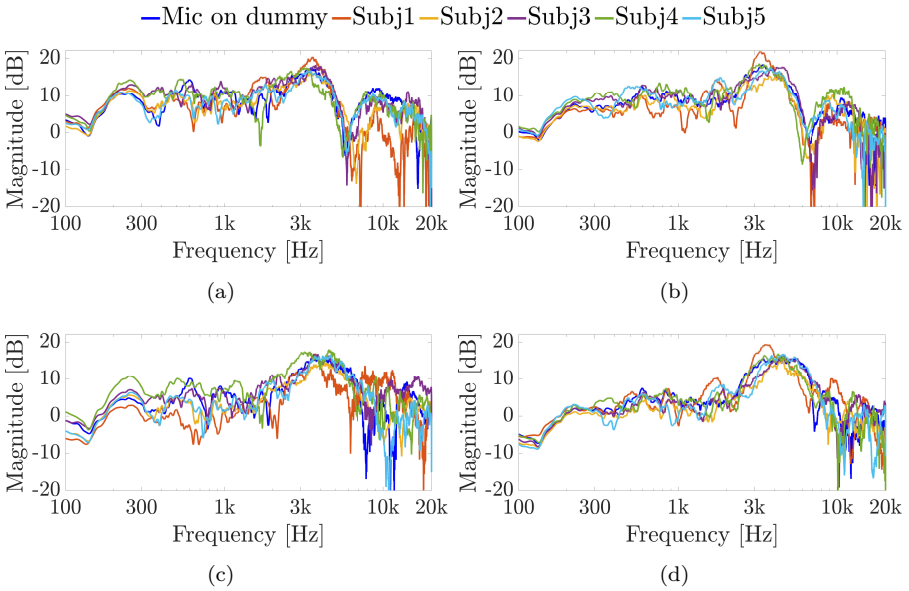


Figure 2.11: Experiment 2: HRTFs comparison of five different subjects with (a)-(b) the Knowles microphone and (c)-(d) the Sennheiser microphone for (a)-(c) $\vartheta = 0°$ and (b)-(d) $\vartheta = 45°$ and a fixed elevation of $\varphi = 0°$. The measurements are compared with the same microphone on the dummy ear.

Table 2.2: LSD values (in dB) obtained in experiment 2. For each source position and subject, the lowest value of the LSD is highlighted.

Source position	Subject	Knowles	Sennheiser
$\vartheta = 0°, \varphi = 0°$	Subj1	2.3	**1.9**
	Subj2	1.8	**1.3**
	Subj3	2.0	**1.2**
	Subj4	2.1	**1.3**
	Subj5	1.8	**1.5**
$\vartheta = 45°, \varphi = 0°$	Subj1	2.7	**1.3**
	Subj2	2.0	**1.3**
	Subj3	2.2	**1.3**
	Subj4	2.1	**1.1**
	Subj4	1.6	**1.5**

the two microphones and the dummy ear can be performed. In particular, the Sennheiser microphone exhibits frequency responses more similar to the dummy ear, especially in the case with $\vartheta = 45°$ (cf. Figure 2.11(d)) while Knowles microphone seems to have slight variations in comparison with the dummy ear. These results are confirmed by Table 2.2 that reports the LSD values comparing the HRTFs of the real subject with the HRTF measured with the mannequin at position P1. Sennheiser microphone shows the lowest value in comparison with the Knowles one and thus better performance.

To summarize, the experimental results have proven that the HRTFs are similar at low frequencies when different types of microphones are involved. In addition, the position of the microphone influences the HRTFs above 4 kHz, and the Sennheiser microphone allows to obtain frequency responses more similar to the ones measured by the dummy head. Finally, the individuaHRTFl HRTFs measured on real subjects have shown how the frequency responses change with different ears. Also in this case, the Sennheiser microphone has produced HRTFs more similar to the mannequin and more similar among the five subjects. However, the worst performance of Knowles is compensated by the price that is one order of magnitude lower than Sennheiser microphone. Future works will investigate the effectiveness of the different HRTF measurements through subjective tests, evaluating the immersive perception.

2.2 Binaural headphones rendering

The binaural synthesis is achieved by the real-time convolution of the input signal with the head-related impulse responses. The employed HRIRs must change with the movement of the source. The synthesis of continuous moving sound sources is achieved with a high directional resolution of the HRIRs database, i.e., with a dense number of measurement points [65]. For this rea-

son, the bigger the HRIRs database, the more realistic the binaural rendering. However, big databases are created by long and time-consuming measurements and need a lot of memory. This problem can be avoided by applying impulse response modeling methods, as in [66], or IRs interpolation techniques. The modeling approach aims at calculating the desired impulse response by means of models which take into account the acoustical properties of the environment, while the interpolation starts from a set of measured impulse responses to estimate the IR at any position in space.

2.2.1 Background on HRTFs interpolation

The HRTF interpolation procedure is a widely used method that allows for reducing the measurement points without affecting the 3D audio experience. Many approaches to impulse response interpolation can be found in the literature. A simple and popular method for the HRIRs interpolation is the bilinear technique, applied by Savioja *et al.* in [67]. The interpolated HRIR is calculated as the weighted mean of the measured impulse responses. Biscainho *et al.* [68] extended the bilinear technique by adding a new auxiliary inter-positional transfer function. However, this approach can be used only when the HRTFs of the database are measured for fixed-angle steps, for both azimuth and elevation. A similar method based on the weighted mean was proposed in [69], where the weights of the impulse responses that have to be interpolated are calculated through the vector base amplitude panning (VBAP) algorithm, firstly introduced by Pulkki in [70].

Hartung *et al.* [71] proposed a frequency domain approach that guarantees better performance than the time-domain methods, at the expense of the computational complexity. Another frequency-domain technique was proposed by Gamper in [72]. In this case, the interpolated HRTF is obtained by applying proper gains to the nearby HRTFs of the database. In this way, the interpolation allows obtaining only the magnitude frequency response, and a spherical head model is needed to derive the phase response.

In 2018, Garcia-Gomez and Lopez [5] proposed a new interpolation method applied to binaural room impulse responses that divides the impulse responses into early reflections and reverberant tail. The idea of splitting the impulse responses was already proposed by Kearney *et al.* in [73] for room impulse responses. In that case, the dynamic time warping (DTW) algorithm was applied for the temporal alignment and the reverberant part was synthesized following the approach of [74]. Differently, Garcia-Gomez and Lopez [5] applied the time-splitting to binaural room impulse responses, where the early reflections part is divided into two frequency bands. The low-frequency portions are linearly interpolated, while peak detection and alignment algorithms are reserved for

the high-frequency parts, which contain most of the information. This method presents a lower computational cost than the DTW algorithm of [73]. The time alignment for the interpolation of HRIRs plays an essential role and it is also applied in [75, 76].

2.2.2 Binaural synthesis of moving sound sources over headphones

In this section, a system for binaural synthesis of moving sound sources for reproduction over headphones is presented. The proposed approach is the one presented in [18], based on the HRTFs interpolation algorithm of [17]. Figure 2.12 shows the scheme of the proposed binaural system. The binaural output signals $y_{\mathrm{L}}(n)$ and $y_{\mathrm{R}}(n)$, which depend on the time index n, are reproduced by means of headphones and are obtained by the real-time convolution of the input signals $x_{\mathrm{L}}(n)$ and $x_{\mathrm{R}}(n)$ with the HRIR of the left ear $h_{\mathrm{L}}(n)$ and the HRIR of the right ear $h_{\mathrm{R}}(n)$, respectively. The HRTFs are calculated by the interpolation of measured HRTFs of a given database. In this case, the MIT Media Lab HRTFs database of [1] is employed.

The system considers the 3D setup of Figure 2.13, where the listener is located in the center of a sphere of fixed radius ρ. Hence, the HRIR $h(\vartheta, \varphi)$ depends only on the azimuth ϑ and on the elevation φ. The point $(0°, 0°)$ is in front of the listener and the azimuth ϑ grows clockwise. The interpolation algorithm considers a rectangular grid of measurement points on the surface of a sphere depending on the azimuth ϑ and the elevation φ. The points to be interpolated are identified following the flowchart of Figure 2.14. Starting from the virtual source position $(\vartheta_{\mathrm{v}}, \varphi_{\mathrm{v}})$, the algorithm obtains the related HRIR $h(\vartheta_{\mathrm{v}}, \varphi_{\mathrm{v}})$, employing a limited HRTFs database. Taking into account a generic HRTFs database, the measurement points depend on a set of possible elevation

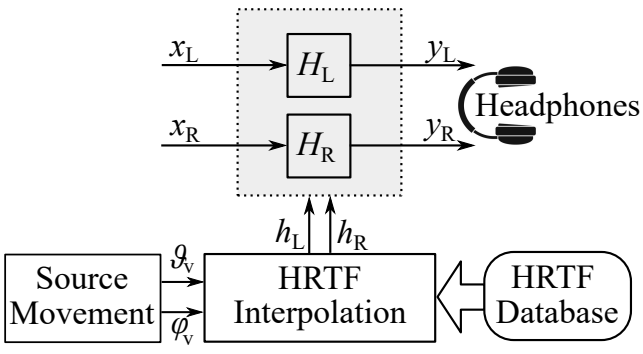


Figure 2.12: Scheme of the proposed binaural system for headphones reproduction.

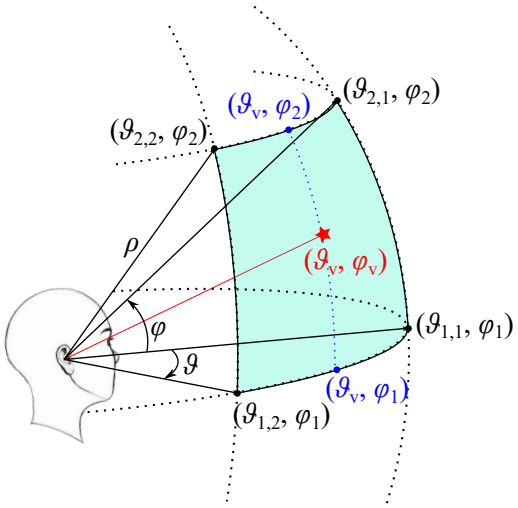


Figure 2.13: 3D setup used for the HRTF interpolation algorithm.

and azimuth angles. All the measurement points are saved and the virtual source position is compared with all the possible positions of the database in terms of elevation φ and azimuth ϑ, according to the scheme of Figure 2.14. If the virtual source position (ϑ_v, φ_v) corresponds to a measurement point, no interpolation is needed, and the respective HRTF is loaded. Differently, if the virtual source position is not a measurement position contained in the database, one, two, or three interpolations are required since the interpolation algorithm can be applied only to two HRTFs. More in detail, five different cases could occur if the desired HRTF $h(\vartheta_v, \varphi_v)$ is not in the database, i.e.,

1) the elevation φ_v is in the database, so two azimuth angles ϑ_1 and ϑ_2 closest to ϑ_v are found in the database, with $\vartheta_1 < \vartheta_v < \vartheta_2$: one interpolation along the azimuth is needed between $h(\vartheta_1, \varphi_v)$ and $h(\vartheta_2, \varphi_v)$ obtaining $h(\vartheta_v, \varphi_v)$;

2) the elevation φ_v is not in the database, and ϑ_v exists for both φ_1 and φ_2, that are the two elevation angles closest to φ_v, with $\varphi_1 < \varphi_v < \varphi_2$: one interpolation along the elevation is needed between $h(\vartheta_v, \varphi_1)$ and $h(\vartheta_v, \varphi_2)$ obtaining $h(\vartheta_v, \varphi_v)$;

3) the elevation φ_v is not in the database, and ϑ_v exists only for φ_1 and not for φ_2: two interpolations are needed, i.e.,
- along the azimuth between $h(\vartheta_{2,1}, \varphi_2)$ and $h(\vartheta_{2,2}, \varphi_2)$ obtaining $h(\vartheta_v, \varphi_2)$,
- along the elevation between $h(\vartheta_v, \varphi_1)$ and $h(\vartheta_v, \varphi_2)$ obtaining $h(\vartheta_v, \varphi_v)$,
with $\vartheta_{2,1} < \vartheta_v < \vartheta_{2,2}$;

4) the elevation φ_v is not in the database, and ϑ_v does not exist for φ_1 but

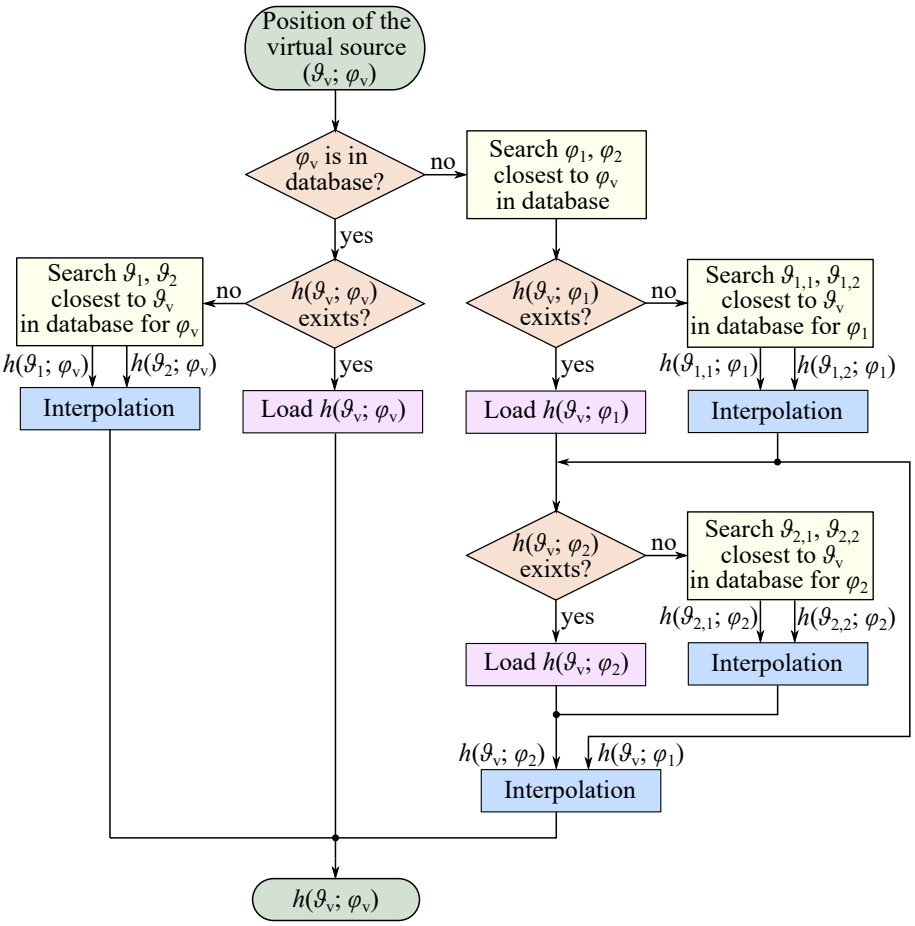


Figure 2.14: Flowchart of the proposed algorithm applied to obtain the HRTF $h(\vartheta_v, \varphi_v)$ related to the virtual source position (ϑ_v, φ_v).

exists for φ_2: two interpolations are needed, i.e.,
- along the azimuth between $h(\vartheta_{2,1}, \varphi_2)$ and $h(\vartheta_{2,2}, \varphi_2)$ obtaining $h(\vartheta_v, \varphi_2)$,
- along the elevation between $h(\vartheta_v, \varphi_1)$ and $h(\vartheta_v, \varphi_2)$ obtaining $h(\vartheta_v, \varphi_v)$,

with $\vartheta_{1,1} < \vartheta_v < \vartheta_{1,2}$;

5) the elevation φ_v is not in the database, and ϑ_v exists neither for φ_1 nor for φ_2: three interpolations are needed, i.e.,
- along the azimuth between $h(\vartheta_{1,1}, \varphi_1)$ and $h(\vartheta_{1,2}, \varphi_1)$ obtaining $h(\vartheta_v, \varphi_1)$,
- along the azimuth between $h(\vartheta_{2,1}, \varphi_2)$ and $h(\vartheta_{2,2}, \varphi_2)$ obtaining $h(\vartheta_v, \varphi_2)$,
- along the elevation between $h(\vartheta_v, \varphi_1)$ and $h(\vartheta_v, \varphi_2)$ obtaining $h(\vartheta_v, \varphi_v)$.

In every case, this procedure is applied for both the left ear and right ear, obtaining $h_L(\vartheta_v, \varphi_v)$ and $h_R(\vartheta_v, \varphi_v)$.

2.2.3 Interpolation algorithm

The single interpolation between two HRIRs $h_1(n)$ and $h_2(n)$ is obtained through the algorithm of [17], shown in Figure 2.15. The two impulse responses are split into two parts: the direct and early reflections $h_{\mathrm{e}i}(n)$ and the late reflections $h_{\mathrm{r}i}(n)$ as follows:

$$\begin{aligned} h_{\mathrm{e}i} &= h_i[1 : n_{\mathrm{t}}], \\ h_{\mathrm{r}i} &= h_i[n_{\mathrm{t}} + 1 : L_{\mathrm{h}}], \end{aligned} \tag{2.3}$$

where $i = 1, 2$, L_{h} is the sample length of the impulse responses, and n_{t} is the transition point between early and late reflections that corresponds to the mixing time t_{m} as $n_{\mathrm{t}} = \lfloor t_{\mathrm{m}} \cdot F_{\mathrm{s}} \rceil$, where F_{s} is the sample rate. Although the mixing time is often defined for room impulse responses, where the late reflections contain the reverberant tail, in this case, the late reflections of the HRIRs mostly include the reflections caused by the pinna, head, and torso and the time splitting is used to focus most of the elaboration on the first part of the HRIRs, leading to a reduced computation. The early reflections $h_{\mathrm{e}i}(n)$ are then divided into two frequency bands using third-order lowpass (LP) and highpass (HP) Butterworth IIR filters with a cutoff frequency of 150 Hz. The low-frequency part of the early reflections $h^{\mathrm{L}}_{\mathrm{e}i}(n)$ and the late reflections $h_{\mathrm{r}i}(n)$ are linearly interpolated, obtaining $h^{\mathrm{L}}_{\mathrm{e_v}}(n)$ and $h_{\mathrm{r_v}}(n)$, respectively. Differently, the high-frequency part of the early reflections $h^{\mathrm{H}}_{\mathrm{e}i}(n)$ is elaborated through a peak detection and alignment algorithm before the linear interpolation. The early reflections of the interpolated impulse response $h_{\mathrm{e_v}}(n)$ is computed as the sum between the high-frequency and the low-frequency part as follows:

$$h_{\mathrm{e_v}}(n) = h^{\mathrm{L}}_{\mathrm{e_v}}(n) + h^{\mathrm{H}}_{\mathrm{e_v}}(n). \tag{2.4}$$

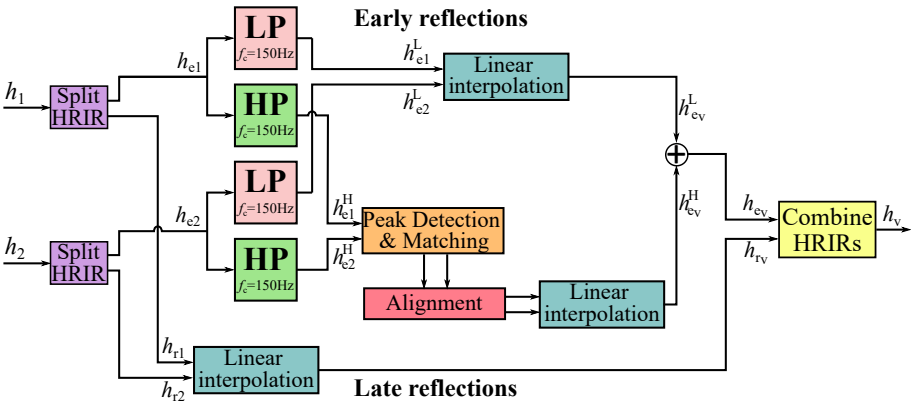


Figure 2.15: Scheme of the HRTFs interpolation algorithm.

Finally, the interpolated impulse response $h_{\mathrm{v}}(n)$ is obtained by concatenating the early and late reflections as

$$h_{\mathrm{v}}(n) = [h_{\mathrm{e_v}}(n);\ h_{\mathrm{r_v}}(n)]. \tag{2.5}$$

The algorithms of linear interpolation, peak detection, and alignment are described in the following.

Linear interpolation

Linear interpolation is the main algorithm of the proposed approach and is obtained by applying the following equation:

$$h_{\mathrm{v}}(n) = h_1(n) + \Big(h_2(n) - h_1(n)\Big)\frac{\xi_{\mathrm{v}} - \xi_1}{\xi_2 - \xi_1}, \tag{2.6}$$

where $h_1(n)$ and $h_2(n)$ are the impulse responses measured at the angles ξ_1 and ξ_2 respectively, and $h_{\mathrm{v}}(n)$ is the estimated IR at position ξ_{v}, with $\xi_1 < \xi_{\mathrm{v}} < \xi_2$. The angle ξ could represent the azimuth ϑ in the case of horizontal interpolation, or the elevation φ in the case of vertical interpolation.

Peak detection and matching

The peak detection algorithm is applied to the high-frequency part of the early reflections because it is the part that contains most of the information on the impulse responses. The aim of this algorithm is to identify S peaks of $h^{\mathrm{H}}_{\mathrm{e1}}(n)$ related to S peaks of $h^{\mathrm{H}}_{\mathrm{e2}}(n)$. The flowchart of the algorithm is shown in Figure 2.16. Firstly, S peaks of $h^{\mathrm{H}}_{\mathrm{e1}}(n)$ are detected considering a minimum peak distance of $d_{\mathrm{p}} = 100$ samples and a minimum peak height of $\gamma_1 = \max(h_1)/20$, chosen after empirical studies. In this way, a vector $\bar{\mathbf{s}}_1 = [s_{1,1}...s_{1,k}...s_{1,S}]$, containing the samples at which the peaks of $h^{\mathrm{H}}_{\mathrm{e1}}(n)$ are located, is obtained. An example is shown at the top of Figure 2.17, where $S = 2$ peaks are found in $h^{\mathrm{H}}_{\mathrm{e1}}(n)$. For each kth peak of $h^{\mathrm{H}}_{\mathrm{e1}}(n)$, S_k temporary peaks of $h^{\mathrm{H}}_{\mathrm{e2}}(n)$ are looked for in the portion $h^{\mathrm{H}}_{\mathrm{e2}}[s_{1,k} - d_{\mathrm{p}}/2 : s_{1,k} + d_{\mathrm{p}}/2]$, as shown in Figure 2.17, with no constraint regarding the distance of the peaks, but with a threshold defined as $\gamma_2 = 0.8\gamma_1$. If no peaks of $h^{\mathrm{H}}_{\mathrm{e2}}(n)$ related to the kth peak of $h^{\mathrm{H}}_{\mathrm{e1}}(n)$ are found, i.e., if $S_k = 0$, the kth peak is deleted. The samples indexes of the temporary peaks of $h^{\mathrm{H}}_{\mathrm{e2}}(n)$ are saved in the vector $\bar{\mathbf{t}}_k = [t_{k,1}...t_{k,j}...t_{k,S_k}]$. In the example of Figure 2.17, two peaks of $h^{\mathrm{H}}_{\mathrm{e1}}(n)$ are detected, while, in the second impulse response, $S_1 = 6$ and $S_2 = 2$ temporary peaks are found. Consequently, the proposed algorithm allows the extraction of one peak for each vector $\bar{\mathbf{t}}_k$ that matches with the kth peak of $h^{\mathrm{H}}_{\mathrm{e1}}(n)$. For each kth peak of $h^{\mathrm{H}}_{\mathrm{e1}}(n)$, a window

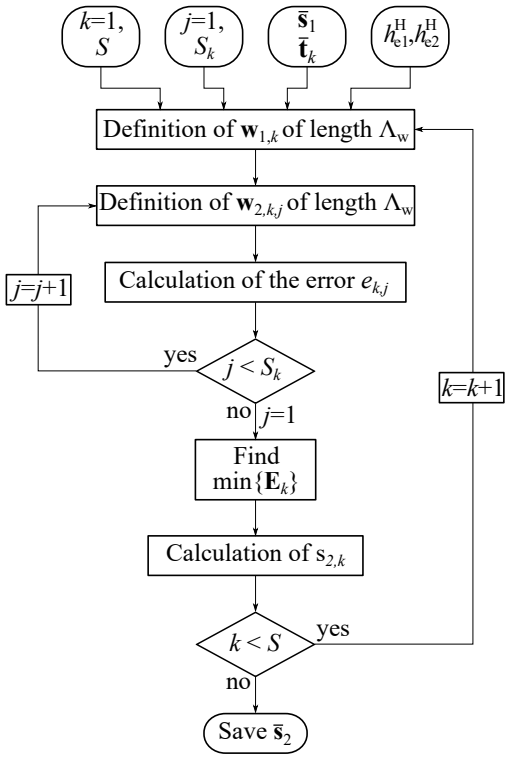


Figure 2.16: Flowchart of the proposed peak detection and matching algorithm.

is defined as follows:

$$\mathbf{w}_{1,k} = h_{\mathrm{e1}}^{\mathrm{H}}\left[\left(s_{1,k} - \frac{\Lambda_{\mathrm{w}}}{2} + 1\right) : \left(s_{1,k} + \frac{\Lambda_{\mathrm{w}}}{2}\right)\right], \tag{2.7}$$

where $k = 1, ..., S$, Λ_{w} is the length of the window $\mathbf{w}_{1,k}$ set to $\Lambda_{\mathrm{w}} = 100$ and $s_{1,k}$ is the sample corresponding to the kth peak of $h_{\mathrm{e1}}^{\mathrm{H}}(n)$. Similarly, for each jth temporary peak of the kth portion of $h_{\mathrm{e2}}^{\mathrm{H}}(n)$, the window $\mathbf{w}_{2,k,j}$ is obtained as follows:

$$\mathbf{w}_{2,k,j} = h_{\mathrm{e2}}^{\mathrm{H}}\left[\left(t_{k,j} - \frac{\Lambda_{\mathrm{w}}}{2} + 1\right) : \left(t_{k,j} + \frac{\Lambda_{\mathrm{w}}}{2}\right)\right], \tag{2.8}$$

where $j = 1, ..., S_k$, Λ_{w} is the length of the window $\mathbf{w}_{2,k,j}$ (that is the same of the window $\mathbf{w}_{1,k}$) and $t_{k,j}$ is the sample corresponding to the jth temporary peak of the kth portion of $h_{\mathrm{e2}}^{\mathrm{H}}(n)$. Then, for each kth peak of $h_{\mathrm{e1}}^{\mathrm{H}}(n)$, an error vector $\mathbf{E}_k = [e_{k,1}...e_{k,j}...e_{k,S_k}]$ of length S_k is defined. Therefore, for each jth

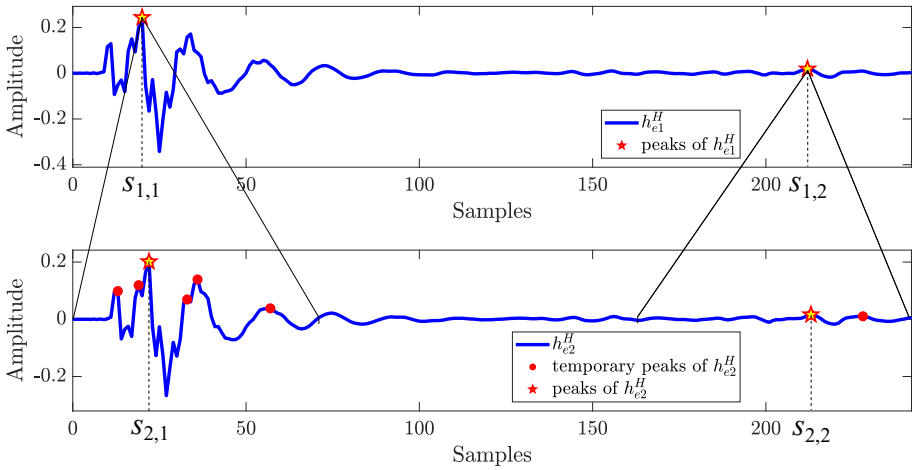


Figure 2.17: Example of the peaks found by the peak detection and matching algorithm applied to two HRIRs related to the angles $\vartheta_1 = 10°$ and $\vartheta_2 = 20°$, respectively, with elevation $\varphi = 0°$.

peak, an element $e_{k,j}$ of the vector $\mathbf{E}_k$ is calculated as follows:

$$e_{k,j} = \sum_{n=1}^{\Lambda_w} \left| \mathbf{w}_{1,k}(n) - \mathbf{w}_{2,k,j}(n) \right|. \tag{2.9}$$

The error function $\mathbf{E}_k$ is subject to a minimization process in order to find the index of the peak of $h_{e2}^H(n)$ that has the best match with the kth peak of $h_{e1}^H(n)$, i.e.,

$$j_{opt} = \arg\min_j \{\mathbf{E}_k\}. \tag{2.10}$$

Finally, the kth peak of $h_{e2}^H(n)$ related to the kth peak of $h_{e1}^H(n)$ is derived as follows:

$$s_{2,k} = t_{j_{opt}}, \tag{2.11}$$

where $s_{2,k}$ is the kth element of the vector $\bar{\mathbf{s}}_2 = [s_{2,1} ... s_{2,k} ... s_{2,S}]$ that is S long and contains the samples of the final peaks of $h_{e2}^H(n)$ that have the best match with the S peaks of $h_{e1}^H(n)$.

Alignment

The alignment algorithm is required after the peak detection to make the S peaks of the two impulse responses aligned. The two IRs are divided into S energy blocks made up around the selected peaks. For each pair of blocks a gravity point (GP), that is the exact sample index where both peaks of the IRs

must coincide, is calculated by the linear interpolation as

$$\mathrm{GP}_k = \left\lfloor s_{1,k} + (s_{2,k} - s_{1,k}) \frac{\xi_\mathrm{v} - \xi_1}{\xi_2 - \xi_1} \right\rceil, \tag{2.12}$$

where $s_{1,k}$ and $s_{2,k}$ are the samples at which the kth peak is located in $h^\mathrm{H}_{\mathrm{e}1}(n)$ and $h^\mathrm{H}_{\mathrm{e}2}(n)$, respectively. Finally, a warping process is used to stretch or compress each kth block signal consistent with the relative position between GP_k and the kth peak. The kth signal block is compressed by deleting the extra samples, and it is stretched by adding new samples following a linear interpolation. After the alignment, the IRs can be interpolated following the Equation (2.6).

2.2.4 Mixing time calculation for BRIRs interpolation

The division into early and late reflections makes the interpolation algorithm suitable also for binaural room impulse responses. In fact, BRIRs are measured in real reverberant environments, so they take into account both HRIR and room impulse response. In this case, the late reflections contain the reverberant tail of the impulse response. Moreover, the mixing time (MT) depends on the source distance ρ and could vary depending on the type of environment (e.g., volume and shape of the room) [77]. For this reason, an automatic calculation of the mixing time is added to the interpolation algorithm in [19]. The mixing time can be calculated in two different ways, the model-based estimator [78,79] and the signal-based estimator [80–83]. The model-based estimators have a simple implementation and define a range of acceptable values for the mixing time. They are not suitable for automatic MT calculation because they require knowledge of room characteristics. Contrarily, signal-model predictors are based on different assumptions about the sound pressure distribution. For this work, the signal-model approach of Primavera *et al.* [83] is applied, based on the Jarque-Bera test [82]. The Jarque-Bera coefficients are calculated inside a sliding window.

More in detail, the impulse response is crossed sample by sample by a fixed-length window and the Jarque-Bera coefficients (JB) are calculated inside the window as follows:

$$\mathrm{JB} = \frac{\Lambda_\mathrm{JB}}{6} \Big(\varsigma^2 + \frac{1}{4} (\kappa - 3)^2 \Big), \tag{2.13}$$

where κ is the kurtosis, defined as the fourth order zero-lag cumulant of a process as

$$\kappa = \frac{\mathrm{E}\Big(h - \mu_\mathrm{h} \Big)^4}{\sigma^4_\mathrm{h}} - 3, \tag{2.14}$$

ς is the skewness of a normal distribution, i.e.,

$$\varsigma = \frac{\mathrm{E}\Big(h - \mu_{\mathrm{h}}\Big)^3}{\sigma_{\mathrm{h}}}, \tag{2.15}$$

Λ_{JB} is the length of the window, μ_{h} and σ_{h} are the mean and the standard deviation of the observed samples of the impulse response h, respectively, and $\mathrm{E}(\cdot)$ represents the expected value. The window size used to analyze the impulse response is $\Lambda_{\mathrm{JB}} = 960$ samples, i.e., 0.02 ms, with a sample rate of $F_{\mathrm{s}} = 48$ kHz. The mixing time t_{m} is selected as the first sample in which the Jarque-Bera coefficients are below 0.005. In Figure 2.18(a), the computation of Jarque-Bera coefficients of an IR measured at the distance of 1 meter is shown. The red line represents the Jarque-Bera coefficients and the blue line is the measured impulse response. In this case, the mixing time is equal to 8.3 ms (i.e., when the Jarque-Bera coefficients decrease below 0.005). The same computation has been applied to a BRIR measured at 2 meters, resulting in a mixing time of 20 ms, as shown in Figure 2.18(b). The difference is due to the highest number of reflections with a larger amplitude and the different distances of the two BRIRs.

2.2.5 Real-time implementation of the binaural system

The proposed algorithm has been implemented as a plugin of the NU-Tech software [84], which is a platform specifically developed to test and tune real-time DSP algorithms through a PC workbench. The developer can write his/her own plugins, called NUTSs (NU-Tech Satellites), in C++ and plug them into the GUI to test the results on a common PC. The internal parameters of every plugin can be adjusted using the RTWatch (RealTime Watch).

Figure 2.19 shows the NU-Tech interface used for experimental tests. The

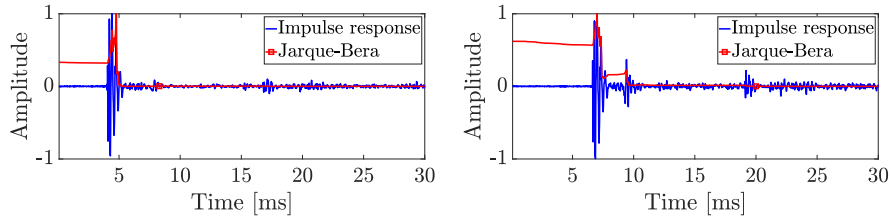


Figure 2.18: Mixing time calculation for BRIRs of the right ear measured at the position ($\vartheta = 30°, \varphi = 0°$) at distance of (a) $\rho = 1$ m, resulting $t_{\mathrm{m}} = 8.3$ ms, and (b) $\rho = 2$ m, resulting $t_{\mathrm{m}} = 20$ ms.

proposed algorithm is implemented by the NUTS "HRTF Interp", realized as a standard C++ dll file. The RTWatch (table at the left bottom of Figure 2.19) shows eight parameters listed and explained below:

- **Bypass**: boolean variable that allows the bypass of the algorithm procedure; in the case of "TRUE", the algorithm is ignored and the input signals are replied at the outputs.
- **Azimuth**: allows to select the azimuth angle of the virtual sound source ϑ_v in degrees from 0° to 359°.
- **Elevation**: allows to set the elevation angle of the virtual sound source φ_v from −40° to 90°, according to the database structure.
- **HRTF length**: reports the length of the used functions; here the length of the impulse responses of the employed database must be declared.
- **File Directory**: contains the path of the employed database in .dat format;
- **Number of interpolation**: shows how many interpolations have been performed.
- **Method**: set to 0 if the proposed algorithm is applied while set to 1 if the reference algorithm of [5] is executed.
- **Test**: used only to perform the experimental tests. It allows the type of test to be selected in order to consider a reduced database, in which the HRIR of the desired position is excluded. This is necessary to compare

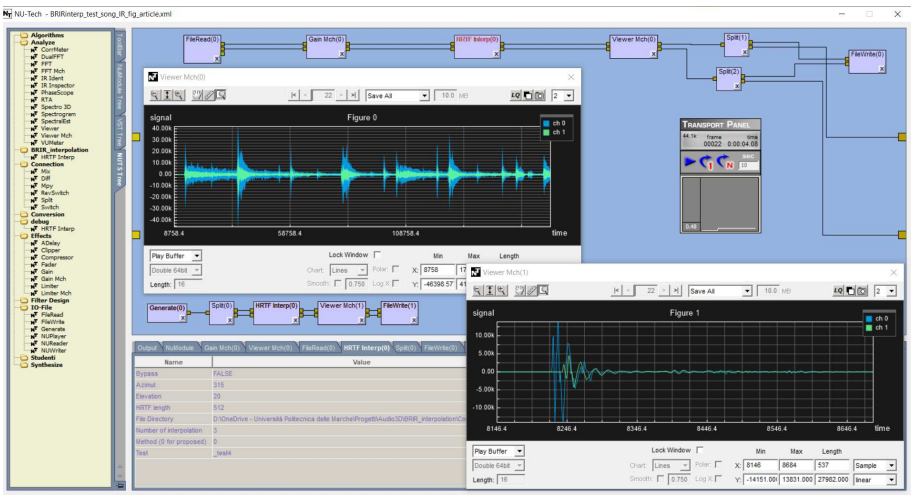


Figure 2.19: NU-Tech setup for listening tests.

the interpolated IRs with the measured ones. If this field is empty, the whole database is considered.

When all the parameters are set, it is sufficient to press the play button on the Transport Panel to start the real-time computation. Starting from the desired elevation and azimuth of the virtual sound source, which can be modified during the reproduction, the NUTS loads the impulse responses that have to be interpolated from the database and calculates the desired HRIR (if the selected angles are already included in the database, the HRIR is simply loaded). Then, the NUTS allows the filtering of the input signal with the desired HRIR using the "Overlap and Save" method [85]. For efficient computation, Intel Integrated Performance Primitives (IPP) libraries have been used [86].

2.2.6 Experimental results on HRTFs

The interpolation algorithm described in Section 2.2.3 has been tested using HRTFs. Taking into account the fact that HRTFs are measured inside an anechoic environment always at the same distance, a fixed mixing time of $t_{\mathrm{m}} = 5\,\mathrm{ms}$ has been used to include the main peak and the first reflections of the impulse response. The experimental tests were carried out employing the MIT Media Lab database [1], where the HRIRs are measured with a KEMAR dummy head microphone considering a fixed distance between the listener and the source equal to $\rho = 1.4$ m in 710 different positions, and a sample rate of 44.1 kHz. Each impulse response has a length of 512 samples, so an order of $N = 511$, and depends on the azimuth ϑ and the elevation φ, with $0° \leq \vartheta \leq 359°$ and $-40° \leq \varphi \leq 90°$. The interpolated impulse response is obtained by the interpolation of the two impulse responses related to the two adjacent measurement points, defined by the employed database. To perform experimental tests, four scenarios have been considered:

- **scenario 1 (S1):** $\vartheta_{\mathrm{v}} = 15°$ and $\varphi_{\mathrm{v}} = -20°$, interpolation between $h(10°, -20°)$ and $h(20°, -20°)$;
- **scenario 2 (S2):** $\vartheta_{\mathrm{v}} = 144°$ and $\varphi_{\mathrm{v}} = 30°$, interpolation between $h(138°, 30°)$ and $h(150°, 30°)$;
- **scenario 3 (S3):** $\vartheta_{\mathrm{v}} = 225°$ and $\varphi_{\mathrm{v}} = -20°$, interpolation between $h(220°, -20°)$ and $h(230°, -20°)$;
- **scenario 4 (S4):** $\vartheta_{\mathrm{v}} = 315°$ and $\varphi_{\mathrm{v}} = 20°$, interpolation between $h(310°, 20°)$ and $h(320°, 20°)$.

The interpolation algorithm has been validated by comparing its performance with those obtained using the method of Garcia-Gomez and Lopez [5], and with the measured HRIRs, in terms of objective and subjective results.

Objective results

Objective tests have been carried out in order to evaluate the effectiveness of the interpolation algorithm. In particular, Figure 2.20 shows the interpolated impulse responses for the four scenarios, in comparison to the measured impulse responses and the HRIRs interpolated with the Garcia-Gomez algorithm [5]. The same results are shown in the frequency domain in Figure 2.21. Looking at the impulse responses of Figure 2.20, the proposed algorithm shows an improvement in the peak detection and alignment algorithms. In fact, some HRIRs interpolated with the method of [5] are delayed in respect to the measured ones (cf. Figures 2.20(d)-2.20(f)), especially when $90° < \vartheta < 270°$, i.e., when the virtual source is behind the listener (scenarios 2 and 3). This mismatching affect also the responses in the frequency domain, as it can be seen in Figures 2.21(d) and 2.21(f), where the magnitude frequency responses of the HRIR interpolated with Garcia-Gomez algorithm are completely different from the measured ones, while the HRTF interpolated by the proposed method is more similar to the measured HRTF. Regarding scenario 1, the HRTF interpolated with the proposed technique fits perfectly with the measured HRTF, in terms of both impulse and frequency response, while the reference algorithm shows slightly worse results. Finally, for scenario 4, the two methods behave in a similar way and the interpolated HRTFs are well-estimated.

For a more detailed assessment, a comparison between the interpolated frequency response and the measured one has been executed, considering a full turn of 360° along the azimuth for different values of elevation. In particular, the mean squared error (MSE), reported in Table 2.3, between the measured magnitude frequency response H_{m} and the interpolated one H_{I} is calculated

Table 2.3: Mean squared error between the measured and the interpolated HRTFs, comparing the proposed algorithm with the one of Garcia-Gomez [5]. The bold numbers are the lowest MSE values.

