Signal Processing 234 (2025) 109992

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

# A state-of-the-art review on acoustic preservation of historical worship spaces through auralization

Hannes Rosseel *, Toon van Waterschoot[1]

*KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Kasteelpark Arenberg 10, Leuven, B-3001, Vlaams-Brabant, Belgium*

ARTICLE INFO

*Keywords:*
Acoustic preservation
Historical worship spaces
Room acoustics
Auralization

ABSTRACT

Historical Worship Spaces (HWS) are significant architectural landmarks which hold both cultural and spiritual value. The acoustic properties of these spaces play a crucial role in historical and contemporary religious liturgies, rituals, and ceremonies, as well as in the performance of sacred music. However, the original acoustic characteristics of these spaces are often at risk due to repurposing, renovations, natural disasters, or deterioration over time. This paper presents a comprehensive review of the current state of research on the acquisition, analysis, and synthesis of acoustics, with a focus on HWS. An example case study of the Nassau chapel in Brussels, Belgium, is presented to demonstrate the application of these techniques for the preservation and auralization of historical worship space acoustics. The paper concludes with a discussion of the challenges and opportunities in the field, and outlines future research directions.

## 1. Introduction

HWS form a significant part of the architectural heritage of many cultures, and are often considered to be architectural landmarks. These spaces provide a unique acoustic environment, mainly characterized by long reverberation times, that is integral to the performance of religious rituals, ceremonies, and the performance of sacred music [1]. The original acoustic properties of HWS are often at risk due to repurposing, renovations, natural disasters, or deterioration over time [2,3]. It is therefore important to the cultural and spiritual communities, as well as to the wider public, to safeguard and preserve the aural heritage that these spaces provide. The safeguarding of intangible cultural heritage has been recognized by UNESCO [4].

Existing review papers on the acoustics of HWS have focused mainly on the historical development of these spaces, the acoustic properties of HWS, and the impact of architectural design decisions on the acoustics of these spaces [1,5]. However, there is a lack of comprehensive reviews on the state of research on the preservation, analysis, and reproduction of HWS acoustics. To address this literature gap, this paper presents a comprehensive review of the current state of research on acoustic preservation and acoustic reproduction of HWS through auralization. Auralization is the process of rendering an acoustic environment audible through the use of sound reproduction techniques [6]. Auralization has been widely used in the field of architectural acoustics to simulate the acoustic properties of spaces, and to evaluate the impact of design decisions on the acoustic performance of buildings [6]. In the context of HWS, auralization can be used to safeguard and recreate the original acoustic characteristics of a space, and to assess the impact of alterations or renovations on the acoustic environment [2].

An overview of the key components in the auralization process for HWS is provided in Fig. 1. This process can be divided into five main components: room Acoustic Acquisition (AA), Sound Field Analysis (SFA), Sound Field Synthesis (SFS), Real-Time Auralization (RTA), and Perceptual Evaluation (PE). The dark-colored boxes represent the relevant methods for each component. The components and their corresponding methods can be utilized either in combination or individually as tools to preserve and reproduce the acoustics of HWS.

The structure of this paper is organized as follows. In Section 2, a literature review on the acoustic preservation of HWS through auralization is presented, situating the relevant research within the framework of the five main components of the auralization process. Following this, each individual component is discussed in detail from Sections 3 to 7, with a focus on the current state of research and its application to the preservation and reproduction of HWS. Section 8 presents a case study on the preservation and auralization of the Nassau Chapel in Brussels, Belgium. Finally, the paper concludes in Section 9 with a discussion of the challenges and opportunities in the field, and an outline of future research directions.

* Corresponding author.
*E-mail address:* hannes.rosseel@esat.kuleuven.be (H. Rosseel).
[1] EURASIP Member.

https://doi.org/10.1016/j.sigpro.2025.109992
Received 22 November 2024; Received in revised form 28 February 2025; Accepted 6 March 2025
Available online 15 March 2025
0165-1684/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

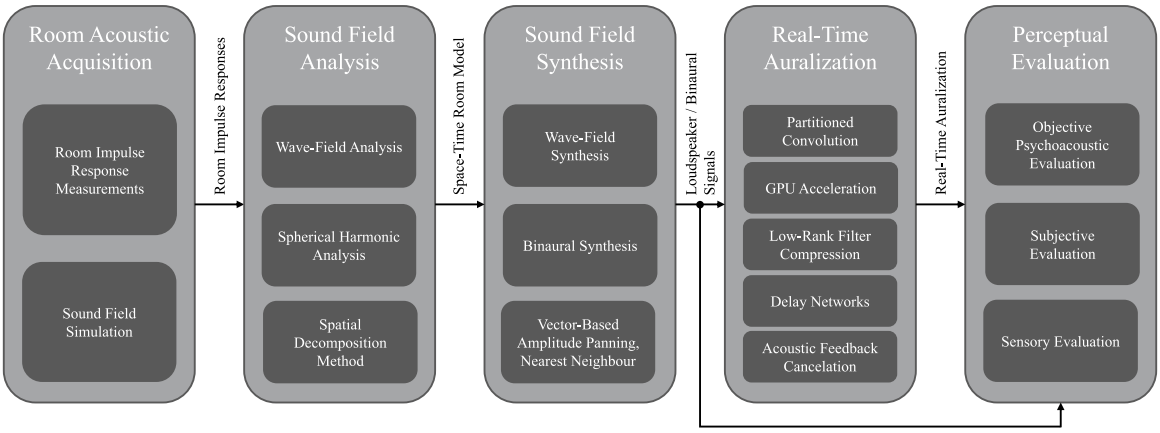



**Fig. 1.** Overview of an auralization process for the preservation and reproduction of HWS. The process consists of five main components: room acoustic acquisition, sound field analysis, sound field synthesis, real-time auralization, and perceptual evaluation. The dark-colored boxes represent the relevant methods for each component. The components and their corresponding methods can be utilized either in combination or individually as tools to preserve and reproduce the acoustics of HWS. All components and methods, together with their application to the preservation of historical worship space acoustics, are discussed in detail in the following sections.

## 2. Acoustic preservation of historical worship spaces through auralization

In this section, an overview of literature relating to the preservation through auralization of HWS is provided. The overview provides a summary of research conducted in HWS, focusing on key components of the auralization process: room acoustic acquisition, sound field analysis, sound field synthesis, real-time auralization, and perceptual evaluation.

The results of the literature search are summarized in Table 1. This table provides an overview of research conducted in HWS where the focus was on the analysis, preservation and reproduction of the acoustics of the spaces through auralization. The table includes the author(s) and publication date, type of worship space, and the employed methods: AA, SFA, SFS, RTA, and PE. The studies are ordered by publication date. A brief summary of each study is provided below.

The Conservation of the Acoustical Heritage and Revival of Sinan's mosques Acoustics (CAHRISMA) project was a collaborative research project that aimed to study the acoustics of historical mosques and Byzantine churches built by the architect Sinan in Istanbul, Turkey. The project aimed to study the acoustics of the spaces using a combination of measurement and simulation techniques. Acoustic measurements were conducted in three mosques and four Byzantine churches by capturing the Room Impulse Responses (RIRs) using both a binaural and B-format probe at various locations in the spaces using the Exponential Sine Sweep (ESS) method. In a different study from the same project, the previously mentioned in-situ recordings of three spaces were compared to auralizations obtained from geometrical models of the spaces using the ODEON software [7]. The comparison was made by convolving the recorded binaural RIRs of a space with anechoic recordings and comparing the results to the auralizations obtained from the ODEON model. The study found that the auralizations were able to reproduce the acoustics of the spaces with a good resemblance to the measured data [7].

Koutsivitis et al. [8] presented a methodology for real-time interactive auralization and visualization techniques for reproducing real world, real-time events in open and closed spaces in 2005. The methodology was applied to the real-time reproduction of the acoustics of the ancient theater of Epidauros, and the ancient temple of Zeus, which are both located in Greece. The acoustics of the spaces were simulated from a geometrical model using Acoustic Ray Tracing (ART) to obtain sets of binaural RIRs for each space. By using readily available Head-Related Transfer Functions (HRTFs) and combining them with the simulated binaural RIRs, the acoustics of the spaces were auralized by tracking the position of the listener in the virtual space and updating the binaural RIRs accordingly. The auralization was performed in real-time by calculating the convolutions of the binaural RIRs with the input signal using a partitioned Overlap-Save (OLS) algorithm [8].

A subjective study on preferred listening conditions in Italian Catholic churches was performed by Martellotta in 2008 [9]. The acoustics of nine Italian Catholic churches were measured using the ESS method as excitation signal and a B-format microphone and a binaural head and torso. The measured RIRs were convolved with anechoic musical excerpts and played back to subjects over loudspeakers in a 'stereo dipole' configuration in a purpose-built dry room. A total of 143 subjects took part in the listening tests, in which they were presented with paired comparisons of the different recordings and asked to state their preference. The results of the listening tests were then analyzed using factor analysis to determine the key acoustic parameters that influenced the subjects' preferences and to determine optimal acoustic conditions for different types of music [9].

Fazenda and Drumm [10] presented a study on the recreation of the sound of Stonehenge using Wave Field Synthesis (WFS) in 2013. The study first measured the acoustics of the original Stonehenge monument in Wiltshire, England using balloon bursts as the excitation signal. These measurements were compared to measurements made with an omnidirectional and a first-order Ambisonic microphone at a restored replica of Stonehenge located in Maryhill, Washington, USA. The acoustics measured at the replica were then used to recreate the acoustics of the original Stonehenge monument using a combination of WFS and First-Order Ambisonic synthesis [10].

Pedrero et al. presented a study in 2014 on the virtual recovery of the sound of the Hispanic rite [11]. The study created geometric models of five pre-Romanesque churches in the Iberian Peninsula using historical sources to reconstruct the buildings in their primitive state. The acoustic simulation software RAVEN [28] was used to simulate the acoustics of the spaces for static and moving source and receiver positions. A total of eight musical pieces were recorded under anechoic conditions and auralized in the acoustic models of the churches [11].