Mean Squared Error MSE [dB]				
Elevation	**Garcia-Gomez [5]**		**Proposed**	
	Left	**Right**	**Left**	**Right**
−30°	-18	-25	**-24**	**-27**
−20°	-17	-23	**-23**	**-26**
−10°	-19	-24	**-22**	**-26**
0°	-16	-20	**-25**	**-26**
10°	-16	-16	**-24**	**-25**
20°	-14	-18	**-26**	**-28**
30°	-15	-16	**-27**	**-28**

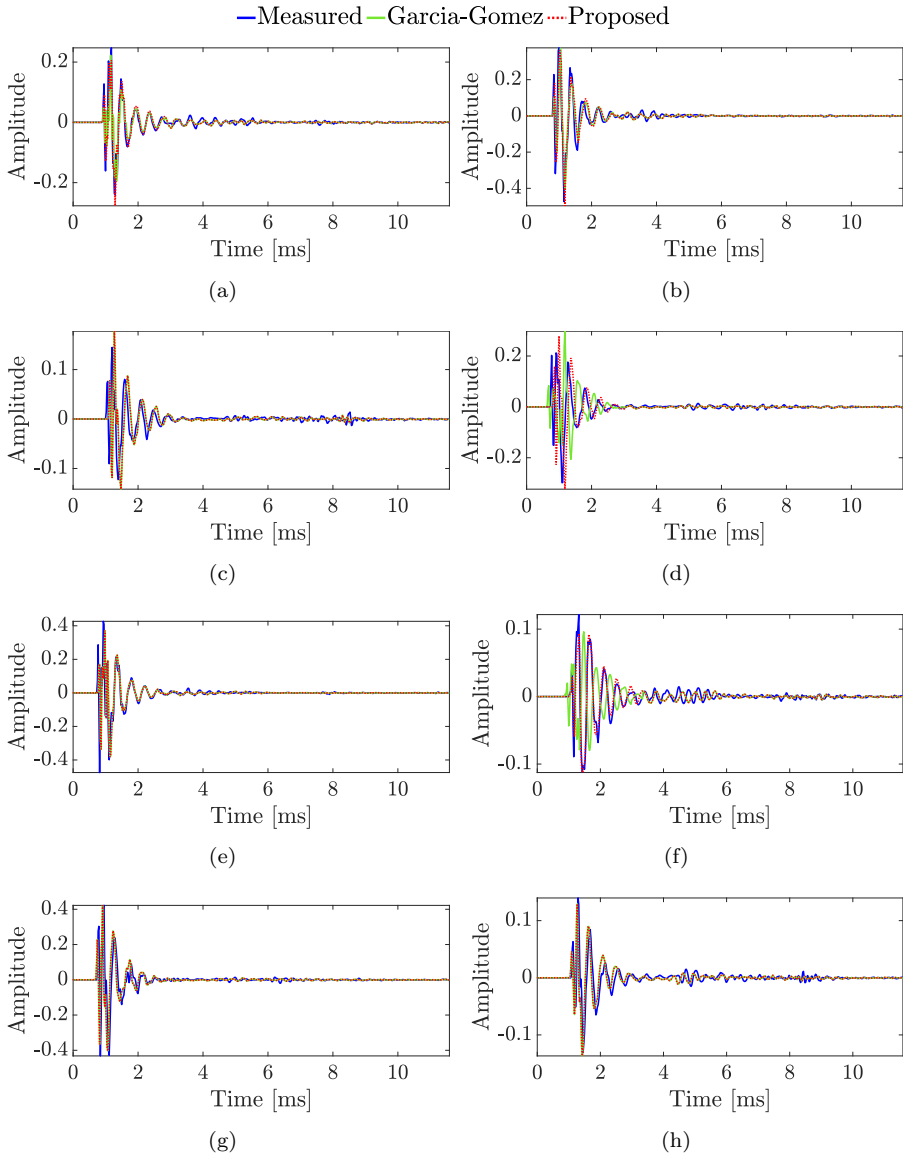


Figure 2.20: Comparison between the measured impulse response, the interpolated impulse response following the Garcia-Gomez algorithm [5], and the interpolated impulse responses with the proposed algorithm considering (a),(c),(e),(g) left ear, and (b),(d),(f),(h) right ear for the scenario 1 (first row), scenario 2 (second row), scenario 3 (third row), scenario 4 (fourth row).

for a fixed elevation angle φ as

$$\mathrm{MSE}(\varphi) = \frac{1}{Q}\sum_{q=1}^{Q}\left[\frac{1}{K}\sum_{k=0}^{K-1}\left(\left|H_{\mathrm{m}}(\varphi,\vartheta_q,k)\right| - \left|H_{\mathrm{I}}(\varphi,\vartheta_q,k)\right|\right)^2\right], \qquad (2.16)$$

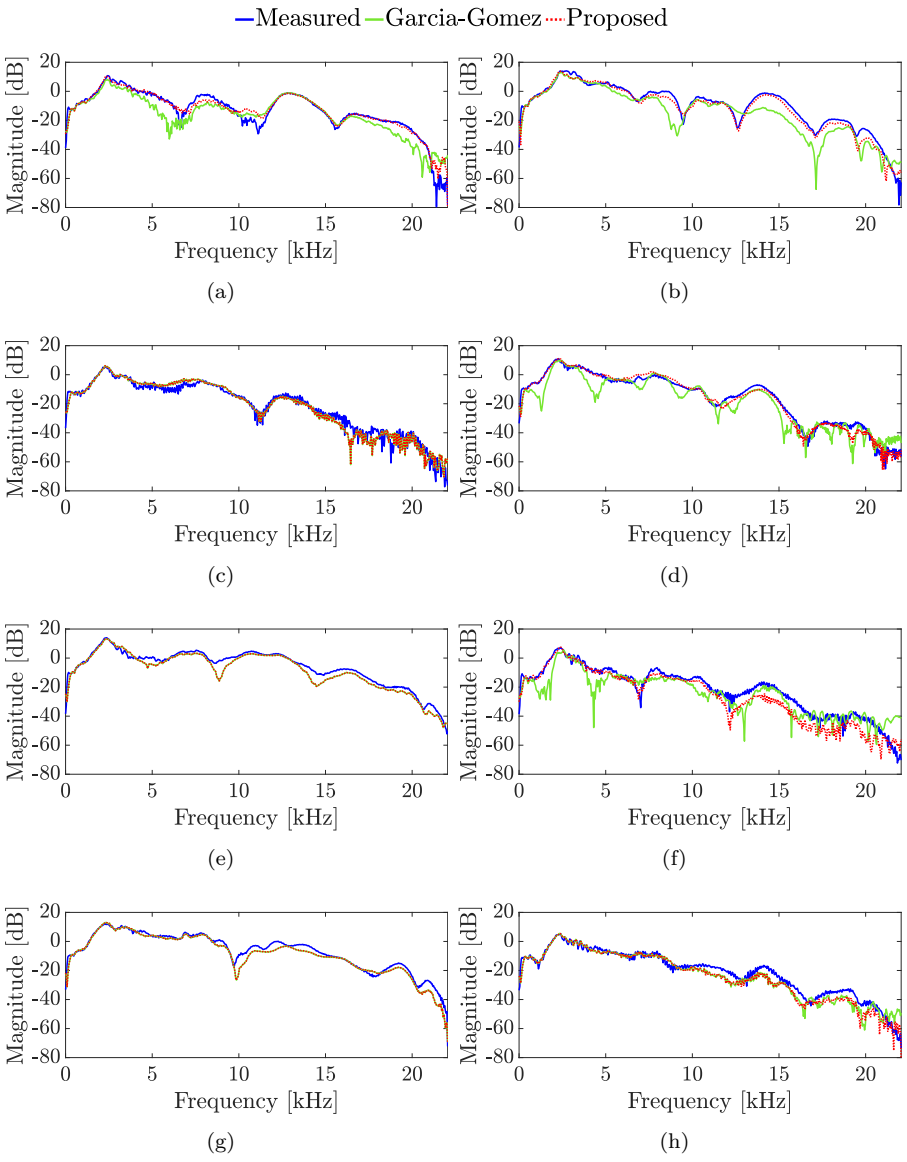


Figure 2.21: Comparison between the measured frequency response, the interpolated frequency response following the Garcia-Gomez algorithm [5], and the interpolated frequency responses with the proposed algorithm considering (a),(c),(e),(g) left ear, and (b),(d),(f),(h) right ear for the scenario 1 (first row), scenario 2 (second row), scenario 3 (third row), scenario 4 (fourth row).

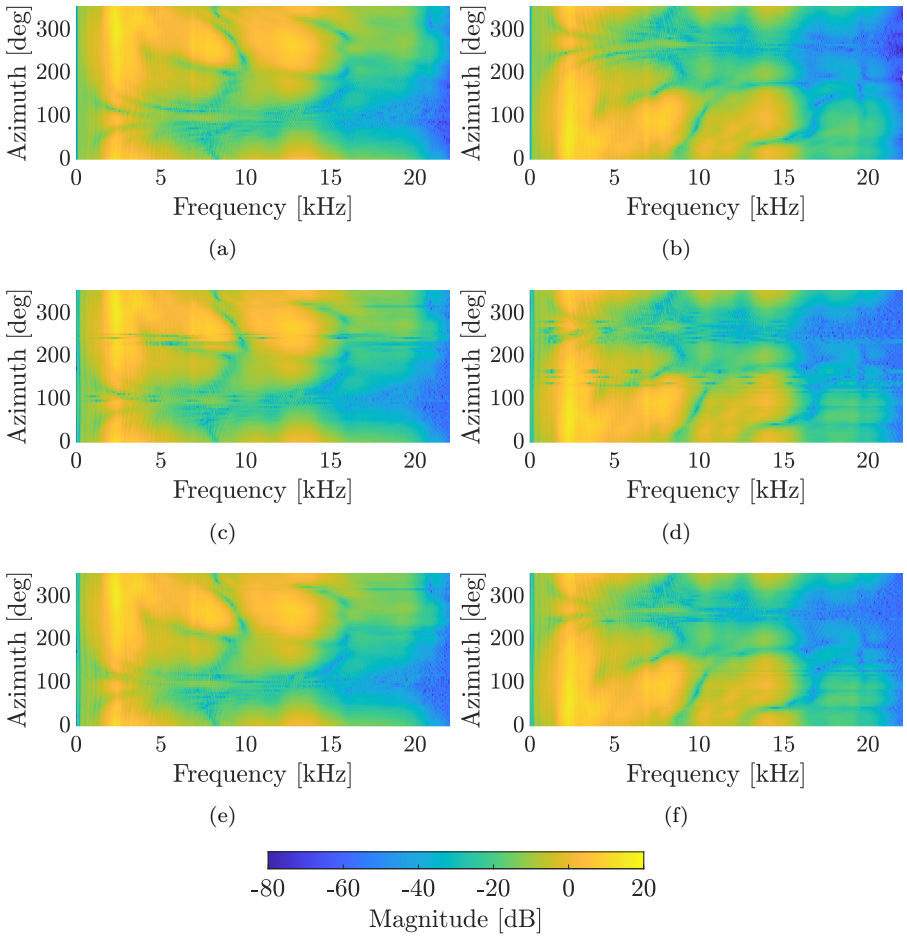


Figure 2.22: Comparison between (a)-(b) the measured frequency responses, (c)-(d) the frequency responses interpolated with the Garcia-Gomez algorithm [5], and (e)-(f) the frequency responses interpolated with the proposed algorithm for the left and the right ears, respectively, varying the azimuth ϑ and considering an elevation of $\varphi = 0°$.

where $K = 1024$, that is the number of frequency bins, and Q is the number of considered angles ϑ_q along the azimuth for a full 360° turn for a certain elevation φ. To perform the comparison with the measured responses, only the azimuth angles provided by the database are considered. For each azimuth angle, the interpolated impulse response is obtained by the interpolation of the two impulse responses related to the two adjacent positions. In Table 2.3, different angles of elevation (i.e., $\varphi = -30°, -20°, -10°, 0°, 10°, 20°, 30°$) are considered for the MSE calculation. The proposed algorithm exhibits the lowest values of error, proving its effectiveness. This result is confirmed in Figure 2.22, where a magnitude frequency responses comparison is shown, varying the azimuth and considering an elevation of $\varphi = 0°$. The proposed technique produces a uniform, similar to the one obtained with the measured HRTFs, while the reference interpolation shows discontinuities, especially for azimuth angles between 200° and 300°.

Subjective results

The interpolated HRTFs have been used to filter the stereo soundtracks in real time using the NU-Tech software [84], following the scheme of Figure 2.12. The filtered tracks have been saved and evaluated by a listening panel and have been reproduced over headphones. The listening panel was composed of 26 listeners, 12 men and 14 women aged between 21 and 56, of which 15 were expert listeners and 11 were not. Listeners are defined as experts if they have already practiced subjective listening tests and are able to perceive relatively subtle degradations, as declared in [87]. Each test consisted of 3 tracks lasting 30 s, with the same song filtered by: the measured HRTFs, the interpolated ones with the reference algorithm of [5], and the interpolated HRTFs with the proposed method. According to ITU-R BS.1289-2 [87], for each track the listeners have to evaluate:

- **Source localization**: the listeners can specify the source localization by choosing among 4 possibilities: northeast (scenario 1 with $\vartheta_v = 15°$ and $\varphi_v = -20°$), southeast (scenario 2 with $\vartheta_v = 144°$ and $\varphi_v = 30°$), southwest (scenario 3 with $\vartheta_v = 225°$ and $\varphi_v = -20°$), and northwest (scenario 4 with $\vartheta_v = 315°$ and $\varphi_v = 20°$).
- **Spatial Impression**: the listeners must evaluate how much the performance appears to take place in an appropriate spatial environment.
- **Transparency**: the listeners must judge if all details of the performance can be clearly perceived.

Both spatial impression and transparency are evaluated with a score between 1 (bad) and 5 (excellent), according to the unipolar discrete five-grade scale

Table 2.4: List of soundtracks used for the listening tests

Genre	Author	Sound Track	Scenarios
Pop	Daft Punk	Get Lucky	1,4
Rock	Pink Floyd	Money	2,4
Jazz	Sarah Vaughan	Lullaby of Birdland	1,3
Classical	Tchaikovsky	The Nutcracker Op. 71 Act I	2,3

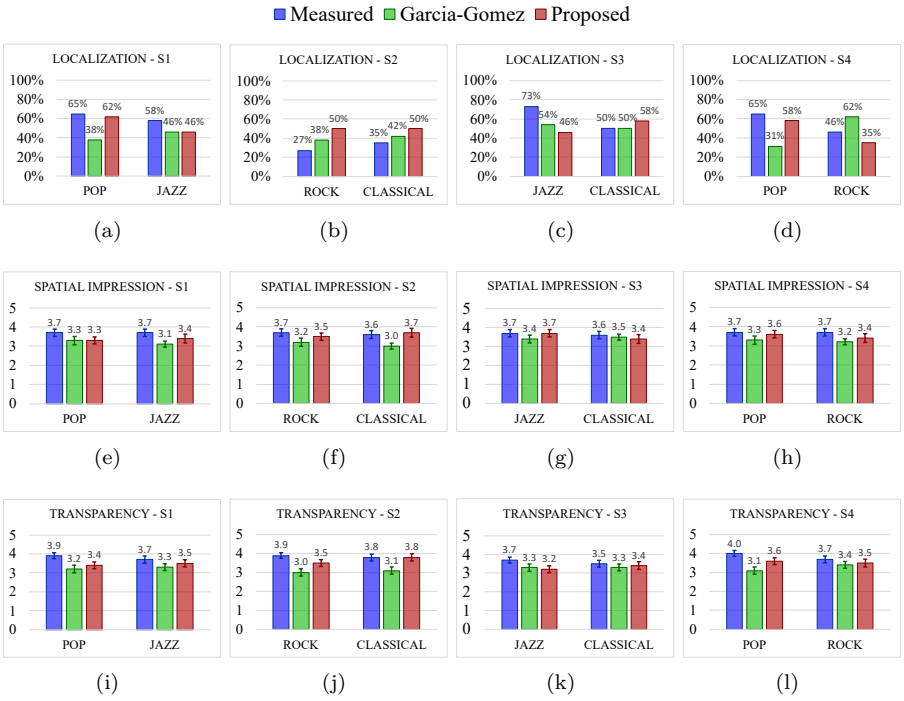


Figure 2.23: Subjective results of the interpolation algorithm applied to HRTFs considering scenario 1 (S1, first column), scenario 2 (S2, second column), scenario 3 (S3, third column) and scenario 4 (S4, fourth column), evaluating (a), (b), (c), (d) the percentage of right responses on the source localization, (e), (f), (g), (h) the spatial impression, and (i), (j), (k), (l) transparency.

of [87]. Four different songs have been employed using four music genres and for each song, only two scenarios have been considered, as shown in Table 2.4, to reduce the duration of the overall listening test. In this way, for each scenario, two music genres are evaluated.

The results are reported in Figure 2.23. The source localization (shown in the first row of Figure 2.23) is evaluated in terms of the percentage of right responses. Regarding source localization, the proposed method shows much

better results in the case of pop music (cf. Figures 2.23(a)-(d)), reaching around 60%. Instead, the algorithm of [5] seems to be more accurate with jazz music for scenario 3 (cf. Figure 2.23(c)) and for scenario 4 with rock music (cf. Figure 2.23(d)). For scenario 2, the measured HRTFs produce the lowest localization accuracy, especially with rock music, which exhibits a percentage of 27%. This value indicates the difficulty of detecting the position ($\vartheta = 144°, \varphi = 30°$) and makes the results obtained by the two interpolation methods less significant. The spatial impression and transparency scores reach values between 3 (Fair) and 4 (Good) for all three cases. The proposed algorithm reaches higher scores than the algorithm of [5] in most of the cases, especially in scenario 2 (cf. Figures 2.23(f)-(j)) and scenario 4 (cf. Figures 2.23(h)-(l)) in terms of both spatial impression and transparency.

2.2.7 Experimental results on BRIRs

The proposed interpolation algorithm presented in Section 2.2.3 has been also tested with BRIRs, where the room reverberation is considered. In this case, the automatic calculation of the mixing time, described in Section 2.2.4, has been added and BRIRs have been measured at different distances. A dataset of BRIRs recorded in a room with a rectangular shape of dimensions 4.5m x 7m x 2.8m has been created. The reverberation time of the room is $T_{60} = 390\,\text{ms}$. The BRIRs have been measured with a sampling frequency of $F_\text{s} = 48\,\text{kHz}$ and a length of 8192 samples (i.e., an order of $N = 8191$) using the Brüel & Kjær head and torso simulator type 4128C and the soundcard Scarlett Focusrite 2i2 2nd generation. The experimental results are divided into objective results and subjective results.

Objective results

Objective results are obtained by evaluating the interpolation algorithm considering two different values of the distance, i.e., $\rho = 1\,\text{m}$ and $\rho = 2\,\text{m}$. As described in Section 2.2.4, the mixing time is equal to 8.3 ms when $\rho = 1\,\text{m}$, while it is equal to 20 ms when $\rho = 2\,\text{m}$ (cf. Figure 2.18). The BRIRs are measured along the azimuth every 5° and, for each position of the virtual sound source ϑ_v, the interpolated BRIR is calculated by interpolating the two measured BRIRs at the positions $\vartheta_1 = \vartheta_\text{v} - 5$ and $\vartheta_2 = \vartheta_\text{v} + 5$. In this case, the analysis is different from the one carried out for HRIRs, since the aim is to evaluate the effectiveness of the mixing time calculation in the interpolation algorithm. Figure 2.24 shows the results obtained for BRIRs at a distance of $\rho = 1\,\text{m}$ using the value of mixing time that is given by the automatic algorithm and a smaller mixing time to show how the algorithm behaves with different t_m. It can be seen that the impulse responses are more reverberant than the

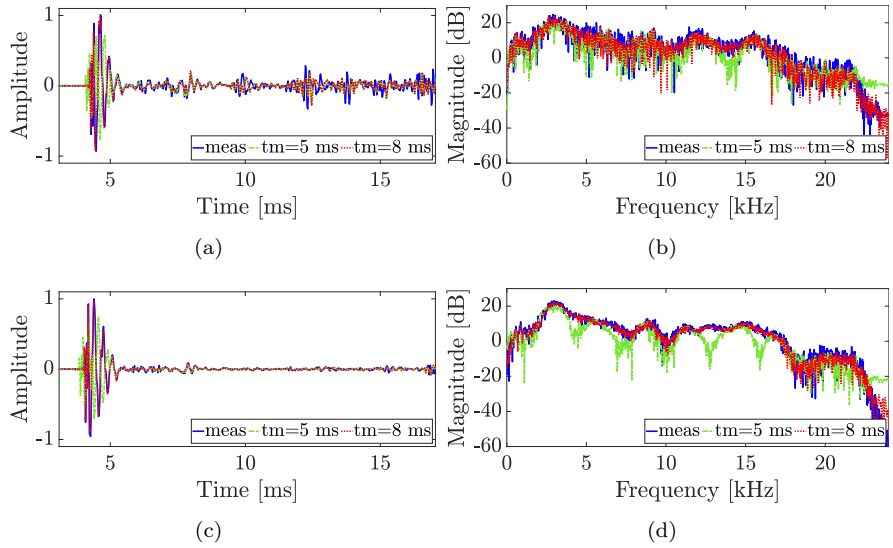


Figure 2.24: Comparison between the measured impulse response and the interpolated one at the azimuth $\vartheta_v = 30°$ and distance $\rho = 1\,\text{m}$ in (a)-(c) the time domain and in (b)-(d) the frequency domain of the (a)-(b) left ear and the (c)-(d) right ear, using two different values of the mixing time: $t_m = 5\,\text{ms}$ and $t_m = 8\,\text{ms}$.

HRIRs shown in Section 2.2.6 due to the reverberation introduced by the room. In particular, in Figures 2.24(a) and 2.24(c), the interpolated impulse response in the time domain is shown, in comparison with the measured BRIR. The blue line is the measured BRIR, the red line is the interpolated BRIR with the automatic t_m and the green line is the BRIR with a smaller t_m. Figure 2.24(a) refers to the left ear and Figure 2.24(c) to the right ear. This case shows that the mixing time is very important to achieve a good performance in interpolation because, if the mixing time is too low, the peak detection and matching cannot properly work, as shown in Figures 2.24(b) and 2.24(d), where the BRIRs are plotted in the frequency domain. As shown in the Figures, using a low t_m produces notches at different frequencies. In Figures 2.25(a) and 2.25(c), the interpolated impulse responses at a distance of $\rho = 2\,\text{m}$ are shown using the mixing time previously calculated at 1 meter and the mixing time recalculated using the impulse response at 2 meters. This case shows the usefulness of an automatic approach to calculate the mixing time. In fact, the two cases involve two BRIRs measured at the same azimuth $\vartheta_v = 30°$, but two different distances. The results show that the increase in the distance changes enough the value of the mixing time, proving that a wrong prediction of the t_m could make the interpolation algorithm unreliable.

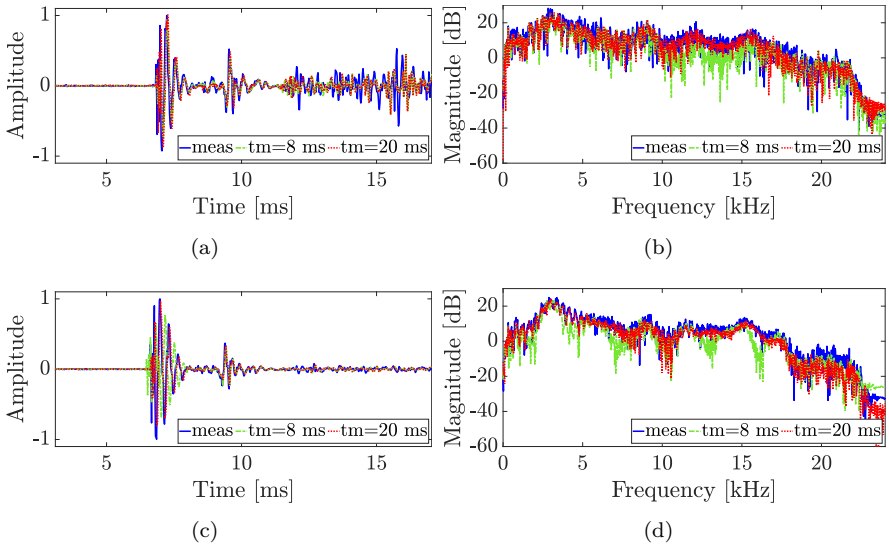


Figure 2.25: Comparison between the measured impulse response and the interpolated one at the azimuth $\vartheta_v = 30°$ and distance $\rho = 2\,\text{m}$ in (a)-(c) the time domain and in (b)-(d) the frequency domain of the (a)-(b) left ear and the (c)-(d) right ear, using two different values of the mixing time: $t_m = 8\,\text{ms}$ and $t_m = 20\,\text{ms}$.

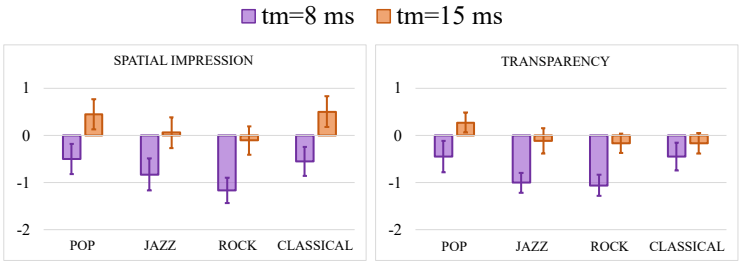
Subjective results

The proposed algorithm has been perceptually evaluated through listening tests. The involved listeners were asked to evaluate two attributes using a pair of headphones as a playback device. The listening panel was composed of 18 listeners, of which 12 were men and 6 were women. Each test consisted of comparing two tracks with a reference track. The tracks reproduced a moving sound source from 30° to 60° with a step of 5° along the azimuth. The source moves 5° every 3 seconds, obtaining a total track duration of 21 seconds. The reference track was obtained by the convolution of the input signal with the measured BRIRs at the distance $\rho = 1.5$ m, while the two test tracks are obtained by the convolution with:

I) the interpolated BRIRs imposing a fixed mixing time of $t_m = 8$ ms arbitrarily chosen;

II) the interpolated BRIRs with the proposed method with the automatic calculation of the mixing time, resulting in $t_m = 15$ ms.

For each position of the virtual sound source ϑ_v, the interpolated impulse response is calculated by interpolating the two measured BRIRs at the positions



(a) (b)

Figure 2.26: Results of the subjective tests evaluating the interpolation algorithm applied to BRIRs with two values of the mixing time (i.e., $t_\mathrm{m} = 8\,\mathrm{ms}$ and $t_\mathrm{m} = 15\,\mathrm{ms}$) in terms of (a) the spatial impression and (b) the transparency.

$\vartheta_1 = \vartheta_\mathrm{v} - 5$ and $\vartheta_2 = \vartheta_\mathrm{v} + 5$, respectively. The listening tests were carried out according to ITU-R BS.1284-2 [87] evaluating the spatial impression (i.e., if the performance seems to take place in an appropriate spatial environment), and the transparency (i.e., if all the details of the performance are clearly perceived). The score may assume values from -3 (much worse) to 3 (much better), using a bipolar discrete seven-grade scale where the 0 means that the test track is perceived as the same as the reference one. The same soundtracks of the experiments described in Section 2.2.6, shown in Table 2.4, are employed to evaluate the algorithm changing the spectral contents of the reproduced song.

Figure 2.26 shows the results obtained through the listening tests, which have been analyzed considering a 95% confidence interval. The subjective results show the effectiveness of the proposed algorithm, proving that a wrong mixing time could damage the listening experience, in terms of both spatial impression and transparency. In fact, the first case that uses a small mixing time of $t_\mathrm{m} = 8\,\mathrm{ms}$ exhibits the lowest scores for all the employed tracks. In particular, it shows the worst performance with rock music, reaching a score of -1.2 on spatial impression and of -1 on transparency. The second case, which employs the automatic calculation of the exact mixing time (resulting in $t_\mathrm{m} = 15\,\mathrm{ms}$), shows good performance. The obtained scores are around 0 in most of the cases and this means that the perceived sound is similar to the one obtained with the measured BRIRs. In particular, the spatial impression seems even better than the one obtained with the measured BRIRs, since the scores are all positive except in the case of rock music and the highest result is 0.5, reached with the classical music. The results obtained with $t_\mathrm{m} = 15\,\mathrm{ms}$ for the transparency attribute are a bit lower than zero except for pop music, but they are still better than the case where $t_\mathrm{m} = 15\,\mathrm{ms}$, as expected.

2.3 Immersive audio rendering over loudspeakers

When the reproduction of immersive audio systems is achieved through two or more loudspeakers, the crosstalk phenomenon occurs. In a stereophonic system (i.e., two loudspeakers), the crosstalk signals are represented by the sound reproduced by the left loudspeaker that reaches the right ear and the sound reproduced by the right loudspeaker that reaches the left ear. A reliable immersive audio rendering, comparable to headphones reproduction, can be obtained using a crosstalk cancellation algorithm that attenuates or eliminates the unwanted signals. In this section, a binaural system for the reproduction over loudspeakers and an adaptive CTC algorithm are presented.

2.3.1 Background on binaural loudspeakers rendering

While the binaural synthesis over headphones can be easily obtained by HRTFs interpolation, the reproduction over loudspeakers does not guarantee the channel separation, so it can be affected by the crosstalk phenomenon. For this reason, crosstalk cancellation algorithms are required when two or more loudspeakers are involved. The simplest crosstalk cancellation procedure consists of the HRTF inversion. It was introduced by Bauer in [13] and employed by Schroeder and Atal for concert hall recordings [88, 89]. Despite the simplicity of this approach, the transfer function inversion is not always possible due to the non-minimum-phase characteristics of most systems. For this reason, more efficient CTC algorithms have been studied over the years. The least mean square (LMS) algorithm is one of the most employed thanks to its simplicity and robustness [90, 91]. Since LMS has a low convergence rate in the case of colored noise as input, it can be improved by applying subband adaptive structures [92]. However, the CTC techniques discussed above necessitate the knowledge of the HRTFs and are sensitive to listener head movements. In this context, Glasgal [93] proposed the recursive ambiophonic crosstalk elimination (RACE) algorithm that is based on the inversion and attenuation of unwanted signals and does not require the HRTFs knowledge. In [94], a CTC system that applies the RACE algorithm to a linear loudspeaker array in combination with least squares frequency invariant (LSFI) beamforming is proposed. The system is then used in [95] and compared to the pressure matching (PM) beamforming.

2.3.2 Binaural synthesis over loudspeakers

The binaural system described in Section 2.2 is adapted for the reproduction over loudspeakers, by adding the RACE algorithm of [93] to reduce the crosstalk signals, as shown in Figure 2.27. The RACE has been chosen for the proposed system because it allows the reduction of the crosstalk signals without know-

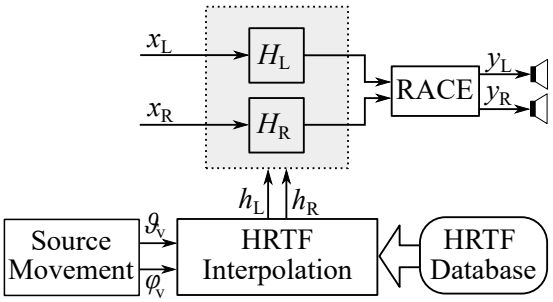


Figure 2.27: Total scheme of the proposed system for immersive audio rendering over loudspeakers.

ing the impulse responses which characterize the acoustic path between the loudspeakers and the listener, avoiding further filtering process. In fact, the algorithm requires only the knowledge of the distance between the loudspeakers and the listener (d_u), the distance between the two loudspeakers (d), and the listener head radius (r_head), considering the generic setup of Figure 2.28(a), in which the listener is located in the middle of the two loudspeakers. Figure 2.28(b) shows the scheme of the RACE. It is a recursive algorithm, where the undesired signal is inverted, properly attenuated, delayed, and added to the other channel. However, the added signal could, in turn, introduce an unwanted contribution that has to be deleted: this creates a recursive "ping-pong" correction between the left and right channels. The RACE algorithm is applied to guarantee an accurate 3D audio experience over loudspeakers. The attenuation typically assumes values between -2 dB and -3 dB, while the delay depends on the value of ITD, computed as [96]

$$\mathrm{ITD} = \frac{r_\mathrm{head}}{c}(\theta_\mathrm{u} + \sin\theta_\mathrm{u}), \tag{2.17}$$

where $c = 343\,\mathrm{m/s}$ is the sound speed and $\theta_\mathrm{u} = \arctan\left[d/(2d_\mathrm{u})\right]$ is the angle that defines the loudspeaker direction (cf. Figure 2.28(a)). Typical values of the delay are between 60 µs and 100 µs. The proposed system has been evaluated using the NU-Tech software [84], by performing subjective tests. The effectiveness of the interpolation algorithm has been already proved and discussed in Section 2.2.6. Therefore, in this case, the study is focused on evaluating the performance of the RACE. The reproduction over headphones, which uses only the interpolation algorithm, is compared with the reproduction over loudspeakers, which implements the whole system (interpolation and RACE), carried out in a normal living environment. Listening tests have been carried out with the purpose of verifying that the spatialization introduced by the moving sound source and perceived with headphones is maintained also with loudspeakers,

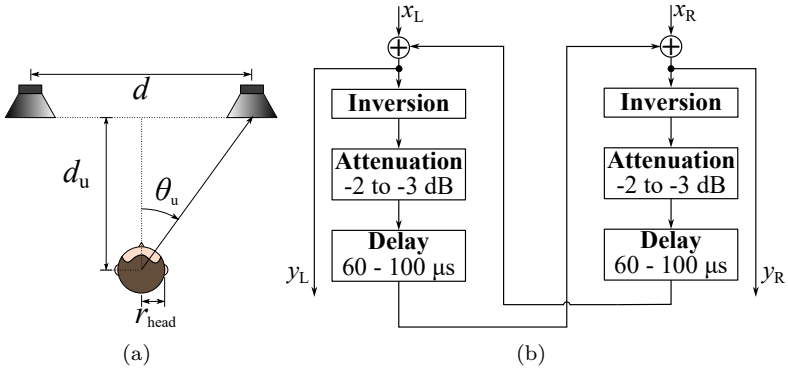


Figure 2.28: (a) setup parameters and (b) scheme of the RACE algorithm.

adding the RACE algorithm. For the reproduction over headphones, the listening panel was composed of 14 expert listeners, 9 men and 5 women aged between 21 and 50. 10 of those expert listeners (4 men and 6 women aged between 21 and 41) have carried out also subjective tests over loudspeakers to evaluate the whole system. The listeners were asked to maintain a fixed position as much as possible. The employed tracks had a length of 20 s and reproduce a moving sound source riding along the following possible paths:

- from right to left in front of the listener;
- from left to right in front of the listener;
- from right to left behind the listener (back);
- from left to right behind the listener (back);

where right corresponds to the azimuth $\vartheta = 90°$ and left to the azimuth $\vartheta = -90°$. The HRTFs have been taken from the MIT Media Lab database [1] without any down-sampling on the measurement grid. The sound source moves along a semicircle and the movement has been produced by the HRTFs interpolation with steps of 9°. Thus, the HRTF interpolation is always carried out between two nearby measurement points, whose distance depends on the database. For example, for $\varphi = 0°$, the measurement step along the azimuth is 5°, so the first HRTF related to $\vartheta_v = 9°$ is calculated by interpolating the HRTFs at the positions $\vartheta_1 = 5°$ and $\vartheta = 10°$. Moreover, three different values for the elevation have been considered: $\varphi = -35°$, $\varphi = 0°$, and $\varphi = 35°$. For subjective tests, the HRTFs have been taken from the MIT Media Lab database [1] without any down-sampling on the measurement grid. The listening panel was asked to detect the left/right movement (i.e., if the sound source is moving from left to right or vice versa), the front/back sound location (i.e., if the sound comes from the front or the back), and the elevation, choosing between −35°, 0°, and 35°. Moreover, listeners had to judge the spatial

impression and the transparency of the performance [87]. Both spatial impression and transparency could be quantified with a score between 1 (bad) and 5 (excellent), according to the unipolar discrete five-grade scale of [87]. For the experiments, two different tracks have been engaged: pop music and speech. The results are reported in Figure 2.29. In particular, the left/right confusion, the front/back confusion, and the elevation accuracy are evaluated in terms of the percentage of right responses and are shown in Figures 2.29(a)-(b)-(c), respectively. The results on spatial impression and transparency have been analyzed by mean with a 95% confidence interval and are shown in Figures 2.29(d)-(e), respectively. The subjective tests have produced excellent results evaluating the left/right source position, in fact, the percentage of right responses is around 100% with both headphones and loudspeakers (cf. Figure 2.29(a)). The front/back positions are more discernible with headphones (cf. Figure 2.29(b)), while the elevation seems to be difficult to be detected in both cases (cf. Figure 2.29(c)). Instead, the spatial impression and the transparency show good scores. In this case, the headphones reproduction seems to perform better when music is considered, while the loudspeakers reproduction reaches higher values with speech (Figure 2.29(d)-(e)).

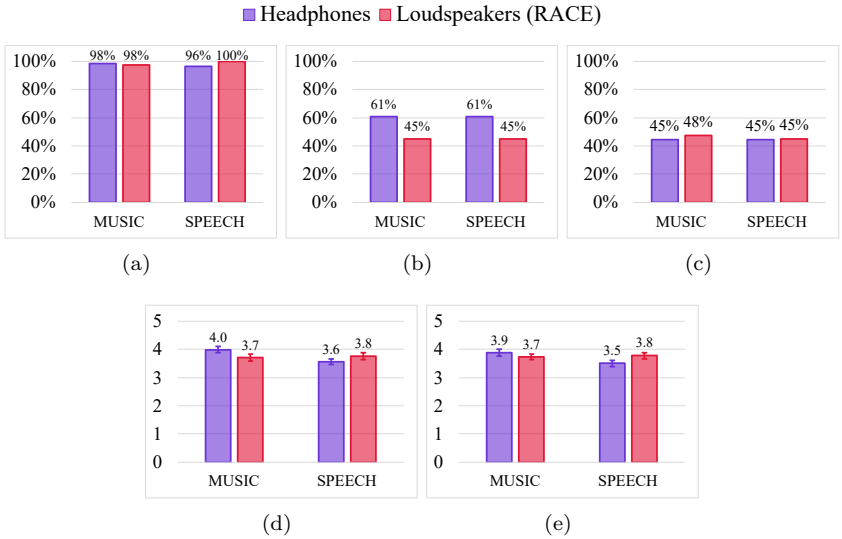


Figure 2.29: Results of the listening tests evaluating the percentage of right responses on (a) left/right sound movement (from right to left or vice versa), (b) front/back sound location and (c) the elevation of the sound source, and testing (d) the spatial impression and (e) the transparency, comparing only the interpolation algorithm through headphones reproduction with the whole proposed system (with RACE) through loudspeakers reproduction.

2.3.3 Subband adaptive crosstalk canceller

The main problem of fixed CTC algorithms, like RACE, is the lack of robustness towards the listener's head movements. This section presents the subband adaptive crosstalk canceler system of [21] that is based on the approach of [92] with the addition of a head tracking system and HRTFs interpolation.

The proposed system is shown in Figure 2.30. In this case, the CTC requires the knowledge of the HRTFs which define the four acoustic paths between the two loudspeakers and the listener ears, i.e., H_{ll}, H_{lr}, H_{rl}, and H_{rr}, as shown in Figure 2.30(a). The scheme of the proposed system is shown in Figure 2.30(b) and is composed of three main blocks: the first block is the head tracker that allows obtaining the x-y coordinates p_{x} and p_{y} of the listener's head. The coordinates are sent to the second block which is the HRTFs interpolator. The interpolation algorithm allows loading the HRTFs from the database just if the detected position corresponds to a measurement point, otherwise, it interpolates the impulse responses of the database to obtain the HRTFs related to the exact position of the listener. The interpolation procedure is applied for all the four HRTFs (H_{ll}, H_{lr}, H_{rl}, H_{rr}), that describe the four paths of the signal. The interpolation algorithm is the same as described in Section 2.2.3, considering the x-y coordinates instead of the spherical ones. Finally, the found HRTFs are used from the adaptive subband crosstalk canceller, which elaborates the input signals X_{L} and X_{R} and obtains the loudspeakers outputs Y_{L} and Y_{R}.

The head tracking is achieved by means of Microsoft Kinect 2.0. The Kinect is composed of an RGB camera with a resolution of 1920 × 1080 pixels, a depth camera with a resolution of 320×240 pixels, and an array of four microphones. The proposed system exploits the head tracking functionalities of the

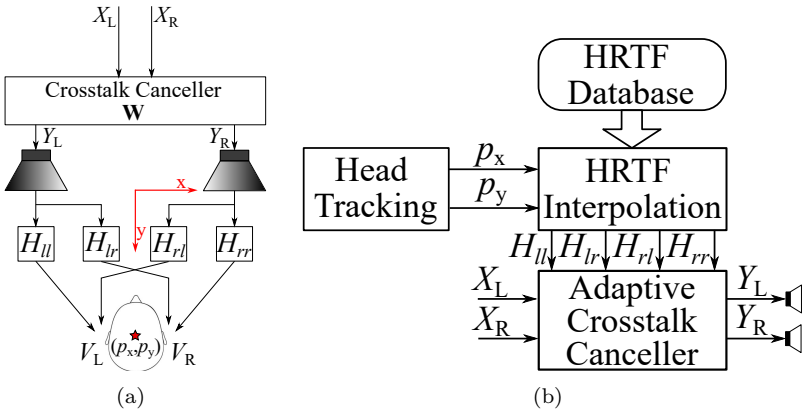


Figure 2.30: (a) setup and (b) scheme of the proposed adaptive CTC system, based on listener head tracking and HRTF interpolation.

Microsoft Kinect software development kit (SDK) 2.0. The toolkit offers several functions to detect the listener's face in real-time, obtaining the distance from the sensor and the x-y coordinates of the listener's head in the 3D space. The coordinates are then used by the interpolation algorithm to calculate the interpolated HRTFs related to the detected position.

Subband crosstalk canceller

Figure 2.30(a) shows a typical two-loudspeaker listening setup, where X_{L} and X_{R} are the binaural signals sent to the loudspeakers. V_{L} and V_{R} are the signals perceived at the listener's ears and can be obtained, in the frequency domain, as follows:

$$\begin{bmatrix} V_{\mathrm{L}} \\ V_{\mathrm{R}} \end{bmatrix} = \mathbf{H} \cdot \mathbf{W} \cdot \begin{bmatrix} X_{\mathrm{L}} \\ X_{\mathrm{R}} \end{bmatrix} \simeq \begin{bmatrix} X_{\mathrm{L}} \\ X_{\mathrm{R}} \end{bmatrix}, \tag{2.18}$$

where $\mathbf{H}$ is the HRTFs matrix, i.e.,

$$\mathbf{H} = \begin{bmatrix} H_{ll} & H_{lr} \\ H_{rl} & H_{rr} \end{bmatrix}, \tag{2.19}$$

and $\mathbf{W}$ is the crosstalk canceller matrix. In the optimal case, the matrix product $\mathbf{H} \cdot \mathbf{W}$ of Equation (2.18) should be an identity matrix, so the crosstalk canceller filters can be obtained by the HRTFs inversion [13]. In general, the crosstalk canceller matrix can be defined as

$$\mathbf{W} = \begin{bmatrix} W_1 & W_3 \\ W_2 & W_4 \end{bmatrix}. \tag{2.20}$$

Equation (2.18) can be rearranged as follows:

$$\begin{aligned} \begin{bmatrix} e_{\mathrm{L}} \\ e_{\mathrm{R}} \end{bmatrix} &= \begin{bmatrix} H_{ll}W_1 + H_{lr}W_2 & H_{ll}W_3 + H_{lr}W_4 \\ H_{rl}W_1 + H_{rr}W_2 & H_{rl}W_3 + H_{rr}W_4 \end{bmatrix} \begin{bmatrix} x_{\mathrm{L}} \\ x_{\mathrm{R}} \end{bmatrix} \\ &= \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} x_{\mathrm{L}} \\ x_{\mathrm{R}} \end{bmatrix}, \end{aligned} \tag{2.21}$$

where C_{11} and C_{22} are the ipsilateral transfer functions and C_{12} and C_{21} are the contralateral transfer functions. In Equation (2.21), the crosstalk canceller filters product can be separated as follows:

$$\begin{bmatrix} e_{\mathrm{L}} \\ e_{\mathrm{R}} \end{bmatrix} = \begin{bmatrix} H_{ll}x_{\mathrm{L}} & H_{lr}x_{\mathrm{L}} & H_{ll}x_{\mathrm{R}} & H_{lr}x_{\mathrm{R}} \\ H_{rl}x_{\mathrm{L}} & H_{rr}x_{\mathrm{L}} & H_{rl}x_{\mathrm{R}} & H_{rr}x_{\mathrm{R}} \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \end{bmatrix}. \tag{2.22}$$

The crosstalk canceller filters $\mathbf{W}$ are obtained by implementing the M-subband adaptive structure presented in [92] and shown in Figure 2.31. The structure is based on the uniform cosine modulated filterbank of [97]. Starting from the prototype filter $p(n)$ of order N_p, the analysis and synthesis filters that build the analysis and synthesis filterbanks $\mathbf{G}$ and $\mathbf{F}$, respectively, are obtained as follows,

$$g_k(n) = 2p(n)\cos\left[\frac{\pi}{M}(k+0.5)\Big(n-\frac{N_\mathrm{p}}{2}\Big)+\gamma_k\right], \tag{2.23}$$

$$f_k(n) = 2p(n)\cos\left[\frac{\pi}{M}(k+0.5)\Big(n-\frac{N_\mathrm{p}}{2}\Big)-\gamma_k\right], \tag{2.24}$$

where $\gamma_k = (-1)^k\frac{\pi}{4}$, for $0 \le k \le M-1$ and $0 \le n \le N_\mathrm{p}$. The double analysis filter-bank $\mathbf{GG}$ is composed by the M filters $G_k(z)G_k(z)$ for $k = 0,\ldots,M-1$, and the $M-1$ filters $G_k(z)G_{k+1}(z)$ for $k = 0,\ldots,M-2$, as depicted in Figure

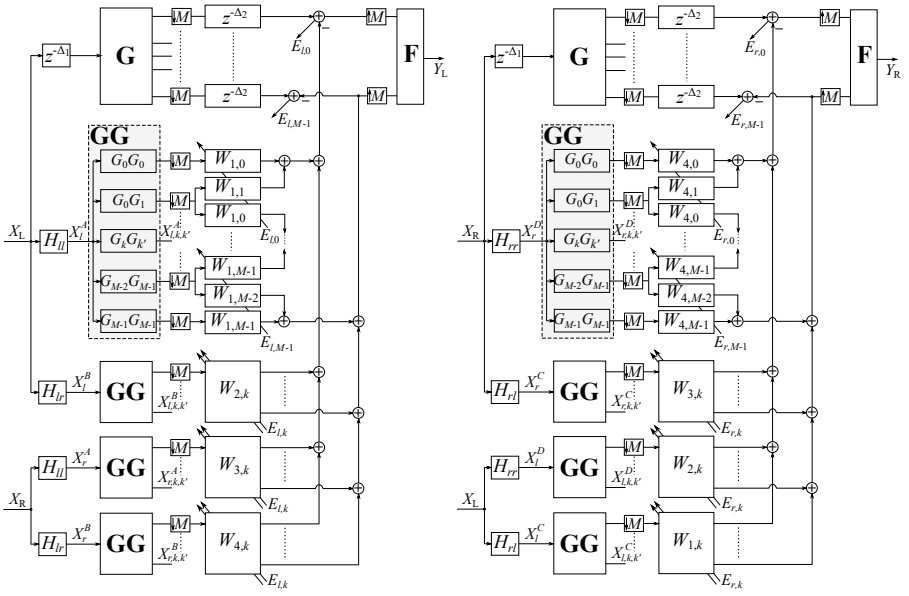


Figure 2.31: Scheme of the subband adaptive crosstalk canceller. Each block $W_{m,k}$ (with $m = 1,...,4$ and $k = 0,...,M-1$) has the same implementation. X_L and X_R describe the input stereo signal and Y_L and Y_R are the loudspeakers outputs.

2.31, with the following impulse responses [98]:

$$g_k(n) * g_k(n) \approx 2[p(n) * p(n)] \cos\left[\frac{\pi}{M}(k+0.5)\Big(n - \frac{N_\mathrm{p}}{2}\Big) + 2\gamma_k\right], \tag{2.25}$$

$$g_k(n) * g_{k+1}(n) \approx 2q_0(n) \cos\left[\frac{\pi}{M}(k+0.5)\Big(n - \frac{N_\mathrm{p}}{2}\Big)\right], \tag{2.26}$$

where

$$q_0(n) = \Big[p(n)e^{j\frac{\pi}{2M}n}\Big] * \Big[p(n)e^{-j\frac{\pi}{2M}n}\Big]. \tag{2.27}$$

The polyphase decomposition [99] is then applied to the double filterbank **GG** in order to improve the computational performance. Referring to Figure 2.31, the signals $X_i^j(z)$, with $i = l, r$, $j = A, B, C, D$, are obtained by filtering the input signals X_L and X_R with the HRTFs H_{ll}, H_{lr}, H_{rl}, and H_{rr} that define the four acoustic paths. The double analysis filterbank produces $2M-1$ output signals that, after the downsampling, are defined as

$$X_{i,k,k}^j = X_i^j(z^{\frac{1}{M}})G_k(z^{\frac{1}{M}})G_k(z^{\frac{1}{M}}), \tag{2.28}$$

for $k = 0, \ldots, M-1$, and as

$$X_{i,k,k+1}^j = X_i^j(z^{\frac{1}{M}})G_k(z^{\frac{1}{M}})G_{k+1}(z^{\frac{1}{M}}), \tag{2.29}$$

for $k = 0, \ldots, M-2$. The signals $X_{i,k,k}^j$ and $X_{i,k,k+1}^j$ are used to update the adaptation subfilters $W_{m,k}$, with $m = 1, \ldots, 4$ and $k = 0, \ldots, M-1$, through the least mean square (LMS) algorithm, i.e.,

$$\begin{aligned} W_{1,k}(n+1) = W_{1,k}(n) &+ \mu_{k,l,A}[X_{l,k,k}^A(n)E_{l,k}(n) + \\ &+ X_{l,k-1,k}^A(n)E_{l,k-1}(n) + X_{l,k,k+1}^A(n)E_{l,k+1}(n)] + \\ &+ \mu_{k,l,C}[X_{l,k,k}^C(n)E_{r,k}(n) + X_{l,k-1,k}^C(n)E_{r,k-1}(n) + \\ &+ X_{l,k,k+1}^C(n)E_{r,k+1}(n)], \end{aligned} \tag{2.30}$$

$$\begin{aligned} W_{2,k}(n+1) = W_{2,k}(n) &+ \mu_{k,l,B}[X_{l,k,k}^B(n)E_{l,k}(n) + \\ &+ X_{l,k-1,k}^B(n)E_{l,k-1}(n) + X_{l,k,k+1}^B(n)E_{l,k+1}(n)] + \\ &+ \mu_{k,l,D}[X_{l,k,k}^D(n)E_{r,k}(n) + X_{l,k-1,k}^D(n)E_{r,k-1}(n) + \\ &+ X_{l,k,k+1}^D(n)E_{r,k+1}(n)], \end{aligned} \tag{2.31}$$

$$\begin{aligned} W_{3,k}(n+1) = W_{3,k}(n) &+ \mu_{k,r,A}[X_{r,k,k}^A(n)E_{l,k}(n) + \\ &+ X_{r,k-1,k}^A(n)E_{l,k-1}(n) + X_{r,k,k+1}^A(n)E_{l,k+1}(n)] + \\ &+ \mu_{k,r,C}[X_{r,k,k}^C(n)E_{r,k}(n) + X_{r,k-1,k}^C(n)E_{r,k-1}(n) + \\ &+ X_{r,k,k+1}^C(n)E_{r,k+1}(n)], \end{aligned} \tag{2.32}$$

$$\begin{aligned} W_{4,k}(n+1) = W_{4,k}(n) + \mu_{k,r,B}[&X^B_{r,k,k}(n)E_{l,k}(n)+ \\ &+ X^B_{r,k-1,k}(n)E_{l,k-1}(n) + X^B_{r,k,k+1}(n)E_{l,k+1}(n)]+ \\ &+ \mu_{k,r,D}[X^D_{r,k,k}(n)E_{r,k}(n) + X^D_{r,k-1,k}(n)E_{r,k-1}(n)+ \\ &+ X^D_{r,k,k+1}(n)E_{r,k+1}(n)]. \end{aligned} \tag{2.33}$$

The error signals are calculated by the following equations:

$$\begin{aligned} E_{l,k}(n) = X_{L,k}(n-\Delta_2 - M\Delta_1) - \Big\{ &\big[W_{1,k}(n)X^A_{l,k,k}(n)+ \\ &+ W_{1,k-1}(n)X^A_{l,k-1,k}(n) + W_{1,k+1}(n)X^A_{l,k,k+1}(n)\big]+ \\ &+ \big[W_{2,k}(n)X^B_{l,k,k}(n) + W_{2,k-1}(n)X^B_{l,k-1,k}(n)+ \\ &+ W_{2,k+1}(n)X^B_{l,k,k+1}(n)\big] + \big[W_{3,k}(n)X^A_{r,k,k}(n)+ \\ &+ W_{3,k-1}(n)X^A_{r,k-1,k}(n) + W_{3,k+1}(n)X^A_{r,k,k+1}(n)\big]+ \\ &+ \big[W_{4,k}(n)X^B_{r,k,k}(n) + W_{4,k-1}(n)X^B_{r,k-1,k}(n)+ \\ &+ W_{4,k+1}(n)X^B_{r,k,k+1}(n)\big]\Big\}, \end{aligned} \tag{2.34}$$

$$\begin{aligned} E_{r,k}(n) = X_{R,k}(n-\Delta_2 - M\Delta_1) - \Big\{ &\big[W_{1,k}(n)X^C_{l,k,k}(n)+ \\ &+ W_{1,k-1}(n)X^C_{l,k-1,k}(n) + W_{1,k+1}(n)X^C_{l,k,k+1}(n)\big]+ \\ &+ \big[W_{2,k}(n)X^D_{l,k,k}(n) + W_{2,k-1}(n)X^D_{l,k-1,k}(n)+ \\ &+ W_{2,k+1}(n)X^D_{l,k,k+1}(n)\big] + \big[W_{3,k}(n)X^C_{r,k,k}(n)+ \\ &+ W_{3,k-1}(n)X^C_{r,k-1,k}(n) + W_{3,k+1}(n)X^C_{r,k,k+1}(n)\big]+ \\ &+ \big[W_{4,k}(n)X^D_{r,k,k}(n) + W_{4,k-1}(n)X^D_{r,k-1,k}(n)+ \\ &+ W_{4,k+1}(n)X^D_{r,k,k+1}(n)\big]\Big\}, \end{aligned} \tag{2.35}$$

where Δ_1 is a modeling delay that considers the HRTFs matrix $\mathbf{H}$ and Δ_2 is the analysis/synthesis filter-bank delay. The step sizes $\mu_{k,i,j}$ are normalized by the sum of instantaneous powers of the signals as follows:

$$\mu_{k,i,j} = \frac{\mu}{\epsilon + P^j_{i,k,k} + P^j_{i,k-1,k} + P^j_{i,k,k+1}}, \tag{2.36}$$

where ϵ is a small coefficient, $j = A, B, C, D$, $i = l, r$, $k = 1, ..., M$ and the power is calculated as

$$P^j_{i,k,k}(n+1) = \eta P^j_{i,k,k}(n) + (1-\eta)[X^j_{i,k,k}(n)]^2, \tag{2.37}$$

and

$$P^j_{i,k,k+1}(n+1) = \eta P^j_{i,k,k+1}(n) + (1-\eta)[X^j_{i,k,k+1}(n)]^2, \tag{2.38}$$

where η is a constant that can range from 0 to 1, $i = l, r$ and $j = A, B, C, D$.

The CTC filters of the matrix $\mathbf{W}$ are calculated from their respective subfilters $W_{m,k}$ as

$$W_m(z) = \sum_{k=0}^{M-1} W_{m,k}(z^M) F_k(z), \tag{2.39}$$

where $F_k(z)$ is the kth filter of the synthesis filterbank, and $m = 1, \ldots, 4$.

Experimental setup

Before implementing the blocks of the proposed system, a database of HRTFs has been created. The setup used for the measurements is shown in Figure 2.32(a). The HRTFs have been measured using a Brüel & Kjær mannequin, that is connected to the input of a Focusrite sound card. Two Genelec loudspeakers are connected to the sound card as output, and the sound card is connected to the PC via USB. The Kinect allows the detection of the dummy head position and is connected to the PC. For the acquisitions, the maximum length pseudo-random binary sequence (MLS), a sample rate of $F_s = 44.1$ kHz, and a HRTFs length of 1024 samples have been applied. The impulse responses have been measured in nine positions, reported in Figure 2.32(b), considering that (0;0) is the Kinect position. The HRTFs acquisitions have been carried out in a semi-anechoic chamber and for each position, four HRTFs have been measured, corresponding to h_{ll}, h_{lr}, h_{rl} and h_{rr}, using the NU-Tech software. Figure 2.33 shows the NU-Tech board for the real-time implementation of the whole system, in which three plugins have been realized: the head tracker, which detects the x-y coordinates of the listener's head, the HRTFs interpolator, that calculates the interpolated head-related impulse responses referred to

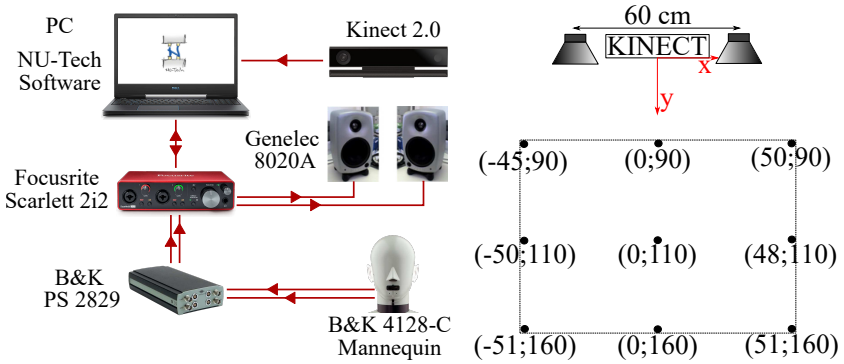


Figure 2.32: (a) scheme of the setup used for HRTFs measurement and (b) x-y positions (in cm) of the measured HRTFs.

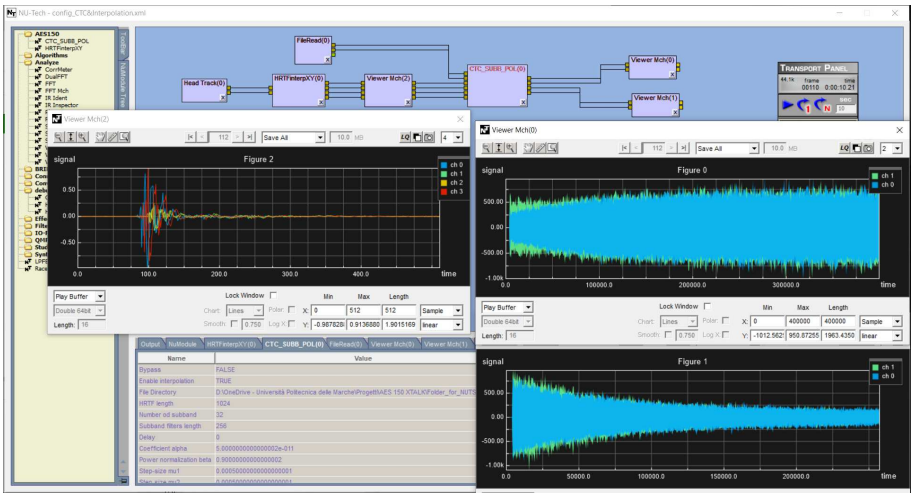


Figure 2.33: NU-Tech implementation of the proposed adaptive CTC system.

the detected position and is implemented as explained in Section 2.2.5, and the crosstalk canceller, that evaluates the canceller filters depending on the HRIRs and elaborates the binaural input signal obtaining the output signals and the error signals.

Experimental results

The proposed system has been tested using the NU-Tech software, imposing a sample rate of $F_s = 44.1\,\text{kHz}$ and a frame size of 4096 samples. Two tests have been carried out applying the following parameters:

- **Test 1**: uncorrelated white noise as input, a step size $\mu = 5\text{e}{-4}$, a coefficient $\epsilon = 5\text{e}{-11}$ and a power normalization of $\eta = 0.9$;
- **Test 2**: two different songs as inputs, a step size $\mu = 5\text{e}{-4}$, a coefficient $\epsilon = 5\text{e}{-5}$, and a power normalization of $\eta = 0.9$.

The experimental results are evaluated in terms of ipsilateral and contralateral transfer functions C_{11} and C_{12}, respectively. The inversion is achieved when the ipsilateral response converges to identity, i.e.,

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} AW_1 + BW_2 & AW_3 + BW_4 \\ CW_1 + DW_2 & CW_3 + DW_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{2.40}$$

Figure 2.34 shows the responses C_{11} and C_{12} obtained in the first test, with white noise as input. Moreover, the mean squared error (MSE) is calculated to evaluate the convergence rate of the algorithm. The MSE of the ith channel is

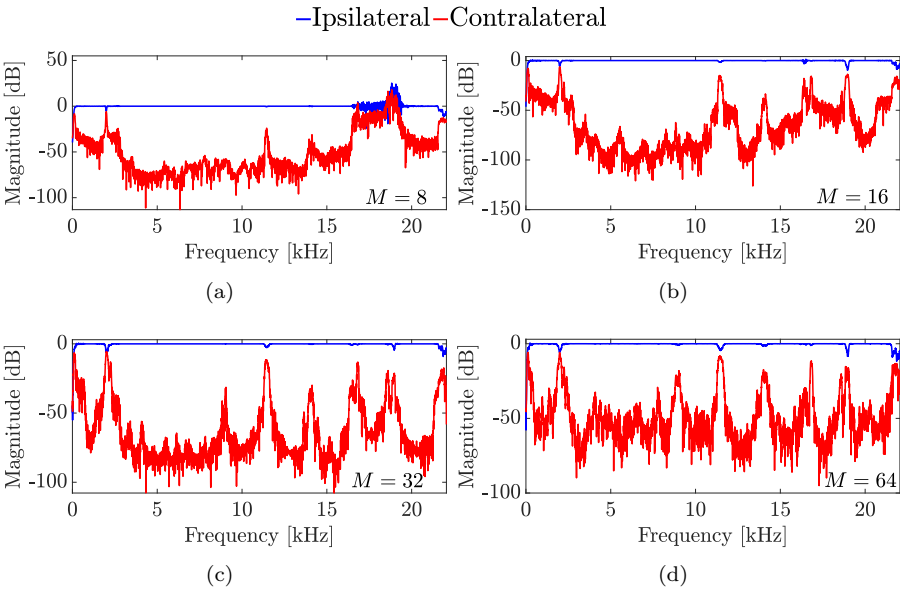


Figure 2.34: Ipsilateral and contralateral frequency response considering (a) $M = 8$ bands, (b) $M = 16$ bands, (c) $M = 32$ bands, (d) $M = 64$ bands, and uncorrelated white noise as input.

obtained as

$$\mathrm{MSE}_i(n) = \mathrm{E}\left[\sum_{k=0}^{M-1} E_{i,k}^2(n)\right], \tag{2.41}$$

where $\mathrm{E}[\cdot]$ represents the expected value and $i = l, r$. Figure 2.35 shows the MSE of the left channel for a different number of subbands M. These results have proved the effectiveness of the algorithm showing an improvement with the increase of the subbands number, both in terms of channel separation and error-signal evolution. Moreover, the convergence rate increases with the number of bands, making the CTC algorithm faster and more suitable for real-time communication with the head-tracker.

The second test evaluates the CTC algorithm with songs as input and Figure 2.36 shows the results in terms of ipsilateral and contralateral responses. Also in this case, the channel separation increases with the number of subbands and this is more relevant at the low frequencies, due to the spectrum of the employed songs that is concentrated at the low frequencies. Informal listening tests have been performed in order to evaluate the crosstalk cancellation in terms of subjective acoustic perception. The involved listeners have verified the algorithm's effectiveness reporting a good channel separation and sound quality.

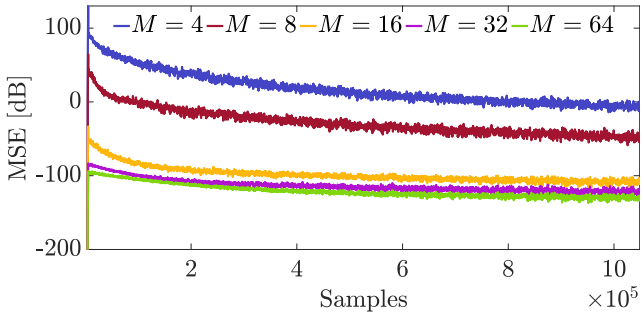


Figure 2.35: MSE of the adaptive CTC algorithm varying the number of sub-bands M.

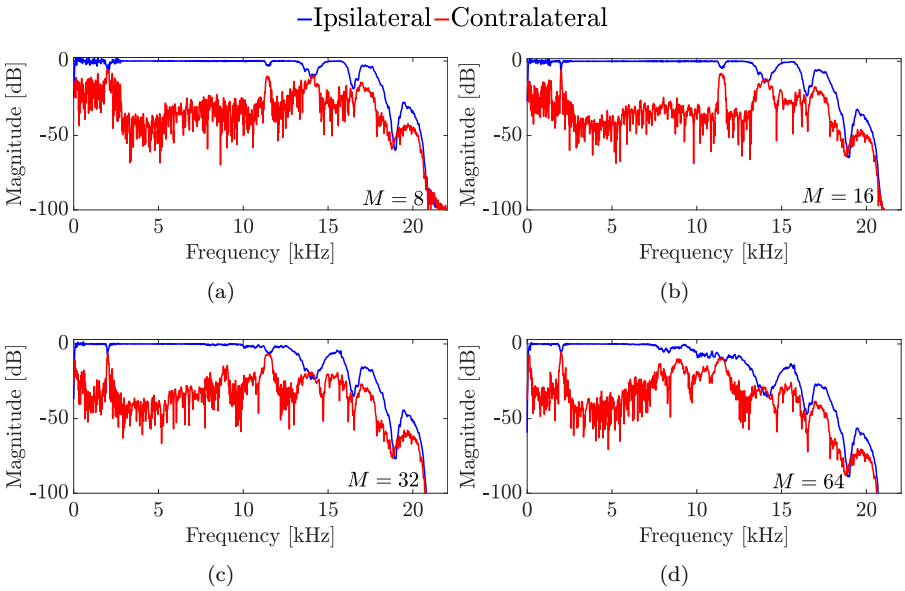


Figure 2.36: Ipsilateral and contralateral frequency response considering (a) $M = 8$ bands, (b) $M = 16$ bands, (c) $M = 32$ bands, (d) $M = 64$ bands, and two different songs as inputs.

2.4 Conclusions of immersive audio rendering

This chapter has presented effective algorithms for immersive audio rendering. The importance of HRTFs in this type of system has been underlined and analyzed. The dependence of the HRTF on the type and the position of the microphone has been examined through a comparative analysis of HRTF measurements. Successively, a system for binaural synthesis over headphones [18] has been presented. The proposed system is based on a HRTF interpolation

algorithm, which splits the impulse responses into early reflections and reverberant tails and separately elaborates the two parts. The proposed interpolation algorithm has been evaluated through objective and subjective tests in comparison with a previous technique, used as a reference. The experimental results have shown good performances of the presented system in terms of estimated HRTFs, mean squared error, and sound quality in listening tests. The interpolation algorithm has been applied also to real reverberant environments, where BRIRs are involved, and an automatic procedure for the mixing time calculation [19] has been added to the system. The results on BRIRs have proved the effectiveness of the interpolation algorithm and the importance of guaranteeing a precise estimation of the mixing time. Then, the binaural system has been adapted to loudspeakers reproduction by adding a fixed crosstalk cancellation algorithm, i.e., RACE [20]. The reproduction over loudspeakers has been compared to the headphones' reproduction through listening tests, obtaining great results. However, fixed CTC solutions are too sensitive to the listener's head movements. Therefore, an adaptive crosstalk canceller based on a subband structure [21] has been presented. The system is capable to detect the position of the listener and adapt the crosstalk cancellation filters. Experimental results have proved that the algorithm is more effective for a greater number of subbands, showing a better channel separation.

Chapter 3

Equalization and Multichannel Systems for Audio Enhancement

Audio equalization is a DSP procedure that aims at reducing the errors caused in the listening experience by the environment or the reproduction system. Equalizers can be classified into manual and automatic EQs [14]. The manual EQs do not need a microphone and the user can adjust the parameters according to his/her equalization preference. The graphic equalizer is an example of manual equalization and allows modifying the gains of different frequency bands [100–102]. The automatic procedure needs a microphone to measure the room impulse response that must be equalized [103]. In this case, adaptive solutions can be employed to update the equalization curve with the possible changes in the environment. Figure 3.1 shows a classification of the main equalization procedures that can be found in the literature. This chapter presents effective techniques for the development of audio equalizers and is organized as follows. Section 3.1 proposes efficient designs for graphic equalizers. Section 3.2 describes a subband adaptive room response equalization method for multichannel systems. Section 3.3 presents a linear-phase crossover network used

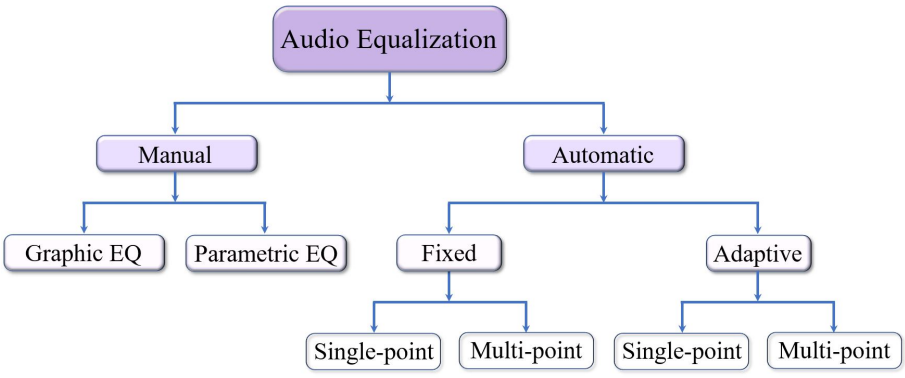


Figure 3.1: Classification of the audio equalization procedure.

in multichannel reproduction. Finally, Section 3.4 concludes the part of audio equalization.

3.1 Graphic equalizers

Graphic equalizers allow the user to adjust the gain of each frequency band, so they are designed through filterbanks which split the input signal into M bands. Three efficient implementations of graphic equalizers are proposed in this section. They are all based on interpolated FIR filters in order to guarantee a linear phase response and a reduced computational complexity. The first GEQ is a linear-phase uniform GEQ, where all the bands have the same width. To further reduce the computational complexity and obtain a logarithmic band division, a linear-phase octave GEQ is successively proposed. Finally, a quasi-linear-phase octave GEQ is presented to lower the latency of the system.

3.1.1 Background on graphic equalizers

Manual equalization consists of parametric equalizers and graphic equalizers. Parametric equalizers give more possibilities to the user that can modify the gain, center frequency, and quality factor Q (or bandwidth) of each filter [101, 104]. Alternatively, graphic equalizers can be seen as specific filterbanks where the user can adjust only the gain of different frequency bands, so the controls define a graph of the magnitude response [14, 100, 105]. Graphic equalizers can be minimum-phase or linear-phase. Minimum-phase GEQs present the smallest latency and are not affected by pre-ringing effects, so they are suitable for live music applications. Differently, linear-phase GEQs preserve the original phase of the signal avoiding audible phase distortion effects [106], so they are preferred for certain applications, such as multichannel equalization [107], speech processing [108], parallel processing, phase compatibility of audio equipment, and crossover network design. Focusing on GEQs design, minimum-phase GEQs are traditionally realized by a set of infinite impulse response (IIR) filters connected either in cascade [100, 109–111] or parallel [100, 112–114]. In [111, 114, 115], minimum-phase GEQs are implemented by second-order IIR sections (a.k.a. biquads) guaranteeing good accuracy and low computational cost. On the other side, linear-phase GEQs are developed by FIR filters [14] that can exhibit an arbitrary phase response and do not suffer from numerical problems, which may occur when IIR filters are involved. The earliest implementations of FIR GEQs date to the 1980s [116–118]. FIR GEQs can be realized by parallel structure [116, 117, 119] or as a single high-order filter that is used to approximate the target frequency response specified by the user [119]. The definition of a target curve is not a trivial task, because it

can be obtained by the interpolation [118,120], but the EQ target curve is not well-defined between the command-gain points. Moreover, a filter length of at least several thousand is required to suit the target response at low frequencies [118,121–123]. In addition, the FIR filter should be completely redesigned whenever a gain is modified, increasing the computational load and making it unsuitable for real-time applications. Aiming at reducing the computational complexity of FIR GEQs, frequency-warped FIR filters [124,125] allow to shorten filter lengths, but show a non-linear phase response. Another solution for computational cost reduction is the fast convolution [126–129], where, for every input frame, the discrete Fourier transform of the signal is multiplied by the filter's impulse response, and the result is inverse transformed. Applying the fast Fourier transform (FFT), a good computational efficiency can be reached. Although the frame-based processing causes much latency, the FFT-based processing allows for a linear phase response [127]. Other linear-phase FIR GEQ implementations are based on multirate approaches [117,118,122,130]. In this case, every band works at different sample rates (e.g., the lowest frequencies use the slowest rate), and after the filtering, all the bands are upsampled to the original sample rate and summed. In [131], a FIR GEQ design, based on interpolated FIR (IFIR) filters [132], is proposed. However, the GEQ of [131] is a uniform equalizer (i.e., the audio frequencies are divided into equal bands), differently from standard graphic equalizers, which use a logarithmic band division [14].

3.1.2 Linear-phase uniform graphic equalizer

The filterbank for the development of the linear-phase uniform GEQ, proposed in [22], is built applying the definition of IFIR filters. IFIR filters are composed of a cascade of two FIR filters [132], as shown in Figure 3.2, and the overall frequency response of the IFIR structure is computed as

$$H_{\mathrm{IFIR}}(z) = F(z^L)G(z), \tag{3.1}$$

where the first FIR is designed from the model filter $F(z)$ applying an upsampling by a factor L, while the second FIR $G(z)$ is called interpolator which is designed to attenuate the unwanted copies of $F(z)$, due to the interpolation

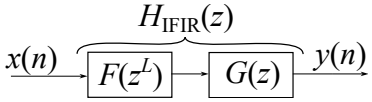


Figure 3.2: Cascade of two FIR filters which represents the IFIR implementation.

procedure. In fact, the cutoff frequency of the model filter $F(z)$ is L times greater than the cutoff frequency of the desired filter, so $F(z)$ can be designed using a lower order N^{F}. The interpolation procedure consists of adding $L-1$ zeros after each sample of the impulse response of $F(z)$. The upsampling allows obtaining the desired cutoff frequency and generates unwanted copies, which are deleted by the interpolator $G(z)$.

In Figure 3.3, the complete IFIR filterbank scheme for the development of a uniform GEQ is reported. $F_i(z)$ is the model filter, L_i is the interpolation factor, $G_m(z)$ is the interpolator, g_m is the assigned gain, and Δ_m is the synchronization delay of the mth band, with $m = 1, ..., \Upsilon, ..., M$, $i = m$ for $m \leq \Upsilon$, where M is the number of bands and must be an odd number, while Υ is the central band and is calculated as the next half-integer of M, i.e., $\Upsilon = (M+1)/2$. For the uniform equalizer, the normalized digital center frequency of the mth band $\omega_{\mathrm{c},m}$ is calculated as

$$\omega_{\mathrm{c},m} = \frac{\pi(2m-1)}{2M}. \tag{3.2}$$

In the following, normalized digital frequencies ω are taken into account and they are linked to the respective analog frequencies f as $\omega = 2\pi f/F_{\mathrm{s}}$, where F_{s} is the sampling frequency. The filters are designed using the Parks-McClellan algorithm [133], which allows determining the order of the filter, knowing the specifications on the cutoff frequencies and ripple amplitude. The specifications on the ripple are the same for all the bands. In particular, starting from the attenuation (in dB) of the passband A_{p} and of the stopband A_{s}, the respective ripple amplitudes can be obtained as

$$\delta_{\mathrm{p}} = \frac{10^{A_{\mathrm{p}}/20} - 1}{10^{A_{\mathrm{p}}/20} + 1}, \tag{3.3}$$

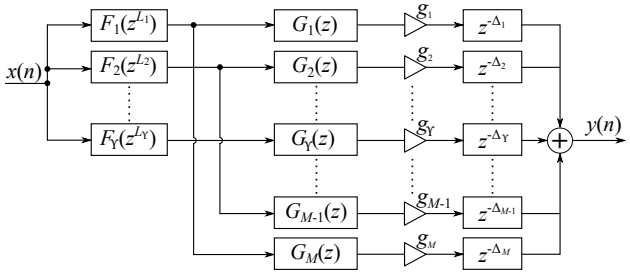


Figure 3.3: Scheme of the linear-phase uniform equalizer based on IFIR filters. The filterbank is designed with $\Upsilon = (M+1)/2$ model filters $F_m(z)$ and M interpolator filters $G_m(z)$.

for the passband, and

$$\delta_{\mathrm{s}} = 10^{A_{\mathrm{s}}/20}, \tag{3.4}$$

for the stopband. The filterbank bands are designed in pairs (the first with the last, the second with the second-to-last, and so on), except for the central one. In fact, for each pair only a filter $F(z)$ is employed. The first and the last band consist of lowpass and highpass filters, respectively, while the central bands consist of bandpass filters. In the following, the entire construction of the IFIR filterbank is explained in detail.

First and last band of the uniform equalizer

The first and the last band are obtained from the same filter $F_1(z)$. The passband cutoff frequency $\omega_{\mathrm{p},1}$ and the stopband cutoff frequency $\omega_{\mathrm{s},1}$ of the first lowpass filter are computed as

$$\omega_{\mathrm{p},1} = \frac{\pi}{M} - \eta\Delta\omega, \tag{3.5}$$

$$\omega_{\mathrm{s},1} = \frac{\pi}{M} + (1-\eta)\Delta\omega, \tag{3.6}$$

where M is the total number of subbands, $\Delta\omega$ is the transition band, and η is a parameter that can take values from 0 to 1 and establishes the overlap between nearby bands. Considering Figure 3.3, the filter $F_1(z)$ is a FIR filter with the following specifications:

$$F_1(z) : \begin{cases} \omega^{\mathrm{F}}_{\mathrm{p},1} = L_1\omega_{\mathrm{p},1}, & \delta^{\mathrm{F}}_{p} = \frac{\delta_{\mathrm{p}}}{2} \\ \omega^{\mathrm{F}}_{\mathrm{s},1} = L_1\omega_{\mathrm{s},1}, & \delta^{\mathrm{F}}_{\mathrm{s}} = \delta_{\mathrm{s}} \end{cases}, \tag{3.7}$$

where δ_{p} and δ_{s} are the ripples in passband and stopband, respectively, and the interpolation factor L_1 is the even value closest to L_1^{opt}, obtained as follows:

$$L_1^{\mathrm{opt}} = \left\lfloor \frac{2\pi}{\omega_{\mathrm{s},1} + \omega_{\mathrm{p},1} + \sqrt{2\pi(\omega_{\mathrm{s},1} - \omega_{\mathrm{p},1})}} \right\rfloor. \tag{3.8}$$

The interpolated filter $F_1(z^{L_1})$ narrows the bandwidth of the lowpass filter and creates several images including a highpass one, that defines the last band of the filterbank, with the following cutoff frequencies:

$$\omega_{\mathrm{p},M} = \frac{\pi(M-1)}{M} + \eta\Delta\omega, \tag{3.9}$$

$$\omega_{\mathrm{s},M} = \frac{\pi(M-1)}{M} - (1-\eta)\Delta\omega. \tag{3.10}$$

Finally, the lowpass filter $G_1(z)$ is computed as

$$G_1(z) : \begin{cases} \omega_{\mathrm{p},1}^{\mathrm{G}} = \omega_{\mathrm{p},1}, & \delta_{\mathrm{p}}^{\mathrm{G}} = \frac{\delta_{\mathrm{p}}}{2} \\ \omega_{\mathrm{s},1}^{\mathrm{G}} = \frac{2\pi}{L_1} - \omega_{\mathrm{s},1}, & \delta_{\mathrm{s}}^{\mathrm{G}} = \delta_{\mathrm{s}} \end{cases}, \tag{3.11}$$

while the highpass filter $G_M(z)$ is designed as

$$G_M(z) : \begin{cases} \omega_{\mathrm{p},M}^{\mathrm{G}} = \omega_{\mathrm{p},M}, & \delta_{\mathrm{p}}^{\mathrm{G}} = \frac{\delta_{\mathrm{p}}}{2} \\ \omega_{\mathrm{s},M}^{\mathrm{G}} = \pi - \frac{2\pi}{L_1} + \omega_{\mathrm{s},1}, & \delta_{\mathrm{s}}^{\mathrm{G}} = \delta_{\mathrm{s}} \end{cases}. \tag{3.12}$$

Intermediate bands of the uniform equalizer

The intermediate bands are characterized by bandpass filters with the following normalized cutoff frequencies:

$$\omega_{\mathrm{s}j,m} = \frac{\pi(m-2+j)}{M} + (-1)^j(1-\eta)\Delta\omega, \tag{3.13}$$

$$\omega_{\mathrm{p}j,m} = \frac{\pi(m-2+j)}{M} + (-1)^{j+1}\eta\Delta\omega, \tag{3.14}$$

where $m = 2, ..., M-1$ and $j = 1, 2$ indicates the first and the second transition band, respectively. Also in this case the interpolation factor L_m, for $m < \Upsilon$, is the even value closest to L_m^{opt}, obtained by the Equation (3.8) considering the frequencies $\omega_{\mathrm{p}2,m}$ and $\omega_{\mathrm{s}2,m}$. As mentioned before, the GEQ bands are designed in pairs (i.e., the first with the last, the second with the second-to-last, and so on) and each pair shares the same model filter that, for intermediate bands, is a bandpass filter $F_i(z)$, defined as

$$F_i(z) : \begin{cases} \omega_{\mathrm{s}1,i}^{\mathrm{F}} = L_i\omega_{\mathrm{s}1,i}, & \delta_{\mathrm{s}}^{\mathrm{F}} = \delta_{\mathrm{s}} \\ \omega_{\mathrm{p}1,i}^{\mathrm{F}} = L_i\omega_{\mathrm{p}1,i}, & \delta_{p}^{\mathrm{F}} = \frac{\delta_p}{2} \\ \omega_{\mathrm{p}2,i}^{\mathrm{F}} = L_i\omega_{\mathrm{p}2,i}, & \delta_{\mathrm{p}}^{\mathrm{F}} = \frac{\delta_{\mathrm{p}}}{2} \\ \omega_{\mathrm{s}2,i}^{\mathrm{F}} = L_i\omega_{\mathrm{s}2,i}, & \delta_{\mathrm{s}}^{\mathrm{F}} = \delta_{\mathrm{s}} \end{cases}, \tag{3.15}$$

where $i = 2, ..., \Upsilon - 1$. For each band a filter $G_m(z)$ is applied, as shown in Figure 3.3. These filters are designed as lowpass filters for the first half of the filterbank and as highpass filters for the second half, i.e.,

$$G_i(z) : \begin{cases} \omega_{\mathrm{p},i}^{\mathrm{G}} = \omega_{\mathrm{p}2,i}, & \delta_{p}^{\mathrm{G}} = \frac{\delta_{\mathrm{p}}}{2} \\ \omega_{\mathrm{s},i}^{\mathrm{G}} = \frac{2\pi}{L_i} - \omega_{\mathrm{s}2,i}, & \delta_{\mathrm{s}}^{\mathrm{G}} = \delta_{\mathrm{s}} \end{cases}, \tag{3.16}$$

$$G_k(z) : \begin{cases} \omega_{\mathrm{p},k}^{\mathrm{G}} = \pi - \omega_{p2,i}, & \delta_{p}^{\mathrm{G}} = \frac{\delta_{\mathrm{p}}}{2} \\ \omega_{\mathrm{s},k}^{\mathrm{G}} = \pi - \frac{2\pi}{L_i} + \omega_{\mathrm{s}2,i}, & \delta_{\mathrm{s}}^{\mathrm{G}} = \delta_{\mathrm{s}} \end{cases}, \tag{3.17}$$

where $i = 2, ..., \Upsilon - 1$ and $k = M - i + 1$.