Postma and Katz [12] developed a method for calibrating geometrical acoustic models with acoustical measurements. The method was applied to a geometric model of the abbey church Saint-Germain-des-Prés in Paris, France. The acoustics of the space was simulated using the CATT-Acoustic software, and the simulated acoustics was calibrated by altering the absorption and scattering coefficients of the surfaces in the model to match the acoustical measurements made in the spaces. The calibrated acoustic model was then used as a basis for simulating the acoustics of the space as it would have sounded in the 17th century. Historical sources were used to recreate a configuration of the space that was historically accurate in the context of liturgical practices of

**Table 1**
Summary of research relating to preservation through auralization of historical worship space acoustics. The table includes the author(s) and publication date, type of worship space, Acoustic Acquisition (AA) method, Sound Field Analysis (SFA) method, Sound Field Synthesis (SFS) method, Real-Time Auralization (RTA) method, and Perceptual Evaluation (PE) method. The acronyms used in this table are defined in the glossary at the end of the paper. Studies are ordered by publication date.

| Author(s) | Worship Space | AA | SFA | SFS | RTA | PE |
|---|---|---|---|---|---|---|
| Weitze et al. (2003) [7] | Mosque, Byzantine church | ODEON | ✗ | Binaural | ✗ | ✗ |
| Koutsivitis et al. (2005) [8] | Greek temple | ART | ✗ | Binaural | partitioned OLS | ✗ |
| Martellotta (2008) [9] | Catholic church | ESS | ✗ | Stereo loudspeaker | ✗ | Preference-based listening test |
| Fazenda and Drumm (2013) [10] | Stonehenge | Balloon bursts, ESS | ✗ | WFS, FOA | ✗ | ✗ |
| Pedrero et al. (2014) [11] | Pre-Romanesque churches | RAVEN | ✗ | Binaural | ✗ | ✗ |
| Postma and Katz (2015) [12] | Abbey church | CATT-Acoustic, ESS | ✗ | ✗ | ✗ | Objective parameters |
| Elicio and Martellotta (2015) [13] | Orthodox church | ESS | SHD | ✗ | ✗ | Objective parameters |
| Frank and Zotter (2016) [14] | Church | ✗ | SDM, SHD | Ambisonics, Loudspeaker array | ✗ | Listening tests |
| Postma et al. (2016) [15] | Abbey church, cathedral | CATT-Acoustic | SHD | Ambisonics, Binaural | Head Tracking, Max/MSP | Objective parameters |
| Postma and Katz (2016) [16] | Abbey church | ESS, CATT-Acoustic | ✗ | Binaural | ✗ | Comparative listening tests |
| Alvarez-Morales et al. (2017) [17] | Cathedral | ESS, CATT-Acoustic | ✗ | Binaural | ✗ | Perceptual attribute ranking |
| Martellotta et al. (2018) [18] | Cathedral and Crypt | ESS | SHD | ✗ | ✗ | Objective parameters |
| Suárez et al. (2018) [19] | Hypostyle mosque | ESS, CATT-Acoustic | ✗ | Binaural | ✗ | Objective parameters |
| Canfield-Dafilou et al. (2019) [20] | Catholic church | All-pass chirp | ✗ | Loudspeaker array | FCR | Interactive evaluation |
| Katz et al. (2020) [21] | Cathedral | ESS, Starter pistol, Balloon bursts | SDM | ✗ | ✗ | ✗ |
| Álvarez-Morales et al. (2020) [22] | Cathedral | ESS, ODEON | SHD | ✗ | ✗ | Objective parameters |
| Autio et al. (2021) [23] | Abbey | ODEON | ✗ | Binaural | ✗ | Objective parameters |
| Eley et al. (2021) [24] | Cathedral | CATT-Acoustic | SHD | Ambisonics, Binaural, Loudspeaker array | FDN, NUPC, Max/MSP | Interactive evaluation |
| Warusfel and Emerit (2021) [25] | Egyptian temple | ESS | SHD | Ambisonics, Binaural | Max/MSP | ✗ |
| Berger et al. (2023) [26] | Catholic church and mosque | ✗ | SHD | Ambisonics, Binaural | VR 6-DOF Head-tracking | ✗ |
| Mullins and Katz (2023) [27] | Cathedral | CATT-Acoustic | ✗ | Binaural | Partitioned convolution, Max/MSP | Interactive, Preference |

the time. The study found that the calibrated acoustic model provided a plausible representation of the acoustics of the space as it would have sounded in the 17th century [12].

The acoustics of the Orthodox church of San Nicola in Bari, Italy, was studied by Elicio and Martellotta [13]. The acoustics were measured at different positions in the church using the ESS method with a dodecahedron loudspeaker and an *Eigenmike 32* spherical microphone array. The measurements were analyzed in terms of key acoustic parameters. A spatial analysis of the sound field was performed using higher-order Spherical Harmonic Decomposition (SHD) to identify the spatial distribution of the sound energy in the church, and creating a directional reflection map of the church [13].

Frank and Zotter presented a study in 2016 that examines spatial impression and directional resolution of synthesized sound fields reproduced with an Ambisonic loudspeaker playback system [14]. The RIRs utilized in this study were measured in St. Andrew's church, Lyddington, UK, and sourced from the openAIR database [29]. The study employed the Spatial Decomposition Method (SDM) to enhance the first-order B-format RIRs to higher spherical harmonic orders. Acoustic reproduction was achieved through Ambisonic decoding of the spherical harmonics to a loudspeaker array. The reproduction was conducted for the original first-order, the SDM-enhanced first-order, the SDM-enhanced third-order, and the SDM-enhanced fifth-order representation of the sound field. A perceptual evaluation of the auralizations using the different representations was performed with two listening tests. The results suggest that higher-order Ambisonics playback provides a larger sweet spot with a more convincing reverberant sound field in terms of envelopment and depth [14].

A study by Postma et al. [15] presented a performance by the Conservatoire de Paris using virtually reconstructed acoustics of the Cathédrale Notre-Dame de Paris. The virtual acoustics were reconstructed based on a geometrical model of the cathedral using the CATT-Acoustic software, and calibrated using acoustic measurements made in the cathedral. The performance was auralized using binaural Ambisonic synthesis and head tracking to provide an immersive experience of the acoustics of the cathedral. The study found that the virtual acoustics provided a realistic representation of the acoustics of the cathedral [15].

The achievable perceptual quality of room acoustic simulations and auralizations has been studied by Postma and Katz [16]. Binaural auralizations were created from simulations of the acoustics of an abbey church Saint-Germain-des-Prés, the Cathédrale Notre-Dame de Paris, and the Théâtre de l'Athénée in Paris, France. The acoustic simulations were calibrated from ESS measurements of the spaces. The study carried out subjective listening tests to compare the perceptual similarity of simulated and measured auralizations. The participants were asked to rate the similarity of the auralizations using eight predefined perceptual attributes.

Álvarez-Morales et al. explored the use of Virtual Reality (VR) to create an immersive acoustic reconstruction of the Royal Chapel of the Gothic Cathedral of Seville, Spain [17]. A simplified geometric model of the Cathedral of Seville and the Royal Chapel was created in Computer Aided Design (CAD) software, and used to simulate the acoustics of the space using CATT-Acoustic. The acoustic simulation was calibrated using acoustical measurements made in the Royal Chapel using the ESS method. A VR auralization system was developed that combined the acoustic auralization with a static visual representation of the space. The acoustic auralizations were performed using binaural synthesis, and the visual representation of the space was displayed using a projector. A preliminary listening test was carried out to perceptually evaluate the experience of the VR auralization system. The perceptual evaluation was conducted using a questionnaire that evaluated the realism of the acoustic auralization and the visual representation of the space. The participants had to rank the acoustic experience for different stimuli [17].

In Martellotta et al. [18], an acoustic analysis of the Cathedral of Cadiz and the underlying crypt, which is acoustically connected to the main cathedral, is presented. The crypt, with its central vaulted rotunda connected to radial galleries and openings to the main cathedral, exhibits a combination of acoustic phenomena like flutter echoes and non-diffuse sound propagation [18]. The acoustics of the cathedral and the crypt were measured using a first-order spherical microphone with the ESS method. The measurements were analyzed by calculating the reverberation time and Early Decay Time (EDT) for different octave bands, and Bayesian analysis was performed to determine the decay rate of the RIRs. The sound field distribution was visualized using directional intensity maps. The study found that the crypt's geometry leads to a non-diffuse sound field in the rotunda, with reflections between the floor and dome causing flutter echoes and longer reverberation times [18].

Suárez et al. presented a study on the virtual acoustic reconstruction of the hypostyle mosque of Córdoba, located in Spain [19]. An acoustic simulation was made from geometric models of the mosque using CATT-Acoustic software. The acoustic simulation was calibrated using acoustical measurements made in the mosque using the ESS method. The calibrated acoustic model was used to simulate binaural auralizations of the mosque [19].

A method for studying the relationship between musical composition and room acoustics of the Chiesa di Sant'Aniceto in Rome, Italy, was presented by Canfield-Dafilou et al. [20]. The study used a six-second all-pass chirp signal to measure a set of RIRs of the Catholic church at positions where the singers were likely to have been located. The processing of the RIRs consisted of removing the direct sound, and generating sets of RIRs with different reverberation times. An interactive auralization system was developed using a loudspeaker array and the Feedback Canceling Reverberator (FCR) method [30] to reproduce the acoustics of the church in real-time with different reverberation times. Participants evaluated the auralizations by first interacting with the auralization system by performing a piece of vocal music for all reverberation times, and then by evaluating the auralizations using questionnaires, which evaluated the preference of the participants for the different reverberation times [20].

Following the destructive fire that occurred in Cathédrale Notre-Dame de Paris in 2019, the acoustic properties of the cathedral were found to have been altered by comparing pre-fire acoustic measurements from 1987 and 2015 to post-fire measurements made on the construction site in 2020 [21]. The 1987 measurements were recorded using balloon bursts as the excitation signal, while the 2015 measurements were recorded at roughly the same location using sine sweeps as the excitation signal [21]. The novel post-fire measurements were made using a 20-second-long sine sweep, which was played back through a loudspeaker and recorded at the several receiver positions using a microphone, as detailed in [21]. These measurements were complemented by measurements made using starter pistol gun shots and balloon bursts. After the measurements were made, a spatial analysis of the sound field was performed using SDM to identify the spatial distribution of the sound energy in the cathedral [21]. The study found that there was a significant decrease in the reverberation time of the cathedral after the fire.

Álvarez-Morales et al. presented an acoustic survey of the York Minster's Chapter House, located in York, UK. The acoustics of the Chapter House were measured at several positions using the ESS method and captured using a Soundfield ST450 microphone, and a binaural dummy head. A geometric model of the space was created using CAD software, and the acoustics of the space were simulated using ODEON software. The simulated acoustics were calibrated using the acoustical measurements made in the Chapter House. A perceptual analysis of the simulations was performed by calculating objective acoustic parameters [22].

Autio et al. [23] performed a historically accurate virtual acoustic reconstruction of the Vadstena Abbey church during a 15th-century service. The church's original acoustics were lost due to renovations. A three-dimensional model of the church was created based on historical sources, and ART was used to simulate the acoustic model in ODEON software. To improve the quality of the simulated acoustics, the absorption coefficients of several surfaces were calibrated using acoustic measurements made in the church. The virtual acoustic auralization was performed using binaural synthesis [23]. The study found that the virtual acoustic reconstruction provided a historically accurate representation of the acoustics of the church during a 15th-century service [23].

A study by Eley et al. presented an interactive real-time auralization system for reproducing Cathédrale Notre-Dame de Paris' acoustics using real-time binaural and multichannel loudspeaker synthesis [24]. The aim of the study was to perceptually validate the virtual reconstruction of the cathedral's acoustics prior to the 2019 fire. Real-time auralization of acoustics was achieved by reproducing the cathedral's acoustic RIRs in real-time using a hybrid approach which combined non-uniform partitioned convolution for reproducing the early reflections, and Feedback Delay Network (FDN) for reproducing the late reverberation. Four singers, members of a medieval ensemble with experience performing in the Cathédrale Notre-Dame de Paris, were asked to validate the reproduction system by performing excerpts of medieval repertoire in the virtual reproduction space in both binaural and multichannel loudspeaker configurations. The singers' feedback was collected using questionnaires which evaluated the similarity of the virtual acoustics to the real acoustics of the cathedral [24]. It should be noted that the participants could only base their comparison on their memory of the cathedral's acoustics [24], which affects the reliability of the results.