Central band of the uniform equalizer

The central Υth band is obtained from a lowpass filter $F_\Upsilon(z)$ and an interpolation factor divisible by four. The cutoff frequencies of the central filter are obtained following the Equations (3.13)-(3.14) imposing $m = \Upsilon$. The interpolation factor L_Υ is the multiple of 4 closest to L_Υ^{opt} and less than L_Υ^{opt}, obtained following the Equation (3.8) using the frequencies $\omega_{\text{p2},\Upsilon}$ and $\omega_{\text{s2},\Upsilon}$. The lowpass filter $F_\Upsilon(z)$ is designed as

$$F_\Upsilon(z) : \begin{cases} \omega_{\text{p},\Upsilon}^{\text{F}} = L_\Upsilon(\frac{\pi}{2} - \omega_{\text{p1},\Upsilon}), & \delta_{\text{p}}^{\text{F}} = \frac{\delta_{\text{p}}}{2} \\ \omega_{\text{s},\Upsilon}^{\text{F}} = L_\Upsilon(\frac{\pi}{2} - \omega_{\text{s1},\Upsilon}), & \delta_{\text{s}}^{\text{F}} = \delta_{\text{s}} \end{cases}. \tag{3.18}$$

Finally, the bandpass filter $G_\Upsilon(z)$ is designed as follows:

$$G_\Upsilon(z) : \begin{cases} \omega_{\text{s1},\Upsilon}^{\text{G}} = \pi - \frac{2\pi}{L_\Upsilon} - \omega_{\text{s1},\Upsilon}, & \delta_{\text{s}}^{\text{G}} = \delta_{\text{s}} \\ \omega_{\text{p1},\Upsilon}^{\text{G}} = \omega_{\text{p1},\Upsilon}, & \delta_{\text{p}}^{\text{G}} = \frac{\delta_{\text{p}}}{2} \\ \omega_{\text{p2},\Upsilon}^{\text{G}} = \omega_{\text{p2},\Upsilon}, & \delta_{\text{p}}^{\text{G}} = \frac{\delta_{\text{p}}}{2} \\ \omega_{\text{s2},\Upsilon}^{\text{G}} = \frac{2\pi}{L_\Upsilon} + \omega_{\text{s1},\Upsilon}, & \delta_{\text{s}}^{\text{G}} = \delta_{\text{s}} \end{cases}. \tag{3.19}$$

Delay computation

The filters of the IFIR filterbank have different lengths, so a synchronization delay must be applied to each band to synchronize the output signals. The delay τ_m of the mth band introduced by the filtering process can be calculated as

$$\tau_m = \frac{N_m^{\text{F}} L_m + N_m^{\text{G}}}{2}, \tag{3.20}$$

where $m = 1, ..., M$, N_m^{F} is the order of the mth filter $F(z)$, N_m^{G} is the order of the mth filter $G(z)$, and L_m is the mth interpolation factor. It is worth noting that, for $m > \Upsilon$, the filters $F_m(z)$ and the respective interpolation factor L_m are the same as those defined for $m < \Upsilon$, in accordance with the scheme of Figure 3.3. Since the GEQ implements a parallel structure, the total delay τ of the uniform GEQ is defined by the maximum delay among all bands, i.e.,

$$\tau = \max\{\tau_m : m = 1, \ldots, M\}. \tag{3.21}$$

Therefore, the synchronization delay that must be applied to the mth band is computed as $\Delta_m = \tau - \tau_m$.

Experimental results of the linear-phase uniform graphic equalizer

For the experimental tests, three different filterbanks have been designed varying the number of bands M, i.e., $M = 9$, $M = 21$, and $M = 31$. For all three configurations, the parameters chosen for the proposed IFIR filterbank are the following:

- transition band $\Delta\omega = 0.005\pi$,
- coefficient $\eta = 0.53$,
- passband attenuation of $A_{\mathrm{p}} = 0.01\,\mathrm{dB}$, i.e., a ripple of $\delta_{\mathrm{p}} = 0.000575$,
- stopband attenuation $A_{\mathrm{s}} = -60\,\mathrm{dB}$, i.e., a ripple of $\delta_{\mathrm{s}} = 0.001$.

Table 3.1 reports the final filter orders, interpolation factors, and delays of the designed IFIR filterbank, for three different configurations. For each band, the delay τ_m is computed following Equation (3.20). The proposed uniform GEQ has been compared with other linear-phase uniform implementations, i.e., the FFT-based equalizer [134], the multirate equalizer of [130], and the equalizer proposed by Hergum in [131], also based on IFIR filters.

The FFT-based equalizer [127] is obtained by filtering the input signal with a target function calculated as the sum of the filter response of the proposed uniform IFIR equalizer. The filtering procedure is achieved by the overlap and add method with an overlap of 50% [135]. A FFT length of 4096 samples and

Table 3.1: Filters length for the uniform IFIR graphic equalizer.

Band	Filter order of $F_m(z)$ N_m^{F}			Filter order of $G_m(z)$ N_m^{G}			Interpolation factor L_m			Delay τ_m		
m	$M=9$	$M=21$	$M=31$	$M=9$	$M=21$	$M=31$	$M=9$	$M=21$	$M=31$	$M=9$	$M=21$	$M=31$
#1	243	147	123	64	68	66	6	10	12	761	769	771
#2	725	242	186	12	50	58	2	6	8	731	751	773
#3	724	361	247	22	34	52	2	4	6	735	739	767
#4	729	363	370	62	60	28	2	4	4	760	756	754
#5	183	725	369	52	14	40	8	2	4	758	732	758
#6	-	722	369	62	18	60	-	2	4	760	731	768
#7	-	723	738	22	22	14	-	2	2	735	734	745
#8	-	723	736	12	30	14	-	2	2	731	738	743
#9	-	725	737	64	50	18	-	2	2	761	750	746
#10	-	725	737	-	146	20	-	2	2	-	798	747
#11	-	121	738	-	60	26	-	12	2	-	756	751
#12	-	-	738	-	146	32	-	-	2	-	798	754
#13	-	-	737	-	50	44	-	-	2	-	750	759
#14	-	-	739	-	30	72	-	-	2	-	738	775
#15	-	-	744	-	22	210	-	-	2	-	734	849
#16	-	-	123	-	18	48	-	-	12	-	731	762
#17	-	-	-	-	14	210	-	-	-	-	732	849
#18	-	-	-	-	60	72	-	-	-	-	756	775
#19	-	-	-	-	34	44	-	-	-	-	739	759
#20	-	-	-	-	50	30	-	-	-	-	751	753
#21	-	-	-	-	68	26	-	-	-	-	769	751
#22	-	-	-	-	-	20	-	-	-	-	-	747
#23	-	-	-	-	-	18	-	-	-	-	-	746
#24	-	-	-	-	-	14	-	-	-	-	-	743
#25	-	-	-	-	-	14	-	-	-	-	-	745
#26	-	-	-	-	-	60	-	-	-	-	-	768
#27	-	-	-	-	-	40	-	-	-	-	-	758
#28	-	-	-	-	-	28	-	-	-	-	-	754
#29	-	-	-	-	-	52	-	-	-	-	-	767
#30	-	-	-	-	-	58	-	-	-	-	-	773
#31	-	-	-	-	-	66	-	-	-	-	-	771

a frame size of 8192 samples have been chosen to ensure an error comparable to the proposed GEQ. The multirate equalizer of [130] implements a subband structure derived from a prototype filter of order N_p. Different filter orders have been imposed for the three configurations to obtain an acceptable error, i.e., $N_\mathrm{p} = 1295$ when $M = 9$, $N_\mathrm{p} = 1343$ when $M = 21$ and $N_\mathrm{p} = 1363$ when $M = 31$. In addition, also the 9-band uniform equalizer of [131] has been considered since it is based on IFIR filters. Hergum's approach of [131] is different from the proposed one and it is based on a tree structure built starting from a model filter of order N_h and appropriate interpolation factors. For the comparison, an order of $N_h = 102$ has been imposed to obtain performances similar to the proposed method.

The comparison has been carried out using a sampling frequency of $F_\mathrm{s} = 48\,\mathrm{kHz}$. Figure 3.4 shows the frequency magnitude responses of the three cases with different values of M, imposing random gain settings, and comparing the proposed equalizer with the multirate GEQ. In the magnitude response comparison, only the multirate GEQ is reported because the FFT-based GEQ is derived from the proposed one, so it exhibits the same frequency response, while Hergum GEQ has different center frequency bands. Looking at the frequency

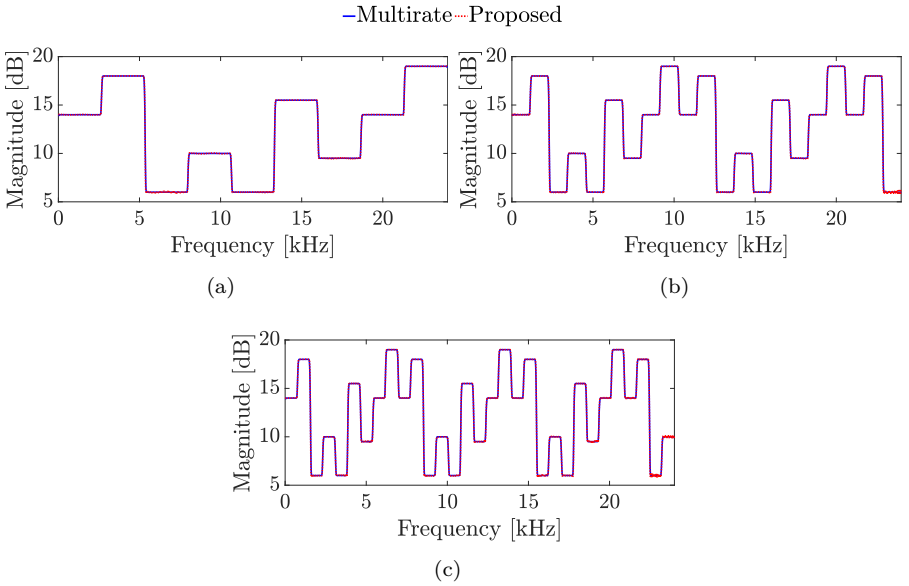


Figure 3.4: Magnitude response comparison between the multirate GEQ and the proposed uniform GEQ, with (a) $M = 9$, (b) $M = 21$, and (c) $M = 31$, and random gains $[14\ 18\ 6\ 10\ 6\ 15.5\ 9.5\ 14\ 19]\,\mathrm{dB}$. For $M = 21$ and $M = 31$ the gains are repeated until the number of bands.

responses of Figure 3.4, it can be seen as the proposed GEQ has a frequency response similar to the multirate approach, which is the one with the lowest error. In particular, both equalizers are characterized by very steep transition bands, as shown in Figure 3.5(a), and linear phase, as proved by the phase responses of Figure 3.5(b).

Table 3.2 shows the comparison results in terms of maximum error, computational complexity, and latency for the configuration with $M = 9$. The error is calculated as the maximum difference, in dB, between the desired and the obtained gains at the center frequencies, defined in Equation (3.2), considering all the possible configurations with ± 12 dB [136], which leads to 512 cases in total when $M = 9$. The error is considered acceptable when it is below 1 dB, according to previous publications that applied the same method to have a quantitative estimation of the GEQ accuracy [136, 137]. In addition, in [138, 139], listening experiments have proven that the audible peak level for octave filters is below 1 dB when white noise is considered as input, while the just noticeable difference in the deviation of the magnitude response is higher than ± 1 dB with other signals, as declared also in [140]. The computational cost of the proposed equalizer is evaluated in terms of the number of multiplications, computed as

$$\text{n°mult.} = \frac{3M+1}{2} + \sum_{m=1}^{(M+1)/2} N_m^{\mathrm{F}} + \sum_{m=1}^{M} N_m^{\mathrm{G}}, \tag{3.22}$$

and in terms of the number of additions, i.e.,

$$\text{n°add.} = \sum_{m=1}^{(M+1)/2} N_m^{\mathrm{F}} + \sum_{m=1}^{M} N_m^{\mathrm{G}}, \tag{3.23}$$

where N_m^{F} and N_m^{G} are the order of the mth filters $F(z)$ and $G(z)$, respectively.

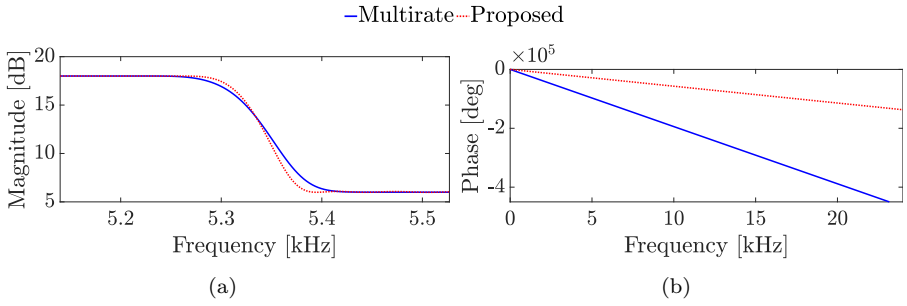


Figure 3.5: (a) zoom of the magnitude response and (b) phase response of the multirate and the proposed GEQs shown in Figure 3.4(a).

Table 3.2: Performance of the proposed uniform GEQ in terms of maximum error, number of multiplications and additions per output sample, and latency (in samples), in comparison with other linear-phase GEQs, considering $M = 9$ bands.

Equalizer	Error [dB]	Mul	Add	Latency
FFT-based	0.47	**100**	**144**	4 857
Multirate	**0.07**	25 209	25 189	2 686
Hergum IFIR	0.40	2 987	2 958	765
Uniform IFIR (proposed)	0.47	2 990	2 976	**761**

As shown by the Equation (3.22), the filter $F(z)$ is considered only in the first half of the filterbank. Finally, latency is given by the maximum delay calculated following Equations (3.21)-(3.20), i.e., the maximum value of τ_m shown in Table 3.1.

Analyzing the obtained results in Table 3.2, all the GEQs present an error below 0.5 dB, broadly below the acceptability limit of 1 dB. The multirate approach reaches the lowest error of 0.07 dB, but it has a computational complexity too large for real-time applications. The FFT-based equalizer uses a single filter obtained from the proposed structure, so it has the same performances in terms of frequency response, and error, while the computational load and the latency are not the same due to the different implementations. In fact, the FFT-based equalizer is the less expensive in terms of the number of operations (with only 100 multiplications and 144 additions per output sample), while the proposed method needs 2 990 multiplications, and 2 976 additions per output sample, similar to the equalizer of Hergum [131]. Moreover, the proposed GEQ offers the lowest latency of 761 samples, i.e., 16 ms, with a sampling frequency of $F_s = 48$ kHz. The latency of the FFT-based approach is the highest since it is given by the sum of the FFT-length of 4096 with the latency of the final filter of the proposed filterbank, i.e., 761, resulting in 4 857 samples (or 101 ms).

The main drawbacks of the proposed uniform GEQ are the uniform band division and the high computational cost. In fact, a logarithmic band division is preferred in GEQs and the elevated computational load is due to the use of brick-wall filters (i.e., very steep transition bands), that are not necessary for GEQ filterbanks.

3.1.3 Linear-phase octave graphic equalizer

As said, uniform GEQs are not much employed in audio applications since a logarithmic band division is preferred due to the human perception of sound and the nature of music. To face the drawbacks of the uniform GEQ described in the previous section, this section proposes the new design of [23] for an octave

GEQ with the following ten band center frequencies, or command frequencies: 31.25 Hz, 62.5 Hz, 125 Hz, 250 Hz, 500 Hz, 1.0 kHz, 2.0 kHz, 4.0 kHz, 8.0 kHz, and 16.0 kHz. The bands are numbered from lowest to highest using index $m = 1, 2, 3, ..., 10$. This design uses the sample rate of $F_\mathrm{s} = 48$ kHz, which is common in professional and mobile audio.

Filter structure

The overall scheme of the proposed linear-phase octave-band equalizer is shown in Figure 3.6. The proposed structure is a tree structure that reminds the one obtained by classic wavelet transformation [141]. The highest band is obtained by the signal path at the top of the figure, while the lowest one by the path at the bottom. The final output is computed by summing the branch outputs.

The filterbank of Figure. 3.6 is designed starting from a half-band lowpass prototype FIR filter $H_\mathrm{LP}(z)$ of even order N and odd length $N + 1$. The impulse response of the prototype filter must be symmetric (i.e., have a linear phase) by the definition of FIR filters, and the delay D to the center point of the prototype filter, in samples, is $D = N/2$. The highest band (i.e., the tenth band) of the equalizer $H_{10}(z)$ is designed as a complementary highpass filter $H_\mathrm{HP}(z)$ of the prototype filter, so $H_{10}(z) = H_\mathrm{HP}(z)$, where

$$H_\mathrm{HP}(z) = z^{-D} - H_\mathrm{LP}(z). \tag{3.24}$$

Due to the fact that integer interpolation factors are used, the cutoff frequency of the lowpass filter is $f_\mathrm{c} = 12$ kHz, which is half of the Nyquist limit 24 kHz. In the proposed design, this corresponds to the cutoff frequency of the highest band, which, in an octave GEQ, is the band edge between the 8-kHz and 16-kHz

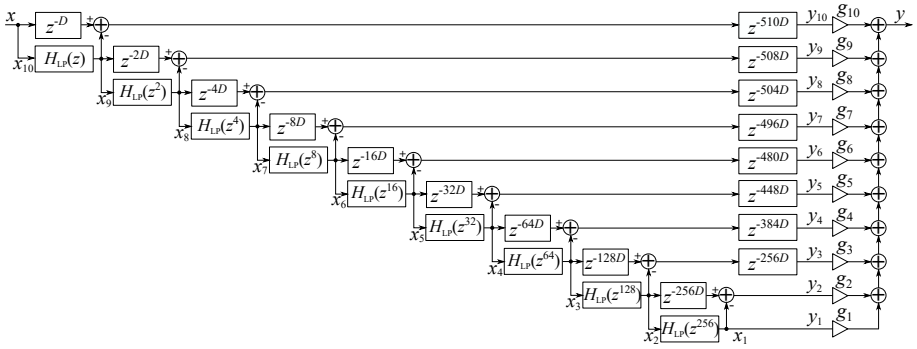


Figure 3.6: Block diagram of the proposed parallel graphic equalizer for ten octave bands. The signal path at the top produces the highest band (16 kHz) whereas the bottom one produces the lowest band (31.25 Hz).

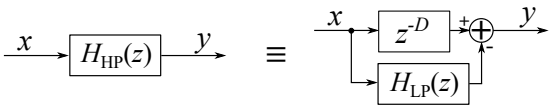


Figure 3.7: Scheme of the complementary filter.

octave bands equaling to $\sqrt{8 \times 16} \approx 11.3\,\text{kHz}$. Therefore, the cutoff frequency of the proposed highest band filter is slightly shifted with respect to the usual octave-band filterbanks. This does not affect the final realization or accuracy of the graphic equalizer, as it is only required to monitor the magnitude error at the command frequencies.

According to Equation (3.24), the filter $H_{\text{HP}}(z)$ can be implemented using a delay line and a subtraction, once the lowpass filtered signal going to the lower bands has been computed using $H_{\text{LP}}(z)$, as shown in Figure 3.7. In addition to computational efficiency, another advantage of complementary filters is the fact that the total response is completely flat, when the neighboring band filters have the same gain, as shown in Figure 3.8. Hergum also pointed out this advantage in his study [131].

The rest of the bands of the filterbank are obtained with stretched versions of the prototype filter, such as $H_{\text{LP}}(z^2)$ and $H_{\text{LP}}(z^4)$, which are prepared by inserting one or three zero samples between every two coefficients of the prototype FIR filter, respectively [142]. The general scheme of delay and filtering operations for the mth band is presented in Figure 3.9(a). The $\mathcal{Z}$ transform of the mth band output signal $Y_m(z)$ is obtained from the input signal $X(z)$ as follows:

$$Y_m(z) = H_m(z)X(z), \tag{3.25}$$

where $H_m(z)$ is the transfer function of the mth band and is computed as

$$H_m(z) = z^{-\Delta_m}[z^{-DL_m} - H_{\text{LP}}(z^{L_m})]G_m(z), \tag{3.26}$$

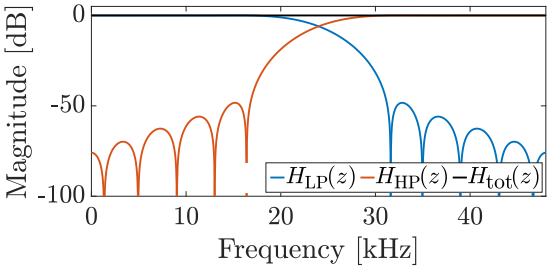


Figure 3.8: Magnitude response of the prototype lowpass filter, its complementary highpass filter, and the total response of their sum.

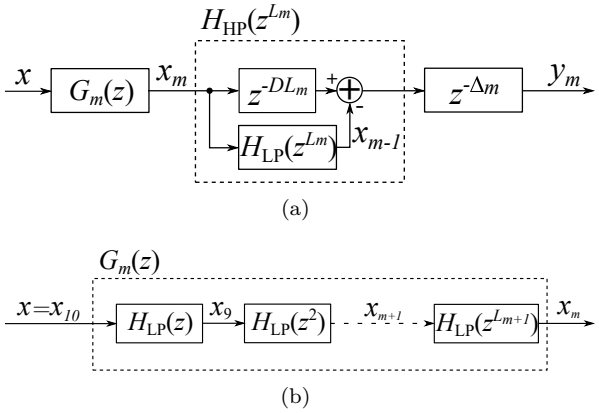


Figure 3.9: (a) filters and delay lines associated with a single band for $m = 2, 3, ..., M$, cf. Fig. 3.6, and (b) details of the transfer function $G_m(z)$.

and the mth interpolation factor L_m is calculated as

$$L_m = 2^{(M-m)} = 2^{(10-m)}, \tag{3.27}$$

and the transfer function $G_m(z)$, which is shown in detail in Figure 3.9(b), is composed of the cascade of all previous band filters:

$$G_m(z) = H_{\mathrm{LP}}(z) \prod_{k=m+1}^{M-1} H_{\mathrm{LP}}(z^{L_k}), \tag{3.28}$$

with $m = 2, 3, ..., M$ and $M = 10$. Looking at Figure 3.9(a), the input signal $x(n)$ is first filtered by the filter $G_m(z)$ and the resulting intermediate signal $x_m(n)$, shown for each band in Figure 3.6, is then filtered by $H_{\mathrm{HP}}(z^{L_m})$ that is implemented through a delay line and a subtraction, according to Equation (3.24). Note that in Figure 3.9(a), when $m = 9$, the signal $x_{10}(n)$ corresponds to the input signal $x(n)$, which is also seen in the top left corner in Figure 3.6. Figure 3.10 shows a design example of the sixth band, with a center frequency of 1 kHz. In this case, the transfer function of the sixth band $H_6(z)$ is obtained by the concatenation of the filter $G_6(z)$ and the filter $H_{\mathrm{HP}}(z^{L_6}) = z^{-DL_6} - H_{\mathrm{LP}}(z^{L_6})$.

A synchronization delay Δ_m, also shown in Figure 3.9(a), must be applied in order to align all the band outputs, and is determined as

$$\Delta_m = \tau - [2^{(M+1-m)} - 1]D = \tau - [2^{(11-m)} - 1]D, \tag{3.29}$$

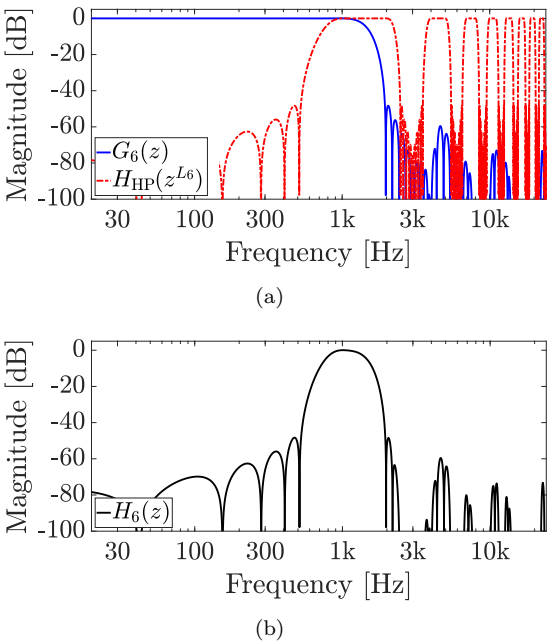


Figure 3.10: Example of the design of the magnitude response of the band filter centered at 1 kHz. Cascading the filters (a) $G_6(z)$ and $H_{\mathrm{HP}}(z^{L_6}) = z^{-DL_6} - H_{\mathrm{LP}}(z^{L_6})$ results in (b) the band filter $H_6(z)$.

where τ is the total delay of the equalizer in samples:

$$\tau = [2^{(M-1)} - 1]D = 511D. \tag{3.30}$$

In Figure 3.6, the synchronization delays Δ_m are shown one upon the other on the right-hand side, next to the command gain factors g_m. In the highest band (the top signal path in Figure 3.6), the total delay of $511D$ samples is formed by the cascade of the delay line z^{-D} and the synchronization delay z^{-510D}. In the lowest band, the synchronization delay is formed by the cascade of all the delay lines between the input (top left corner in Figure 3.6) and the output y_1 (bottom right corner in Figure 3.6), which have the lengths D, $2D$, $4D$, $8D$, $16D$, $32D$, $64D$, $128D$, and $256D$. This adds up to $511D$ samples of delay.

The lowest band filter of the equalizer is obtained as a byproduct when the signal $x_2(n)$ is filtered with the prototype filter stretched by a factor of 2^8, or 256, as shown in Figure 3.6. The resulting signal $x_1(n)$ does not require further processing, as it is the output signal $y_1(n)$ of the lowest band filter. The filter $H_{\mathrm{LP}}(z^{256})$ also implements the largest input-output delay, so a synchronization delay is unnecessary in the two lowest bands, as seen in Figure 3.6.

Finally, as presented in Figure 3.6, the desired gain factor g_m of each band is

applied and the total response of the equalizer $y(n)$ is obtained as a weighted sum of all band output signals:

$$y(n) = \sum_{m=1}^{M} g_m y_m(n). \tag{3.31}$$

Since the band filters determine the gain on their own band very accurately, optimization of filter gains is unnecessary, and command gains can directly be used as weights g_m. This is an advantage with respect to recursive GEQs, for those applications where command gains are varying constantly, such as unmasking EQs for ambient noise [143].

Prototype Filter Design

The overall performance of the proposed GEQ depends on the prototype filter $H_{\mathrm{LP}}(z)$, which is imposed to be a half-band lowpass filter. A peculiarity of half-band filters is that every second sample of the impulse response is zero by definition, except for the middle coefficient [142]. In this way, a half-band filter of order N actually contains only $N_{\mathrm{nz}} = N/2 + 2$ non-zero coefficients. This characteristic allows for reducing the computational cost by avoiding multiplications with zero coefficients during the filtering computation. Moreover, the linear phase characteristic implies that the impulse responses are symmetric, approximately halving the number of multiplications.

The FIR filter could be designed by optimization methods, as the least squares or the Remez algorithm [144], or by other efficient possibilities, such as a method based on iterated sine [145]. In contrast, in this work the filter is designed using the windowing technique [144], which is the simplest method, but effective for the proposed system. Starting from the cutoff frequency $f_{\mathrm{c}} = 12$

Table 3.3: Window function tested for the design of the prototype filter.

Window	Equation
Rectangular	$\mathrm{w}(n) = 1$
Bartlett	$\mathrm{w}(n) = \frac{2}{N}\left[\frac{N}{2} - \left\lvert n - \frac{N}{2}\right\rvert\right]$
Hamming	$\mathrm{w}(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N}\right)$
Hanning	$\mathrm{w}(n) = 0.5\left[1 - \cos\left(\frac{2\pi n}{N}\right)\right]$
Blackman	$\mathrm{w}(n) = 0.42 - 0.5\cos\left(\frac{2\pi n}{N}\right) + 0.08\cos\left(\frac{4\pi n}{N}\right)$
Kaiser	$\mathrm{w}(n) = \frac{I_0\left(\beta\sqrt{1-\left(\frac{2n}{N}-1\right)^2}\right)}{I_0(\beta)}$

kHz, the prototype filter coefficients are computed as

$$h_{\mathrm{LP}}(n) = \mathrm{w}(n)\left[\frac{\sin\big(\omega_c(n-D)\big)}{\pi(n-D)}\right], \tag{3.32}$$

where $\omega_\mathrm{c} = 2\pi f_\mathrm{c}/F_\mathrm{s}$, $\mathrm{w}(n)$ is the window function applied, and $D = N/2$ is the delay of the filter, with N the filter order, so the filter length is $N+1$.

In this study, several window functions $\mathrm{w}(n)$ have been tested for the design of the prototype filter: rectangular, Bartlett, Hamming, Hanning, Blackman, and Kaiser windows, reported in Table 3.3. The rectangular and the Bartlett windows are the simplest ones, but they are characterized by a modest attenuation in the stopband. For this reason, they do not guarantee an acceptable performance and are not involved in the following. The Hamming and Hanning windows have similar properties in terms of transition band and attenuation. The Blackman method ensures the largest attenuation but has a wide transition band. Regarding the equation of the Kaiser window [146], I_0 is the zeroth-order modified Bessel function of the first kind, and the attenuation depends on the

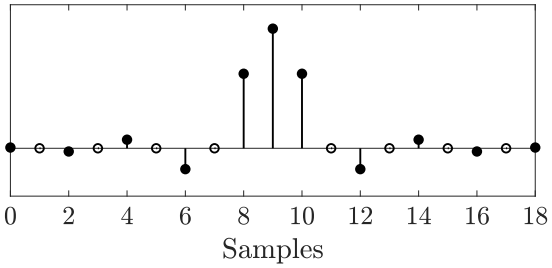


Figure 3.11: Design of the prototype filter with the Kaiser window with $\beta = 4$. The filter order is $N = 18$ (i.e., 19 samples long), but it has only $N_\mathrm{nz} = 11$ non-zeros coefficients (shown with black dots).

Table 3.4: Coefficients of the FIR prototype filter of Fig. 3.11.

Index	Value	Index	Value
0	0.00313	10	0.31158
1	0	11	0
2	-0.01338	12	-0.08718
3	0	13	0
4	0.03593	14	0.03593
5	0	15	0
6	-0.08718	16	-0.01338
7	0	17	0
8	0.31158	18	0.00313
9	0.5		

parameter β. In general, a bigger β guarantees a higher attenuation in the stopband of the filter [146]. In view of this, only Hamming, Hanning, Blackman, and Kaiser methods are considered in the following. Figure 3.11 and Table 3.4 show an example filter design, using the Kaiser window with a length of 19 samples and a coefficient $\beta = 4$. Thus, the final filter has an order of $N = 18$ and the number of non-zeros elements is $N_{\text{nz}} = 11$.

Experimental results of the linear-phase octave graphic equalizer

The proposed equalizer is first evaluated in terms of error of the magnitude response, comparing different orders and four windowing methods for the prototype filter design: Hamming, Hanning, Blackman, and Kaiser. Table 3.5 shows the performance of the proposed octave equalizer varying the orders N and the design of the prototype lowpass filter. The latency and the computational complexity are proportional to the filter order. For the experiments, a sampling frequency of 48 kHz has been used. For the Kaiser window, the parameter value $\beta = 4$ is chosen for all the simulations after empirical studies, since it ensures the lowest error when using low filter orders, i.e., when having as low computational cost as possible. Lower values of β do not guarantee sufficient attenuation to obtain acceptable accuracy. Instead, higher values of β ensure a better attenuation, but the error of the total equalizer becomes acceptable only by increasing the filter order.

Table 3.5: Performance of the proposed equalizer with varying orders N and designs of the prototype lowpass filter. The designs having their maximum error below 1 dB are highlighted.

N (N_{nz})	Latency τ	Mul (sym)	Add	Window	Error [dB]
18 (11)	4599	109 (64)	108	**Kaiser**	**0.79**
				Hamming	2.44
				Hanning	**0.99**
				Blackman	5.92
26 (15)	6643	145 (82)	144	Kaiser	1.03
				Hamming	1.99
				Hanning	1.13
				Blackman	**0.76**
54 (29)	13797	271 (145)	270	**Kaiser**	**0.82**
				Hamming	**0.96**
				Hanning	**0.91**
				Blackman	**0.05**

Similarly to the uniform equalizer of the previous section, the error is calculated as the maximum difference between the desired and the obtained gains at the octave center frequencies, considering all the possible configurations with ± 12 dB, which leads to 1024 cases in total [136]. Moreover, when two adjacent bands have the same gain, the error is computed as the maximum deviation from the straight line that connects the two gains at the center frequencies. As explained in the experimental results of Section 3.1.2, the error is considered acceptable when it is below 1 dB, according to [136, 137], in which the GEQ accuracy is calculated in the same way.

In Table 3.5, $N = 18$ is the lowest filter order considered, since it uses the shortest window that leads to a 1-dB accuracy. Empirical tests proved that shorter windows lead to larger errors. Looking at Table 3.5, it is worth noting that sometimes the error increases with the increase of the filter order. In particular, this happens with Kaiser and Hanning windows which are characterized by a lower attenuation. In fact, the increase of the filter order N makes the transition band steeper but produces more lobes in the stopband maintaining the same attenuation, as shown in Figure 3.12. These lobes can cause a wider ripple on the total response of the GEQ that may make the error exceed 1 dB, especially when the command gain is -12 dB. The latency τ is the delay in samples of the total equalizer and is computed following Equation (3.30).

The filtering is implemented by avoiding the operations with zero elements of the filter, so the number of multiplications for each output sample is calculated as

$$\text{n}^\circ \text{ mult.} = (M-1)N_{\text{nz}} + M, \tag{3.33}$$

where N_{nz} is the number of non-zero elements of the prototype filter and $M = 10$ is the number of bands. The number of multiplications can be further

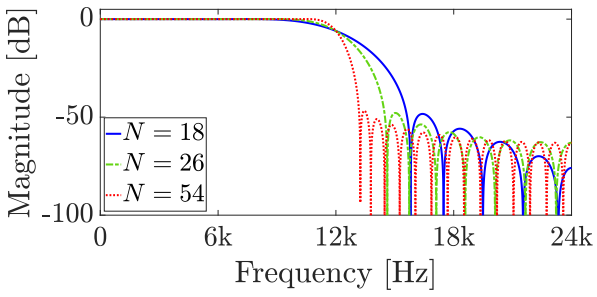


Figure 3.12: Design of the prototype lowpass filter varying the order N, using a Kaiser window function with $\beta = 4$.

reduced by accounting for the symmetry of the impulse responses as

$$\text{n}^\circ \text{ mult. (sym)} = (M-1)\frac{N_{\text{nz}}+1}{2} + M. \tag{3.34}$$

Finally, the number of additions is computed as follows:

$$\text{n}^\circ \text{ add.} = (M-1)N_{\text{nz}} + M - 1. \tag{3.35}$$

Analyzing the results of Table 3.5, the Blackman technique with $L_{\text{win}} = 57$ shows the lowest error (0.05 dB), but the computational cost (541 operations per sample) and the latency (13 797 samples, or 287 ms) are the highest. A latency that large is unacceptable for some applications, such as live sound or sound with moving images; however, for audio playback, without visual or other references, even such a latency may be acceptable. The Hamming window has the worst performance in terms of both computational cost and error. Finally, the Hanning method with $N = 20$ and the Kaiser method with $N = 18$ both guarantee an acceptable error (below 1 dB) with the lowest computational cost (64 multiplications and 108 additions, or 172 operations per output sample). The Kaiser technique shows a lower error equal to 0.79 dB, which is thus considered the best design and is used in the comparison with the other methods. The total latency of the equalizer is $\tau = 4599$ samples, or 95.8 ms at the sample rate of 48 kHz.

Figure 3.13 shows the output signals of each band of the proposed equalizer

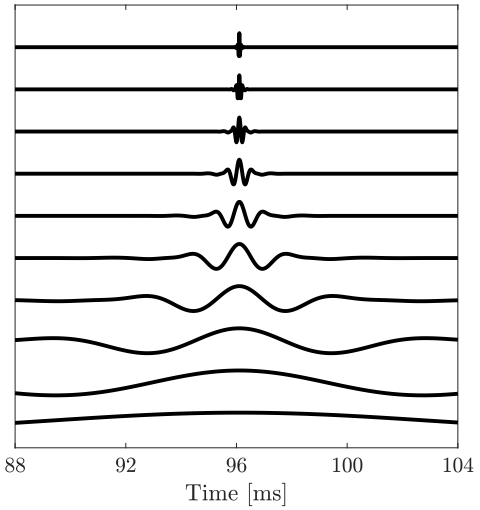


Figure 3.13: Band filter impulse responses of the proposed GEQ, using the Kaiser window, from the highest band (top) to the lowest one (bottom).

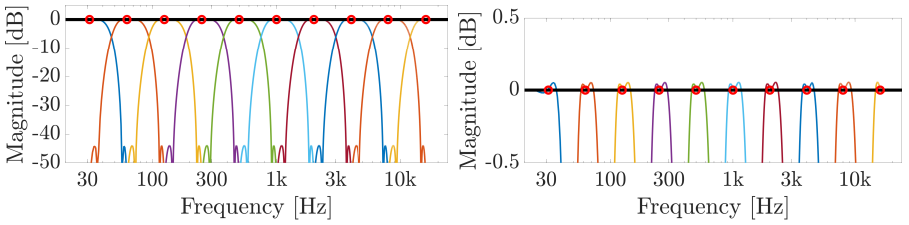


Figure 3.14: (a) Magnitude responses of the band filters with all the command gains (circles) at 0 dB and (b) its details between −0.5 dB and 0.5 dB. The solid black line shows the total response.

as a response to a unit impulse from the highest band (on the top of the figure) to the lowest one (on the bottom). All band filters are seen to be symmetric, which implies a linear phase response.

Figure 3.14 shows the magnitude frequency response of each band and the total frequency response of the equalizer when all the gains are set to the same value of 0 dB. The use of complementary filters guarantees a completely flat total response. Even if the single band presents a ripple, the total response is flat thanks to the compensation of the stopband ripples of the adjacent filters, as shown in Fig. 3.14(b).

Figure 3.15 shows example magnitude frequency responses of three different test configurations:

a) the zigzag command settings (±12 dB);

b) the special zigzag setting: [12 -12 -12 12 -12 -12 12 -12 -12 12] dB, which is declared the most difficult case for the equalizer of [147];

c) an arbitrary setting [8 10 -9 10 3 -10 -6 1 11 12] dB.

In Figure 3.15, the response obtained by applying the Blackman window with $N = 54$ and the one obtained with the Kaiser window with $N = 18$ are reported. Although the Blackman window with $N = 54$ guarantees the lowest error (0.05 dB), the final equalizer shows steeper transition bands. However, sharp transitions lead to a lengthening of the impulse response, and, thus, more audible pre-ringing for linear-phase filters, which can ruin the important transients of musical instrument sounds.

Figure 3.16 shows total impulse responses of the proposed GEQ for the first configurations of Figure 3.15 comparing the Blackman widow and the Kaiser window. All the impulse responses in Figure 3.16 are symmetric, which also proves the linear phase of each band filter and the total response of the equalizer. The proposed system has been tested also varying the sampling frequency

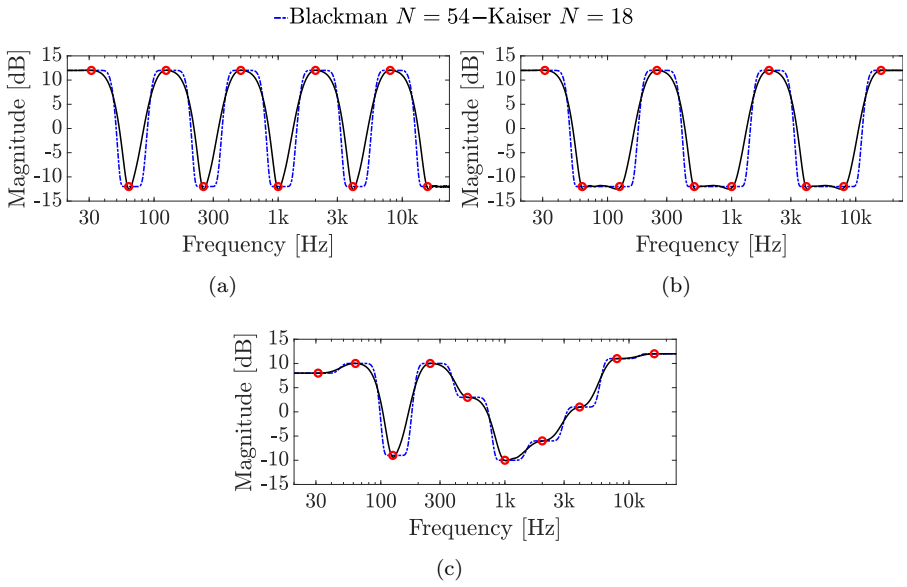


Figure 3.15: Magnitude response of the proposed equalizer for two different prototype filters, considering (a) the zigzag configuration (±12 dB), (b) the gains [12 -12 -12 12 -12 -12 12 -12 -12 12] dB, and (c) the arbitrary gains [8 10 -9 10 3 -10 -6 1 11 12] dB.

F_s. A sampling frequency of 44.1 kHz produces a slight decrease of the center frequencies with the ratio of $44.1/48 = 0.918$, but otherwise, the same performance is obtained using the same filter coefficients. However, higher sampling rates, such as 88.2 kHz and 96 kHz, would require a larger prototype filter order N to guarantee an acceptable error.

The proposed octave equalizer is also compared with previous linear-phase FIR octave GEQ designs in terms of error, computational cost, and latency. The FFT-based equalizer of Schöpp and Hetze [127] and a single FIR GEQ [119] obtained from the proposed structure are included in the comparison. Other linear-phase multirate state-of-the-art approaches, such as the multirate GEQ of [130], have not been considered in the comparison, since they have a very large latency and computational cost not competitive with the proposed method, as shown for the uniform GEQ in Table 3.2.

The FFT-based equalizer of [127] consists in the design of a target frequency response of the equalizer that depends on the desired gains and on the filtering of the input signal with that frequency response using the overlap and add method with an overlap of 50% [135]. Here, the target frequency response is calculated by summing the filter responses of the proposed IFIR structure. The FFT length of 16384 is chosen to obtain the same response and the same

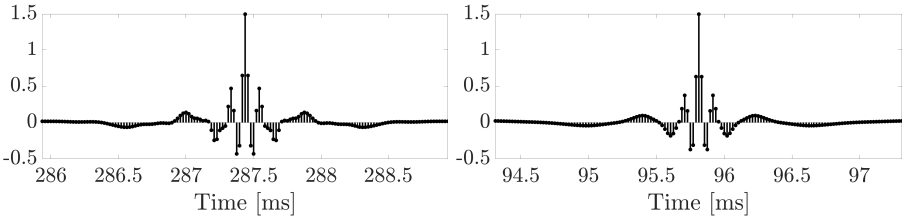


Figure 3.16: Impulse response of the proposed equalizer designed using (a) the Blackman window with an order of $N = 54$ and (b) the Kaiser window with an order of $N = 18$ for the configuration of Fig. 3.15(a).

error as the proposed implementation, so the frame size of the overlap and add method has a length of 8192 samples.

The single FIR method, similarly to the FFT-based one, is formed as the sum of the filter responses of the proposed IFIR structure and executes the time-domain convolution. The length of the single FIR filter is 9199, and it produces the same error as the proposed equalizer.

Table 3.6 compares the proposed equalizer and the other two methods in terms of latency and computational cost. For each method, the table shows the latency in samples of the total equalizer and the number of multiplications and additions per output sample. All three methods have exactly the same transfer function, and thus, the same error equal to 0.79 dB, as shown in Table 3.5.

The FFT-based and the single FIR methods use the same filters, but apply different implementations. Table 3.6 shows that the FFT method presents the largest latency because of the frame-based FFT processing that introduces an algorithmic delay of 16384 samples in addition to the filter group delay of 4599 samples. The latency could be reduced using the zero-latency partitioned convolution [148, 149]. In that case, the latency would be the same as that of

Table 3.6: Performance of the proposed octave equalizer (with the Kaiser window design) compared with other linear-phase octave GEQs. The symmetry has been accounted for in the number of multiplications. The best result in each column is highlighted.

Method	Error [dB]	Mul	Add	Latency
Single FIR	**0.79**	4 600	9 198	**4 599**
FFT-based	**0.79**	116	168	20 983
Octave IFIR (proposed)	**0.79**	**64**	**108**	**4 599**

the proposed method but the computational cost would be larger. Table 3.6 also shows that the FFT GEQ needs considerably more multiplications and additions (284 operations, in total) than the proposed method (172 in total). The proposed method thus requires 39% less operations per sample than the FFT method.

The time-domain filtering carried out with the single FIR presents the same latency as the proposed method, but 80 times larger computational complexity, which is seen by comparing the number of multiplications and additions in Table 3.6. The proposed method shows the best performance in terms of latency (4599 samples or 95.8 ms, which is 78% less than the FFT method) and computational complexity (172 operations per sample, of which 64 multiplications and 108 additions).

The computational complexity of the proposed equalizer is competitive even with IIR filters. The state-of-the-art IIR octave GEQ uses 50 multiplications per sample [150], that is 78% of the multiplications needed by the proposed GEQ (64 multiplications per sample). The required delay memory is much larger in the proposed method than in IIR equalizers, however. Additionally, the proposed GEQ does not require any operations, when the command gains are changed, whereas, in IIR-based GEQs, the filter gains must be optimized, e.g., using a neural network [111, 151].

The fairly large latency of the proposed method, almost 100 ms, seems large but it is acceptable in audio playback. It still raises the question of whether this much latency could cause a synchronization problem when sound is associated with the video. However, an ITU recommendation states that the detection threshold for latency of sound with respect to vision is 125 ms and the acceptability threshold is 185 ms [152]. This implies that the latency of the proposed method by itself does not exceed the detection threshold in audiovisual synchronization.

3.1.4 Low-latency quasi-linear-phase octave graphic equalizer

Aiming at reducing the latency of the linear-phase octave GEQ proposed in the previous section, a low-latency quasi-linear-phase GEQ [24] is presented in this section. The scheme of the proposed equalizer is shown in Figure 3.17. The total structure is a hybrid implementation obtained combining IIR and

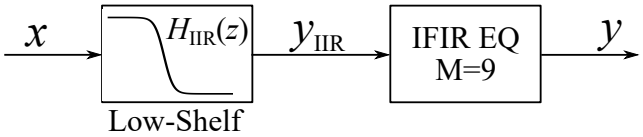


Figure 3.17: Scheme of the proposed hybrid graphic equalizer.

FIR filters. This structure introduces a zero-latency IIR filter to reduce the number of FIR filters that cause the latency increase. In particular, the input signal $x(n)$ is filtered by an eighth-order IIR low-shelving filter that designs the first band of the GEQ. Then, the output of the IIR part $y_{\mathrm{IIR}}(n)$ is filtered by the octave IFIR GEQ that implements $M = 9$ bands, from the second to the tenth, using the structure of [23], described in Section 3.1.3. The design of the IIR filter and the IFIR filterbank is explained below.

IIR equalizer

The first band of the equalizer is obtained through an 8th-order lowpass IIR shelving filter, that is designed following the implementation of [137]. The order of the IIR shelving filter is chosen as $N_{\mathrm{I}} = 8$ in order to guarantee a transition band steep enough not to affect the subsequent frequency bands. The dB-gain of the IIR filter is set to

$$\Gamma = \Gamma_1 - \Gamma_2, \tag{3.36}$$

where Γ_1 and Γ_2 are the desired dB-gains of the first and the second band, respectively. The gain of the second band is subtracted to restore the low-frequencies gains to zero since the second band is designed as a lowpass filter in the IFIR structure. The linear value of the IIR filter gain is defined as $g = 10^{\Gamma/20}$.

For the design of the shelving filter, the normalized digital cutoff frequency is set to the geometric mean of the neighboring center frequencies, that is,

$$\omega_{\mathrm{c}} = \sqrt{\omega_1 \omega_2}, \tag{3.37}$$

where $\omega_1 = 2\pi \cdot 31.25/F_{\mathrm{s}}$ and $\omega_2 = 2\pi \cdot 62.5/F_{\mathrm{s}}$, with 31.25 Hz and 62.5 Hz the center frequencies of the first and the second band, respectively. This results in a cutoff frequency in Hertz of approximately 44 Hz that corresponds to the center point between the two first center frequencies on a logarithmic axis. The transfer function of the low-shelving IIR filter $H_{\mathrm{IIR}}(z)$ comprises cascade second-order sections, and it is obtained as

$$\begin{aligned} &H_{\mathrm{IIR}}(z) = \\ &\prod_{i=1}^{N_{\mathrm{I}}/2} \Bigg(1 + 2V \frac{\zeta\big(\zeta + \gamma_i + 2\zeta z^{-1} + (\zeta - \gamma_i) z^{-2}\big)}{1 + 2\zeta\gamma_i + \zeta^2 + (2\zeta^2 - 2) z^{-1} + (1 - 2\zeta\gamma_i + \zeta^2) z^{-2}} + \\ &+ V^2 \frac{\zeta^2\big(1 + 2z^{-1} + z^{-2}\big)}{1 + 2\zeta\gamma_i + \zeta^2 + (2\zeta^2 - 2) z^{-1} + (1 - 2\zeta\gamma_i + \zeta^2) z^{-2}} \Bigg), \end{aligned} \tag{3.38}$$

where $V = \sqrt[N_\mathrm{I}]{g} - 1$ and $\gamma_i = \cos(\alpha_i)$, with

$$\alpha_i = \left(\frac{1}{2} - \frac{2i-1}{2N_\mathrm{I}}\right)\pi. \tag{3.39}$$

Looking at Equation (3.38), the three terms of the product describe a unique second-order section, since the denominators of the second-order terms are identical. The constant ζ is used to map the desired digital cutoff frequency ω_c to the analog one $\Omega_\mathrm{c} = \sqrt[2N_\mathrm{I}]{g}$ and it is computed as

$$\zeta = \frac{1}{\sqrt[2N_\mathrm{I}]{g}} \tan\left(\frac{\Omega_\mathrm{c}}{2}\right). \tag{3.40}$$

Once the IIR filter is designed, the output signal of the IIR equalizer $y_\mathrm{IIR}(n)$ is obtained by filtering the input signal $x(n)$ with the obtained filter as shown in Figure 3.17, so the $\mathcal{Z}$ transform of the output $Y_\mathrm{IIR}(z)$ is obtained as follows:

$$Y_\mathrm{IIR}(z) = H_\mathrm{IIR}(z)X(z). \tag{3.41}$$

IFIR equalizer

The IFIR structure is a modification of the one of [23], described in Section 3.1.3. The scheme of the IFIR part is shown in Figure 3.18 and it is derived from a half-band lowpass prototype FIR filter $H_\mathrm{LP}(z)$ of even order N, and delay $D = N/2$, as for the octave GEQ of Section 3.1.3, where the design of the prototype filter is analyzed. The difference is that, in this case, the IFIR structure designs only nine bands from the second to the tenth. In fact, in comparison with the previous structure of Figure 3.6, the last branch at the bottom, which produces the output $y_1(n)$ of the first band, is deleted in Figure 3.18 and the synchronization delays of the branches are reduced by half.

Similarly to the linear-phase octave GEQ of Section 3.1.3, the mth band output signal $Y_m(z)$ is obtained as

$$Y_m(z) = H_m(z)Y_\mathrm{IIR}(z), \tag{3.42}$$

where $Y_\mathrm{IIR}(z)$ is the output of the IIR filter, described above, and $H_m(z)$ is the transfer function of the mth band, computed following Equations (3.26)-(3.28), with $m = 2, 3, \ldots, M+1$, and $M = 9$.

The synchronization delay of the mth band Δ_m, which is needed to align all the band outputs, is determined as

$$\Delta_m = \tau - [2^{(M+2-m)} - 1]D = \tau - [2^{(11-m)} - 1]D, \tag{3.43}$$

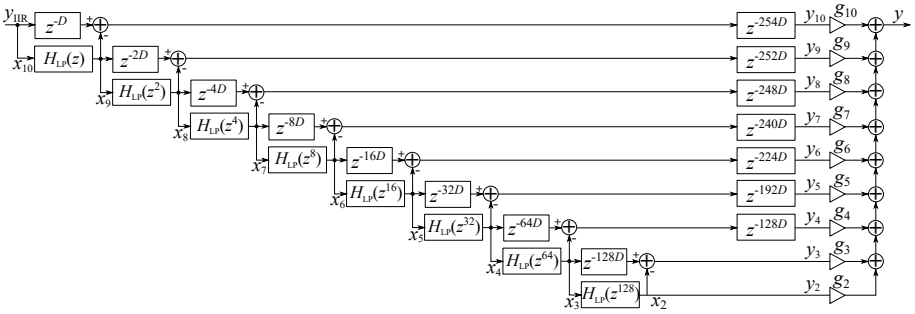


Figure 3.18: IFIR GEQ structure implementing the bands from the second to the tenth ($M = 9$). It is a modification of a previous 10-band GEQ, shown in Fig. 3.6.

where τ is the total delay of the equalizer in samples:

$$\tau = [2^{(M-1)} - 1]D = 255D, \tag{3.44}$$

which is half of the delay of the linear-phase octave GEQ of Equation (3.30).

In Figure 3.18, the synchronization delays Δ_m are shown one upon the other on the right-hand side, next to the command gain factors g_m. In the highest band (the top signal path in Figure 3.18), the total delay of $255D$ samples is formed by the cascade of the delay line z^{-D} and the synchronization delay z^{-254D}. In the second band (the lowest of the IFIR equalizer), the synchronization delay is formed by the cascade of all the delay lines between the input (top left corner in Figure 3.18) and the output y_2 (bottom right corner in Figure 3.18), which have the lengths D, $2D$, $4D$, $8D$, $16D$, $32D$, $64D$ and $128D$. This adds up to $255D$ samples of delay.

The second band filter of the equalizer is obtained as a byproduct, when the signal $x_3(n)$ is filtered with the prototype filter stretched by a factor of 2^7, or 128, as shown in Figure 3.18. The resulting signal $x_2(n)$ does not require further processing, as it is the output signal $y_2(n)$ of the second band filter. The filter $H_{\mathrm{LP}}(z^{128})$ also implements the largest input-output delay, so a synchronization delay is unnecessary in the two lowest bands, as seen in Figure 3.18. Finally, gain factor g_m of each band is applied and the total response of the equalizer $y(n)$ is obtained as a weighted sum of all band output signals from the second band to the tenth:

$$y(n) = \sum_{m=2}^{M+1} g_m y_m(n), \tag{3.45}$$

where $y_m(n)$ is the output of the mth band, calculated by Equation (3.42).

Performance of the hybrid octave graphic equalizer

The performance of the hybrid octave GEQ has been evaluated by implementing a lowpass shelving IIR filter of order $N_{\mathrm{I}} = 8$ and an IFIR filterbank of $M = 9$ bands based on a half-band lowpass prototype FIR filter $H_{\mathrm{LP}}(z)$ of order $N = 18$, designed with the Kaiser window with $\beta = 4$. A sample rate of $F_{\mathrm{s}} = 48\,\mathrm{kHz}$ has been used. Figure 3.19 shows the magnitude response of the IIR filter $H_{\mathrm{IIR}}(z)$, of the IFIR filterbank with $M = 9$ and of the total equalizer with a zigzag command setting (± 12 dB). In the figure, the red circles correspond to the desired gain of each mth band g_m at the related center frequency. The gain Γ of the IIR filter, depicted by the blue curve in Figure 3.19(a), is set to $24\,\mathrm{dB}$ ($12 + 12\,\mathrm{dB}$), according to Equation (3.36).

In the same way as the previous GEQ, the error in the frequency response can be calculated as the maximum difference between the desired and the obtained gains at the octave center frequencies, including all the possible configurations with $\pm 12\,\mathrm{dB}$, which leads to 1024 cases in total [136]. For the hybrid GEQ, the resulting error is $0.76\,\mathrm{dB}$, lower than the acceptability limit of $1\,\mathrm{dB}$ [136, 137]. The computational cost, evaluated by the number of operations per output sample, depends on the filtering with the IFIR filterbank and on the IIR filtering. The number of operations of the IFIR part is obtained by Equations

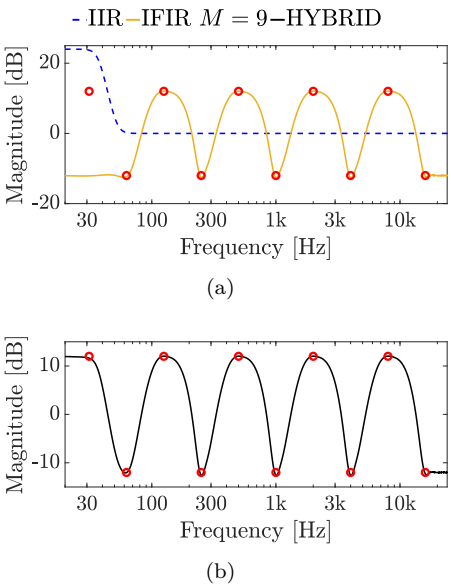


Figure 3.19: Magnitude responses of (a) the IIR part of the equalizer and the IFIR part and (b) the total proposed hybrid equalizer after the gains computation with the zigzag configuration ($\pm 12\,\mathrm{dB}$).

(3.34) and (3.35), where $N_{\text{nz}} = N/2 + 2$ is the number of non-zero elements of the half-band filter of order $N = 18$, and $M = 9$ is the number of IFIR bands. Moreover, the 8th-order shelving IIR filter ($N_{\text{I}} = 8$) adds $5N_{\text{I}}/2 = 20$ multiplications and $4N_{\text{I}}/2 = 16$ additions to the total number of operations, considering a cascade of four (corresponding to $N_{\text{I}}/2$) second-order sections implemented with the direct form II. The resulting number of multiplication of the hybrid equalizer is 77 multiplications and 112 additions per output sample. Finally, the latency of the system is introduced only by the FIR part of the equalizer and it is computed following Equation (3.44), where $D = N/2 = 9$ is the delay in samples of the prototype filter. The resulting latency is 2 295 samples, corresponding to 48 ms with a sampling frequency of 48 kHz. Further experimental results of the hybrid GEQ are reported in the following section, in comparison with the previously presented GEQs.

3.1.5 Comparison of the proposed graphic equalizers

In this section, the performances of the three presented graphic equalizers are compared in terms of error, computational complexity, and latency. Table 3.7 summarizes the results discussed in the previous sections. Although the uniform GEQ presents the lowest error of 0.47 dB and latency of 761 (or 16 ms), the computational cost is too high and not competitive with the other structures. Moreover, as underlined above, uniform GEQs are not much employed in audio applications, and a logarithmic band division is preferable.

Comparing the two octave GEQs, the error of the hybrid solution of 0.76 dB is slightly lower than the error introduced by the IFIR equalizer of 0.79 dB, but both the implementations guarantee an acceptable error below the limit of ±1 dB. The computational complexity of the hybrid equalizer is a bit higher than the octave IFIR GEQ. In fact, the octave IFIR equalizer needs a total of 172 operations per input sample (64 multiplications and 108 additions), while the hybrid method needs 189 operations (77 multiplications and 112 additions). Finally, the latency of the hybrid equalizer is reduced by 50% in comparison with the octave IFIR GEQ. In fact, in the hybrid equalizer, only the last nine bands are designed with FIR filters (i.e., $M = 9$), so the total delay is $255D$ (cf.

Table 3.7: Comparison of the performances of the three proposed GEQs. The best result for each column is highlighted.

Method	Error [dB]	Mul	Add	Latency
Uniform IFIR	**0.47**	2 990	2 976	**761**
Octave IFIR	0.79	**64**	**108**	4 599
Octave hybrid IIR/IFIR	0.76	77	112	2 295

Equation (3.44)). On the contrary, in the linear-phase octave equalizer, all the ten bands are designed with FIR filters (i.e., $M = 10$) and the resulting delay is doubled obtaining a value of $511D$ (cf. Equation (3.30)). Using a prototype filter of order $N = 18$, the delay is $D = 9$, so the IFIR equalizer introduces a latency of 4 599 samples (or 96 ms), while the hybrid GEQ shows a latency of only 2 295 samples (or 48 ms). This reduction allows the equalizer to be more competitive in real-time applications, where a big latency is not tolerated.

The two proposed octave GEQs are also compared in terms of magnitude response, impulse response, and group delay. Figure 3.20 shows example magnitude frequency responses of the three different test configurations used also in Section 3.1.3, which considers a zigzag command setting (±12 dB), a special zigzag setting, and an arbitrary setting. As can be seen in Figure 3.20, the magnitude response of the hybrid equalizer perfectly overlaps the response of the IFIR equalizer except in the transition of the first band, where the shelving filter is applied. In fact, in the hybrid GEQ, the first band is narrower than the first band of the IFIR GEQ. However, this characteristic does not affect the performance of the final GEQ, since the desired gain at the first center frequency is always reached.

Figure 3.21 shows the total impulse response of the hybrid equalizer com-

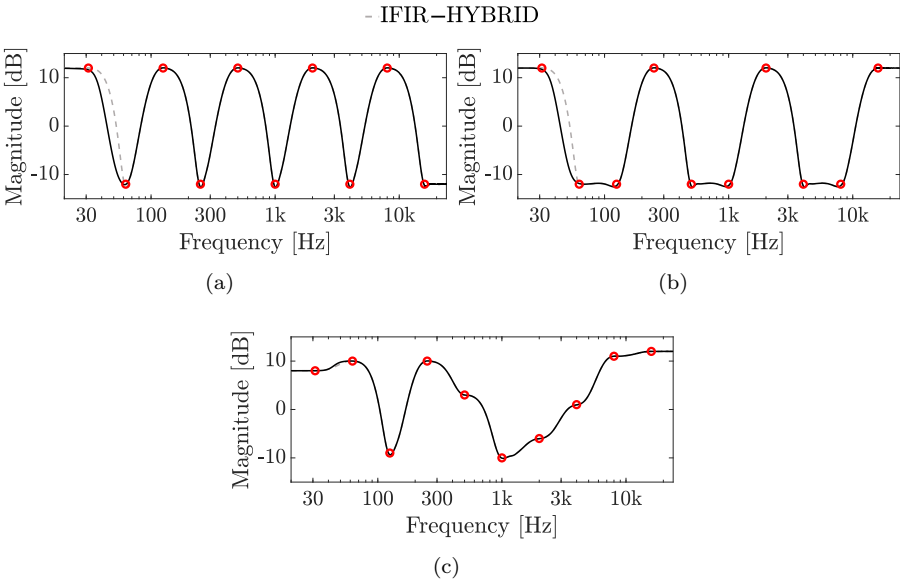


Figure 3.20: Magnitude response of the hybrid equalizer compared with the total IFIR equalizer, with (a) the zigzag configuration (±12 dB), (b) the gains [12 -12 -12 12 -12 -12 12 -12 -12 12] dB, and (c) the arbitrary gains [8 10 -9 10 3 -10 -6 1 11 12] dB.

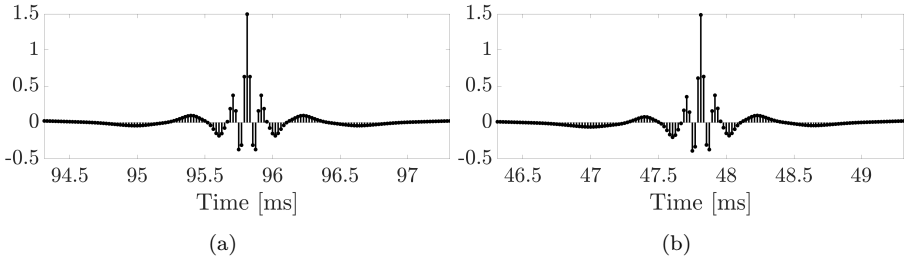


Figure 3.21: Impulse response of (a) the linear-phase octave IFIR equalizer and (b) the hybrid equalizer with the configuration of Fig. 3.20(a), which is slightly asymmetric.

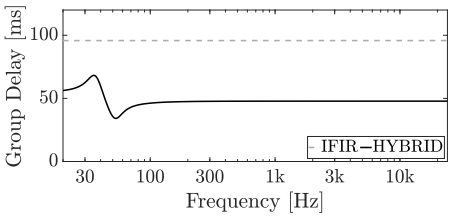


Figure 3.22: Comparison between the group delay functions of the hybrid equalizer and the linear-phase octave IFIR equalizer with the configuration of Fig. 3.20(a).

pared with the IFIR equalizer. This comparison highlights that the nonlinear-phase IIR filters barely affect the symmetry of the total impulse response, but allow to reduce the delay by half. The latency reduction introduced by the hybrid solution is shown also in Figure 3.22, where the group delays of the two implementations are compared. The IFIR equalizer presents a constant group delay of 95.8 ms, which means it has a linear phase. The hybrid equalizer introduces a nonlinearity in the phase at the lower frequencies up to 100 Hz, but at higher frequencies, the group delay assumes a constant value of 47.8 ms.

3.2 Room response equalization

Room response equalization aims at improving the quality of sound reproduction by compensating the undesired effects introduced by the room environment and the reproduction system. The basic idea of RRE systems is to measure the room impulse response and obtain the equalization curve through its inversion. However, several methods for equalization curve calculation exist in the literature.

3.2.1 Background on room response equalization

Automatic EQs are usually applied for room response equalization (RRE), so they require the measurement of the RIR. They can be classified into fixed and adaptive approaches. Fixed EQs are based on a priori design of the equalization curve before the filtering procedure [7, 153–156]. However, a real environment is a time-varying, weakly non-stationary system [154], so the RIR changes with the position [153] and with time [154]. For this reason, adaptive solutions can be found in the literature to solve the problems linked to environmental variations. Both fixed and adaptive EQs can be applied to single-point systems (i.e., single-input single-output SISO and multiple-input single-output MISO) [155] or multi-point systems (i.e., single-input multiple-output SIMO, multiple-input multiple-output MIMO) [7, 157]. The single-point equalization works only in a reduced area around the measurement point since the RIR is measured in a single position. Differently, the multi-point EQ is effective in a wider area because it is obtained from several measurements of the RIRs in different positions. One of the first adaptive RRE approaches was proposed by Elliot *et al.* in [158], where the equalizer is obtained from the minimization of the sum of the squared errors between the equalized signal and the delayed input signal. Ferreira *et al.* [159] have introduced the subbands division of the signal to design a single-point equalizer that is obtained from the update of the subband filter weights. In this way, the system is robust towards peaks and notches of the room transfer function (RTF). In [160, 161], the approach of [159] has been extended to multi-point systems by considering the warped frequency domain. However, all the adaptive equalization methods described above are applied to a single sound source. When multichannel systems are involved, the non-uniqueness problem occurs. A possible way to solve this problem is to minimize the inter-channel coherence [162,163], but this procedure could introduce significant distortions [164]. In [165, 166], a technique for reducing the inter-channel coherence has been applied without modifying the audio quality. The RRE of [165,166] is obtained through the design of a prototype in the warped frequency domain. This method guarantees good performance at low frequencies and a reduced computational cost. In [167], the inter-channel coherence

is minimized by a technique based on the missing fundamental phenomenon, and RIR estimation is obtained by applying a normalized least mean square (NLMS) optimization. Finally, in [25] a subband adaptive structure, based on [97], has been introduced for the impulse response identification to develop a multichannel and multiple position adaptive room response equalizer.

3.2.2 Adaptive multichannel equalization system

This section presents a subband implementation of a multichannel and multiple position adaptive room response equalizer (RRE), proposed in [25]. The proposed system provides an iterative estimation of the room impulse responses and, at the same time, a multipoint equalization. The system repeats the approach presented in [167] with a variation in the room response identification that is performed considering a multirate subband approach in order to reduce the computational cost and to improve the convergence speed. The block diagram of the proposed approach is shown in Figure 3.23. In the case of more than one loudspeaker, convergence problems can occur due to the channel correlation. These problems are very common in multichannel acoustic echo cancellation because of the possible errors in the identification of the acoustic paths [168, 169]. In this context, a method to reduce the inter-channel coherence must be exploited, as described in [170]. Furthermore, the identification of room responses is achieved by subband adaptive filtering using the structure presented in [97]. In this way, $P \cdot Q$ room responses between the P loudspeakers and the Q microphones are estimated and then exploited for the equalizer design. The outputs y_p of the system are obtained from the inputs x_p, with $p = 1, ..., P$, as

$$\begin{bmatrix} y_1 \\ \vdots \\ y_P \end{bmatrix} = \left[\mathbf{H}_{\mathrm{EQ}}\right] \cdot \left[\mathbf{H}_{\mathrm{D}}\right] \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_P \end{bmatrix}, \tag{3.46}$$

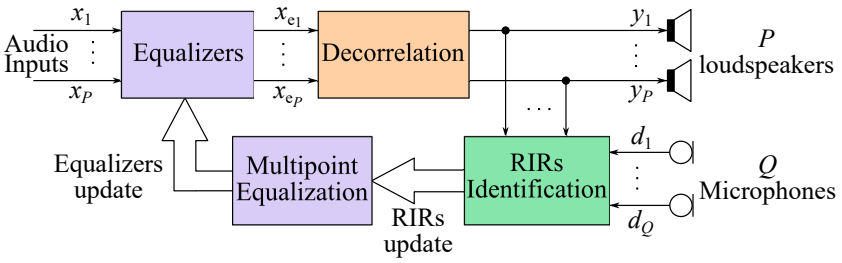


Figure 3.23: Block diagram of the proposed multichannel and multipoint adaptive equalizer.

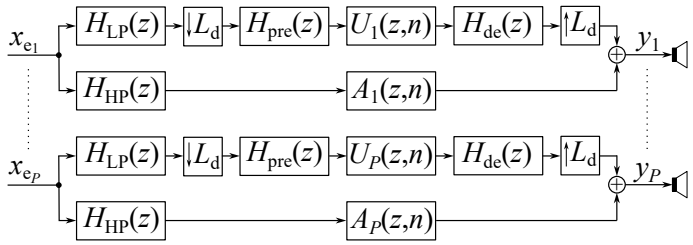


Figure 3.24: Multichannel decorrelation procedure, where $H_{\mathrm{LP}}(z)$ and $H_{\mathrm{HP}}(z)$ are the lowpass and highpass filters, respectively, $U_p(z,n)$ is the adaptive notch filter, $A_p(z,n)$ is the time-varying allpass filter for the pth channel, and $H_{\mathrm{pre}}(z)$ and $H_{\mathrm{de}}(z)$ are the pre-emphasis and de-emphasis filters, respectively.

where $\mathbf{H}_{\mathrm{D}}$ characterizes the decorrelation process, and $\mathbf{H}_{\mathrm{EQ}}$ contains the equalization curves calculated after the room impulse response identification procedure. The three steps of the algorithm, i.e., input signals decorrelation, subband room response identification, and multipoint equalizer design are described in the following.

3.2.3 Input decorrelation

The multichannel input signals decorrelation is obtained by the psychoacoustic criteria of the missing fundamental [171], as reported in [172]. This psychoacoustic phenomenon is associated with the human capability of perceiving the fundamental frequency although it is not actually present in the signal. Moreover, second-order time-varying allpass filters are included in this approach to expand the solution to the entire frequency spectrum [172]. The scheme of the correlation algorithm is reported in Figure 3.24, in which each input channel $x_{\mathrm{e}_p}(n)$ of the decorrelation block, being $p = 1, \cdots, P$, is divided into two subbands [167].

Hence, after a decimation operation of L_{d}, an adaptive notch filter $U_p(z,n)$ is applied in the lower frequencies while a second-order time-varying allpass filter $A_p(z,n)$ is applied in the high-frequency spectrum causing an alteration of the signal phase. This method allows us to accurately identify each processed channel $y_p(n)$.

In the low-frequency band, the notch filters are created by P second-order lattice structures in order to remove the P fundamental frequencies. The pth notch filter is described as

$$U_p(z,n) = \frac{1 + 2k_p(n)z^{-1} + z^{-2}}{1 + k_p(n)[1 + \alpha_p(n)]z^{-1} + \alpha_p(n)z^{-2}}, \tag{3.47}$$

where $p = 1, \cdots, P$ represents the channel index, $k_p(n)$ is the adaptive coefficient connected to the tracked frequency $f_p(n)$ and $\alpha_p(n)$ is the pole-zero contraction factor that controls the bandwidth of the filter [173]. The central frequency of the notch filter can change at each new sample to track the time-varying fundamental frequency that must be deleted. Thus, decorrelation is guaranteed in the whole low-frequency band providing results similar to a set of time-varying allpass filters [174].

The contraction factor $\alpha_p(n)$ is related to each channel and the time-varying vector $\boldsymbol{\alpha}(n) = [\alpha_1(n), \cdots, \alpha_P(n)]$ provides disparity among channels even if the fundamental frequency is the same. In particular, a right circular shift of one sample $s[\boldsymbol{\alpha}(n), 1] = [\alpha_P(n), \alpha_1(n), \cdots, \alpha_{P-1}(n)]$ is applied to the vector $\boldsymbol{\alpha}(n)$ every K samples [172], as explained in the following equation:

$$\boldsymbol{\alpha}(n) = \begin{cases} s[\boldsymbol{\alpha}(n-1), 1] & \text{if } \left(n - K\left\lfloor \frac{n}{K} \right\rfloor\right) = 0 \\ \boldsymbol{\alpha}(n-1) & \text{otherwise,} \end{cases} \tag{3.48}$$

with the vector $\boldsymbol{\alpha}(0) = [0.95\ 0.55]$ at time instant $n = 0$. The adaptive coefficient $k_p(n)$ is included in the interval $(-1, 1)$ to avoid the divergence of the filter and is represented by the following sigmoid function:

$$k_p(n) = \frac{2}{1 + e^{-\gamma_p(n)}} - 1, \tag{3.49}$$

being $\gamma_p(n) \in \mathbb{R}$. The tracking of the fundamental frequency is obtained by finding $\gamma_p(n)$ that minimizes the output energy of the filter in (3.47), as described in [175]. In this way, the filter of (3.47) is completely determined and it is capable of removing the fundamental frequency. It is possible to derive this fundamental frequency $f_p(n)$ with the knowledge of the sampling frequency F_s and of the downsampling factor L_d, considering the following equation:

$$f_p(n) = \frac{F_\mathrm{s}}{L_\mathrm{d}} \cdot \frac{1}{2\pi} \cos^{-1}[-k_p(n)]. \tag{3.50}$$

Moreover, the low-frequency band could contain also some harmonics, so a pre-emphasis filter $H_\mathrm{pre}(z)$ is used in order to improve the method in the low frequencies, and a de-emphasis filter $H_\mathrm{de}(z)$ is applied to annul the effect of the first. These filters are described as follows:

$$H_\mathrm{pre}(z) = \frac{1}{1 - \nu z^{-1}} \tag{3.51}$$

$$H_\mathrm{de}(z) = 1 - \nu z^{-1}, \tag{3.52}$$

being $0 < \nu < 1$.

Considering the high-frequency range, the phase of the input channels is

changed by the application in each channel of P second-order time-varying allpass filters. The transfer function of the p-th allpass filter is described by the following equation [176]:

$$A_p(z,n) = \frac{k_p^2(n) - 2k_p(n)z^{-1} + z^{-2}}{1 - 2k_p(n)z^{-1} + k_p^2(n)z^{-2}}, \tag{3.53}$$

so, it is identified by a pole with a multiplicity of 2 connected to the coefficient $k_p(n)$ of Equation (3.49). This characterization of the allpass filter allows us to maintain the spatial perception of the speech [174] because the restriction $|k_p(n)| < 1$ guarantees stability and causality of the filter and ensures that the inter-aural time delay difference between the two ears is lower than the well-known "just noticeable inter-aural delay" [177]. As described in [172], the alteration in sound direction is negligible as the maximum variation in the time of arrival is about $40\,\mu$s for all frequencies.

3.2.4 Subband room response identification

The architecture used for the room response identification is based on the subband adaptive filtering structure with critical sampling of [97]. Figure 3.25 shows the subband structure considering $P = 2$ loudspeakers and one of the microphones (i.e., $Q = 1$). The structure is the same as described in Section 2.3.3, where it is used for the adaptive crosstalk cancellation algorithm. Therefore, starting from a prototype filter $p(n)$ of order N_p, cosine modulated analysis and synthesis filterbanks $\mathbf{G}$ and $\mathbf{F}$, respectively, are designed considering M subbands. The double analysis filterbank $\mathbf{GG}$ is derived from $\mathbf{G}$ and consists of M filters $G_k(z)G_k(z)$ for $k = 0, \ldots, M-1$, and $M-1$ filters $G_k(z)G_{k+1}(z)$ for $k = 0, \ldots, M-2$ [98]. As a consequence, considering as input signal the signal of the pth loudspeaker $y_p(n)$ with $p = 1, ..., P$, the outputs of the filterbank after the downsampling operation are $2M-1$ signals derived as follows in $\mathcal{Z}$-domain:

$$Y_{p,k,k}(z) = Y_p(z^{\frac{1}{M}})G_k(z^{\frac{1}{M}})G_k(z^{\frac{1}{M}}) \tag{3.54}$$

for $k = 1, \ldots, M-1$, and

$$Y_{p,k,k+1}(z) = Y_p(z^{\frac{1}{M}})G_k(z^{\frac{1}{M}})G_{k+1}(z^{\frac{1}{M}}) \tag{3.55}$$

for $k = 1, \ldots, M-2$.

These signals constitute the inputs of a bank of adaptive filters. The coefficients of the adaptive sub-filters from the pth input to the qth microphone are collected in the vector $W_{q,p,k}$. The order of each adaptive sub-filter should be at least $N_\mathrm{w} = (N_\mathrm{s} + N_\mathrm{p} + 2)/M$ [97], where N_s is the order of the entire system that must be identified and N_p is the order of the prototype filter $p(n)$. The

sub-filters are characterized by a uniform frequency bandwidth of π/M and a center frequency of $\pi/(2M)$ and are obtained through the minimization of the sum of the instantaneous subband squared errors, represented by

$$J_q(n) = \sum_{k=0}^{M-1} E_{q,k}^2(n), \tag{3.56}$$

where $q = 1, ..., Q$ and Q is the number of microphones. The error signals are obtained by the following relation

$$E_{q,k}(n) = D_{q,k}(n - \Delta) - \sum_{p=1}^{P} Z_{q,p,k}(n), \tag{3.57}$$

where $D_{q,k}(n)$ is the desired signal of the kth subband for the qth microphone, $\Delta = \frac{N_{\mathrm{P}}+1}{M}$ is the delay introduced by the filterbank $\mathbf{G}$, and $Z_{q,p,k}(n)$ is defined

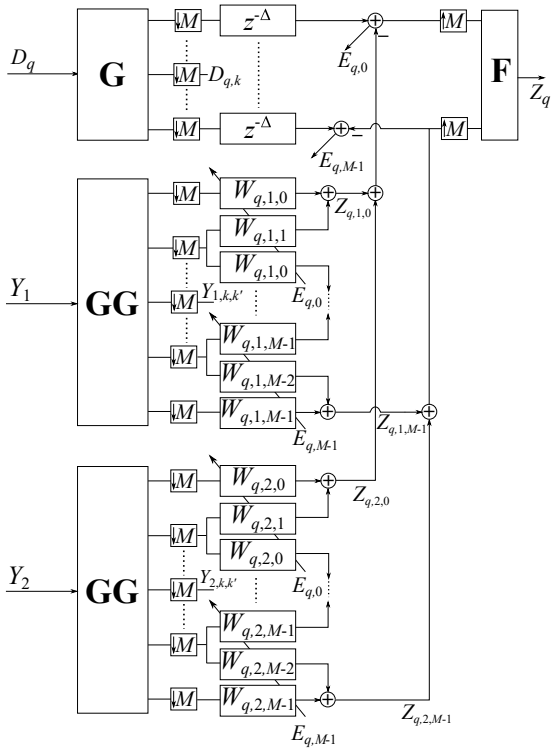


Figure 3.25: Subband RIRs identification procedure with $P = 2$ for the qth microphone.

as

$$\begin{aligned} Z_{q,p,k}(n) = {} & Y_{q,k,k}(n)W_{q,p,k}(n) + Y_{p,k-1,k}(n)W_{q,p,k-1}(n) \\ & + Y_{p,k,k+1}(n)W_{q,p,k+1}(n). \end{aligned} \tag{3.58}$$

The filters $W_{q,p,k}$ are adapted according to the following equation:

$$\begin{aligned} W_{q,p,k}(n+1) = {} & W_{q,p,k}(n) + \mu_{p,k}\big[Y_{p,k,k}(n)E_{q,k}(n) \\ & + Y_{p,k-1,k}(n)E_{q,k-1}(n) + Y_{p,k,k+1}(n)E_{q,k+1}(n)\big]. \end{aligned} \tag{3.59}$$

The step sizes $\mu_{p,k}$ are normalized by the sum of instantaneous powers of the signals involved in the adaptation of the coefficients, i.e.,

$$\mu_{p,k} = \frac{\mu}{\epsilon + P_{p,k,k} + P_{p,k-1,k} + P_{p,k,k+1}}, \tag{3.60}$$

where ϵ is a small positive constant to avoid division by zero and the power is estimated as

$$P_{p,k,k'}(n+1) = \eta P_{p,k,k'}(n) + (1-\eta)Y^2_{p,k,k'}(n), \tag{3.61}$$

with η a constant in the range $(0,1)$. Finally, the reconstructed frequency response between the pth loudspeaker and the qth microphone $\widehat{H}_{q,p}(z)$ is computed as the sum of the interpolated sub-filters $W_{q,p,k}(z^M)$, filtered by the synthesis filterbank $\mathbf{F}$, i.e.,

$$\widehat{H}_{q,p}(z) = \sum_{k=0}^{M-1} W_{q,p,k}(z^M)F_k(z), \tag{3.62}$$

with $F_k(z)$ the kth filter of the synthesis filterbank. The computational complexity of the subband identification algorithm is lower than a full band LMS approach as reported in [97]. In particular, the overall number of multiplications per input sample required for the filtering and adaptation of the sub-filters $W_{q,p,k}(z^M)$ is computed as

$$\frac{2(3M-2)(N_\mathrm{s}+1)}{M^2} + \frac{2(3M-2)(N_\mathrm{p}+1)}{M^2}, \tag{3.63}$$

where the first term corresponds to the filtering operations and the second term corresponds to the adaptation procedure, with N_s the length of the full band system. For high-order adaptive filters, the dominant term in the above expression is $6N_\mathrm{s}/M$, which is about $M/3$ times smaller than the number of multiplications required by the fullband LMS algorithm ($2N_\mathrm{s}$) as reported in [97].