Warusfel and Emerit [25] presented a study on the acoustic assessment of the Dendara temple, located in Egypt. The acoustics of the temple were measured at several positions using the ESS method, and captured using an *Eigenmike 32* spherical microphone array. A spatial analysis was performed using fourth-order SHD to identify the energy decay deviation of the sound energy in the temple. The measurements were used to auralize the acoustics of the edifice using real-time binaural Higher-Order Ambisonic (HOA) synthesis using the Max/MSP software. In future research, the measurements will be used to calibrate a geometrical acoustic model of the temple to simulate the acoustics of the space as it would have sounded in antiquity [25].

In [26], Berger et al. created a VR auralization system for exploring the acoustics of the Jaame-Abbasi mosque (Shah mosque) in Isfahan, Iran, and the Santa Maria dei Derelitti church in Venice, Italy. The acoustics of the spaces were measured using a first-order Ambisonic microphone. These measurements, together with a three-dimensional model of the spaces, were used to create a VR auralization system that allowed users to explore the acoustics of the spaces in real-time using a six-Degrees of Freedom (DOF) head-tracking VR system. The study found that the VR auralization system provided an immersive experience of the acoustics of the spaces, and allowed users to explore the acoustics of the spaces from different perspectives [26].

To better understand the relationship between ancient music performance practices and the acoustics of the performance space, an immersive binaural auralization system was developed and presented by Mullins and Katz in 2023 [27]. The system was designed to provide a binaural auralization of the Cathédrale Notre-Dame de Paris' acoustics during choral performances. A set of binaural RIRs, which model the acoustics of the cathedral, were simulated using CATT-Acoustic. The binaural synthesis used a set of HRTFs to achieve realistic spatialization of the reproduced sound. The auralization system was designed to be used in real-time by efficiently convolving the binaural synthesis filters with the input signal using partitioned convolution methods. A calibration process ensured that the latency and energy ratios of the auralization system were within acceptable limits. A subjective analysis in the form of an interactive choir performance was conducted to evaluate the perceptual quality of the auralization [27].

In the following sections, we provide a detailed overview of the research conducted in HWS, focusing on key components of the auralization process: room acoustic acquisition, sound field analysis, sound field synthesis, real-time auralization and perceptual evaluation. Each section provides a general overview of the research methods, after which a detailed focus is given to research that is relevant or has been applied to the acoustic preservation and reproduction of HWS.

## 3. Room acoustic acquisition

### 3.1. Methods

The acoustic properties of a room are determined by its geometry, surface materials, and the presence of sound-absorbing objects. From a system identification perspective, the acoustics between a sound source and a receiver positioned within the enclosure can be modeled as a Linear Time-Invariant (LTI) system, assuming geometric and atmospheric conditions in the room remain constant over time [31,32]. The behavior of an LTI system is characterized entirely by its Impulse Response (IR). For acoustics in a room, this is referred to as the Room Impulse Response (RIR), which represents the acoustic response of the room to an impulsive sound source, captured at a receiver position. A RIR $h(t, \mathbf{x}, \mathbf{r})$ is a function of time $t$, the source position $\mathbf{x}$, and the receiver position $\mathbf{r}$, and can be modeled as the sum of all acoustic events in the room [33]:

$$h(t, \mathbf{x}, \mathbf{r}) = \sum_{p=0}^{\infty} h_p(t, \mathbf{x}_p, \mathbf{r}), \tag{1}$$

where $h_p(t, \mathbf{x}_p, \mathbf{r})$ is the impulse response of the $p$th event, where $\mathbf{x}_p$ denotes the $p$th event's position.

In order to measure the acoustics of a given space, it is necessary to obtain the RIRs between relevant source and a receiver positions within the space. In spaces that are accessible to researchers, the RIR between a source and receiver position can be acquired by measuring the response of the room to an impulsive sound source, following ISO 3382 [34,35]. For inaccessible or acoustically altered spaces, the RIR can be estimated from the room's geometry and surface material properties using simulation methods, such as the Image-Source Method (ISM), Acoustic Ray Tracing (ART), numerical wave-based methods, statistical methods, or using a combination of those. This section provides an overview of the methods used to acquire RIRs in accessible and inaccessible spaces, and discuss the challenges associated with acquiring RIRs in HWS.

#### 3.1.1. Room impulse response measurements in accessible spaces

The RIR in spaces that are still accessible can be acquired by exciting the room with an omnidirectional impulsive sound source, and capturing the response at a receiver position [36]. Prior to the advent of digital signal processing, the most common methods for acquiring RIRs were typically conducted through the use of hand claps, balloon pops [37–39], high-voltage sparks [40], or by firing a starter pistol [37,40] in the room, and recording the acoustic response using a tape recorder. However, these methods lack repeatability and exhibit poor omnidirectionality, which can lead to inaccuracies in the measured RIRs [37].

To achieve accurate and repeatable RIR measurements, a space can be excited with a known signal $s(t)$, that is played back through a loudspeaker system. The acoustic response $y(t, \mathbf{x}, \mathbf{r})$ can then be captured at a receiver position $\mathbf{r}$ using a microphone setup [36]. Crucially, the excitation signal $s(t)$ should exhibit an autocorrelation function that is a Dirac delta function, i.e., $s(t) * s(-t) = \delta(t)$, where $\delta(t)$ is the Dirac delta function and $(*)$ denotes the convolution operator. The RIR can then be estimated by convolving the response signal with the time-reversed excitation signal, as

$$h(t, \mathbf{x}, \mathbf{r}) = y(t, \mathbf{x}, \mathbf{r}) * s(-t). \tag{2}$$

A common excitation signal that meets the above criteria is the Maximum Length Sequence (MLS) signal, proposed for measuring room acoustics by Schroeder [41], and further researched in [42–46]. MLS is a pseudo-random binary sequence that is generated using maximal linear-feedback shift registers. MLS signals are periodic, have a flat frequency spectrum, and a circular autocorrelation equal to a Kronecker delta function [47], which is the discretized version of the Dirac delta function, making it a suitable excitation signal for RIR measurements when the MLS period is longer than the reverberation time of the room [41]. The correlation between the MLS signal and the response signal can be efficiently computed in the frequency domain using the Fast Walsh–Hadamard Transform (FWHT) algorithm [31,42]. The main advantage of MLS is the high Signal-to-Noise Ratio (SNR) and robustness to impulsive and non-stationary noise sources [46,48]. However, the main disadvantage of MLS signals is the significant non-linear distortion that occurs when the signal is played back through a loudspeaker system. This distortion is caused by loudspeaker non-linearities, which can affect the accuracy of the RIR measurements [31,44,49].

To mitigate the distortion caused by the non-linear characteristics of loudspeakers when playing back MLS signals, sine sweeping excitation signals have been proposed. Sine sweep signals offer a higher SNR as compared to MLS signals and a complete rejection of harmonic distortion coming from the loudspeaker system [37,50–53]. However, sine sweep measurements are sensitive to impulsive noise sources, which, when present, introduce artifacts in the measured RIRs [37,50–53]. Sweeping sines are generated by modulating a sine wave with a time-varying phase $\varphi(t)$ [37,50,51,54], i.e.,

$$s_{\mathrm{sw}}(t) = \sin(\varphi(t)). \tag{3}$$

The derivative of the phase denotes the instantaneous frequency of the sweeping sine signal, $f_{\mathrm{i}}(t) = \frac{d\varphi(t)}{dt}$, which, for room acoustic measurements, can be either linearly increasing [51,54] or exponentially increasing over time [50,51,55]. The latter is known as the Exponential Sine Sweep (ESS) signal, which has been shown to have a higher SNR in the lower frequency range as compared to linearly increasing sine sweeps [50], a desirable property for acoustic measurements. The length of the sine sweep signal should be chosen such that the risk of impulsive noise being present in the recording is minimized, while ensuring that a desired SNR is achieved [53,55]. Therefore, the choice of excitation signal for RIR measurements depends on specific requirements, such as the desired SNR, the mitigation of impulsive noise sources, and the desired accuracy of the measured RIRs.

A recently proposed method, called Mosaic, uses across-measurement median filtering to recover an impulsive noise-free sweep signal when the sine sweep signal is recorded multiple times [56]. The Mosaic method is based on the assumption that the impulsive noise sources are uncorrelated across measurements, and that the RIR is stationary across measurements. The method has been shown to be effective in removing impulsive noise sources from RIR measurements, and can be used to improve the accuracy of RIR measurements contaminated with impulsive noise [56].

When information about the noise sources in the room is available, [57] proposed a method for designing an all-pass chirp signal that allows for the estimation of the RIR with a constant SNR across the frequency range of interest. By designing the all-pass chirp signal such that more energy is present at noisy frequency bands, the SNR remains constant [57].

In [58], the authors provide guidelines for the measurement of RIRs in historical buildings, specifically churches. The aim of the study was to simplify and normalize the measurement process of RIRs in historical buildings, and to improve the comparability of RIR measurements in different buildings. The authors discuss the choice of source positions, receiver positions, and measurement equipment, which are important factors that affect the accuracy of the measured RIRs.

*3.1.2. Sound field reconstruction in inaccessible spaces*

In spaces that are inaccessible, partially accessible, or acoustically altered, acoustic measurements are not feasible or desired. In such cases, a 3D digital model of the room can be created in Computer Aided Design (CAD) software from the room's geometry and surface materials. The geometry of the room can be obtained from architectural drawings, paintings, historical documents [59], or from a 3D laser scan in case the room is still partially accessible [60,61]. In order to create a realistic acoustic model of the room, the acoustic properties of the room's surfaces need to be modeled by assigning absorption and scattering coefficients [6]. This information can be obtained from tables for common building materials or from free-field measurements [6]. However, the accuracy of this data is limited, leading to inaccuracies in the simulated RIRs [6]. For accessible or partially accessible spaces, the surfaces of a 3D model are assigned a rough approximation of the absorption and scattering coefficients. During the calibration phase, these parameters are fine-tuned to match measured acoustic properties of the room to within one Just Noticeable Difference (JND) [62].

The RIR between a source and receiver position can be estimated from the room's geometry and surface material properties using simulation techniques. In literature, these techniques can be divided into two main categories, namely: wave-based methods and geometrical acoustics methods [63]. Wave-based methods, such as the Finite-Difference Time-Domain method, the Finite-Element method, and the Boundary Element Method, numerically solve the wave equation in the room to estimate the RIR. These methods provide the most accurate results, but are computationally expensive, limiting their use for large and complex geometries [64]. However, as Graphics Processing Unit (GPU)-based implementations of wave-based methods can significantly reduce the computation time required for simulating large room acoustics [63,65,66], these methods are becoming more accessible for simulating room acoustics in large and complex geometries. Geometrical acoustics methods, such as the ISM and ART, are based on the assumption that sound waves propagate in straight lines and are reflected off the room's surfaces. In [64], an extensive overview of geometrical acoustics methods for simulating room acoustics is provided. In the remainder of this section, we will focus on the image-source and ray-tracing methods, which are commonly used for estimating RIRs in simulated environments.