3.2.5 Multipoint equalizer design

The proposed identification procedure has been tested with a multipoint equalization technique [156]. This approach allows for enlarging the listening sweet spot taking into consideration different microphone positions. A quasi-anechoic environment has been employed to obtain also a general equalization of the used loudspeakers [178]. The prototype function is derived from the combination of quasi-anechoic IRs, derived from a gated version (up to the first reflection) of the responses, with the IRs recorded in the real environment. Figure 3.26 shows the equalization approach used in the presented work that follows these steps:

1. For each loudspeaker, Q impulse responses of order N_s are measured in the zone that must be equalized.

2. A pre-processing is applied exploiting the quasi-anechoic IR spectrum for frequency greater than a certain transition frequency and the original (ungated) IR spectrum below the same transition frequency. This operation is performed by applying the following equation to each RIR, in the frequency domain:

$$H_{q,p}(e^{j\omega}) = \widehat{H}_{q,p}(e^{j\omega}) \cdot w_\mathrm{lf}(e^{j\omega}) + \widetilde{H}_{q,p}(e^{j\omega}) \cdot w_\mathrm{hf}(e^{j\omega}), \tag{3.64}$$

   where $\widehat{H}_{q,p}(e^{j\omega})$ is the frequency response of the original estimated RIR, $\widetilde{H}_{q,p}(e^{j\omega})$ is the frequency response of the gated RIR, $w_\mathrm{lf}(e^{j\omega})$ and $w_\mathrm{hf}(e^{j\omega})$ are the half Hann windows used for selecting the lowpass and highpass frequency bands, respectively. The linear combination in Equation (3.64) is used to equalize only the direct sound in the mid- and high-frequency range, which determines localization and most of the timbre perception, while full equalization is applied in the modal frequency range [165].

3. The magnitude frequency responses of the measured IRs are estimated through Q FFTs of N_FFT samples.

4. A smoothing operation is applied to the magnitude frequency responses, simulating the poorer frequency resolution at higher frequencies of the

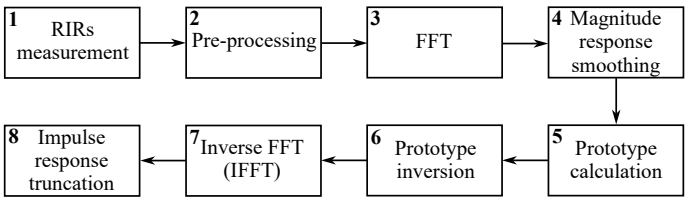


Figure 3.26: Multi-point room response equalization procedure.

human auditory system. The approach of [179] has been used to obtain a nonuniform frequency magnitude spectrum smoothing of the frequency response $H_{q,p}(e^{j\omega})$, resulting in the smoothed frequency response $H_{\mathrm{sm}_{q,p}}(e^{j\omega})$, with $q = 1, \ldots, Q$ and $p = 1, \ldots, P$. In this way, a broader equalized zone is achieved by exploiting a less precise equalization at higher frequencies resulting from the nonuniform resolution, which decreases with increasing the frequency.

5. A prototype response of the involved acoustic environment is derived taking into account all smoothed IRs. The prototype frequency response $H_{\mathrm{pr}_p}(e^{j\omega})$ of the pth loudspeaker is obtained using an arithmetic mean of the zero-phase smoothed frequency responses as

$$H_{\mathrm{pr}_p}(e^{j\omega}) = \frac{1}{Q} \sum_{q=1}^{Q} H_{\mathrm{sm}_{q,p}}(e^{j\omega}), \tag{3.65}$$

with $H_{\mathrm{sm}_{q,p}}(e^{j\omega})$ the smoothed transfer function from the pth loudspeaker to the qth microphone.

6. A frequency domain inverse filter is obtained through the use of a frequency deconvolution with regularization technique [180] applied to the prototype as

$$H_{\mathrm{inv}_p}\left(e^{j\omega}\right) = \frac{H^*_{\mathrm{pr}_p}\left(e^{j\omega}\right)}{\left|H_{\mathrm{pr}_p}\left(e^{j\omega}\right)\right|^2 + \chi}, \tag{3.66}$$

being $H^*_{\mathrm{pr}_p}(e^{j\omega})$ the complex conjugate of $H_{\mathrm{pr}_p}(e^{j\omega})$, and χ the regularization factor, which avoids excessive gains often appearing at high frequencies. In the experiments, a small regularization factor with value $\chi = 10^{-5}$ is applied. The equalization filter frequency response of length K is computed in the unwarped domain by interpolating with a cubic spline [161] the values of $H_{\mathrm{inv}_p}(e^{j\omega})$.

7. The inverse FFT of the interpolated frequency response is performed, obtaining the time domain equalization curve.

8. Finally, the resulting sequence is truncated to determine the final equalization filter for the pth loudspeaker.

This equalization technique is an effective approach that guarantees a reduced computational complexity, making it suitable for real-time applications. In fact, the computational load mainly depends on the inverse FFT operation, which is characterized by a computational complexity of the order of $W \log W$, where W is the number of frequency bins of the final equalization filter magnitude response [180].

3.2.6 Experimental results of multichannel equalization

For the experiments, two real rooms are employed. Figure 3.27 shows the size of the room with loudspeakers and microphone positions. The microphones are placed distance was set to 30 cm and they were placed at 1.2 m height. Professional equipment was used following the procedure described in [167]. More in detail, measurements have been performed using a professional ASIO sound card and microphones with an omnidirectional response. A personal computer running NU-Tech platform has been used to manage all I/Os [84]. The impulse responses have been derived using a logarithmic sweep signal excitation [181] at 48 kHz sampling frequency. These responses are then used as terms of comparison in the identification procedure.

For the adaptation procedure, a filter length of 4096 samples (i.e., an order of $N_s = 4095$) has been considered, working on frames of 8192 samples, with $M = 256$ subbands and a sampling frequency of $F_s = 48$ kHz.

The equalizer is designed in the warped domain considering $W = 8192$ frequency points and the final length of the equalizer is 1024 samples. The equalized frequency range goes from 10 Hz to 20 kHz with the same sampling frequency of 48 kHz. The stereo input signals used for the presented results are the following soundtracks:

I) "International Geophysical Year" from Donald Fagen (for experiments 1 and 3),

II) "I Sat by the Ocean" from Queens Of The Stone Age (for experiments 2 and 4).

Two songs have been chosen as input signals in order to evaluate the algorithm performance with variable inputs in a real scenario. This is a very important

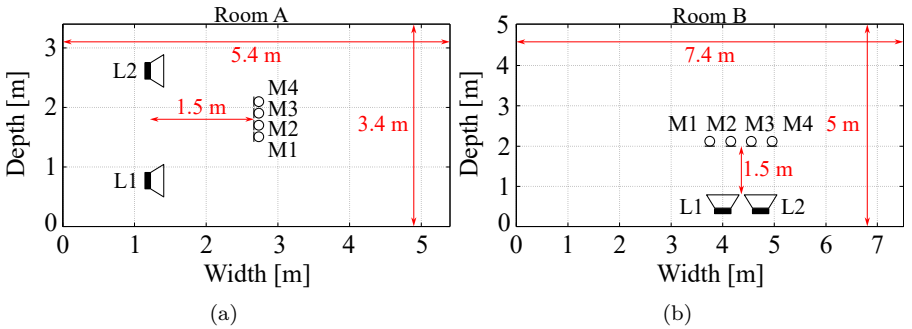


Figure 3.27: Loudspeakers and microphones positions (a) in room A, for experiments 1 and 2, and (b) in room B, for experiments 3 and 4.

aspect since the identification algorithm must work during sound reproduction without altering the sound perception.

Four experiments have been carried out employing different soundtracks in the two rooms, i.e.,

- **Experiment 1**: soundtrack (I) in room A,
- **Experiment 2**: soundtrack (II) in room A,
- **Experiment 3**: soundtrack (I) in room B,
- **Experiment 4**: soundtrack (II) in room B,

where room A is referred to Figure 3.27(a), and room B is shown in Figure 3.27(b). Figures 3.28, 3.29, 3.30, and 3.31 show the four magnitude responses relative to the four paths between the one loudspeaker, and the four microphones, considering the four experiments, respectively. The Figures report the magnitude responses related to the right loudspeaker for experiments 1 and 3 (cf. Figures 3.28 and 3.30, respectively) and to the left loudspeaker for experiments 2 and 4 (cf. Figures 3.29 and 3.31, respectively). The impulse responses identified by the proposed algorithm imposing $M = 1$ and $M = 256$ subbands are compared with the ones measured with the logarithmic sweep signal procedure. The good performance obtained in terms of identification is also confirmed by the results reported in Figure 3.32, where the difference between the magnitude frequency response of the identified IR and the magnitude response of the measured IR is shown.

It is evident that the subband structure is capable of identifying the impulse responses and, increasing the number of subbands, it is possible to obtain a very accurate estimation of the responses. This is confirmed by all experiments thus considering different inputs. Furthermore, the use of the subband structure allows us to obtain a higher convergence rate with the increase of the number of subbands M, as verified by the mean squared error (MSE) evolution, shown in Figure 3.33.

This is due to the fact that increasing the number of subbands, the signal is divided into small frequencies parts where it is more stationary and where it is possible to use a more suitable stepsize exploiting Equations (3.60) and (3.61). However, increasing the number of subbands could lead to an increase in the memory usage of the hardware system, increasing the parallel computational load. A compromise between identification performance and hardware capability should be evaluated.

Figure 3.34 shows the results for the four experiments and for each channel in terms of smoothed real room magnitude responses $H_{\mathrm{sm}_q}(e^{j\omega})$ (with $q = 1, 2, 3, 4$) identified with the subband procedure, the prototype response $H_{\mathrm{pr}}(e^{j\omega})$ calculated with $M = 1$ and the one determined with $M = 256$, and

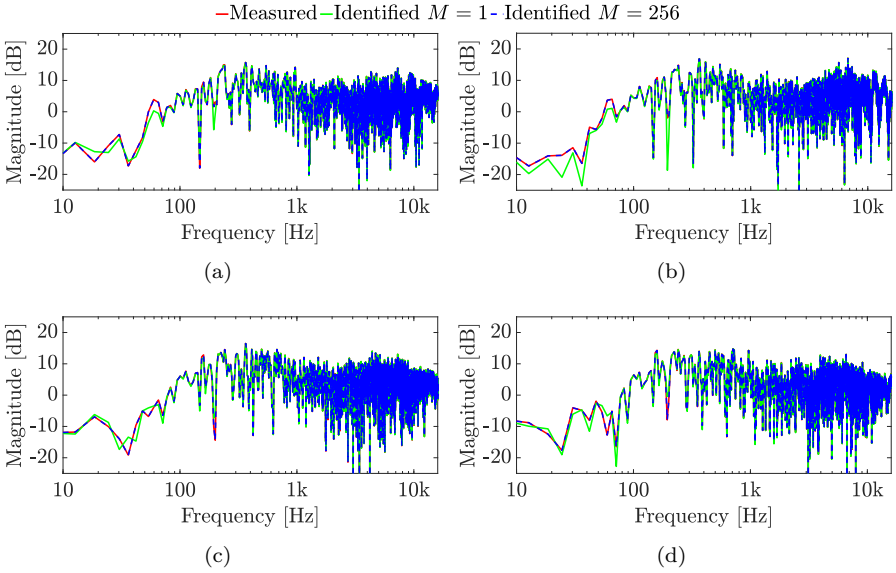


Figure 3.28: Experiment 1: Magnitude frequency responses comparison between the measured IRs, the IRs identified with $M = 1$, and the IRs identified with $M = 256$ for the right loudspeaker channel.

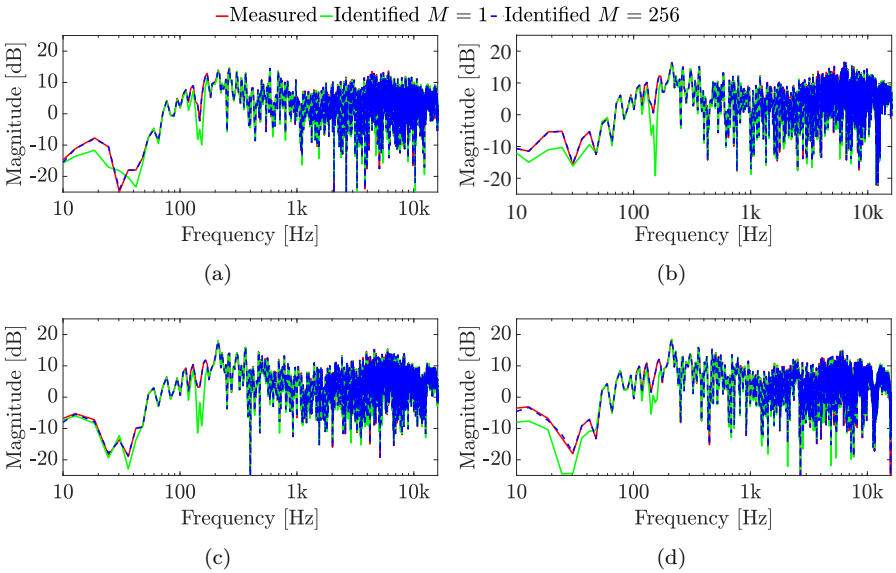


Figure 3.29: Experiment 2: Magnitude frequency responses comparison between the measured IRs, the IRs identified with $M = 1$, and the IRs identified with $M = 256$ for the left loudspeaker channel.

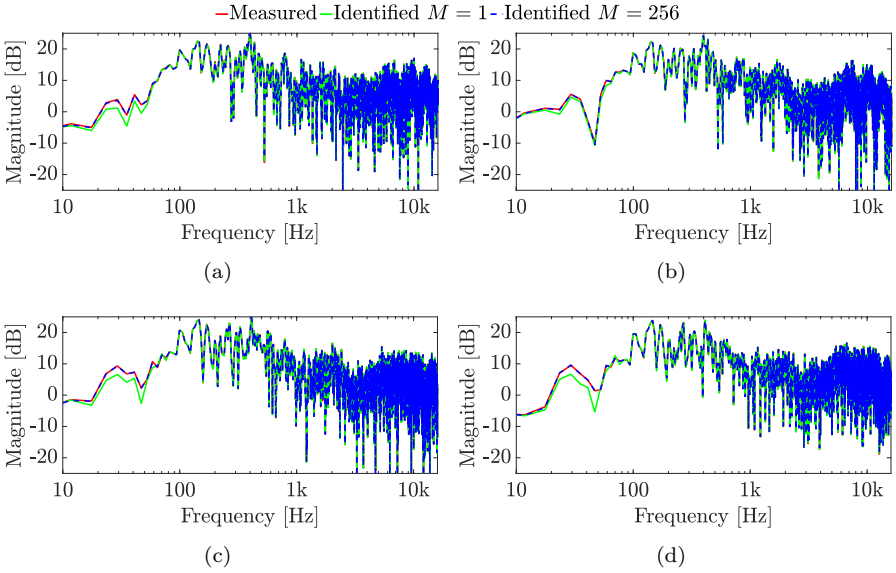


Figure 3.30: Experiment 3: Magnitude frequency responses comparison between the measured IRs, the IRs identified with $M = 1$, and the IRs identified with $M = 256$ for the right loudspeaker channel.

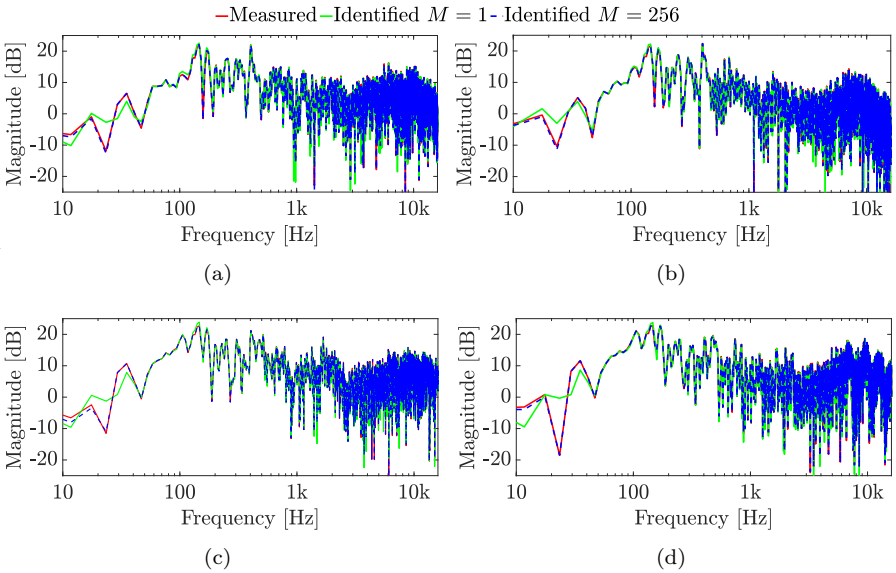


Figure 3.31: Experiment 4: Magnitude frequency responses comparison between the measured IRs, the IRs identified with $M = 1$, and the IRs identified with $M = 256$ for the left loudspeaker channel.

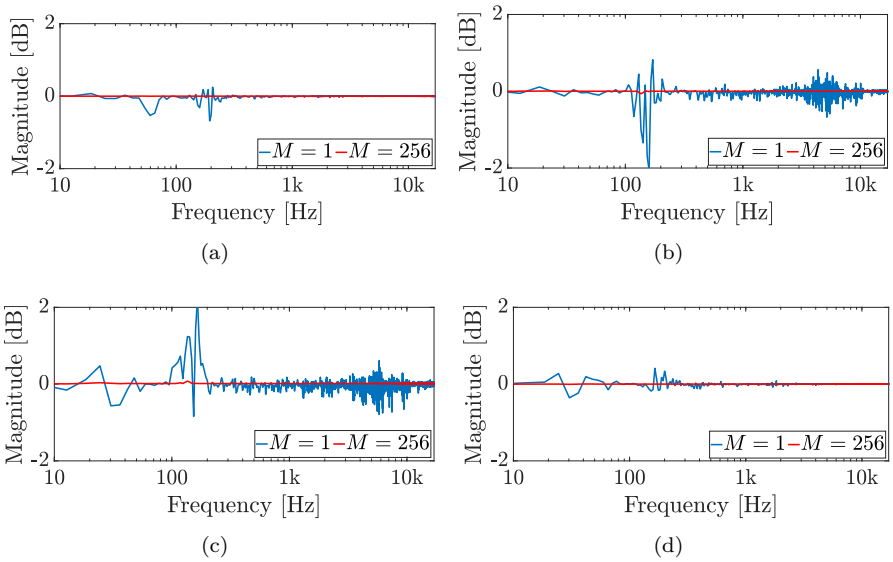


Figure 3.32: Difference between the real room magnitude responses and the identified room magnitude responses for $M = 1$ and for $M = 256$ considering one microphone with reference to the right channel, in the case of (a) experiment 1, (b) experiment 2, (c) experiment 3, and (d) experiment 4.

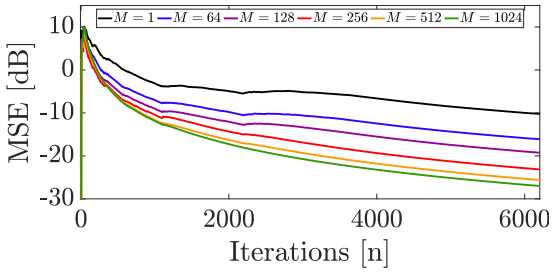


Figure 3.33: MSE for the subband identification with M =1, 64, 128, 256, 512, 1024, considering white noise as the input signal.

the equalization curves $H_{\text{inv}}(e^{j\omega})$ obtained from the single band identification (i.e., $M = 1$) and the subband identification (i.e., $M = 256$). The subband identification implies a different resolution of the equalization curve that can increase the quality of the final equalization procedure with respect to the single-band identification. This is also demonstrated by the spectral deviation (SD) measure reported in Tables 3.8 and 3.9. The SD gives a measure of the deviation of the magnitude response from a flat one [182], considering each IR before (initial SD) and after the equalization (final SD), as also reported

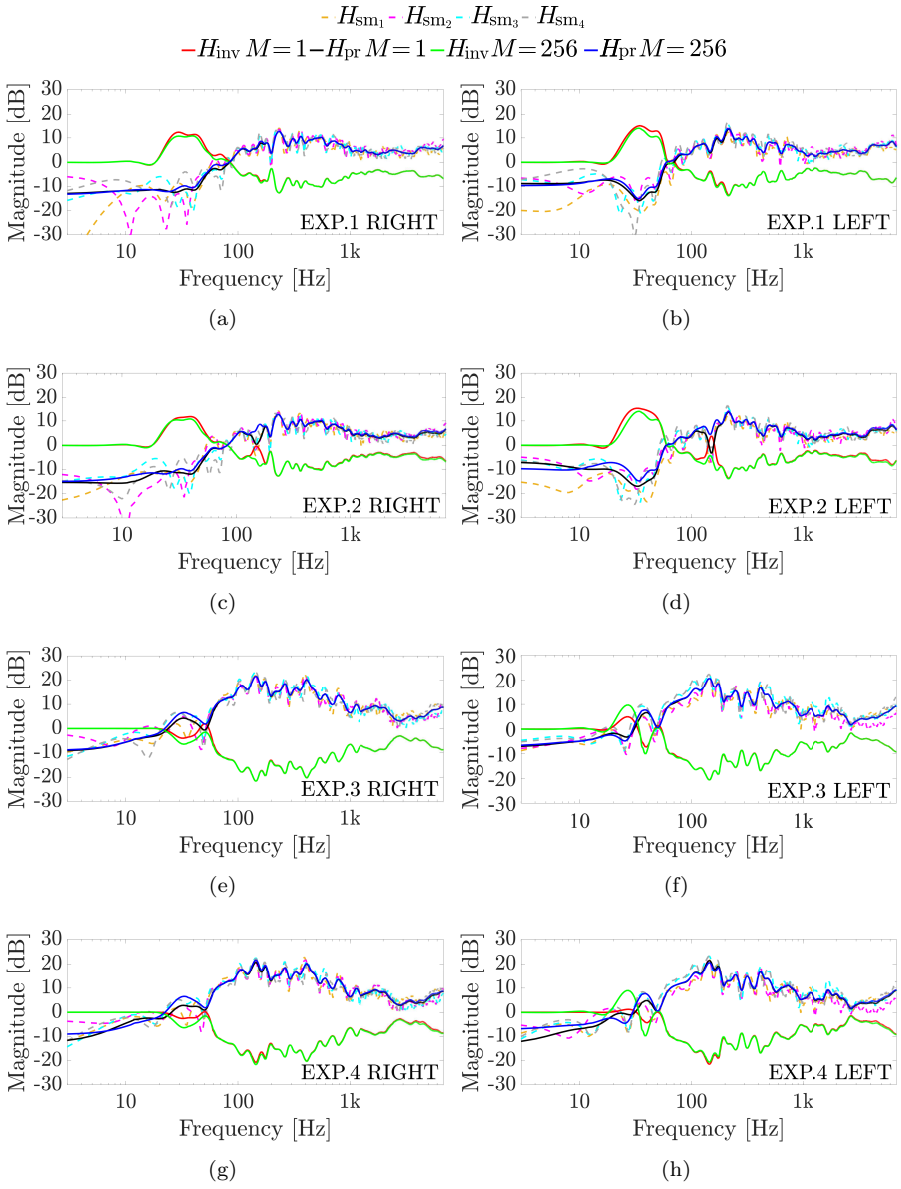


Figure 3.34: Comparison between the four room magnitude responses $H_{\mathrm{sm}_{q,p}}$, the equalization curve H_{inv} for $M = 1$, the prototype response H_{pr} for $M = 1$, the equalization curve H_{inv} for $M = 256$, the prototype response H_{pr} for $M = 256$ for the right channel (first column) and left channel (second column), in the case of (a)-(b) experiment 1, (c)-(d) experiment 2, (e)-(f) experiment 3, and (g)-(h) experiment 4. A smoothing factor of 1/12 has been applied.

Table 3.8: SD evaluation considering the single band identification ($M = 1$) and the subband identification ($M = 256$) for all four experiments with a frequency range of 10Hz-20kHz.

Experiments	Initial SD	Final SD $M = 1$	Final SD $M = 256$
EX1 - right channel	10.96	2.60	**2.59**
EX1 - left channel	9.50	2.66	**2.65**
EX2 - right channel	10.96	2.60	**2.59**
EX2 - left channel	9.50	2.66	**2.65**
EX3 - right channel	10.41	2.65	**2.64**
EX3 - left channel	12.17	2.59	**2.58**
EX4 - right channel	10.41	2.65	**2.64**
EX4 - left channel	12.17	2.59	**2.58**

in [103]. The SD of the frequency response $H(e^{j\omega})$ is calculated as [182]

$$\mathrm{SD} = \sqrt{\frac{1}{k_2 - k_1 + 1} \sum_{k=k_1}^{k_2} \Big(10 \log_{10} \big|H(e^{j\omega_k})\big| - D\Big)^2}, \tag{3.67}$$

where

$$D = \frac{1}{k_2 - k_1 + 1} \sum_{k=k_1}^{k_2} 10 \log_{10} \big|H(e^{j\omega_k})\big|, \tag{3.68}$$

where k_1 and k_2 are the lowest and highest frequency indexes, respectively, of the considered band. The initial SD is computed imposing $H(e^{j\omega}) = H_{\mathrm{pr}}(e^{j\omega})$, while the final SD is calculated with $H(e^{j\omega}) = H_{\mathrm{pr}}(e^{j\omega}) \cdot H_{\mathrm{inv}}(e^{j\omega})$. In Tables 3.8 and 3.9, the initial and the final SD values are reported for each experiment and for each channel, comparing the single band with the subband structure. In particular, Table 3.8 shows the results obtained in the full frequency range between 10 Hz and 20 kHz, while Table 3.9 takes into account only the low-frequency range from 10 Hz and 200 Hz. It is evident that the subband identification procedure allows for a reduction of the SD in comparison with the case with $M = 1$. This reduction is even more emphasized in the low frequencies, as shown in Table 3.9, while the improvement in the broadband case is marginal, as shown in Table 3.8.

Table 3.9: SD evaluation considering the single band identification ($M = 1$) and the subband identification ($M = 256$) for all four experiments with a frequency range of 10Hz-200Hz.

Experiments	Initial SD	Final SD $M = 1$	Final SD $M = 256$
EX1 - right channel	4.09	3.98	**3.79**
EX1 - left channel	3.91	2.81	**2.64**
EX2 - right channel	4.09	3.90	**3.79**
EX2 - left channel	3.91	2.65	**2.63**
EX3 - right channel	3.80	3.43	**3.20**
EX3 - left channel	3.78	3.25	**2.95**
EX4 - right channel	3.80	3.36	**3.20**
EX4 - left channel	3.78	3.33	**3.00**

3.3 Crossover network for multichannel systems

In multichannel systems, crossover networks are used to split the input signal into different frequency ranges that are reproduced by different drivers of the same loudspeaker or different loudspeakers of a multichannel system. Figure 3.35 shows a two-way loudspeaker system composed of a tweeter (for higher frequencies) and a woofer (for lower frequencies). Conventional high-quality crossover networks are designed in order to verify the following four requirements [183]:

I. flatness in the magnitude of the combined outputs,

II. adequate steep cutoff rates of the individual filters in their stop bands,

III. symmetric and uniform polar response for the combined output,

IV. acceptable phase response for the combined output, the most desirable characteristic being phase linearity.

The effectiveness of a crossover network is based on the fulfillment of these requirements, along with the performance in terms of latency and computational cost [184].

3.3.1 Background on crossover networks

Crossover networks are employed to divide the signal into two or more frequency ranges that are reproduced by different loudspeakers [15, 185–187]: sub-woofer (for frequencies lower than 100 Hz), woofer (between 100 Hz and 300 Hz), mid-range (300 Hz - 3 kHz) and tweeter (3 kHz - 20 kHz). Similar to graphic equalizers, crossover networks can be considered as filterbanks [188] and are classified into minimum-phase and linear-phase approaches. The main difference is that the main function of GEQs is to modify the spectral balance,

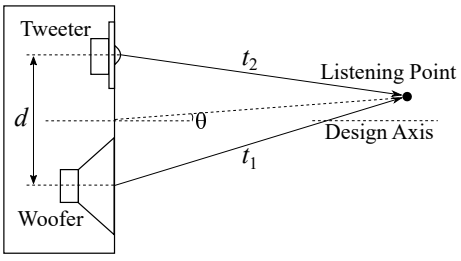


Figure 3.35: Two-way loudspeaker system diagram for calculating the polar response, when the driver distance is d and the flight times from the two drivers to the listening point are t_1 and t_2.

whereas a crossover network must minimize the band leakage, precisely segregating each band into its own signal, which is consequently processed and played separately. Conventional high-quality crossover networks are designed in order to have a flat magnitude of the combined outputs, steep cutoff rates, symmetric polar response, and acceptable phase response (i.e., linear at least in the crossover region) [183]. In the literature, many approaches for the design of minimum-phase crossover networks can be found, e.g., networks derived from analog models [189, 190], allpass-based techniques [183, 191, 192], polynomial-based approaches [193–195], and systems based on hybrid IIR/FIR digital filters [184, 196, 197]. Considering analog models, IIR Linkwitz-Riley filters [190] are the most used since they guarantee all the requirements ensuring a reduced computational cost. They are derived from the cascade of two identical Butterworth filters [189]. This configuration allows obtaining a flat magnitude response of the combined outputs that is not achieved with a Butterworth crossover network. Although IIR filters are widely used, they do not ensure a completely linear phase, so linear-phase crossover networks can be developed using FIR filters. Linear-phase solutions can be categorized into time-domain approaches [198], multirate approaches [199], and frequency-domain approaches [200]. The main drawback of FIR filters is the high computational cost.

3.3.2 Linear-phase crossover network

In this section, the multi-way crossover network of [26] is presented. It is based on IFIR filters, described in Section 3.1.2, where they are used for the deployment of graphic equalizers. Similarly to a GEQ, the crossover network splits the signal into P frequency bands, but without applying gains to the band signals. Moreover, crossovers provide P outputs that are reproduced by different drivers, while a GEQ has a single output obtained by the band signals sum.

Filter design

The scheme of the proposed P-way crossover network is shown in Figure 3.36. The cutoff frequencies of the P bands of the crossover are f_{c_1}, f_{c_2}, ..., $f_{c_{P-1}}$, where f_{c_1} is the cutoff frequency of the first lowpass filter and $f_{c_{P-1}}$ is the cutoff frequency of the last highpass filter. Starting from $P-1$ basis lowpass filters $H_1(z)$, $H_2(z)$, ..., $H_{P-1}(z)$ with cutoff frequencies of f_{c_1}, f_{c_2}, ..., $f_{c_{P-1}}$, respectively, the combination of these lowpass filters and their highpass complementary filters allows to obtain the P outputs of the crossover network. The basis lowpass filters $H_i(z)$ are IFIR filters obtained, as shown in Figure 3.37,

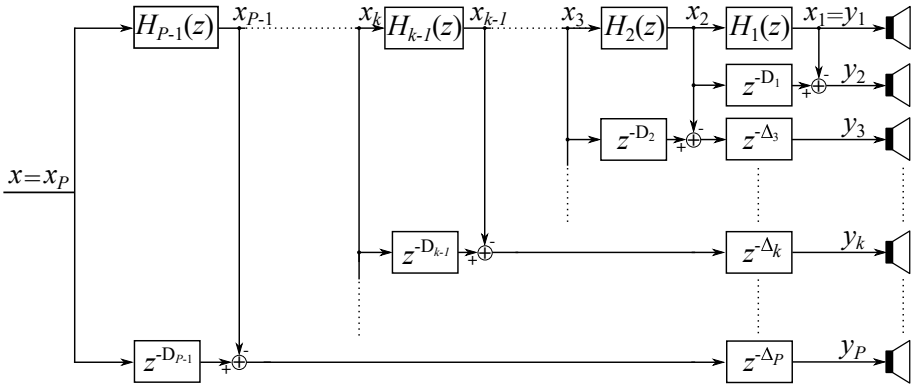


Figure 3.36: Scheme of the proposed P-way crossover network. $H_1(z)$, $H_2(z)$, ..., $H_{P-1}(z)$ are the IFIR basis lowpass filters with cutoff frequencies of f_{c_1}, f_{c_2}, ..., $f_{c_{P-1}}$, respectively.

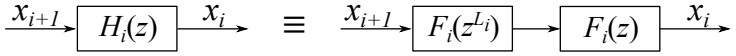


Figure 3.37: Design of the ith basis lowpass filter of the proposed crossover network using IFIR method, with $i = 1, ..., P-1$.

as

$$H_i(z) = F_i(z^{L_i})F_i(z), \tag{3.69}$$

where $F_i(z)$ is the ith model filter, with $i = 1, 2, ..., P-1$. The model filter is designed with the windowing method using the Kaiser window with a shape parameter $\beta = 10$ [201]. The design of the filter $F_i(z)$ is achieved considering a cutoff frequency of $f_{c_i}^{\mathrm{F}} = L_i f_{c_i}$, where f_{c_i} is the cutoff frequency of the desired lowpass filter and L_i is the interpolation factor. In the proposed system, the order N_i of the filter $F_i(z)$ must be an even value, because the filter delay, that is $N_i/2$ must be an integer value to allow the time synchronization of the crossover outputs. For this reason, N_i is chosen as the even number closest to the optimum order $N_{i_{\mathrm{opt}}}$, as follows:

$$N_i = 2N_{i_{\mathrm{opt}}} - 2\left\lfloor \frac{N_{i_{\mathrm{opt}}}}{2} \right\rfloor, \tag{3.70}$$

and $N_{i_{\mathrm{opt}}}$ is calculated by the following Equation given by [201], i.e.,

$$N_{i_{\mathrm{opt}}} = \left\lfloor \frac{A_{\mathrm{SB}} - 8}{2.285\Delta\omega_i} \right\rfloor, \tag{3.71}$$

where the brackets $\lfloor \cdot \rceil$ denote the rounding to the closest integer value, A_{SB} is the stopband attenuation and $\Delta\omega_i$ is the width of the transition band, which is imposed to be twice the cutoff frequency of the model filter $f_{c_i}^{\mathrm{F}}$, e.g.,

$$\Delta\omega = \frac{4\pi f_{c_i}^{\mathrm{F}}}{F_{\mathrm{s}}} = \frac{4\pi L_i f_{c_i}}{F_{\mathrm{s}}}, \tag{3.72}$$

where F_{s} is the sampling frequency. To further reduce the computational complexity and the memory allocation, in this work the interpolator filter $G(z)$ is imposed equal to the model filter $F(z)$, e.g.,

$$G(z) = F(z). \tag{3.73}$$

Equation (3.73) can be applied when the filter $F(z)$ is designed in order to eliminate the images of the interpolated version $F(z^L)$. Empirical tests proved that this characteristic is achieved when the equation $2Lf_c = F_{\mathrm{s}}/L - 2f_c$ is satisfied, which means computing the interpolation factor as

$$L_i = \left\lfloor \frac{-f_c + \sqrt{f_c^2 + 2f_c F_{\mathrm{s}}}}{2f_c} \right\rceil. \tag{3.74}$$

In the case of $L_i = 1$ the Equation (3.69) is not applied and the filter $H_i(z)$ is designed as a single FIR filter, so it is equal to the filter $F_i(z)$, e.g., $H_i(z) = F_i(z)$. The respective highpass filter $H_i^{\mathrm{H}}(z)$ with cutoff frequency f_{c_i} is obtained as the complementary filter of $H_i(z)$ as follows:

$$H_i^{\mathrm{H}}(z) = z^{-D_i} - H_i(z), \tag{3.75}$$

where D_i is the delay calculated as

$$D_i = \frac{N_i L_i + N_i}{2}. \tag{3.76}$$

The use of complementary filters allows reducing the computational complexity and guarantees a flat magnitude response of the combined outputs, verifying the requirement I on magnitude flatness. Taking into account Figure 3.36, the kth output of the crossover network $Y_k(z)$ is computed as

$$Y_k(z) = [X_k(z)z^{-D_{k-1}} - X_{k-1}(z)]z^{-\Delta_k}, \tag{3.77}$$

where X_{k-1} is obtained as

$$X_{k-1} = X(z) \prod_{i=k-1}^{P-1} H_i(z), \tag{3.78}$$

with $k = 2, ..., P$ and considering $X_P(z) = X(z)$. The synchronization delay Δ_k is applied starting from the third band and is defined as

$$\Delta_k = \sum_{i=1}^{k-2} D_i, \tag{3.79}$$

with $k = 3, ..., P$ and D_i is the delay introduced by the ith basis filter and it is calculated following Equation (3.76). The output of the first band $Y_1(z)$ is simply equal to $X_1(z)$ that is obtained by Equation (3.78).

Finally, the total delay of the crossover network τ is computed as

$$\tau = \sum_{i=1}^{P-1} D_i. \tag{3.80}$$

The computational complexity is given by the number of operations per output sample. The number of multiplications of the proposed crossover network is computed as

$$\text{n}^\circ \text{ Mul} = \sum_{i=1}^{P-1} b_i N_i + b_i, \tag{3.81}$$

and the number of additions is calculated as follows:

$$\text{n}^\circ \text{ Add} = \sum_{i=1}^{P-1} b_i N_i + 1, \tag{3.82}$$

where N_i is the order of the ith model filter $F_i(z)$, N is the number of ways of the crossover and b_i is a parameter that depends on the value of the ith interpolation factor L_i as follows:

$$b_i = \begin{cases} 2, & \text{if } L_i > 1, \\ 1, & \text{if } L_i = 1. \end{cases} \tag{3.83}$$

Experimental results of the crossover network

The proposed crossover has been tested considering a 4-way configuration (i.e., $P = 4$) with the following cutoff frequencies: $f_{c_1} = 120\,\text{Hz}$, $f_{c_2} = 1\,\text{kHz}$, and $f_{c_3} = 8\,\text{kHz}$ and a sampling frequency of $F_\text{s} = 48\,\text{kHz}$. In this case, three basis lowpass filters $H_1(z)$, $H_2(z)$, and $H_3(z)$ have been designed using the IFIR technique, as explained above, obtaining the following interpolation factors: $L_1 = 14$, $L_2 = 4$ and $L_3 = 1$, and the following filter orders: $N_1 = 92$, $N_2 = 38$, and $N_3 = 20$. The evaluation has been carried out by examining the four requirements listed at the beginning of this section, comparing the proposed crossover with the Linkwitz-Riley approach [190], with the time filtering

of equivalent FIR filters and with the FFT implementation. The Linkwitz-Riley crossover network [190] is obtained considering 4th-order filters. The FIR crossover is obtained by implementing the same scheme of Figure 3.36, but the basis filters $H_i(z)$ are designed as normal FIR filters with the Kaiser window with a shape parameter of $\beta = 10$ and the following orders: $N_1 = 1282$, $N_2 = 154$, $N_3 = 20$. The FFT method is obtained by calculating the frequency response of each band of the FIR crossover and applying the overlap and save algorithm with an FFT length of 1024 samples. Table 3.10 shows the results obtained by the experimental tests. In the table, the checkmark means the verification of the requirement, while the distortion index (DI) quantifies the level of distortion and it is calculated as

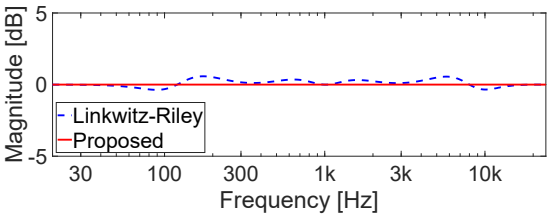
$$\mathrm{DI} = \frac{\max |T(e^{j\omega})|_{\mathrm{dB}} + \min |T(e^{j\omega})|_{\mathrm{dB}}}{2}, \tag{3.84}$$

where $T(z)$ is the sum of all the bands' frequency responses of the crossover. The DI should take values close to 0 dB to have a flat response. The Linkwitz-Riley crossover guarantees only requirements II and III.