The ISM was first proposed in the context of simulating rectangular room acoustics by Allen and Berkley [67], and later extended to spaces with arbitrary polyhedral geometry [68]. ISM assumes ideal specular reflections from the room's surfaces, and models the room as an infinite number of image sources, which are virtual sources that are mirrored across the room's surfaces. The RIR is then estimated as the sum of the contributions from all image sources in the room. ISM can be used to accurately estimate RIRs in small to medium-sized rooms with simple geometries [67]. For rectangular shaped rooms with ideal specular reflections, ISM is a solution to the wave equation [67]. However, when modeling highly symmetrical spaces using ISM, such as perfectly rectangular rooms, care should be taken to avoid sweeping echoes [69] that occur due to the regularity with which sound waves are reflected in the room. This effect is known to occur even in real rooms, such as at the staircase of the El Castillo pyramid [70,71]. To prevent this unwanted effect from occurring in room acoustics simulations, a small random offset can be introduced to each image source position [69].

When modeling the acoustics of arbitrary geometries using ISM, the same principle is applied as for rectangular rooms, but only image sources that are visible from the receiver position are considered in the RIR estimation [68]. Visibility checks need to be performed for each image source in the room, which can be computationally expensive for complex geometries. In [72], binary space partitioning trees are introduced as a method to efficiently determine the visibility of image sources in arbitrary geometries. The method is shown to be effective in reducing the computational cost of ISM simulations for complex geometries. However, the method is limited to static setups, and does not account for moving sound sources or receivers in the room.

Hybrid methods for simulating room acoustics exist which combine different simulation techniques to estimate RIRs for complex geometries [64]. For example, the ISM and ART, which will be further discussed below, was combined by Vorländer in 1989 [73], and this approach is used in the commercially available ODEON software package [74]. The hybrid method uses ART to discover the location of image-sources that are visible from the receiver position, as these are the image-sources that contribute to the RIR. For complex geometries, this leads to a significant reduction in the number of image-sources that need to be considered, as the ratio of non-visible to visible image-sources is typically high [73].

The ISM is a deterministic approach to simulating room acoustics that employs only specular reflection paths. An alternative method for simulating room acoustics based on a room's geometry, is stochastic ray tracing, a popular method for simulating light particle propagation in computer graphics. In the context of room acoustics, ray tracing is used to simulate the propagation of sound waves in an enclosure. This approach was first described for simulating room acoustics by Krokstad et al. in 1968 [75,76]. In this method, which we will from now on refer to as ART, sound rays are assigned an initial energy value and are radiated from a source position in the room. The absorption and scattering coefficients are assigned to all surfaces within the room, thereby determining the energy loss and direction of the rays upon reflection off the room's boundaries. The trajectory of the rays is then traced until they reach a desired receiver position within the room. At this position, the energy of each ray is recorded as a function of time. The RIR is then estimated as the sum of the energy contributions from all rays that reach the receiver position. The main advantage of ART over the ISM is the ability to model sound source characteristics, diffuse reflections, and acoustic scattering in complex geometries [6,64]. The accuracy of ART increases with the number of rays that are employed. However, determining the optimal number of rays to use in a simulation is not trivial as it depends on the room's geometry, the desired accuracy of the RIR, and the available computational resources [6,64]. For complex geometries, the number of rays required to accurately simulate the room's acoustics used to be prohibitively high, which leads to long simulation times. However, with the advent of modern computing hardware, such as GPUs, the computational time required for ART simulations has been significantly reduced [66]. Under real-time conditions, ART can be used at the cost of reduced accuracy by using fewer rays in the simulation [6]. In [77], an iterative ART method was proposed for simulating room acoustics in real-time. The method uses a reduced number of rays in the simulation and refines the simulation iteratively by adding more rays until a desired accuracy is achieved. Preliminary results show the method to be effective in simulating room acoustics in real-time, with a reduced computational cost as compared to traditional ART methods [77].

As previously stated, ART requires a high number of individual rays to accurately simulate the acoustics of a space. To reduce the computational cost further and to increase the efficiency of ART simulations, volumetric rays, also called beam tracing, can be used [64,78]. The main principle of beam tracing is to expand the rays into volumetric beams that are traced through the room until they reach the receiver position. The beams are then split into smaller beams when they encounter two or more surfaces in the room, which allows for a more efficient simulation of the room's acoustics as compared to traditional ART methods. For an overview of various ray and beam tracing methods and their application to room acoustics, we refer the reader to [64]. Additionally, a comparison of acoustic simulation and measurement methods for HWS can be found in [79].

Commercially available software to simulate room acoustics using ISM and ART include Epidaure [80,81], ODEON [74], CATT-Acoustic [82], Ramsete [83], EASE [84], and Treble [85]. These software packages are widely used in the field of room acoustics for

simulating room acoustics in complex geometries, and are used in the design of concert halls, theaters, and other performance spaces.

A notable recent development in room acoustics estimation is the use of machine learning and deep learning methods to predict RIRs [86–88]. These methods rely on large datasets of measured RIRs to train models that predict the RIR from a room's geometry and surface material properties. An advantage of machine learning methods is that they can be used to predict additional room acoustics parameters directly from the room's geometry [89] or captured audio [90]. Additionally, the prediction of RIRs and sound fields by means of generative artificial intelligence methods is an emerging field, in which the two most promising models are Generative Adversarial Networks (GANs), e.g. as used in Mesh2IR [91], and Neural Acoustic Fields (NAFs) [92]. However, the accuracy of machine-learning methods is limited by the quality and quantity of the available training data. To the best of our knowledge, these methods have not been applied yet to the prediction of HWS acoustics.

### 3.2. Acoustic acquisition of historical worship spaces

In HWS, the acoustics of the space are of particular interest due to the unique architectural design of the spaces, which can have a significant impact on the sound field within the space. In spaces that remain unaltered and accessible, acoustic measurements can be conducted to preserve and analyze the acoustics of the space [93,94]. Before current methods for measuring RIRs were developed, the reverberation time was measured by exciting the space with 1/6 octave band filtered white noise signals, and measuring the acoustic decay pattern [95,96]. A study from 1971 measured the reverberation time of two Basilicas in Rome at single frequencies using an organ pipe as a sound source [97]. In [98], RIR measurements of a mosque in Ankara, Turkey, were conducted by exciting the space with broadband white noise.

The MLS method has been used to measure the acoustics of various HWS, including Catholic churches in Poland [99,100], 21 mosques in Saudi Arabia [101], nine Romanesque churches in Southern Italy [102,103], 12 Mudejar-Gothic churches in Seville, Spain [104,105], six churches in Switzerland and Portugal [106], and the Cathedral of Benevento in Italy [107].

Sine sweep methods, which offer a higher SNR and complete rejection of harmonic distortion from the loudspeaker system, have been used extensively to measure the acoustics at various positions in HWS. The method has been used to measure the acoustics in churches [9,17,108–117], cathedrals [18,21,22,118–120], Buddhist temples [121], Egyptian temples [25], mosques [19,122–124], and historic monuments such as Stonehenge [10,125].

Alternative methods, such as balloon bursts and starter pistol shots, have been used less frequently due to the lack of repeatability and the potential for inaccuracies in the measured RIRs. However, in spaces with no access to mains power, the use of balloon bursts or starter pistol shots may be the only viable option for measuring the acoustics of HWS [21,39,126–131].

When the acoustics of HWS have been altered over time, when the space is no longer accessible, or when a complete acoustic reconstruction of the space is desired, the methods described in Section 3.1.2 can be used to reconstruct the acoustic sound field of the space from a geometrical model of the room and its surface material properties. This has been done for various HWS including mosques [7,19,98,123,124,132–134], ancient Greek spaces [8,135], churches [23,105,112,114–117,136–144], cathedrals [15,17,22,24,27,107,118–120,145–151], and Buddhist temples [121].

Due to imperfections in the modeled room geometry and surface material properties, the geometrical model only approximates the true space, limiting the accuracy of the simulated RIR. When the HWS is still or partially accessible, the accuracy of the simulated RIRs can be improved by calibrating the model to measured RIRs from the space [12,15–17,19,22–24,107,112,114,117–121,123,124,138,143,146,150–152].

**Table 2**
Available software packages for room acoustic simulation and their use in HWS.

| Software | License | Use in HWS |
|---|---|---|
| ODEON [74] | Commercial | [7,22,23,98,107,114,117,121,123,124,134,135,142,144] |
| CATT-Acoustic [82] | Commercial | [12,15–17,19,24,27,79,105,112,116,118–120,136–139,141,143,145–153] |
| Epidaure [80,81] | Commercial | [79] |
| Ramsete [83] | Commercial | [115] |
| EASE [84] | Commercial | [133] |
| Treble [85] | Commercial | |

In [79], a comparison of various room simulation and measurements techniques for HWS was conducted. The study compared the acoustic parameters of a Basilica in Valencia, Spain, measured using MLS and sine sweep methods, with the simulated acoustics of the space using simulation software CATT-acoustic and Epidaure.

A methodology for studying the acoustic environment of cathedrals was presented in [153]. The authors used both experimental measurements and computer simulations to analyze the cathedral of Malaga as a case study. The acoustic behavior of the cathedral at relevant positions for liturgical activities, including the altar, choir, pulpit, organ, and retrochoir, and for different listener zones within the cathedral were investigated. This study aimed to answer how the sound field is influenced by architectural design of the cathedral and how it impacted various liturgical and cultural activities, including music and speech intelligibility [153].

It should be noted that acoustic measurements of historically valuable spaces other than worship spaces have been carried out in the context of archaeoacoustics research at various locations around the world [154].

Commercially available software packages for room acoustic simulation and their use in simulating the acoustics of HWS are listed in Table 2. The software packages are widely used in the field of room acoustics for simulating room acoustics in complex geometries, and are used in the design of concert halls, theaters, and other performance spaces.

## 4. Sound field analysis

As discussed in the previous section, the RIR fully describes the acoustic wave propagation between a sound source and receiver in an enclosed space. However, the RIR does not provide a spatial representation of the larger sound field in the room. The sound field for a given sound source positioned in the space is the three-dimensional representation of sound pressure and pressure velocity around a sensor position in the room. For simulated room acoustics, the analysis of the sound field is trivial, as the entire sound field can be computed at every point in space. For real-world measurements however, the sound field is not directly accessible and must be inferred from acoustic measurements. The sound field in a space can be analyzed from microphone array RIR measurements, which carry the spatial information of the sound field. Sound Field Analysis (SFA) aims to extract the spatio-temporal characteristics of sound propagation around a sensor array from measurements made in the sound field. This spatial information can be used for room acoustic analysis and design, validation of room acoustic simulations, as well as for auralization of the room acoustics.