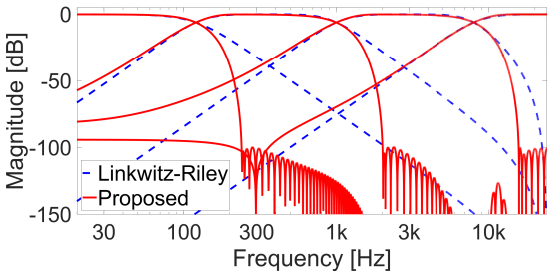
Regarding the magnitude flatness, Figure 3.38(a) shows the magnitude frequency response of the combined outputs comparing Linkwitz-Riley with the proposed system, confirming the results obtained for the distortion index. In fact, Linkwitz-Riley presents a distortion of 0.5 dB, while the proposed crossover shows a completely flat response. Figure 3.38(b) shows the comparison in terms of the magnitude frequency response of the four bands. In the proposed approach, the stopbands at the low frequencies have a smaller attenuation than the Linkwitz-Riley crossover, while the high frequencies are

Table 3.10: Comparison between crossovers, evaluating the requirements, the distortion index, the latency, and the computational cost

Crossover	Magnitude flatness I	Cutoff rate II	Polar response III	Phase response IV	DI (desired 0 dB)	Latency	Number of Add	Number of Mul	Total Operations
Linkwitz-Riley		✓	✓		0.5 dB	**185**	**32**	**36**	**68**
FIR	✓	✓	✓	✓	**0 dB**	724	1 459	1 459	2 918
FFT	✓	✓	✓	✓	**0 dB**	1 748	360	244	604
Proposed	✓	✓	✓	✓	**0 dB**	795	283	285	568



(a)



(b)

Figure 3.38: Comparison between 4th order Linkwitz-Riley crossover with the proposed IFIR crossover considering the following 4 bands: <120Hz, 120Hz-1kHz, 1kHz-8kHz, >8kHz. Fig. (a) is the total magnitude frequency response of the combined outputs of the crossover, while Fig. (b) shows the magnitude frequency response of each band.

more attenuated. However, a good suppression of the low frequencies that reach the last driver (generally a tweeter) and could damage the loudspeaker is obtained. For this reason, requirement II on the cutoff rate is verified by both techniques. Figure 3.39 shows the polar diagrams corresponding to the considered cutoff frequencies of the 4-way crossover, comparing the proposed IFIR crossover with the Linkwitz-Riley method. The figure is obtained taking into account the total response $H_{\mathrm{T}}(\omega, \theta)$ of the system at frequency ω and angle θ (cf. Figure 3.35), computed as [202]

$$H_{\mathrm{T}}(\omega, \theta) = \sum_{k=1}^{4} C_k(e^{j\omega}) e^{-j\omega(t_k - t_1)}, \tag{3.85}$$

where $C_k(e^{j\omega})$ is the frequency response of the kth band, and t_k is the flight times from the kth driver to the listener position, with $k = 1, ..., 4$. The two methods produce almost identical polar responses, which verify requirement III. Figure 3.40 shows the group delay of the total 4-way crossover, comparing the Linkwitz-Riley method with the proposed one. As expected, the proposed

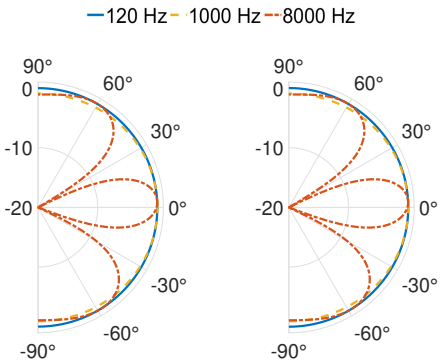


Figure 3.39: Polar plot of the considered 4-way crossover network using (a) the Linkwitz-Riley method and (b) the proposed IFIR method for the design of the filters, considering a distance between loudspeakers of 5 cm and distance from the origin of 1 m.

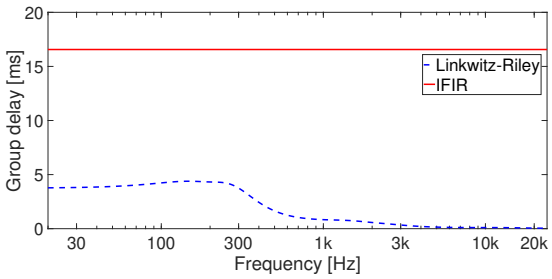


Figure 3.40: Group delay of the 4-way crossover network, comparing 4th order Linkwitz-Riley crossover with the proposed IFIR crossover.

crossover presents a constant group delay, i.e., a linear phase and this means a symmetric time response, satisfying requirement IV. Regarding the latency and computational cost, the Linkwitz-Riley method presents the lowest computational cost and the lowest latency, as expected, with only a total of 68 operations per output sample and a latency of 185 samples (i.e., 3.9 ms). All the other linear-phase methods (i.e., FIR and FFT) are based on the proposed system changing the implementation, so they verify all four requirements but they differ in computational cost and latency. The FIR method is the most expensive in terms of the number of operations reaching a total of 2 918 operations per output sample, while the FFT implementation shows the highest latency of 1 748 samples (i.e., 36.4 ms). The proposed method has a latency of 795 samples (i.e., 16.6 ms) similar to the FIR method and requires a total of 568 operations per output sample (of which 285 multiplications and 283 additions), which is smaller than both the FIR method and the FFT implementation.

3.4 Conclusions of audio equalization

Audio equalization techniques for audio rendering enhancement have been analyzed in this chapter. The importance of ensuring a linear phase in graphic equalizers has been underlined and, based on this assumption, three linear-phase graphic equalizers have been presented. They are all based on IFIR filters to reduce the computational complexity. The first GEQ uses a uniform filterbank in which every band has the same width. Experimental results have proven the effectiveness of the IFIR approach in GEQs development, however, the uniform band division is not suitable for graphic equalizers, and the computational cost of the uniform GEQ is still too high for real-time audio applications. Therefore, an octave band GEQ has been developed by exploiting IFIR and complementary filters. The octave GEQ implements a tree structure derived from a half-band prototype filter, so the performance of the complete equalizer depends on the design of the prototype filter. The windowing method has been applied for the prototype design and the Kaiser window shows the best results. The final GEQ has exhibited excellent performance, especially in terms of computational complexity. However, the latency is quite high (just below 100 ms). To reduce the latency, a hybrid FIR/IIR GEQ has been developed starting from the octave graphic equalizer. With the hybrid GEQ, the latency is reduced by half by designing the first band with a lowpass shelving IIR filter and relaxing the constraint on phase linearity. In fact, the linear phase is guaranteed above 100 Hz, while the computational cost remains considerably low. Regarding room response equalization, a subband implementation of a multichannel and multiple-position adaptive equalizer has been presented in this chapter. A subband structure is used for the identification of the room impulse responses in two real rooms considering two loudspeaker channels and four microphones. Experimental results have shown that the subband structure can identify more accurately the RIRs with a higher number of subbands, increasing the convergence rate. Then, the identified RIRs are used to obtain a prototype curve that has to be equalized in the frequency domain. The equalizer design is performed in the warped frequency domain to guarantee computational saving. Four experiments have been carried out employing different soundtracks as input signals and different environments, proving that the identification with the subband structure can produce better performances also in terms of equalization in comparison with the single-band approach. When multichannel systems are involved, loudspeakers may reproduce different frequency bands and crossover networks can be applied to divide the input signal spectrum. In this context, the implementation of a linear-phase crossover network based on IFIR filters has been presented. The proposed crossover has been compared with the Linkwitz-Riley approach and with other linear-phase

implementations and has exhibited lower computational complexity than other linear-phase techniques. Experimental results have shown that the proposed crossover network satisfies all the requirements in terms of flatness of the magnitude response of the combined outputs, steep cutoff rates, symmetric polar response, and linear phase response.

Chapter 4

Active Noise Control for Audio Enhancement

Active noise control is a technology that can improve the listening experience. It works by producing sound waves that are the exact opposite of the noise that is interfering with the sound of the desired audio. This process is known as destructive interference and it cancels out the unwanted noise, allowing the desired audio to be heard more clearly. ANC can be used for several applications, such as reducing noise from an engine or other machinery or reducing external noise from a noisy environment. With the reduction of the noise produced by external sources, the quality of the audio rendering is greatly enhanced. In this chapter, an innovative subband active noise control system is presented. The chapter is organized as follows. Section 4.1 presents the background on ANC systems that can be found in the literature. Section 4.2 explains the proposed subband active noise control system. Section 4.3 shows the experimental results of the proposed approach considering different types of noise. Finally, Section 4.4 concludes the chapter.

4.1 Background of active noise control

Active noise control systems can be classified into feedforward and feedback structures, as shown in Figure 4.1. A tutorial review of ANC systems is presented in [16]. Feedforward structures use a reference microphone (in addition to the error microphone) to capture the primary noise signal [203], as shown in Figure 4.2(a), and can be divided into narrowband and broadband systems, according to the characteristics of the primary noise. Feedback systems do not include a reference microphone, but use only the signal measured by the error microphone [204], as shown in Figure 4.2(b). In the literature, both fixed and adaptive solutions can be found. However, adaptive filters are more suitable because they follow the variations of the acoustic path. One of the most used adaptive algorithms in ANC systems is the least mean square (LMS) algorithm, but usually, it generates instability [205]. The secondary source, that repro-

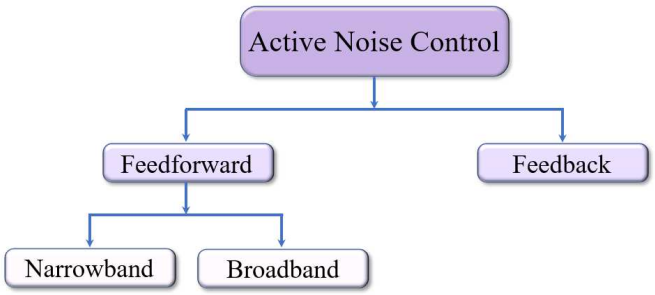


Figure 4.1: Active noise control systems classification.

duces the "antinoise" signal, introduces a secondary path between the control source and the error microphone which has to be evaluated. In particular, the ANC system is usually built on filtered-x least mean square (FxLMS) [206–208] algorithm. In [207, 208], the FxLMS approach is applied for the reduction of snoring noise. Snoring noise can reach a volume of 90 dB and can cause several problems, such as loss of productivity, reduction of attention, and unsafely driving [209, 210]. Recently, several studies have highlighted the strong similarity between the snoring activity and the vocal signal [211, 212]. In fact, the snoring signal is characterized by a fundamental frequency followed by high-order harmonics [212], like the vocal signal, and most of the power is concentrated at lower frequencies. In particular, the inspiration produces a signal between 100 Hz and 200 Hz, while the expiration is focused between 200 Hz and 300 Hz. Hence, the fundamental frequency to be eliminated resides between 100 Hz and 300 Hz.

Figure 4.3 shows a block diagram of a feedforward ANC system based on FxLMS algorithm, in which $x(n)$ represents the primary noise, $s(n)$ is the impulse response of the secondary path and $w(n)$ is the filter to be adapted with an LMS algorithm, controlling the residual noise $e(n)$ captured by the error microphone. The system of Figure 4.3 can be improved with the introduction

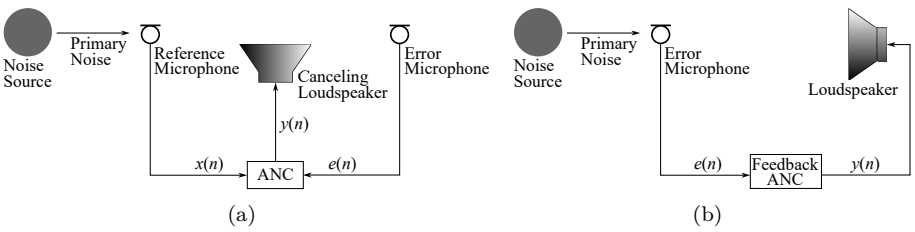


Figure 4.2: Single-channel (a) feedforward and (b) feedback ANC systems.

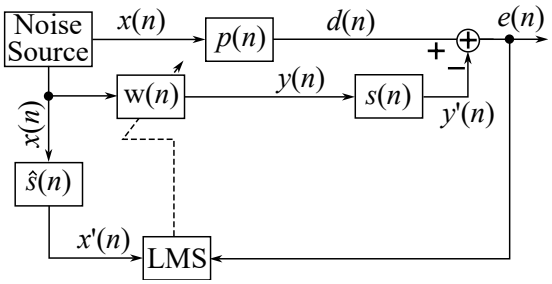


Figure 4.3: Scheme of a simple ANC system based on FxLMS.

of an online estimation of the secondary path $s(n)$. In fact, the secondary path is time-varying, as the primary path, due to atmospheric changes or loudspeakers and microphones damage. The literature offers two techniques for online secondary path modeling. The first is based on the injection of additional random noise $v(n)$ [213], while the second uses the output $y(n)$ to estimate the secondary path [214]. The first solution presents a better performance in terms of convergence rate, speed of response to variations of the primary noise, updating duration, and computational cost [215]. In particular, the system proposed by Eriksson in [213] involves also a supplementary adaptive filter to model $s(n)$, in addition to the introduction of a white noise uncorrelated with the primary noise $x(n)$. However, the additional noise could perturb the secondary path calculation and reduce the convergence rate. To solve this problem, two alternatives are proposed in [216] and [217], respectively. The first method, proposed by Bao *et al.*, aims at canceling the interference of the additional noise by introducing another adaptive filter [216]. However, the noise $v(n)$ degrades the convergence on the adaptation of the new filter. The second method, proposed by Kuo *et al.*, tries to reduce the noise interference by adding an adaptive error filter [217]. However, the effect of the additional noise on the filter $\mathrm{w}(n)$ is examined neither in [216] nor [217]. Contrarily, this aspect is investigated by Zhang *et. al.* in [6], where three cross-updated filters are employed. This system is improved in [27] by introducing the delayless subband structure of [218] in the primary path estimation in order to increase the convergence rate and reduce the error. The same approach is then implemented in real-time in [28].

4.2 Subband active noise control system

This section presents the subband active noise control system with online secondary path estimation proposed in [27,28], where snoring noise has been used for experiments. The scheme of the proposed ANC algorithm is shown in Figure 4.4. Starting from the approach of [6], where an ANC system with online

secondary path estimation is proposed, a subband adaptive filtering (SAF) structure has been added in the primary path estimation. The SAF technique implements a delayless subband structure [218], which allows the improvement of the entire ANC system.

4.2.1 ANC with online secondary path estimation

The proposed ANC approach is based on the filtered-x LMS technique with the online secondary path estimation [6]. The scheme of the proposed algorithm is shown in Figure 4.4. In this section, the SAF structure is not taken into account, so w(n) is a single filter updated through a LMS algorithm, as proposed in [6]. The input signal $x(n)$ is filtered by the estimated secondary path $\hat{s}(n)$. The secondary path estimation is achieved by additional uncorrelated noise injection to the output of ANC controller that may perturb the primary path estimation. A solution to this problem has been proposed in [6], by calculating the error for the primary path estimation $e'(n)$ as

$$e'(n) = e(n) - \hat{s}(n) * v(n), \tag{4.1}$$

where $v(n)$ is the injected uncorrelated noise, $\hat{s}(n)$ is the estimated secondary path and $e(n)$ is calculated as

$$e(n) = d(n) - s(n) * y(n) + s(n) * v(n), \tag{4.2}$$

where $d(n)$ is the desired signal, $y(n)$ is the output of the ANC controller and $s(n)$ is the real secondary path. In the ideal case, when $\hat{s}(n) = s(n)$, the

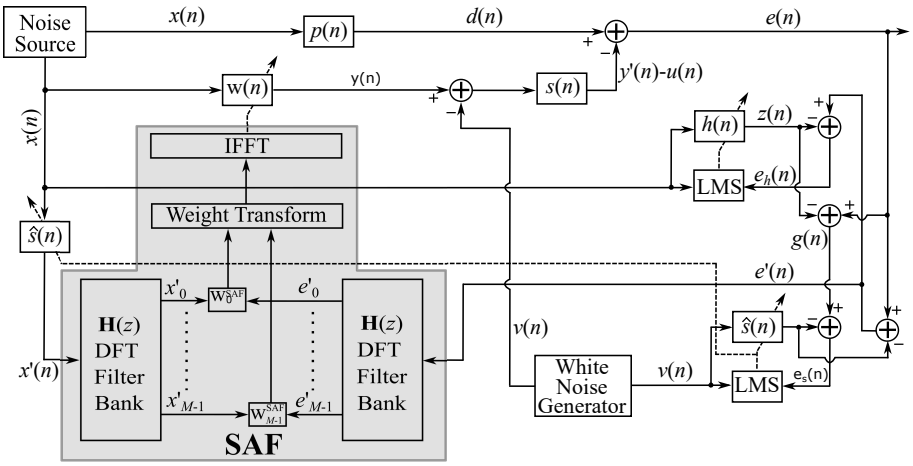


Figure 4.4: Scheme of the proposed ANC system with secondary path modeling and delayless subband algorithm.

error becomes $e'(n) = d(n) - s(n) * y(n)$ and the perturbation caused by the additional noise is removed. The error $e'(n)$ is used in the main adaptive filter as the error signal for $\mathrm{w}(n)$ and as the desired signal for the additional adaptive filter $h(n)$. In fact, these filters have the following updating equations:

$$\mathrm{w}(n+1) = \mathrm{w}(n) + \mu_{\mathrm{w}} x'(n) e'(n), \tag{4.3}$$

and

$$h(n+1) = h(n) + \mu_h x(n)[e'(n) - z(n)], \tag{4.4}$$

where $z(n) = h(n) * x(n)$ is the output of filter $h(n)$, $x(n)$ is the noise signal, $x'(n)$ is the noise signal filtered with secondary path and μ_{w} and μ_h are the step size of the filters $\mathrm{w}(n)$ and $h(n)$, respectively. The update equation of the secondary path estimation uses $v(n)$ as input signal and $e_s(n)$ as error signal, as follows:

$$\hat{s}(n+1) = \hat{s}(n) + \mu_s v(n) e_s(n), \tag{4.5}$$

where μ_s is the step size of the filter $\hat{s}(n)$ and $e_s(n)$ is computed as follows:

$$e_s(n) = g(n) - \hat{u}(n), \tag{4.6}$$

where $g(n) = e(n) - z(n)$ and $\hat{u}(n)$ is noise injected filtered by $\hat{s}(n)$.

4.2.2 Subband adaptive filtering

The proposed system introduces a delayless subband adaptive filtering (SAF) structure in the primary path estimation, as shown in Figure 4.4. The SAF approach improves the convergence rate of the ANC controller, but also the convergence rate of the whole system, due to the dependence among the three adaptive filters, i.e., $\mathrm{w}(n)$, $\hat{s}(n)$, and $h(n)$. The delayless subband adaptive filter algorithm is implemented as proposed in [218]. The signal $x'(n)$, that is the input $x(n)$ filtered by $\hat{s}(n)$, and the error $e'(n)$, obtained by the Equation (4.1), are decomposed in subband by the analysis filterbank $\mathbf{H}(z)$, derived from a prototype filter of order N_{p} and described as follows:

$$\mathbf{H}(z) = [H_0(z), H_1(z), ..., H_{M-1}(z)]^T, \tag{4.7}$$

where $H_k(z)$ is the transfer function of the kth analysis bandpass filter of order N_{p}, with $k = 0, ..., M-1$ and M the number of subbands. The weights of the kth subband $\mathbf{w}_k^{\mathrm{SAF}}(n)$ are updated following the normalized LMS (NLMS) algorithm, i.e.,

$$\mathbf{w}_k^{\mathrm{SAF}}(n+1) = \mathbf{w}_k^{\mathrm{SAF}}(n) + \mu_{\mathrm{w}} \frac{x_k'^*(n) e_k'(n)}{\epsilon + ||x_k'(n)||^2}, \tag{4.8}$$

where $\mathbf{x}_k'^*(n)$ is the complex conjugate of the input signal of the kth subband $x_k'(n)$, μ_{w} is the step size, ϵ is a small number to avoid division by zero, and $e_k'(n)$ is the error of the kth subband. The fullband filter $\mathrm{w}(n)$ is obtained by implementing the following steps:

- the subband weights are transformed by the calculation of M FFT of length $(N+1)/L$, where N is the order of the fullband filter and $L = M/2$ is the decimation factor;
- the complex samples of FFT are stacked to form the first half of the array of the fullband filter;
- to complete the array, the central point is set to zero and the second half of the array is obtained by the complex conjugating and reversing the first half of the array;
- the inverse FFT of length $(N + 1)$ is performed to obtain the fullband filter $\mathrm{w}(n)$.

4.3 Experimental results of the ANC system

The proposed algorithm has been evaluated through several experiments by comparing it with the reference ANC system of [6], which does not use a subband implementation. The primary path and the secondary path are measured from the setup of [208] inside a semi-anechoic chamber and they are modeled as FIR filters of order $N = 255$ (i.e., with a length of 256 samples). The experimental tests have been carried out first with white noise and then with a snoring signal as input, evaluating:

- the estimation of the primary path,
- the estimation of the secondary path,
- the relative error $\varepsilon(n)$ of the primary path estimation,
- the error of the secondary path estimation $e_s(n)$,
- the residual error $e(n)$ of the whole ANC system.

The relative error of the primary path estimation is calculated as

$$\varepsilon(n) = \frac{||\mathrm{w}(n) - p(n)||}{||p(n)||}, \tag{4.9}$$

where $p(n)$ is the impulse response of the primary path. The error of the secondary path estimation $e_s(n)$ is computed by Equation (4.6), and the error

of the total system $e(n)$ is derived from Equation (4.2). For both the algorithms (i.e., the proposed SAF approach and the approach of Zhang [6]), the length of the filters is 512 taps for $w(n)$ and 256 taps for the adaptive filter of secondary path $\hat{s}(n)$ and for $h(n)$. For the subband structure, the order of the prototype filter is $N_\mathrm{p} = 511$ with white noise, and $N_\mathrm{p} = 255$ with snoring noise.

4.3.1 Results on white noise

The proposed algorithm has been compared with the system of Zhang [6], used as a reference, first using white noise as input. The optimal values of the step sizes for both the algorithms (i.e., the proposed and the reference) has been found after a large number of simulations. For the reference algorithm, the optimal values for the step size are the following: $\mu_\mathrm{w} = 0.002$, $\mu_s = 0.002$, $\mu_h = 0.001$, while for the proposed algorithm, the following values have been found: $\mu_\mathrm{w} = 0.008$, $\mu_s = 0.004$, $\mu_h = 0.001$. Moreover, for the proposed SAF structure, a number of subbands of $M = 128$ has been chosen.

In Figure 4.5, the time and frequency responses of the estimated primary path are shown, comparing the measured impulse response $p(n)$ with the weights w(n) estimated with the reference algorithm and with the proposed system. The proposed approach perfectly fits the measured IR, while the reference approach has some difference in the low-frequency range, especially below 200 Hz. At higher frequencies, both approaches correctly adapt to the primary path. The convergence rate is improved too, as shown in Figure 4.7(a), where the

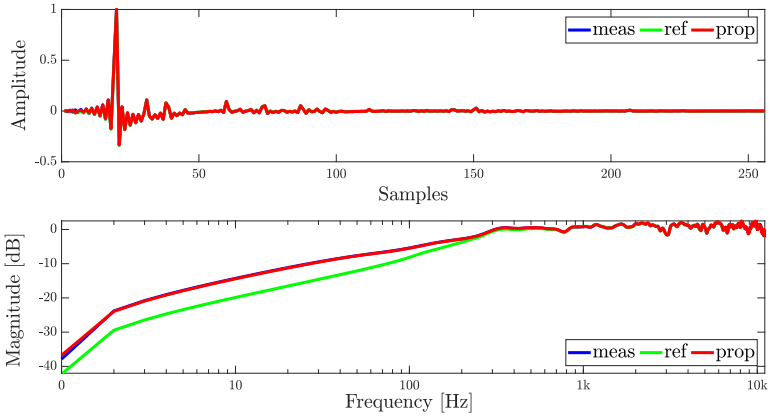


Figure 4.5: Comparison between the measured primary path, the primary path estimated by the reference algorithm of [6], and the primary path estimated by the proposed approach in the time domain (above) and in the frequency domain (below) with white noise as input.

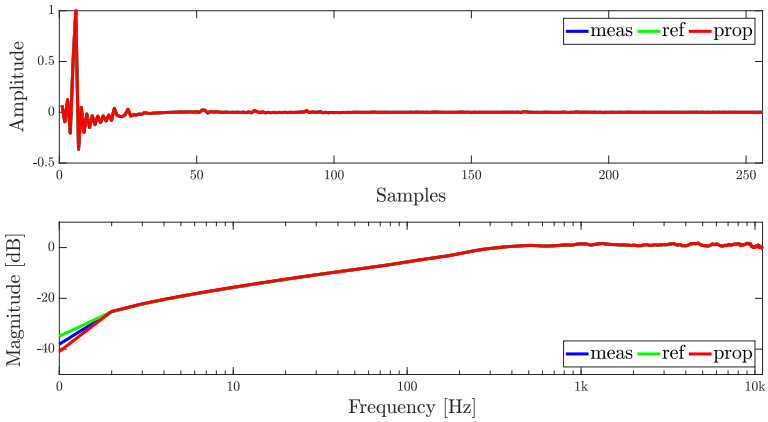


Figure 4.6: Comparison between the measured secondary path, the secondary path estimated by the reference algorithm of [6], and the secondary path estimated by the proposed approach in the time domain (above) and in the frequency domain (below) with white noise as input.

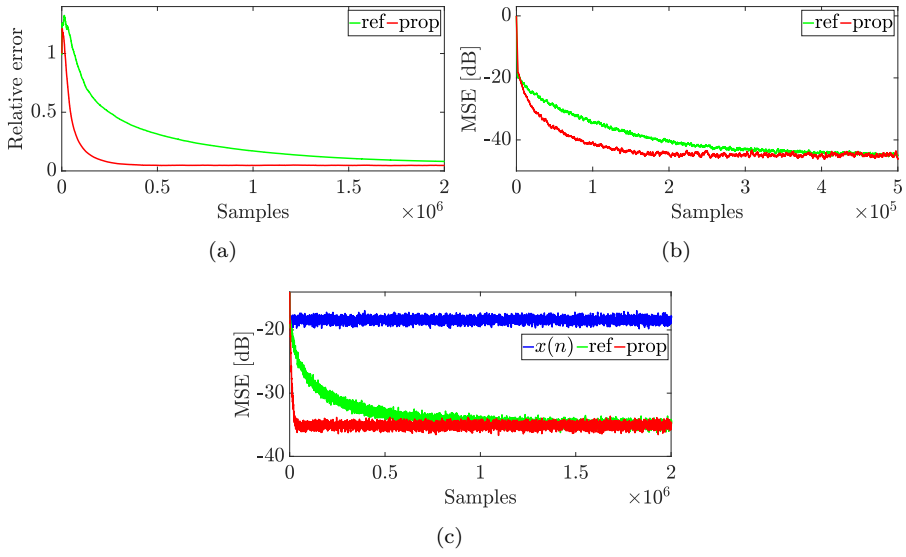


Figure 4.7: Comparison between the reference algorithm of [6] and the proposed algorithm, evaluating (a) the relative error of the primary path estimation, (b) the error of the secondary path estimation, and (c) the MSE in relation to the input signal $x(n)$, with white noise as input.

relative error $\varepsilon(n)$, calculated following the Equation (4.9), is reported.

The comparison between the measured secondary path $s(n)$ and the weights of the adaptive filter $\hat{s}(n)$ of the reference and the proposed algorithms is shown in Figure 4.6. In this case, both methods exhibit a frequency response that perfectly fits with the secondary path. However, also in this case, the proposed approach shows a higher convergence rate, proved by the evolution of the secondary path estimation error $e_s(n)$, shown in Figure 4.7(b). As it can be clearly seen, the application of the SAF structure improves the performance of the secondary path estimation too. Figure 4.7(c) shows the ANC performance in terms of residual noise $e(n)$. The final residual noise of both algorithms is the same and this is due to the injection noise. However, the proposed approach has a better performance in terms of convergence rate as a result of the application of the SAF technique.

4.3.2 Results on snoring noise

After the validation with white noise, the proposed algorithm has been tested considering the snoring noise as input. The snoring signal has been downloaded from [219]. Even in this case, several simulations have been achieved to find the optimal values of step size. For the reference algorithm, the found values are the following: $\mu_{\mathrm{w}} = 0.003$, $\mu_s = 0.0013$, $\mu_h = 0.001$ and for the proposed algorithm are: $\mu_w = 0.014$, $\mu_s = 0.001$, $\mu_h = 0.001$. For the SAF structure, a number of subbands of $M = 64$ has been selected. Considering the primary path estimation, both algorithms perfectly reconstruct the time and frequency

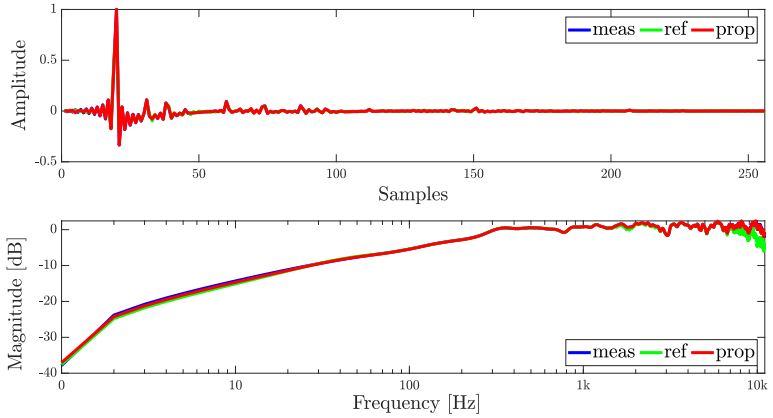


Figure 4.8: Comparison between the measured primary path, the primary path estimated by the reference algorithm of [6], and the primary path estimated by the proposed approach in the time domain (above) and in the frequency domain (below) with snoring noise as input.

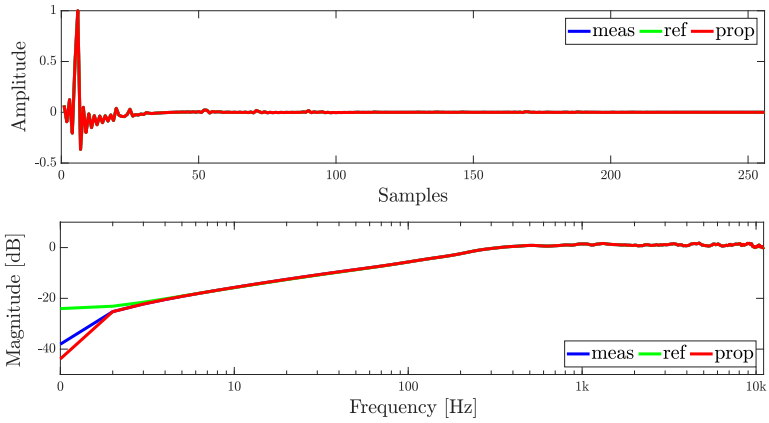


Figure 4.9: Comparison between the measured secondary path, the secondary path estimated by the reference algorithm of [6], and the secondary path estimated by the proposed approach in the time domain (above) and in the frequency domain (below) with snoring noise as input.

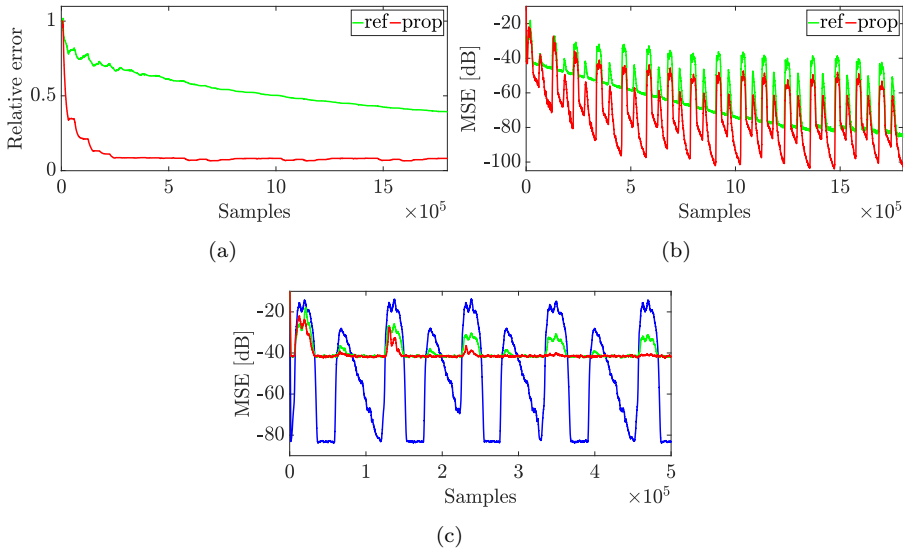


Figure 4.10: Comparison between the reference algorithm of [6] and the proposed algorithm, evaluating (a) the relative error of the primary path estimation, (b) the error of the secondary path estimation, and (c) the MSE in relation to the input signal $x(n)$, with snoring noise as input.

responses, as shown in Figure 4.8. However, evaluating the convergence rate of the primary path, the proposed algorithm exhibits a significant improvement, as shown in Figure 4.10(a), where the error $\varepsilon(n)$, calculated by Equation (4.9) is reported. Regarding the secondary path estimation, both the reference and the proposed algorithms correctly fit the time and the frequency responses, as shown in Figure 4.9. A small difference is visible below 10 Hz, but it it not relevant for snoring. The convergence rate of the secondary path is greatly improved with the proposed algorithm, as reported in Figure 4.10(b), where the error $e_s(n)$, defined by Equation (4.6), is shown. Finally, Figure 4.10(c) shows the residual noise $e(n)$, in comparison with the input snoring noise. The proposed algorithm reaches a residual noise of about 10 dB lower than the reference approach, with a better convergence rate.

4.4 Conclusions of active noise control

In this chapter, an active noise control system with online secondary path estimation has been presented. The system aims at improving a reference algorithm by applying an adaptive subband structure in the primary path estimation. Experiments have been carried out to compare the proposed approach with the reference using white noise and snoring noise as input signals. For both the noises, the primary and the secondary path estimations are evaluated in terms of impulse and frequency responses and errors. Results have shown that the subband structure improves the estimation of the primary path in terms of estimated response, error, and convergence rate. Moreover, also the estimation of the secondary path produces better results than the reference algorithm, proving that the SAF approach, applied only to the primary path, enhances the performance of the whole ANC system. This aspect is also confirmed by the evaluation of the residual noise, which, in the proposed system, is about 10 dB lower than the reference technique.

Chapter 5

Conclusions

In this thesis, innovative systems for immersive audio rendering enhancement have been presented. In particular, a comparative analysis of HRTF measurements has been described in Chapter 2, comparing the transfer functions acquired by different in-ear microphones in different positions with the ones measured by a standard head and torso simulator. Results have shown that the position and the type of the microphone affect the HRTF at higher frequencies. Successively, a binaural synthesis system based on HRTF interpolation has been proposed, aiming at reducing the measurement procedure and improving the spatial resolution. The interpolation algorithm is based on the division of the impulse responses into early reflections and reverberant tails. The transition between the early reflections and the reverberant tail is defined by the mixing time, which can be fixed or calculated. The system has been implemented in real-time and tested in a semi-anechoic environment, considering HRTFs, and in a real environment, employing BRIRs and adding an automatic mixing time calculation based on the Jarque-Bera test. The results on HRTFs have proved the effectiveness of the proposed algorithm in comparison with a reference technique, and the experiments on BRIRs have confirmed the relation between the distance and the mixing time, verifying the necessity of an automatic mixing time calculation.

The binaural system has been also adapted for the reproduction over loudspeakers, by adding the RACE algorithm. The introduction of the crosstalk canceller is needed to reduce the interference signals between loudspeakers with the aim of performing a reproduction similar to the one obtained with headphones. The system has been evaluated through listening tests and has proved the effectiveness of the introduced CTC technique. However, the RACE algorithm is a fixed CTC method which assumes that the listener does not move during the reproduction. Therefore, an adaptive CTC system, based on a subband structure, has been proposed. Differently from RACE, the proposed adaptive CTC algorithm requires the knowledge of the four HRTFs that characterize the four acoustic paths between the loudspeakers and the listener's ears. The proposed system detects the listener's head position by means of a head

tracker and calculates the exact HRTFs by applying the interpolation among a reduced set of HRTFs pre-measured in the listening area. The CTC filters are then calculated through a subband structure. Experiments have shown that a higher number of subbands improves the performance of the crosstalk canceller and reduces the error of the system.

Chapter 3 is focused on audio equalization systems, which are essential for audio rendering enhancement. Equalization can be achieved through several approaches. In this thesis, graphic equalizers and an adaptive multichannel room response equalization (RRE) system have been presented. More in detail, three different GEQs have been presented based on linear-phase interpolated FIR filters. In particular, a linear-phase uniform GEQ, a linear-phase octave GEQ, and a low-latency quasi-linear-phase octave GEQ have been proposed. The uniform equalizer implements a parallel filterbank in which the bands have the same width, while the octave GEQs use a tree structure of interpolated filters that designs a ten-band filterbank with a logarithmic band division, more suitable for audio equalization. The quasi-linear-phase GEQ is designed by a hybrid FIR/IIR structure, derived from the linear phase octave GEQ by designing the first band with a lowpass IIR shelving filter. In this way, the linear phase is guaranteed only above about 100 Hz, but the latency is reduced by half. Experimental results have also shown that the uniform GEQ has the highest computational complexity, while the octave GEQs have proven a greater performance.

Successively, an adaptive multichannel and multiple positions room response equalization system has been presented. The system is capable of adaptively identifying the RIRs of different points in the equalization zone by implementing a subband structure. The equalization is then achieved by inverting a prototype frequency response, calculated by the combination of the smoothed measured frequency responses. Experimental results have evaluated the identification algorithm comparing the subband implementation with the single-band one, proving that a better identification is obtained with the subband structure. In addition, it has been shown that a good RIR estimation is also reflected in the equalization performance. In multichannel systems, several loudspeakers are involved and, generally, they can reproduce different frequency bands, so crossover networks are needed. In this context, a linear-phase digital crossover network has been presented using interpolated FIR filters for multichannel systems. The proposed crossover can split the signal into multiple frequency bands defined by the selected cutoff frequencies. The proposed crossover has been compared with the most used Linkwitz-Riley crossover network and other linear-phase implementations. Experiments have shown that the proposed structure can verify all the requirements needed for a high-quality crossover network and presents a lower computational complexity than other

linear-phase techniques.

Finally, since audio rendering can be improved also by noise reduction, a subband active noise control system with online secondary path estimation has been presented in Chapter 4. The ANC system is based on a previous algorithm, used as a reference, and a subband structure has been applied to the estimation of the primary path. Experiments have been carried out using white noise and snoring noise to evaluate the primary and secondary path estimation, the error, and the convergence rate. It has been shown that the subband implementation improves the performance not only on the primary path estimation but also on the secondary path estimation and on the total system, especially at the low frequencies.

Bibliography

[1] B. Gardner and K. Martin, "HRTF Measurements of a KEMAR Dummy-Head Microphone," *MIT Media Lab Perceptual Computing Technical Report #280*, 1994 (update on October 13, 2006). [Online]. Available: http://www.media.mit.edu

[2] Knowles, "FG-23329-D65 Microphones," https://www.mouser.it/datasheet/2/218/knowles_corporation_04272020_FG-23329-D65-1840470.pdf, 2009.

[3] Sennheiser, "MKE 2-EW GOLD Microphone," https://assets.sennheiser.com/global-downloads/file/5712/MKE2_Gold_Manual_07_2015_EN.pdf.

[4] B. . K. r, "Head and torso simulator (hats) type 4128c," https://www.bksv.com/en/transducers/simulators/head-and-torso/hats-type-4128c, 2010.

[5] V. Garcia-Gomez and J. J. Lopez, "Binaural Room Impulse Responses Interpolation for Multimedia Real-Time Applications," in *AES Convention 144*, May 2018. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=19479

[6] M. Zhang, H. Lan, and W. Ser, "Cross-Updated Active Noise Control System with Online Secondary Path Modeling," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 598–602, Jul. 2001.

[7] S. Bharitkar and C. Kyriakakis, *Immersive Audio Signal Processing*. Springer, 2006.

[8] W. G. Gardner, *3-D Audio using Loudspeakers*. Kluwer Academic Publishers, 1998.

[9] K. Iida, *Head-Related Transfer Function and Acoustic Virtual Reality*. Springer, 2019.

[10] M. Burkhard and R. Sachs, "Anthropometric Manikin for Acoustic Research," *The Journal of the Acoustical Society of America*, vol. 58, no. 1, pp. 214–222, 1975.

[11] G. Yu, R. Wu, Y. Liu, and B. Xie, "Near-Field Head-Related Transfer-Function Measurement and Database of Human Subjects," *The Journal of the Acoustical Society of America*, vol. 143, no. 3, pp. EL194–EL198, 2018.