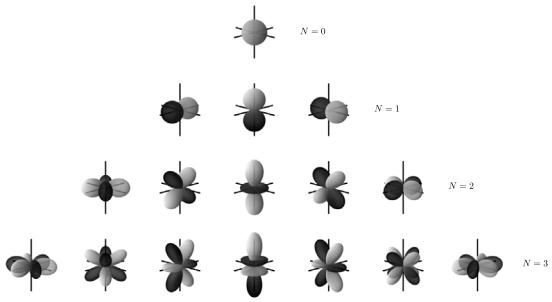



**Fig. 2.** Magnitude of spherical harmonics up to order $N = 3$. Light regions represent positive sign values, while dark regions represent negative sign values. The radius indicates the magnitude of the spherical harmonic.

### 4.1. Methods

SFA is the process of extracting spatial information from the sound field in a space using microphone array measurements. The geometry of the microphone array, the number of microphones, and the microphone spacing are important factors that affect the spatial resolution of the analysis. Depending on the geometry and properties of the microphone array, different spatial analysis methods are used. The most common microphone array geometries for SFA are linear arrays, circular arrays, and spherical arrays. The former two are used for two-dimensional SFA, while the latter can be used for three-dimensional SFA.

One of the earliest methods that enabled the analysis of sound fields in an enclosed space using a linear microphone array is Wave Field Analysis (WFA). This analysis method was presented by Berkhout et al. in 1997 [155]. The method is directly related to Wave Field Synthesis (WFS), a sound field synthesis method that will be further discussed in Section 5. WFA is a spatial analysis method that provides insight into the temporal and spatial structure of the sound field by transforming linear microphone array measurements from the space–time domain to the so-called "ray parameter versus intercept time" domain using the Linear Radon Transform. This transformation decomposes the measured data into a set of traveling plane wave components, which can be used to analyze properties of a sound field, such as diffusivity, lateral energy content, and allows for the discrimination and analysis of reflection and diffraction events in the sound field [156–158]. In [159], WFA was adapted to analyze and simulate bending wave fields of a vibrating steel plate using linear sensor arrays.

An alternative method for analyzing the sound field on a circular or spherical surface within an enclosed space is the Spherical Harmonic Decomposition (SHD) method. The method can be used with circular or spherical microphone arrays for two- and three-dimensional SFA [160, 161]. The SHD method decomposes the sound field captured by the microphone array into spherical harmonic components that form an orthogonal basis on the surface of a sphere. The spherical harmonics can be used to represent the sound field in the room, where the coefficients of the spherical harmonics describe the spatial distribution of the sound field. The SHD method is a powerful tool for SFA and has been used in various applications such as beamforming [162–165], direction of arrival estimation [166,167], plane wave decomposition [168–170], and spatial sound field reproduction through Ambisonics [171,172]. The latter will be further discussed in Section 5. The order of the spherical harmonics determines the spatial resolution of the analysis, with higher-order spherical harmonics providing higher spatial resolution, as shown in Fig. 2 where the magnitude of real spherical harmonics up to order $N = 3$ is visualized. However, higher-order spherical harmonics require a larger number of microphones to be accurately captured, specifically $(N+1)^2$ microphones for analyzing up to order $N$ spherical harmonics [163].

WFA and SHD are wave-based analysis methods which aim to analyze the true physical sound field in the room over an area surrounding the microphone array. In contrast, the Spatial Decomposition Method (SDM) [33] is a parametric method that models the RIR between a source and a microphone array as the sum of a finite number of image sources arriving at the microphone array's geometric center [67]. The method assumes that every sample in the pressure signal recorded at the center of the microphone array, is the result of a single acoustic event, i.e., image source, that can be described by a certain Direction of Arrival (DOA) $\theta(t)$ and reference pressure value $p(t)$. When the DOA is known for every sample in the reference pressure signal, the sound field can be reconstructed by summing the contributions of all localized image sources.

The SDM analyzes the multichannel RIRs captured by the microphone array by applying a sliding short-time windowing function to the measured RIRs. The windowing function is non-zero only for the interval $t \in [\kappa - \frac{L}{2} \dots \kappa + \frac{L}{2}]T$, where $L$ denotes the window length and $T = 1/f_s$ is the sampling period. The window length should be chosen to exceed the time it takes for a wavefront to propagate through the microphone array, and is therefore dependent on the microphone spacing. By sliding the short-time window over the RIRs at every discrete time step $T$, a set of overlapping short-time windows is obtained. The DOA $\theta(t)$ can be estimated for every short-time window position by applying a localization algorithm to each short-time window in the set of RIRs. The localization algorithm can be implemented using the Generalized Cross-Correlation (GCC) [33,173] for open microphone arrays, or using pseudo-intensity vectors for B-format microphone arrays [174–176]. At low sampling rates, the localization performance can be improved using interpolation techniques [177–179]. The reference pressure signal $p(t)$ is measured using an omnidirectional microphone located at the center of the microphone array. In case there is no microphone located in the geometric center of the array, the reference pressure signal can be predicted from the image-source locations [33]. The estimated DOAs and reference pressure values can be used to visualize the directionality of the sound field in the room. This has been shown to be useful for the analysis of sound fields in large spaces, specifically for the analysis of concert halls [180–183], nightclubs [184], music studios [185], car cabins [186], and chamber music hall acoustics [187].

While the SDM is a powerful method for the analysis of sound fields in enclosed spaces, it has some limitations [188]. The main assumption of the SDM is that for every overlapping analysis window, precisely one acoustic reflection, i.e., image source is present. This assumption is not always valid in practice, i.e., when the microphone array is located close to a reflective surface, when the room's geometry is highly symmetrical or when considering the late part of the RIR where the reflection density is high. In these cases, the SDM may not accurately estimate the DOA. In these cases, the analysis window length should be kept to a minimum to reduce the number of overlapping reflections in the analysis window.

### 4.2. Spatial sound field analysis of historical worship space acoustics

It is known that the spatial properties of the sound field in HWS vary significantly across the space due to the complex geometry and surface material properties of these spaces. The use of SFA in HWS allows researchers to analyze the unique spatial properties at different locations within the space, which can be used to better understand the acoustics at choir stalls, organ lofts, and other important locations within the space. The spatial analysis of the sound field in HWS can also be used to evaluate the impact of architectural changes, such as the deployment of sound-absorbing materials, on the spatial distribution of sound energy in the space.

In 2002, de Vries et al. used a circular microphone array for RIR measurements inside a church in Weimar, Germany, to analyze the spatial sound field for the reproduction of choir concert recordings using WFS [108]. The study found that the circular microphone array provided a good spatial resolution of the sound field in the church,

which was essential for the accurate reproduction of the choir concert recordings using WFS.

Spatial SFA based on SHD was used to analyze and visualize the spatial energy distribution of the measured sound field in churches [13,189], cathedrals [18,22], and an ancient Egyptian temple [25]

The spatial energy distribution of a church in Lyddington, UK, was analyzed using the SDM in [14]. In 2017, Amengual et al. used a tetrahedral microphone array to analyze the spatial sound field of the reproduced acoustics of a hall, studio, and church [190] using the SDM. The study compared the spatial resolution of the SDM with IRIS [191], a commercial SFA software, and found that both methods provided similar spatial resolution in simple sound fields. However, in complex and dense sound fields, such as those found in HWS, the higher temporal resolution of the SDM provided a better spatial energy distribution over IRIS. In 2020, Katz and Weber [21] used the SDM to analyze the energy distribution of the sound field in the Cathédrale Notre-Dame de Paris in pre- and post-fire conditions.

## 5. Sound field synthesis

The spatial analysis of the sound field in a room provides valuable information about the spatial distribution of sound pressure and pressure velocity around a sensor position in the room. This information can be used for auralization of room acoustics, as well as for room acoustic analysis and design, as shown in the previous section. Sound Field Synthesis (SFS) is the process of auralizing, i.e., recreating, the sound field in a room such that the listener perceives the sound field as if they were present in the measured space. In this section, we will discuss various SFS methods, including loudspeaker synthesis and headphone synthesis, and how the spatial information obtained from SFA can be used for SFS.

### 5.1. Methods

#### 5.1.1. Loudspeaker synthesis

Loudspeaker synthesis is the process of recreating a sound field of a room using an array of loudspeakers. The loudspeakers are positioned in the listening room such that the sound field is either synthesized at a specific listening position, referred to as the *sweet spot*, or over an extensive listening area which allows for listeners to experience the synthesized sound field as they move around the room. The number of loudspeakers and their positioning in the room are important factors that affect the spatial resolution of the synthesis. While the loudspeaker configuration can be arbitrary, common loudspeaker configurations for SFS are linear arrays, circular arrays, and spherical arrays. The former two are used for two-dimensional SFS, while the latter can be used for three-dimensional SFS.

Loudspeaker synthesis is achieved by constructing synthesis filters from the spatial information of the sound field. The synthesis filters are then convolved with a dry audio signal, such as speech or music to create the synthesized sound field.

A sound field consists of a set of auditory events, see (1), also called virtual sources, that are perceived as coming from a certain direction. Accurate reproduction of sound fields using a set of loudspeakers creates the illusion that the virtual source is coming from a specific direction, regardless of the loudspeaker positioning in the reproduction setup. A common method to achieve this is by mapping the virtual sources to the loudspeaker array in a way that the sound source is perceived as coming from a specific direction. A straightforward method, called Nearest Loudspeaker Synthesis (NLS) achieves this by mapping each virtual source to the loudspeaker that is closest to the desired direction of the virtual source [186]. This method works well for loudspeaker arrays containing many loudspeakers, but is not suitable for loudspeaker arrays with few loudspeakers as the spatial resolution of the synthesis is limited by the loudspeaker spacing. To achieve higher spatial resolution in the synthesis, the virtual source can be positioned between loudspeakers in the array. This can be achieved using amplitude and/or time-delay differences between the loudspeakers, a technique known as panning [192]. Panning is a common technique used in audio production to position a sound source in a stereo or multichannel loudspeaker setup. The simplest form of panning in a stereo loudspeaker setup is amplitude panning, where the amplitude of the sound source is varied between the left and right loudspeakers to create the illusion that the sound source is coming from a specific direction. In a multichannel loudspeaker setup, amplitude panning can be extended to Vector Base Amplitude Panning (VBAP) [193,194] which is a generalized amplitude panning technique for positioning a sound source between apertures of loudspeaker triplets. In VBAP, the loudspeaker directions for a loudspeaker triplet are represented by three-dimensional unit vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$, and the desired sound source direction is represented by a three-dimensional vector $\mathbf{v}_s$. The sound source is then panned to the loudspeaker triplet by calculating the gains $g_1$, $g_2$, and $g_3$ for each loudspeaker in the triplet using the following equation [193],

$$[g_1, g_2, g_3] = \mathbf{v}_s^T \mathbf{V}^{-1}, \tag{4}$$

where $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]^T$ is the matrix of loudspeaker directions. The gains $g_1$, $g_2$, and $g_3$ are then used to drive the loudspeakers in the triplet. VBAP is a flexible and efficient method for panning sound sources to arbitrary loudspeaker layouts. In addition, the loudspeaker gains are normalized to ensure that the sound source is not amplified or attenuated during panning [193,194]. A shortcoming of intensity panning in VBAP is the nonuniform spatial spread of the panned virtual source [195]. To address this, Multiple-Direction Amplitude Panning (MDAP) was proposed by Pulkki in 1999 [195], which is a method that uses multiple loudspeaker triplets to pan a sound source to a specific direction. The method ensures that the panned virtual source is uniformly spread over the loudspeaker array [195].

Panning techniques can be used to position virtual sources around a listener in a room using loudspeaker arrays. In Section 4, the SDM was discussed as a parametric method to analyze the sound field in a room using microphone array measurements by decomposing the sound field into a set of image sources described by their DOA $\theta(t)$ and reference pressure value $p(t)$. From this, the sound field can be reproduced by positioning the virtual sources around the listener using panning techniques such as VBAP or NLS [33,186].

In [196], it was shown that the SDM in combination with amplitude-based panning techniques can decrease the perceived brightness of the reproduced sound field. Therefore, a trade-off between the perceived brightness of the sound field and the spatial resolution of the synthesis must be considered when using panning techniques for SFS.

In 1988, Berkhout introduced the concept of WFS [197], a holographic SFS method which reproduces sound fields that are perceptually indistinguishable from the measured sound field. WFS was introduced as an auralization method for room acoustics in 2004 by de Vries and Hulsebos [198]. The basis of WFS can be found in Huygens' principle, which states that a wavefront can be considered as a continuous distribution of secondary spherical wave fronts. The wavefront can be synthesized during reproduction by summing the contributions of all secondary sources [197,199] on the boundary surface. Mathematically, WFS is based on the Kirchhoff-Helmholtz integral, which states that, when the sound field on the boundary surface $S$ of a closed, source-free volume $V$, is known in terms of pressure and normal particle velocity, the sound pressure can be calculated at any point in the volume $V$ [197,199]. The sound field information on the boundary surface $S$ can be obtained by measuring the sound field using microphone arrays, similarly to WFA. In practice, due to a finite number of microphones and loudspeakers that can be used during the analysis and reproduction, the distribution of secondary sources on the surface is not continuous, but discrete. This discretization of the boundary surface introduces spatial aliasing and truncation artifacts in

the synthesized sound field. The spatial aliasing artifacts are caused by the finite loudspeaker spacing, which limits the frequency range that can be accurately reproduced by WFS, known as the spatial aliasing frequency [200]. The spatial aliasing frequency is dependent on the loudspeaker spacing $\Delta x$ and the speed of sound $c$ in the medium, and is given by,

$$f_{\max} = \frac{c}{2\Delta x}. \tag{5}$$

To avoid spatial aliasing, the loudspeaker spacing must be smaller than half the wavelength of the highest frequency in the sound field that is to be reproduced. The truncation artifacts are caused by the finite size of the loudspeaker array. A discussion and analysis of the spatial aliasing and truncation artifacts in WFS can be found in [201–204].

In [205], three two-dimensional microphone array configurations were considered for WFS, namely: a linear array, a cross array, and a circular array. This study found that the circular array configuration provided the best performance in terms of homogeneous spatial resolution and artifact-free reconstruction area for auralization using WFS [205]. While circular and three-dimensional microphone arrays can be used for WFA, no literature was found that specifically addresses the use of WFS with three-dimensional microphone arrays. A method to extract three-dimensional sound field information from two-dimensional microphone array measurements using WFS was, however, proposed by [206].

Recently, Amengual Garí presented an exploratory study that combined the SDM analysis and WFS for SFS [207]. The study used the SDM to decompose the sound field in a room into a set of image-sources that were reproduced using WFS by modeling the image-sources as either point sources or plane waves. When modeling the image-sources as plane waves, the synthesized sound field showed good agreement with the measured sound field. However, further research is still needed to evaluate the perceptual performance of the combined SDM and WFS method for SFS [207].

In Section 4, the SHD method was discussed as a spatial analysis method to analyze the sound field in a room by representing the sound field as a sum of spherical harmonic components. The spatial representation that is retrieved from SHD can be used for SFS using Ambisonic or Higher-Order Ambisonic (HOA) synthesis [171,172]. By decoding the spherical harmonic components of the sound field to a loudspeaker array using Ambisonic decoding equations [208,209], the sound field can be reconstructed. For first-order Ambisonic reproduction, a well-known method for decoding the spherical harmonic components to arbitrary and irregular loudspeaker layouts, was presented by Gerzon in 1992 and referred to as Vienna decoders [210]. For higher-order Ambisonic decoding, All-Round Ambisonic Decoding (ALLRAD) was presented by Zotter and Frank in 2012 [211]. This method combines Ambisonic decoding with VBAP to provide a flexible and efficient method for decoding higher-order Ambisonic signals to arbitrary loudspeaker layouts by first decoding the spherical harmonic components to a virtual loudspeaker array, after which the virtual loudspeaker signals are panned to the physical loudspeaker array using VBAP.

A study by Frank and Zotter [14] in 2016 investigated the dependency of the spatial impression of synthesized reverberant sound fields on the directional resolution using HOA synthesis [14]. The original sound field was measured using a first-order Ambisonic microphone array, and the SDM was used to create first-, third-, and fifth-order Ambisonic representations of the sound field. The Ambisonic representations were then decoded to a loudspeaker array and the synthesized sound field was presented to listeners in a listening room. The study found that auralizations using higher-order Ambisonic representations increased the perceptual sweet spot of the synthesized sound field, and provided well-focused direct sound [14].

Ambisonic and WFS are both holophonic, wave-based sound field reproduction methods, meaning that they aim to physically reconstruct the measured sound field in a room. While the two methods are based on different principles, it has been shown that WFS and HOA are equivalent under certain conditions as they are both able to provide an exact solution to the wave equation [212,213].

#### 5.1.2. Headphone synthesis

Headphone synthesis, also known as binaural synthesis, is the process of recreating the sound field of a room for an individual listener by means of headphones. The main advantage of using headphones for SFS is that the reproduced and listener's acoustic environment can be decoupled, which allows for a more controlled listening experience. However, headphone synthesis also presents some challenges, such as the feeling of being closed off from the environment, and in-head localization, which is the perception of sound sources coming from inside the listener's head, leading to unnatural listening experiences [6]. To prevent in-head localization and allow for spatialization, the sound field must be synthesized by taking into account the listener's individual Head-Related Transfer Functions (HRTFs). The HRTF is the direction-dependent transfer function describing the filtering effect of the listener's head, shoulders, torso, and pinnae on a sound wave traveling from free field to the listener's eardrum [214]. The HRTF is unique to each listener, and therefore, personalized HRTFs are required for accurate headphone synthesis.

In practice, obtaining personalized HRTFs is done by placing a probe microphone in the listener's ear canal and measuring the acoustic transfer function from a sound source placed in the free field to the probe microphone. This process is repeated for both ears and for a wide range of sound source directions. While these measurements provide the most accurate HRTFs for a listener, the process is costly and time-intensive [6]. Therefore, publicly available HRTFs are often used for headphone synthesis. Commonly used HRTF databases are the *CIPIC* database, which contains individual HRTFs of 45 subjects [215], and the *ARI* database, which contains HRTFs of more than 250 subjects [216]. For a detailed comparison of different HRTF databases, see [217].

In real acoustic environments, the listener's head orientation can change freely, influencing the directional perception of the sound sources in the acoustic environment. To replicate this realism, and enhance the listening experience, the listener's head orientation must be tracked during binaural synthesis. This tracking allows virtual sound sources to adjust their positions relative to the listener's head, ensuring that the spatial audio experience remains consistent and immersive, even as the listener moves their head. This dynamic adjustment effectively fixes the virtual sources around the listener's head, maintaining their perceived locations regardless of the listener's head orientation.

Head tracking in 6-Degrees of Freedom (DOF) can be achieved through various methods, including electromagnetic, camera-based, and optical tracking [218,219], as well as through the use of Inertial Measurement Units (IMUs) that integrate inertial sensors such as accelerometers, gyroscopes, and magnetometers, which are mounted on the listener's head or headphones [220–222]. For a comprehensive review of head tracking techniques for binaural synthesis, the reader is referred to [223]. In [220], it was shown that head tracking for binaural synthesis improved the externalization of virtual sources when using non-individualized HRTFs. Moreover, head tracking reduced the number of front–back confusions in the localization of virtual sources, which is a common issue when using non-individualized HRTFs [224].

Binaural synthesis is achieved by mapping the spatial information of the measured sound field to a set of virtual loudspeakers positioned around the listener's head. Each virtual loudspeaker corresponds with a specific HRTF depending on the relative direction of the virtual loudspeaker to the listener's head. When the listener moves their head, the HRTFs are adjusted accordingly to maintain the illusion that the virtual loudspeaker array is fixed around the listener's head as the HRTFs provide the frequency response characteristics necessary to recreate the correct spatial audio cues, enabling the perception of sound originating from distinct directions relative to the listener's head. The binaural synthesis filters, one for each ear, are created by convolving each virtual loudspeaker, encoded by the spatial information of the measured sound field, with the corresponding HRTF. These convolutions are then superimposed to create binaural synthesis filters that can be used for

auralization of a dry audio signal by convolving this signal with the binaural synthesis filters and presenting the resulting binaural signals to the listener through headphones. The SFS methods used for loudspeaker synthesis can be adapted to binaural synthesis by mapping the spatial information of the sound field to these virtual loudspeakers, instead of physical loudspeakers.

The spatial information obtained from the SDM analysis, i.e., the DOA $\theta(t)$ of the image sources and reference pressure value $p(t)$, can be used for binaural synthesis by panning the image sources to the virtual loudspeakers using VBAP or NLS [175,196,225]. Since the virtual loudspeakers can be positioned in a dense grid around the listener's head, usually with a spatial resolution of 2.5°–5° [217], the achievable spatial resolution of binaural synthesis is significantly higher compared to loudspeaker synthesis, which motivates the use of NLS for binaural synthesis over VBAP [196]. In [176], several improvements to the SDM method for binaural synthesis were proposed, including the use of a spherical microphone array for SFA, the optimal microphone array size and short-time window length for DOA estimation, and an equalization method to reduce spectral whitening in the late reverberation due to rapidly varying DOA estimates. The improvements were evaluated with perceptual experiments. In [226], the perceptual performance of the SDM for binaural synthesis was evaluated using a variety of microphone array geometries. One perceptual study showed that binaural synthesis using the SDM is indistinguishable from loudspeaker synthesis in the absence of visual cues [219]. However, when visual cues were present, listeners could distinguish between binaural and loudspeaker synthesis, indicating that visual cues play a significant role in the perception of SFS methods [219].

Binaural WFS can be achieved by creating a set of virtual loudspeakers around the listener's head and using the WFS method to synthesize the sound field to these virtual loudspeakers [227]. In a binaural listening experiment, Wierstorf et al. showed that acoustic reproduction using WFS suffers from coloration artifacts due to spatial aliasing occurring when the inter-loudspeaker distance is larger than the wavelength of the sound field [228]. The coloration can be reduced by using HRTFs with a higher spatial resolution, which effectively decreases the inter-loudspeaker distance in the binaural synthesis.