[12] T. Hirahara, H. Sagara, I. Toshima, and M. Otani, "Head Movement During Head-Related Transfer Function Measurements," *Acoustical Science and Technology*, vol. 31, no. 2, pp. 165–171, 2010.

[13] B. B. Bauer, "Stereophonic Earphones and Binaural Loudspeakers," *J. Audio Eng. Soc.*, vol. 9, no. 2, pp. 148–151, Apr. 1961.

[14] V. Välimäki and J. D. Reiss, "All about audio equalization: Solutions and frontiers," *Appl. Sci.*, vol. 6, no. 5, May 2016, https://doi.org/10.3390/app6050129.

[15] A. V. Mäkivirta, *Loudspeaker Design and Performance Evaluation.* New York, NY: Springer New York, 2008, pp. 649–667. [Online]. Available: https://doi.org/10.1007/978-0-387-30441-0_33

[16] M. K. Kuo and R. M. Morgan, "Active Noise Control: a Tutorial Review," *Proc. IEEE*, vol. 87, no. 6, pp. 943–973, Jun. 1999.

[17] V. Bruschi, S. Nobili, S. Cecchi, and F. Piazza, "An Innovative Method for Binaural Room Impulse Responses Interpolation," *Presented at the 148th AES Convention, Online*, May 2020.

[18] V. Bruschi, S. Nobili, and S. Cecchi, "A Real-Time Implementation of a 3D Binaural System based on HRIRs Interpolation," in *2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA)*, 2021, pp. 103–108.

[19] V. Bruschi, S. Nobili, A. Terenzi, and S. Cecchi, "An Improved Approach for Binaural Room Impulse Responses Interpolation in Real Environments," in *152nd Convention of the Audio Engineering Society.* Audio Engineering Society, 2022.

[20] V. Bruschi, S. Nobili, and S. Cecchi, "Real Time Binaural Synthesis of Moving Sound Sources over Loudspeakers," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA).* IEEE, 2021, pp. 1–9.

[21] V. Bruschi, S. Nobili, F. Bettarelli, and S. Cecchi, "Listener-position Sub-band Adaptive Crosstalk Canceller using HRTFs Interpolation for Immersive Audio Systems," in *150th Convention of the Audio Engineering Society.* Audio Engineering Society, May 2021.

[22] V. Bruschi, S. Nobili, A. Terenzi, and S. Cecchi, "A Low-Complexity Linear-Phase Graphic Audio Equalizer Based on IFIR Filters," *IEEE Signal Process. Lett.*, vol. 28, pp. 429–433, Feb. 2021, https://doi.org/10.1109/LSP.2021.3057228.

[23] V. Bruschi, V. Välimäki, J. Liski, and S. Cecchi, "Linear-Phase Octave Graphic Equalizer," *J. Audio Eng. Soc., Special Issue on Audio Filter Design*, 2022.

[24] V. Bruschi, V. Välimäki, J. Liski, S. Cecchi *et al.*, "A Low-Latency Quasi-Linear-Phase Octave Graphic Equalizer," in *International Conference on Digital Audio Effects*, 2022, pp. 94–100.

[25] S. Cecchi, A. Terenzi, V. Bruschi, A. Carini, and S. Orcioni, "A Subband Implementation of a Multichannel and Multiple Position Adaptive Room Response Equalizer," *Applied Acoustics*, vol. 173, p. 107702, 2021.

[26] V. Bruschi, S. Nobili, A. Terenzi, and S. Cecchi, "Using Interpolated FIR Technique for Digital Crossover Filters Design," in *Proceedings of the 30th European Signal Processing Conference (EUSIPCO)*, 2022, pp. 214–218.

[27] S. Nobili, V. Bruschi, F. Bettarelli, and S. Cecchi, "An Efficient Active Noise Control System with Online Secondary Path Estimation for Snoring Reduction," in *2021 29th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, p. 7.

[28] ——, "A Real Time Subband Implementation of an Active Noise Control System for Snoring Reduction," in *2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA)*. IEEE, 2021, pp. 109–114.

[29] B. P. Bovbjerg, F. Christensen, P. Minnaar, and X. Chen, "Measuring the Head-Related Transfer Functions of an Artificial Head with a High-Directional Resolution," in *109th Convention of the Audio Engineering Society*, 2000.

[30] J.-G. Richter and J. Fels, "On the Influence of Continuous Subject Rotation During High-Resolution Head-Related Transfer Function Measurements," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 4, pp. 730–741, 2019.

[31] H. Wierstorf, M. Geier, and S. Spors, "A Free Database of Head Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances," in *130th Convention of the Audio Engineering Society*, 2011.

[32] Y. Li, S. Preihs, and J. Peissig, "Acquisition of Continuous-Distance Near-Field Head-Related Transfer Functions on KEMAR Using Adaptive Filtering," in *152nd Convention of the Audio Engineering Society*, 2022.

[33] C. I. Cheng and G. H. Wakefield, "Moving Sound Source Synthesis for Binaural Electroacoustic Music using Interpolated Head-Related Transfer Functions (HRTFs)," *Computer Music Journal*, vol. 25, no. 4, pp. 57–80, 2001.

[34] B. Masiero, M. Pollow, and J. Fels, "Design of a Fast Broadband Individual Head-Related Transfer Function Measurement System," in *Forum Acusticum*, 2011, pp. 2197–2202.

[35] J. G. Bolaños and V. Pulkki, "HRIR Database with Measured Actual Source Direction Data," in *133rd Convention of the Audio Engineering Society*, 2012.

[36] T. Carpentier, H. Bahu, M. Noisternig, and O. Warusfel, "Measurement of a Head-Related Transfer Function Database with High Spatial Resolution," in *7th Forum Acusticum (EAA)*, 2014.

[37] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, "Head-Related Transfer Functions of Human Subjects," *Journal of the Audio Engineering Society*, vol. 43, no. 5, pp. 300–321, May 1995.

[38] Q. Ye, Q. Dong, Y. Zhang, and X. Li, "Fast head-related transfer function measurement in complex environments," in *20th International Congress on Acoustics, Sydney, Australia*, 2010, pp. 23–27.

[39] J. Usher and W. L. Martens, "Perceived Naturalness of Speech Sounds Presented using Personalized Versus Non-Personalized HRTFs," in *Proceedings of the 13th International Conference on Auditory Display*. Montréal, Canada: Georgia Institute of Technology, Jun. 2007.

[40] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural Technique: Do We Need Individual Recordings?" *Journal of the Audio Engineering Society*, vol. 44, no. 6, pp. 451–469, 1996.

[41] C. Armstrong, L. Thresh, D. Murphy, and G. Kearney, "A Perceptual Evaluation of Individual and Non-Individual HRTFs: A Case Study of the SADIE II Database," *Applied Sciences*, vol. 8, no. 11, p. 2029, 2018.

[42] S. Li and J. Peissig, "Measurement of Head-Related Transfer Functions: A Review," *Applied Sciences*, vol. 10, no. 14, p. 5014, 2020.

[43] S. Mehrgardt and V. Mellert, "Transformation Characteristics of the External Human Ear," *The Journal of the Acoustical Society of America*, vol. 61, no. 6, pp. 1567–1576, 1977.

[44] F. M. Wiener and D. A. Ross, "The Pressure Distribution in the Auditory Canal in a Progressive Sound Field," *The Journal of the Acoustical Society of America*, vol. 18, no. 2, pp. 401–408, 1946.

[45] C. Searle, L. Braida, D. Cuddy, and M. Davis, "Binaural Pinna Disparity: Another Auditory Localization Cue," *The Journal of the Acoustical Society of America*, vol. 57, no. 2, pp. 448–455, 1975.

[46] J. C. Middlebrooks, J. C. Makous, and D. M. Green, "Directional Sensitivity of Sound-Pressure Levels in the Human Ear Canal," *The Journal of the Acoustical Society of America*, vol. 86, no. 1, pp. 89–108, 1989.

[47] G. Djupesland and J. Zwislocki, "Sound Pressure Distribution in the Outer Ear," *Acta Oto-Laryngologica*, vol. 75, no. 2-6, pp. 350–352, 1973.

[48] F. M. Wiener, "On the Diffraction of a Progressive Sound Wave by the Human Head," *The Journal of the Acoustical Society of America*, vol. 19, no. 1, pp. 143–146, 1947.

[49] E. A. Shaw, "Earcanal Pressure Generated by a Free Sound Field," *The Journal of the Acoustical Society of America*, vol. 39, no. 3, pp. 465–470, 1966.

[50] J. C. Middlebrooks, "Narrow-Band Sound Localization Related to External Ear Acoustics," *The Journal of the Acoustical Society of America*, vol. 92, no. 5, pp. 2607–2624, 1992.

[51] F. L. Wightman and D. J. Kistler, "Headphone Simulation of Free-Field Listening. I: Stimulus Synthesis," *The Journal of the Acoustical Society of America*, vol. 85, no. 2, pp. 858–867, 1989.

[52] P.-A. Hellstrom and A. Axelsson, "Miniature Microphone Probe Tube Measurements in the External Auditory Canal," *The Journal of the Acoustical Society of America*, vol. 93, no. 2, pp. 907–919, 1993.

[53] H. Møller, "Fundamentals of binaural technology," *Applied acoustics*, vol. 36, no. 3-4, pp. 171–218, 1992.

[54] D. Hammershøi and H. Møller, "Sound Transmission To and Within the Human Ear Canal," *The Journal of the Acoustical Society of America*, vol. 100, no. 1, pp. 408–427, 1996.

[55] V. R. Algazi, C. Avendano, and D. Thompson, "Dependence of Subject and Measurement Position in Binaural Signal Acquisition," *Journal of the Audio Engineering Society*, vol. 47, no. 11, pp. 937–947, 1999.

[56] W. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," *The Journal of the Acoustical Society of America*, vol. 97, no. 6, pp. 3907–3908, 1995.

[57] Q. Ye, Q. Dong, Y. Zhang, and X. Li, "Fast Head-Related Transfer Function Measurement in Complex Environments," in *Proceedings of the 20th International Congress on Acoustics, Sydney, Australia*, 2010, pp. 23–27.

[58] M. Rothbucher, K. Veprek, P. Paukner, T. Habigt, and K. Diepold, "Comparison of head-related impulse response measurement approaches," *The Journal of the Acoustical Society of America*, vol. 134, no. 2, pp. EL223–EL229, 2013.

[59] B. Xie, *Head-related transfer function and virtual auditory display.* J. Ross Publishing, 2013.

[60] A. Carini, S. Cecchi, L. Romoli, and G. L. Sicuranza, "Perfect Periodic Sequences for Legendre Nonlinear Filters," in *Proc. 22nd European Signal Processing Conference*, Lisbon, Portugal, Sep. 2014, pp. 2400–2404.

[61] A. Carini, S. Cecchi, and L. Romoli, "Room Impulse Response Estimation using Perfect sequences for Legendre Nonlinear filters," in *Proc. 23nd European Signal Processing Conference*, Nice, France, Aug. 2015.

[62] A. Carini, L. Romoli, S. Cecchi, and S. Orcioni, "Perfect Periodic Sequences for Nonlinear Wiener Filters," in *Proc. 24th European Signal Processing Conference*, Budapest, Hungary, Aug. 2016.

[63] S. Cecchi, V. Bruschi, S. Nobili, A. Terenzi, and A. Carini, "Using Periodic Sequences for HRTFs Measurement Robust Towards Nonlinearities in Automotive Audio Applications," in *2022 IEEE International Workshop on Metrology for Automotive (MetroAutomotive)*, 2022, pp. 99–104.

[64] "NU-Tech," Leaff Engineering. [Online]. Available: http://www.nu-tech-dsp.com/content.php

[65] P. Minnaar, J. Plogsties, and F. Christensen, "Directional resolution of head-related transfer functions required in binaural synthesis," *Journal of the Audio Engineering Society*, vol. 53, no. 10, pp. 919–929, 2005.

[66] Y. Haneda, Y. Kaneda, and N. Kitawaki, "Common-acoustical-pole and residue model and its application to spatial interpolation and extrapolation of a room transfer function," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 6, pp. 709–717, Nov 1999.

[67] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating interactive virtual acoustic environments," *Journal of the Audio Engineering Society*, vol. 47, pp. 675–705, 09 1999.

[68] L. W. P. Biscainho, F. P. Freeland, and P. S. R. Diniz, "Using inter-positional transfer functions in 3D-sound," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, 2002, pp. II–1961–II–1964. [Online]. Available: https://doi.org/10.1109/ICASSP.2002.5745014

[69] M. Queiroz and G. Sousa, "Efficient Binaural Rendering of Moving Sound Sources using HRTF Interpolation," *Journal of New Music Research*, vol. 40, pp. 239–252, 09 2011.

[70] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466, 2003.

[71] K. Hartung, J. Braasch, and S. J. Sterbing, "Comparison of different methods for the interpolation of head-related transfer functions," in *16th AES Conference: Spatial Sound Reproduction*, Mar 1999. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=8026

[72] H. Gamper, "Head-related transfer function interpolation in azimuth, elevation, and distance," *The Journal of the Acoustical Society of America*, vol. 134, p. EL547, 12 2013. [Online]. Available: https://doi.org/10.1121/1.4828983

[73] G. Kearney, C. Masterson, S. Adams, and F. Boland, "Dynamic time warping for acoustic response interpolation: Possibilities and limitations," in *17th European Signal Processing Conference*, Aug 2009, pp. 705–709.

[74] C. Masterson, G. Kearney, and F. Boland, "Acoustic impulse response interpolation for multichannel systems using dynamic time warping," in *35th AES Conference: Audio for Games*, Feb 2009. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15188

[75] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3616–3627, 2018.

[76] F. Brinkmann and S. Weinzierl, "Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition," in *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*. Audio Engineering Society, 2018.

[77] A. Lindau, L. Kosanke, and S. Weinzierl, "Perceptual evaluation of model- and signal-based predictors of the mixing time in binaural room impulse responses," *J. Audio Eng. Soc.*, vol. 60, no. 11, Nov 2012.

[78] B. Blesser, "An interdisciplinary synthesis of reverberation viewpoints," *J. Audio Eng. Soc.*, vol. 49, no. 10, pp. 867–903, Oct. 2001.

[79] L. Rubak and G. Johansen, "Artificial reverberation based on a pseudo-random impulse response ii," *Presented at the 106th Convention of the Audio Engineering Society*, May 1999.

[80] L. Abel and G. Huang, "A simple, robust measure of reverberation echo density," *Presented at the 106th Convention of the Audio Engineering Society*, Oct. 2006.

[81] R. Stewart and M. Sandler, "Statistical measures of early reflections of room impulse responses," in *Proc. 10th International Conference on Digital Audio Effects*, Bordeaux, France, Sep. 2007.

[82] C. M. Jarque and A. K. Bera, "A test for normality of observations and regression residuals," *Internat. Statist. Rev*, pp. 163–172, 1987.

[83] A. Primavera, S. Cecchi, J. Li, and F. Piazza, "Objective and subjective investigation on a novel method for digital reverberator parameters estimation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 2, pp. 441–452, 2014.

[84] A. Lattanzi, F. Bettarelli, and S. Cecchi, "NU-Tech: The Entry Tool of the hArtes Toolchain for Algorithms Design," in *Proc. of 124th Audio Engineering Society Convention*, Amsterdam, The Netherlands, May 2008, pp. 1–8.

[85] S. J. Orfanidis, *Introduction to Signal Processing*. Prentice Hall, 1995.

[86] "IPP Libraries," Intel Corporation, 2011. [Online]. Available: http://software.intel.com/en-us/intel-ipp/

[87] ITU-R BS. 1284-2, "General methods for the subjective assessment of sound quality ," 2019.

[88] M. Schroeder and B. Atal, "Computer simulation of sound transmission in rooms," *Proceedings of the IEEE*, vol. 51, no. 3, pp. 536–537, Mar. 1963.

[89] M. R. Schroeder, "Computer model for Concert Hall Acoustics," *J. Am. Physics*, vol. 41, pp. 461–471, 1973.

[90] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Prentice-Hall, Inc., Upper Saddle River, NJ, 1985.

[91] L. Lim and C. Kyriakakis, "Multirate Adaptive Filtering for Immersive Audio," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, Salt Lake City, UT, USA, May 2001, pp. 3357–3360.

[92] S. Cecchi, L. Palestini, P. Peretti, F. Piazza, and F. Bettarelli, "Subband Adaptive Crosstalk Cancellation: a novel Approach for Immersive Audio," in *Proc. 124th Audio Engineering Society Convention*, Amsterdam, The Netherlands, May 2008.

[93] R. Glasgal, "360° localization via 4.x RACE processing," in *Proc. of 123rd Audio Engineering Society Convention*, New York, USA, Oct. 2007.

[94] C. Hohnerlein and J. Ahrens, "Perceptual evaluation of a multiband acoustic crosstalk canceler using a linear loudspeaker array," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 96–100.

[95] V. Bruschi, N. Ortolani, S. Cecchi, and F. Piazza, "Immersive sound reproduction in real environments using a linear loudspeaker array," in *147th Convention of the Audio Engineering Society*. Audio Engineering Society, Oct. 2019.

[96] E. Benjamin, "An experimental verification of localization in two-channel stereo," in *121st Convention of the Audio Engineering Society*. Audio Engineering Society, 2006.

[97] M. Petraglia, R. Alves, and P. Diniz, "New Structures for Adaptive Filtering in Subbands with Critical Sampling," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3316–3327, Dec. 2000.

[98] M. Petraglia and P. Batalheiro, "Prototype Filter Design for Subband Adaptive Filtering Structures with Critical Sampling," in *Proc. IEEE International Symposium on Circuits and Systems*, vol. 1, Geneva, Switzerland, May 2000, pp. 543–546.

[99] P. Vaidyanathan, "Multirate digital filters, filter banks, polyphase networks, and applications: a tutorial," *Proceedings of the IEEE*, vol. 78, no. 1, pp. 56–93, 1990.

[100] R. A. Greiner and M. Schoessow, "Design aspects of graphic equalizers," *J. Audio Eng. Soc.*, vol. 31, no. 6, pp. 394–407, Jun. 1983.

[101] J. S. Orfanidis, "High-Order Digital Parametric Equalizer Design," *J. Audio Eng. Soc.*, vol. 53, no. 11, Nov. 2005.

[102] N. Dourou, V. Bruschi, S. Spinsante, and S. Cecchi, "The influence of listeners' mood on equalization-based listening experience," in *Acoustics*, vol. 4, no. 3. MDPI, 2022, pp. 746–763.

[103] S. Cecchi, A. Carini, and S. Spors, "Room Response Equalization - A Review," *Applied Sciences (Switzerland)*, vol. 8, no. 1, 2018.

[104] F. Keiler and U. Zolzer, "Parametric second- and fourth-order shelving filters for audio applications," in *IEEE 6th Workshop on Multimedia Signal Processing*. IEEE, 2004, pp. 231–234.

[105] S. K. Mitra and J. F. Kaiser, *Handbook for Digital Signal Processing*, 1st ed. New York, NY, USA: John Wiley & Sons, Aug. 1993.

[106] A. Favrot and C. Faller, "Wiener-Based Spatial B-Format Equalization," *J. Audio Eng. Soc.*, vol. 68, no. 7/8, pp. 488–494, Jul./Aug. 2020, https://doi.org/10.17743/jaes.2020.0029.

[107] J. Vilkamo, T. Bäckström, and A. Kuntz, "Optimized covariance domain framework for time–frequency processing of spatial audio," *J. Audio Eng. Soc.*, vol. 61, no. 6, pp. 403–411, Jun. 2013.

[108] B. Radlovic and R. Kennedy, "Nonminimum-Phase Equalization and Its Subjective Importance in Room Acoustics," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 6, pp. 728–737, Nov. 2000, https://doi.org/110.1109/89.876311.

[109] Y. Hirata, "Digitalization of conventional analog filters for recording use," *J. Audio Eng. Soc.*, vol. 29, no. 5, pp. 333–337, May 1981.

[110] S. Prince and K. R. S. Kumar, "A novel Nth-order IIR filter-based graphic equalizer optimized through genetic algorithm for computing filter order," *Soft Comput.*, vol. 23, no. 8, pp. 2683–2691, Apr. 2019, https://doi.org/10.1007/s00500-018-3640-9.

[111] J. Rämö, J. Liski, and V. Välimäki, "Third-Octave and Bark Graphic-Equalizer Design with Symmetric Band Filters," *Appl. Sci.*, vol. 10, no. 4, pp. 1–22, Feb. 2020, https://doi.org/10.3390/app10041222.

[112] S. Tassart, "Graphical equalization using interpolated filter banks," *J. Audio Eng. Soc.*, vol. 61, no. 5, pp. 263–279, May 2013.

[113] J. Rämö, V. Välimäki, and B. Bank, "High-precision parallel graphic equalizer," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 12, pp. 1894–1904, Dec. 2014, https://doi.org/10.1109/TASLP.2014.2354241.

[114] J. Liski, B. Bank, J. O. Smith, and V. Välimäki, "Converting series biquad filters into delayed parallel form: Application to graphic equalizers," *IEEE Trans. Signal Process.*, vol. 67, no. 14, pp. 3785–3795, Jul. 2019, https://doi.org/110.1109/TSP.2019.2919419.

[115] V. Välimäki and J. Liski, "Accurate cascade graphic equalizer," *IEEE Signal Process. Lett.*, vol. 24, no. 2, pp. 176–180, Feb. 2017, https://doi.org/10.1109/LSP.2016.2645280.

[116] J. A. Jensen, "A new principle for an all-digital preamplifier and equalizer," *J. Audio Eng. Soc.*, vol. 35, no. 12, pp. 994–1003, Dec. 1987.

[117] J. Henriquez, T. Riemer, and R. Trahan Jr., "A Phase-Linear Audio Equalizer: Design and Implementation," *J. Audio Eng. Soc.*, vol. 38, no. 9, pp. 653–666, Sept. 1990.

[118] M. Waters, M. Sandler, and A. C. Davies, "Low-Order FIR Filters for Audio Equalization," in *91st Convention of the Audio Engineering Society*, Oct. 1991, paper 3188.

[119] D. S. McGrath, "An Efficient 30-Band Graphic Equalizer Implementation for a Low-Cost DSP Processor," in *95th Convention of the Audio Engineering Society*, Oct. 1993, paper 3756.

[120] P. H. Kraght, "A linear-phase digital equalizer with cubic-spline frequency response," *J. Audio Eng. Soc.*, vol. 40, no. 5, pp. 403–414, May 1992.

[121] D. S. McGrath, "A new approach to digital audio equalization," in *97th Convention of the Audio Engineering Society*, San Francisco, CA, Nov. 1994, paper 3899.

[122] R. Väänänen and J. Hiipakka, "Efficient audio equalization using multirate processing," *J. Audio Eng. Soc.*, vol. 56, no. 4, pp. 255–266, Apr. 2008.

[123] M. K. Othman and T. H. Lim, "Run time analysis of an audio graphic equalizer for portable industrial directional sound systems in industrial usage," in *Proceedings of the 14th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, Xi'an, China, Jun. 2019, pp. 2177–2181, https://doi.org/110.1109/ICIEA.2019.8833760.

[124] R. J. Oliver, "Frequency-warped audio equalizer," US Patent 7,764,802 B2, Jul. 2010.

[125] J. Siiskonen, "Graphic Equalization Using Frequency-Warped Digital Filters," Master's thesis, Aalto University School of Electrical Engineering, Espoo, Finland, Jul. 2016.

[126] B. D. Kulp, "Digital Equalization Using Fourier Transform Techniques," in *85th Convention of the Audio Engineering Society*, Los Angeles, CA, Oct. 1988, paper 2694.

[127] H. Schöpp and H. Hetze, "A linear-phase 512-band graphic equalizer using the fast-Fourier transform," in *96th Convention of the Audio Engineering Society*, Feb. 1994, paper 3816.

[128] G. F. P. Fernandes, L. G. P. M. Martins, M. F. M. Sousa, F. S. Pinto, and A. J. S. Ferreira, "Implementation of a new method to digital audio equalization," in *106th Convention of the Audio Engineering Society*, Munich, Germany, May 1999, paper 4895.

[129] S. Ries and G. Frieling, "PC-based equalizer with variable gain and delay in 31 frequency bands," in *108th Convention of the Audio Engineering Society*, Paris, France, Feb. 2000, paper 5173.

[130] S. Cecchi, L. Palestini, E. Moretti, and F. Piazza, "A new approach to digital audio equalization," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, Oct. 2007, pp. 62–65, https://doi.org/110.1109/ASPAA.2007.4393011.

[131] R. Hergum, "A Low Complexity, Linear Phase Graphic Equalizer," in *85th Convention of the Audio Engineering Society*, Nov. 1988, paper 2738.

[132] Y. Neuvo, D. Cheng-Yu, and S. Mitra, "Interpolated finite impulse response filters," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 32, no. 3, pp. 563–570, Jun. 1984, https://doi.org/10.1109/TASSP.1984.1164348.

[133] T. Parks and J. McClellan, "Chebyshev approximation for nonrecursive digital filters with linear phase," *IEEE Transactions on Circuit Theory*, vol. 19, no. 2, pp. 189–194, 1972.

[134] H. Schopp and H. Hetzel, "A Linear Phase 512 Band Graphic Equalizer using the Fast Fourier Transform," in *Proc. 96th Audio Engineering Society Convention*, Amsterdam, The Netherlands, Feb. 1994.

[135] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, no. 1, pp. 14–37, Jan. 1992, https://doi.org/10.1109/79.109205.

[136] J. Liski and V. Välimäki, "The quest for the best graphic equalizer," in *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx)*, Sep. 2017, pp. 95–102.

[137] M. Holters and U. Zölzer, "Graphic equalizer design using higher-order recursive filters," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, Montreal, Canada, Sep. 2006, pp. 37–40.

[138] F. E. Toole and S. E. Olive, "The modification of timbre by resonances: Perception and measurement," *J. Audio Eng. Soc.*, vol. 36, no. 3, pp. 122–142, Mar. 1988.

[139] L. D. Fielder, "Analysis of traditional and reverberation-reducing methods of room equalization," *J. Audio Eng. Soc.*, vol. 51, no. 1/2, pp. 3–26, Feb. 2003.

[140] H. Korhola and M. Karjalainen, "Perceptual study and auditory analysis on digital crossover filters," *J. Audio Eng. Soc.*, vol. 57, no. 6, pp. 413–429, Jun. 2009.

[141] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, "A real-time algorithm for signal analysis with the help of the wavelet transform," in *Wavelets, Time-Frequency Methods and Phase Space*, J. M. Combes, A. Grossmann, and P. Tchamitchian, Eds., 1989, pp. 286–297, https://doi.org/10.1007/978-3-642-75988-8_28.

[142] P. P. Vaidyanathan, *Multirate Systems and Filterbanks*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1993.

[143] J. Rämö, V. Välimäki, and M. Tikander, "Perceptual headphone equalization for mitigation of ambient noise," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, Canada, May 2013, pp. 724–728.

[144] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Prentice-Hall, Inc., Upper Saddle River, NJ, 1999, chapter 7, pp. 465-478.

[145] E. Wesfreid and M. Wickerhauser, "Adapted local trigonometric transforms and speech processing," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3596–3600, Dec. 1993, https://doi.org/10.1109/78.258104.

[146] J. F. Kaiser, "Nonrecursive digital filter design using the I_0-sinh window function," in *Proceedings of the IEEE International Symposium on Circuits & Systems (ISCAS)*, San Francisco, CA, Apr. 1974, pp. 20–23.

[147] R. J. Oliver and J.-M. Jot, "Efficient multi-band digital audio graphic equalizer with accurate frequency response control," in *139th Convention of the Audio Engineering Society*, Oct. 2015, paper 9406.

[148] W. G. Gardner, "Efficient convolution without input/output delay," *J. Audio Eng. Soc.*, vol. 43, no. 3, pp. 127–136, Mar. 1995.

[149] F. Wefers and J. Berg, "High-Performance Real-Time FIR-Filtering using Fast Convolution on Graphics Hardware," in *Proceedings of the 13th International Conference on Digital Audio Effects (DAFx)*, Sep. 2010.

[150] J. Liski, J. Rämö, and V. Välimäki, "Graphic equalizer design with symmetric biquad filters," in *Proc. IEEE Workshop on Applications of Signal Process. to Audio and Acoust. (WASPAA)*, New Paltz, NY, USA, Oct. 2019, pp. 55–59.

[151] V. Välimäki and J. Rämö, "Neurally controlled graphic equalizer," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 27, no. 12, pp. 2140–2149, Dec. 2019, https://doi.org/10.1109/TASLP.2019.2935809.

[152] International Telecommunication Union, "Relative Timing of Sound and Vision for Broadcasting," Recommendation ITU-R BT.1359-1, 1998.

[153] J. Mourjopoulos, "On the variation and invertibility of room impulse response functions," *J. Sound Vibr.*, vol. 102, no. 2, pp. 217–228, Sep. 1985.

[154] P. D. Hatziantoniou and J. N. Mourjopoulos, "Errors in real-time room acoustics dereverberation," *J. Audio Eng. Soc.*, vol. 52, no. 9, pp. 883–899, Sep. 2004.

[155] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Amer.*, vol. 66, no. 1, pp. 165–169, Jul. 1979.

[156] A. Carini, S. Cecchi, F. Piazza, I. Omiciuolo, and G. L. Sicuranza, "Multiple position room response equalization in frequency domain," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 122–135, 2012.

[157] L.-J. Brannmark, A. Bahne, and A. Ahlen, "Compensation of loudspeaker–room responses in a robust mimo control framework," *IEEE transactions on audio, speech, and language processing*, vol. 21, no. 6, pp. 1201–1216, 2013.

[158] S. J. Elliott and P. A. Nelson, "Multiple-Point Equalization in a Room Using Adaptive Digital Filters," *J. Audio Eng. Soc.*, vol. 37, no. 11, pp. 899–907, Nov. 1989.

[159] A. J. S. Ferreira and A. Leite, "An Improved Adaptive Room Equalization in the Frequency Domain," in *Proc. 118th Audio Engineering Society Convention*, Barcelona, Spain, May 2005.

[160] S. Cecchi, A. Primavera, F. Piazza, and A. Carini, "An Adaptive Multiple Position Room Response Equalizer," in *Proc. EUSIPCO 2011*, Barcellona, Spain, May 2011, pp. 1274–1278.

[161] S. Cecchi, A. Carini, A. Primavera, and F. Piazza, "An Adaptive Multiple Position Room Response Equalizer in Warped Domain," in *Proc. EUSIPCO 2012*, Bucharest, Romania, 2012, pp. 1955–1959.

[162] J. Benesty, C. Paleologu, T. Gansler, and S. Ciochina, *A perspective on stereophonic acoustic echo cancellation*. Springer Science & Business Media, 2011, vol. 4.

[163] C. Stanciu, C. Paleologu, J. Benesty, S. Ciochina, and F. Albu, "Variable-forgetting factor rls for stereophonic acoustic echo cancellation with widely linear model," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*. IEEE, 2012, pp. 1960–1964.

[164] L. Romoli, S. Cecchi, P. Peretti, and F. Piazza, "A mixed decorrelation approach for stereo acoustic echo cancellation based on the estimation of the fundamental frequency," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 2, pp. 690–698, 2012.

[165] S. Cecchi, L. Romoli, F. Piazza, and A. Carini, "A Multichannel and Multiple Position Adaptive Room Response Equalizer in Warped Domain," Sep. 2013, Proc. 8th Int'l Symposium on Image and Signal Processing and Analysis.

[166] S. Cecchi, L. Romoli, A. Carini, and F. Piazza, "A Multichannel and Multiple Position Adaptive Room Response Equalizer in Warped Domain: Real-time Implementation and Performance Evaluation," *Applied Acoustics*, vol. 82, pp. 28–37, Aug. 2014.

[167] S. Cecchi, L. Romoli, M. Gasparini, A. Carini, and F. Bettarelli, "An Adaptive Multichannel Identification System for Room Response Equalization," in *Proc. Int. Conf. Electronics, Computers and Artificial Intelligence*, Bucharest, România, Jun. 2015.

[168] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic Acoustic Echo Cancellation - An Overview of the Fundamental Problem," *IEEE Signal processing letters*, vol. 2, no. 8, pp. 148–151, 1995.

[169] S. Shimauchi and S. Makino, "Stereo Projection Echo Canceller with True Echo Path Estimation," in *1995 International Conference on Acoustics, Speech, and Signal Processing*, vol. 5. IEEE, 1995, pp. 3059–3062.

[170] J. Benesty, M. M. Sondhi, and Y. Huang, *Springer Handbook of Speech Processing*. Springer, 2008.

[171] E. Larsen and R. M. Aarts, *Audio bandwidth extension*. J. Wiley & Sons, 2004.

[172] L. Romoli, S. Cecchi, and F. Piazza, "A Novel Decorrelation Approach for Multichannel System Identification," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Florence, Italy, May 2014, pp. 6652–6656.

[173] J. Lee, E. Song, Y. Park, and D. Youn, "Effective Bass Enhancement Using Second-Order Adaptive Notch Filter," *IEEE Trans. Consum. Electron.*, vol. 54, pp. 663–668, May 2008.

[174] M. Ali, "Stereophonic Acoustic Echo Cancellation System Using Time-Varying All-pass Filtering for Signal Decorrelation," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 6, Seattle, WA, USA, May 1998, pp. 3689–3692.

[175] S. Cecchi, L. Romoli, P. Peretti, and F. Piazza, "Low-complexity Implementation of a Real-time Decorrelation Algorithm for Stereophonic Acoustic Echo Cancellation," *Signal Processing*, vol. 92, no. 11, pp. 2668–2675, Nov. 2012.

[176] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*. Prentice Hall International Inc., 1999, pp. 274-279.

[177] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*. Springer-Verlag, 1990.

[178] B. Bank, "Combined quasi-anechoic and in-room equalization of loudspeaker responses," in *Proc. 134th Audio Engineering Society Convention*, May 2013.

[179] P. D. Hatziantoniou and J. N. Mourjopoulos, "Generalized Fractional-Octave Smoothing of Audio and Acoustic Responses," *J. Audio Eng. Soc.*, vol. 48, pp. 259–280, Apr. 2000.

[180] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast Deconvolution of Multichannel Systems using Regularization," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 2, pp. 189–194, Mar. 1998.

[181] S. Cecchi, L. Palestini, P. Peretti, L. Romoli, F. Piazza, and A. Carini, "Evaluation of a multipoint equalization system based on impulse response prototype extraction," *Journal of the Audio Engineering Society*, vol. 59, no. 3, pp. 110–123, 2011.

[182] S. Bharitkar and C. Kyriakakis, *Immersive Audio Signal Processing*. New York: Springer, 2006.

[183] S. P. Lipshitz and J. Vanderkooy, "A Family of Linear-Phase Crossover Networks of High Slope Derived by Time Delay," *J. Audio Eng. Soc.*, vol. 31, no. 1/2, pp. 2–20, Feb. 1983.

[184] *Linear Phase Crossover Filters Advantages in Concert Sound Reinforcement Systems: a practical approach*, San Francisco, CA, USA, Oct. 2006.

[185] M. Colloms, *High Performance Loudspeakers*, 6th ed. John Wiley & Sons, Chichester, UK, 2005.

[186] M. Kleiner, *Electroacoustics*. CRC Press, Boca Raton, FL, 2013.

[187] P. Newell and K. Holland, *Loudspeakers: For Music Recording and Reproduction,*, 2nd ed. New York: Taylor & Francis Ltd, Nov. 2018.

[188] J. D. Johnston and R. E. Crochiere, "An all-digital "commentary grade" subband coder," *J. Audio Eng. Soc.*, vol. 27, no. 11, pp. 855–865, May 1979.

[189] J. R. Ashley, "Butterworth filters as loudspeaker frequency-dividing networks," *Proc. IEEE*, vol. 58, no. 6, pp. 959–960, Jun. 1970.

[190] S. H. Linkwitz, "Active Crossover Networks for Noncoincident Drivers," *J. Audio Eng. Soc.*, vol. 24, no. 1, pp. 2–8, Feb. 1976.

[191] J. Vanderkooy and S. P. Lipshitz, "Is phase linearization of loudspeaker crossover networks possible by time offset and equalization?" *J. Audio Eng. Soc.*, vol. 32, no. 12, pp. 946–955, Dec. 1984.

[192] P. Reviriego, J. Parera, and R. Garcia, "Linear-phase crossover design using digital IIR filters," *J. Audio Eng. Soc.*, vol. 46, no. 5, pp. 406–411, May 1998.

[193] D. G. Fink, "Time offset and crossover design," *J. Audio Eng. Soc.*, vol. 28, no. 9, pp. 601–611, Sep. 1980.

[194] R. Miller, "A Bessel Filter Crossover and its Relation to Other Types," in *Proc. 105th Audio Engineering Society Convention*, San Francisco, CA, USA, Sep. 1998.

[195] K. Hlurprasert, P. Rakluea, N. Wongsin, and V. Pirajnanchai, "Design of 4-way crossover network by using bernstein polynomial," in *Proceedings of the International Electrical Engineering Congress (iEECON)*, Mar. 2014, pp. 1–4.

[196] R. Thaden, S. Müller, G. Behler, A. Goertz, M. Makarski, and J. Kleber, "A Loudspeaker Management System with FIR/IIR Filtering," in *Proceedings of the 32nd International AES Conference: DSP For Loudspeakers*, Jan. 2007.

[197] L. Palestini, P. Peretti, S. Cecchi, F. Piazza, A. Lattanzi, and F. Bettarelli, "Linear Phase Mixed FIR/IIR Crossover Networks: Design and Real-Time Implementation," in *123rd Convention of the Audio Engineering Society*, New York, USA, Oct. 2007, paper 7311.

[198] R. Wilson, G. Adams, and J. Scott, "Application of Digital Filters to Loudspeaker Crossover Networks," *J. Audio Eng. Soc.*, vol. 37, no. 6, pp. 445–464, Jun. 1989.

[199] M. Hämäläinen, "Optimization of multirate crossover filters," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 1999, pp. 63–66.

[200] S. Azizi, H. Hetzel, and H. Schopp, "Design and Implementation of Linear Phase Crossover Filters using the FFT," in *Proc. 98th Audio Engineering Society Convention*, Paris, France, Jan. 1995.

[201] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Prentice-Hall, Inc., Upper Saddle River, NJ, 1999.

[202] R. G. Greenfield, "Polar response errors in digital crossover alignments," in *100th Convention of the Audio Engineering Society*, May 1996, paper 4215.

[203] S. J. Elliott and C. C. Boucher, "Interaction between multiple feedforward active control systems," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 4, pp. 521–530, 1994.

[204] S. M. Kuo, X. Kong, and W. S. Gan, "Applications of adaptive feedback active noise control system," *IEEE Trans. Control Syst. Technol.*, vol. 11, no. 2, pp. 216–220, mar 2003.

[205] S. J. Elliott and P. A. Nelson, "The application of adaptive filtering to the active control of sound and vibration," *NASA STI/Recon Technical Report N*, vol. 86, p. 32628, 1985.

[206] D. Morgan, "An analysis of multiple correlation cancellation loops with a filter in the auxiliary path," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 454–467, 1980.

[207] S. M. Kuo and R. Gireddy, "Real-time experiment of snore active noise control," in *2007 IEEE International Conference on Control Applications*, 2007, pp. 1342–1346.

[208] S. Cecchi, A. Terenzi, P. Peretti, and F. Bettarelli, "Real time implementation of an active noise control for snoring reduction," *Journal of the Audio Engineering Society*, May 2018.

[209] S. Chakravarthy and S. Kuo, "Application of active noise control for reducing snore," in *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*, vol. 5, 06 2006, pp. V – V.

[210] C. Chang, S.-T. Pan, and K. Liao, "Active noise control and its application to snore noise cancellation," *Asian Journal of Control*, vol. 15, 11 2013.

[211] R. Beck, M. Odeh, A. Oliven, and N. Gavriely, "The acoustic properties of snores," *European Respiratory Journal*, vol. 8, no. 12, pp. 2120–2128, 1995.

[212] D. Pevernagie, R. M. Aarts, and M. De Meyer, "The acoustics of snoring," *Sleep Medicine Reviews*, vol. 14, no. 2, pp. 131–144, 2010.

[213] L. J. Eriksson and M. C. Allie, "Use of random noise for on-line transducer modeling in an adaptive active attenuation system," *J. Acoust. Soc. Amer.*, vol. 85, no. 2, pp. 797–802, Feb. 1989.

[214] K. S. M. and W. M., "Parallel adaptive on-line error-path modeling algorithm for active noise control systems," *IEEE Electron. Lett.*, vol. 28, pp. 375–377, Feb 1992. [Online]. Available: https://doi.org/10.1049/el:19920235

[215] C. Bao, P. Sas, and H. V. Brussel, "Comparison of two online identification algorithms for active noise control," in *Proc. Recent Advances in Active Control of Sound Vibration*, 1993, pp. 38–51.

[216] ——, "Adaptive active control of noise in 3-D reverberant enclosure," *J. Sound Vibr.*, vol. 161, no. 3, pp. 501–514, Mar. 1993.

[217] S. M. Kuo and D. Vijayan, "A secondary path modeling technique for active noise control systems," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 4, pp. 374–377, Jul. 1997.

[218] R. M. Morgan and T. J. C., "A Delayless Subband Adaptive Filter Architecture," *IEEE Trans. Signal Process*, vol. 43, no. 8, pp. 1819–1830, Aug. 1995.

[219] Epidemic Sound, https://www.epidemicsound.com/sound-effects/sneezes/, online; accessed 15 May 2021.