For binaural Ambisonic and HOA synthesis, the spherical harmonic components of the sound field are decoded to a set of virtual loudspeakers positioned around the listener's head [175,221,229]. The decoded signals are then convolved with the corresponding HRTF and summed to obtain the binaural signal that is presented to the listener through headphones. When the original sound field was analyzed using first-order SHD, directional sharpening can be achieved by using the SDM to re-encode the sound field into higher-order SHD components, and subsequently decoding the higher-order SHD components to a set of virtual loudspeakers positioned around the listener's head [14,230].

### 5.2. Synthesis of historical worship space acoustics

The reproduction of HWS acoustics using SFS methods provides a unique and immersive experience for listeners and performers of historical music. Binaural synthesis of HWS acoustics has been used to recreate the acoustics of ancient Greek temples [8], cathedrals [15,17,24,27,139], churches [23,26,144,231], ancient Egyptian temples [25], and mosques [19,26].

Binaural synthesis renders a historical sound field for individual listeners. This has the advantage of providing a more controlled listening experience, as the listener is acoustically decoupled from the listening environment. However, for interactive SFS scenarios, such as live performances, binaural synthesis may not be suitable as it does not provide a shared listening experience for multiple listeners, and may lead to a feeling of disconnection amongst musicians. In contrast, loudspeaker synthesis can be used to recreate the historical sound field over an extensive listening area, allowing for multiple listeners to experience the sound field as they move around the room. Loudspeaker synthesis has been used to recreate the acoustics of churches [9], cathedrals [20,24], and the Stonehenge monument [10].

## 6. Real-time auralization

Real-time auralization of acoustics is an essential tool in the field of audio engineering, music production, and virtual and augmented reality applications. The goal of real-time auralization is to reproduce the acoustics of a given environment in real time, allowing the user to experience the sound of the environment as if they were physically present in that space.

In real-time auralization, the processing time of the audio signal must be kept within a desired latency budget to ensure that the synthesized sound field is perceived without noticeable delay. The latency budget can be defined as the maximum allowable delay between the input audio signal and the output sound field, and for block-based processing with a block size of $n_x$ samples, is defined as $n_x/f_s$, where $f_s$ denotes the sampling frequency of the synthesis system. For example, in a real-time auralization system with a block size of $n_x = 256$ samples and $f_s = 48$ kHz, the maximum allowed processing latency would be 5.33 ms.

Environments with rich and complex acoustics, such as concert halls and HWS, are characterized by their long reverberation time. The acoustic synthesis of such spaces requires filters containing many filter taps, which can be computationally expensive to convolve with an audio signal under real-time conditions. To address this issue, various methods have been proposed to accelerate the convolution process, including uniform and non-uniform partitioned convolution, GPU acceleration, and RIR compression techniques. In addition to convolution-based methods, feedback delay networks and scattering delay networks have been proposed as alternative approaches to interactive auralization.

Finally, acoustic feedback cancellation is an important consideration in real-time auralization systems, as it can help to prevent unwanted feedback loops that can occur when the output of a loudspeaker is picked up by a microphone and re-amplified, causing unwanted artifacts in the reproduced sound such as howling or ringing [232]. Acoustic feedback cancellation algorithms are designed to identify and suppress feedback signals in real time, allowing for stable and artifact-free auralization.

### 6.1. Methods

#### 6.1.1. Fast convolution

Convolution-based methods enable real-time auralization of room acoustics by convolving a dry audio signal, such as an instrument recording or a musical performance, with the filters derived from the synthesis methods outlined in the previous section. The resulting audio signal is then played back through a loudspeaker system or headphones, allowing the listener to experience the sound of the synthesized environment. Convolution-based methods are widely used in real-time auralization systems due to their accuracy and flexibility in modeling complex acoustic environments. However, the computational cost of convolving synthesis filters with audio signals in real time can lead to a prohibitively long processing delay, especially when synthesizing sound fields with long reverberation times in multichannel loudspeaker arrays.

To illustrate the computational cost of convolution-based methods in the time-domain, consider the following example: for a room with a reverberation time of 10 seconds, a 10-second-long RIR that is sampled at 48 kHz requires 480,000 filter taps to accurately model the sound field. When synthesizing the sound field in a loudspeaker array containing 24 channels, the time-domain convolution of the audio signal with the synthesis filters requires approximately $11.5 \times 10^6$ multiply-accumulate operations per input sample. At a sampling rate of 48 kHz, this corresponds to a computational load of 1.11 TFLOPS, which is beyond the capabilities of most consumer-grade CPUs.

*6.1.1.1. Frequency-domain convolution.* The computational cost of time-domain convolution can be reduced by performing the convolution in the frequency domain using the block-based Fast Fourier Transform (FFT) algorithm [233]. In frequency-domain convolution, the audio signal is buffered in a block of $n_x$ samples, after which the input block and the synthesis filters are transformed into the frequency domain using an FFT size of $n_f$ frequency bins. When $n_x < n_f$, the input block is zero-padded to match the FFT size. The frequency-domain transformed input block and synthesis filters are then multiplied in the frequency domain, and the resulting product is transformed back to the time domain using the inverse FFT to obtain the output signal. A block-based Overlap-Add (OLA) or Overlap-Save (OLS) method can be used to obtain the final output block of length $n_x$ samples. The latency of frequency-domain convolution is determined by the block size $n_x$, and the computational cost is $O(n_f \log n_f)$ [233]. The computational advantage of frequency-domain convolution over time-domain convolution increases with the block size $n_x$, however, this block size $n_x$ is upper bounded by the latency budget. Therefore, by choosing an appropriate FFT size, the computational cost of frequency-domain convolution can be lower than time-domain convolution. Nevertheless, the computational cost of frequency-domain convolution can still exceed the latency budget when processing large filter lengths, as the FFT size is lower bounded by the sum of the longest filter length and the analysis window length, minus one.

*6.1.1.2. Partitioned convolution.* To address the computational complexity of frequency-domain convolution with long synthesis filters, the filters can be decomposed into a set of shorter sub-filters that allow for frequency-domain convolution with an input signal within the latency budget. This approach, known as partitioned convolution, was first presented by Stockham in 1966 for processing large filters [234] and has been introduced for real-time Finite Impulse Response (FIR) filtering in audio applications by Torger and Farina in 2001 [235]. Two main approaches to partitioned convolution have been proposed: uniform partitioned convolution and non-uniform partitioned convolution.

In uniform partitioned convolution, a filter of length $n_h$ is decomposed into a set of $K$ sub-filters of length $n_k$, such that $n_h \leq K \cdot n_k$. A uniformly-partitioned input signal with block-length $n_x$ is then convolved with each sub-filter in the frequency domain using the FFT algorithm. Each sub-filter is offset by $k \cdot n_k$, where $k$ is the index of the sub-filter. The resulting partial convolutions are then summed using the OLS or OLA method to obtain the final output signal, which is outputted in blocks of length $n_x$ [236]. When $n_x = n_k$, the delays of all sub-filters can be directly realized in the frequency domain [236].

In non-uniform partitioned convolution, the filter is decomposed into a set of $K$ sub-filters of varying lengths, allowing for more efficient use of the available computational resources. Since frequency-domain convolution becomes computationally more efficient as the FFT size increases, non-uniform partitioned convolution can be used to minimize the computational cost of the convolution process by choosing the optimal length of each sub-filter in the frequency domain. Since the latency-budget of each sub-filter is determined by $k \cdot n_x / f_s$, the length of each sub-filter can be adjusted to balance the computational cost of frequency-domain convolution with the latency budget. For later sub-filters, the length can be increased to reduce the computational cost of the convolution process, while for earlier sub-filters, the length can be reduced to ensure that the convolution process remains within the latency budget [236].

In partitioned convolution methods, the transformation from the time-domain to the frequency domain is performed only once for all sub-filters and input blocks, reducing the computational cost of the convolution process. For a comprehensive overview of uniform and non-uniform partitioning methods and their application to real-time auralization systems, the reader is referred to the work of Wefers [236].

Partitioned convolution methods decompose the filter into smaller sub-filters that can be concurrently processed with an input signal. This approach makes uniform and non-uniform partitioned convolution well-suited for parallel processing on GPUs, which are optimized for processing large amounts of data in parallel. By delegating the convolution process to a GPU, the processing time can be significantly reduced [237], enabling real-time auralization with long reverberation times in multichannel loudspeaker arrays.

*6.1.1.3. Computational reduction techniques.* Computational reduction techniques aim to reduce the number of filter taps required to accurately model the sound field while maintaining the perceptual quality of the synthesized sound. Common techniques are thresholding, truncation, and compression of the synthesis filters. In thresholding, the filter taps with magnitudes below a certain threshold $\epsilon$ are set to zero, reducing the number of non-zero filter taps in the filter, and therefore reducing the required amount of addition and multiplication operations in the time-domain convolution process. In truncation, the filters are truncated, so that the convolution with an input signal can be performed within the desired latency budget [238,239]. Compression of the synthesis filters can be achieved by exploiting the underlying structure of the filters. For example, it was shown in [240] that filters modeling room acoustics, i.e., RIRs, exhibit a low-rank structure when rearranged into a matrix or tensor. This property can be used to compress the filters using low-rank matrix factorization techniques, reducing the number of filter taps [240]. Moreover, the compressed filter representation can be directly used for fast convolution in the time domain [241,242]. In [243], this method was extended to multichannel RIRs measured in the same room, showing that the low-rank structure is shared across channels, which can be exploited to further reduce the computational cost of the convolution process. The authors applied this method to compress and achieve multichannel fast convolution of synthesis filters in real-time auralization systems [244]. While low-rank compressed filters offer computational advantages, it is important to note that they can introduce artifacts in the synthesized sound field at high compression rates, specifically in the high-frequency range [244].

*6.1.1.4. Filtered velvet noise.* Computational reduction can also be achieved by separating the late reverberation from the direct and early reflections of the sound field and synthesizing them separately. A computationally efficient method for modeling the late reverberation can be achieved using Filtered Velvet Noise (FVN) [245,246]. Velvet noise is a type of sparse pseudo-random ternary sequence that contains only the values $-1$, 0, and 1. This inherent sparsity allows for efficient convolution, as the convolution process can be implemented using a series of additions and subtractions, rather than multiplications. To model the late reverberation of a sound field using FVN, the late reverberation is first divided into segments, each approximated by an independent FVN sequence. To obtain the synthesized reverberation, an input signal is filtered with the cascading FVN sequences, and a coloration filter and an associated gain factor are applied to approximate the frequency-dependent decay of the reverberation in the time domain. The resulting FVN-filtered sequences are summed to obtain the final output [245,246]. Recent advancements have extended FVN to incorporate parameter estimation using machine learning, enabling automatic design of FVN sequences to match the energy decay profile of a given acoustic environment [247]. While FVN efficiently reproduces an approximation of the late reverberation of a sound field, noticeable differences can still exist between measured and synthesized reverberation [246].

### 6.1.2. Delay networks

Due to the computational complexity of convolution-based methods, alternative approaches to achieve real-time auralization have been proposed. Delay networks are types of recursive filters that can be used to efficiently synthesize artificial reverberation with low computational cost [248–250]. An Feedback Delay Network (FDN) is a recursive filter that consists of a network of delay lines, a set of input and output gains, and a unitary feedback matrix that models the feedback path

of the filter. The feedback matrix determines the frequency response of the FDN and can be designed to model the frequency-dependent sound energy decay of a given acoustic environment [250–255].

The output of the FDN is obtained by summing the delayed and attenuated signals from the delay lines, which creates a dense and diffuse reverberation effect. FDNs are widely used in real-time auralization systems due to their low computational cost and ability to model long reverberation times [250].

The spectrum of artificial reverberation synthesized by an FDN is often colored due to resonating modes. To address this issue, differentiable FDNs have been proposed, which optimize the feedback matrix, as well as the input and output gains, to achieve a near-colorless artificial reverberation [256].

While FDNs offer a computationally efficient method that is simple to design, which can achieve high quality artificial reverberation, they lack the ability to accurately reproduce the direct and early reflections of a space. Therefore, hybrid methods that combine convolution-based methods with FDNs have been proposed. In [257], a hybrid reverberation processor was introduced that reproduces the early reflections of a space using convolution-based methods and synthesizes the late reverberation using an FDN that is automatically adjusted to match the energy decay profile of the space [257].

Another method to synthesize the early reflections and room modes using a recursive filter is the Scattering Delay Network (SDN), which was proposed by De Sena et al. [192]. SDNs are artificial reverberators that model the early reflections and room modes of a space using a network of scattering nodes. To model the acoustic path between a source and receiver in a room using SDNs, each reflective surface in the room is represented by a single scattering node that is placed at the location on the surface corresponding to the first-order reflection point. The scattering nodes are connected with the source, and with each other using bidirectional delay lines that model the sound propagation delay between the nodes and the absorption of the sound energy by the reflective surfaces [192]. The receiver node is connected to each scattering node using a unidirectional delay line that models the sound propagation from each reflective surface to the receiver. The scattering matrix determines how the energy is circulated within the network [192]. By positioning the scatter nodes at the locations of the first-order reflections, higher-order reflections and room modes are implicitly modeled by the scattering network, allowing for the synthesis of a realistic sound field with low computational cost [192].

### 6.1.3. Acoustic feedback cancellation

Real-time interactive auralization systems use microphones to capture sound sources in the environment. These microphone signals are typically processed by a conditioner function **G**, convolved with the synthesis filters **H**, and played back through loudspeakers to create the synthesized sound field. Acoustic feedback occurs when the output of the auralization process is played back through a loudspeaker system and is again picked up by the microphones, creating a feedback loop that can lead to unwanted artifacts. Acoustic feedback cancellation algorithms are designed to identify and suppress feedback signals, allowing for stable and artifact-free auralization [232].

Acoustic feedback cancellation is a well-explored area within public address systems, characterized by unknown and variable feedback paths **F** due to the movement of sound sources and microphones [232]. In this area, several techniques have been proposed for mitigating acoustic feedback, including phase modulation [258–260], and notch filters [261,262]. However, these methods may introduce sound artifacts, including coloration and distortion, which degrade the perceptual quality of the auralization. Alternatively, adaptive filtering techniques can be employed to estimate the multichannel feedback paths **F**. The estimated feedback signal is then subtracted from the microphone signals to obtain the feedback-compensated signal [263–265]. For an exhaustive overview of acoustic feedback cancellation methods in public address systems, the reader is referred to the work of van Waterschoot et al. [232].

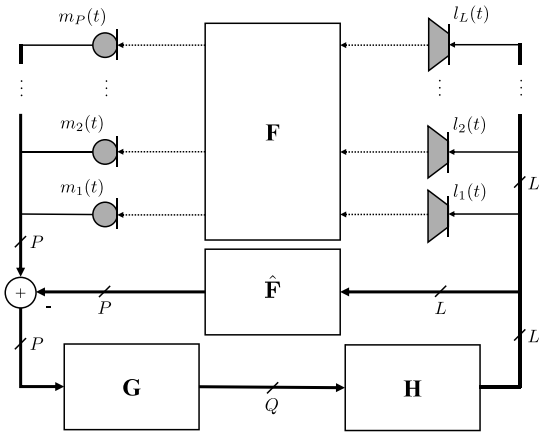



**Fig. 3.** Block diagram of a real-time auralization system containing $L$ loudspeakers and $P$ microphones. The microphone signals are first preprocessed by a conditioner function **G**, after which they are convolved with the synthesis filters **H** to create the synthesized sound field. Acoustic feedback cancellation is implemented using the multichannel feedback path estimate $\hat{\mathbf{F}}$ and subtracting the resulting feedback signals from the microphone signals, to compensate for the acoustic feedback path **F**.

In the context of real-time auralization systems, where the loudspeaker and microphone configurations are typically fixed, the multichannel feedback paths **F** do not rapidly change over time. This allows for the estimation of the feedback paths **F** using system identification techniques in a calibration phase and use the estimated feedback paths to cancel the feedback signals in real time during auralization without altering the perceptual quality of the synthesized sound field. In Fig. 3, the block diagram of a real-time auralization system with acoustic feedback cancellation is shown. The microphone signals are first preprocessed by a conditioner function **G**, after which they are convolved with the synthesis filters **H** to create the synthesized sound field. Acoustic feedback cancellation is implemented by estimating the multichannel feedback path **F** and subtracting the estimated feedback signals from the microphone signals.

In [30], an acoustic feedback cancellation algorithm was proposed for real-time auralization systems. The method designs a multichannel feedback path estimate $\hat{\mathbf{F}}$ that incorporates robustness to measurement errors, model uncertainties, and slight variations in the feedback paths due to moving performers or participants. For a single loudspeaker and microphone, the robust time-domain feedback path estimate $\hat{f}(t)$ is obtained by applying a time-frequency weighting function $w(t, f_b)$ to the measured feedback path $\tilde{f}(t, f_b)$, where $f_b$ denotes the frequency bin in the time-frequency representation, i.e.,

$$\hat{f}(t) = \sum_{b=1}^{B} \tilde{f}(t, f_b) \cdot w(t, f_b), \tag{6}$$

where $B$ represents the number of frequency bins in the time-frequency representation of the feedback path. The weighting function $w(t, f_b)$ is designed to suppress the feedback path estimate in regions where the feedback path is not well known by using a Wiener-like weighting, i.e.,

$$w(t, f_b) = \frac{f^2(t, f_b)}{f^2(t, f_b) + \sigma^2(t, f_b)}, \tag{7}$$

where $f(t, f_b)$ denotes the true feedback path and $\sigma^2(t, f_b)$ denotes the variance of the feedback path estimate [30]. The true feedback path $f(t, f_b)$ and variance $\sigma^2(t, f_b)$ can be estimated by repeated measurements of the feedback path in the calibration phase. The true feedback path $f(t, f_b)$ is estimated by averaging the measured feedback paths over time, while the variance $\sigma^2(t, f_b)$ is estimated by calculating the variance of the measured feedback paths over time. Additionally,

**Table 3**
Software that enables real-time synthesis of room acoustics and their use in reproducing the acoustics of HWS.

| Software | License | Use in HWS |
|---|---|---|
| Max/MSP [273] | Commercial | [15,24,25,27] |
| Pure Data [274] | BSD-3-Clause | |

$f^2(t, f_b)$ is smoothed over time and frequency to obtain a smooth windowing function [30].

The robust feedback path estimate $\hat{f}(t, f_b)$ is then used to design a feedback cancellation filter that is applied to the microphone signals to obtain the feedback-compensated signal. The method was shown to achieve robust feedback suppression greater than 20 dB in real-time multichannel auralization systems [30]. The method has been used to design auralization systems in a variety of applications, including recording studios [266,267], Virtual Reality (VR) for historical preservation [26], and a series of art installations [268].

### 6.2. Real-time auralization of historical worship space acoustics

Real-time auralization of HWS is a challenging task due to the rich acoustics and long reverberation times by which these spaces are characterized. The acoustics of HWS are of particular interest to musicologists, musicians, and audio engineers, as they provide a unique sonic environment in which music and liturgy were composed and performed. Real-time auralization of HWS allows for the interactive exploration of the acoustics of these spaces, enabling musicians to experience the sound of the space as if they were physically present in it, and to investigate the intrinsic relationship between musical performance and the space in which it is performed [269–272].

The real-time auralization of HWS using convolution-based methods has been proposed in several studies. Uniformly partitioned convolution methods have been used to synthesize the acoustics of the Cathédrale Notre-Dame de Paris [27] and the acoustics of ancient Greek spaces [8].

A non-uniform partitioned convolution scheme was used by Eley et al. to synthesize the early reflections of Cathédrale Notre-Dame de Paris in real time [24]. The late and diffuse reverberation of the cathedral was synthesized using an FDN model. The authors showed that the combination of non-uniform partitioned convolution and FDN synthesis can achieve a realistic and immersive auralization of the cathedral in real time [24].

In real-time binaural synthesis of HWS, head tracking techniques have been proposed to enhance the spatial realism of the auralization. Head tracking allows the listener to move their head in the virtual space, changing the perceived direction of the sound sources and reflections. Postma et al. [15] used head tracking in a VR system to enhance the spatial realism of the auralization of the Cathédrale Notre-Dame de Paris. Head tracking was also used by Berger et al. [26] in a VR system which enabled the listener to experience the acoustics of ancient caves, mosques, and churches in 9-DOF VR.

In [20], real-time loudspeaker-based auralization of the Chiesa di Sant'Aniceto, a historical church in Rome, was achieved by implementing the feedback-canceling reverberator.

Real-time auralization of HWS can also be implemented in the Max/MSP software [273]. Max/MSP software is widely used for real-time audio applications, and has been used for the real-time binaural auralization of an ancient Egyptian temple [25], and the Notre-Dame de Paris [27].

A summary of software enabling the real-time synthesis of room acoustics and their use in reproducing the acoustics of HWS is provided in Table 3.

## 7. Perceptual evaluation

Loudspeaker and headphone-based sound field synthesis methods aim to accurately synthesize an acoustic environment in a reproduction room. The original sound field in the room is however, immensely complex, making the true physical reproduction of the sound field using a finite number of microphones and loudspeakers challenging. Luckily, the human auditory system is prone to perceptual illusions, which can be exploited to create a plausible reconstruction of the virtual sound field in the reproduction room. The perceptual evaluation of sound field synthesis methods is crucial to assess the perceived quality of the synthesized sound field. In this section, we will discuss the various perceptual evaluation methods in the context of sound field auralization systems.

### 7.1. Methods

The reproduction quality of auralization systems can be evaluated using objective psychoacoustic-based methods, subjective methods, or sensory evaluation methods.

#### 7.1.1. Objective psychoacoustic-based evaluation methods

Objective psychoacoustic-based evaluation methods are based on the human auditory system's perception of sound. These methods involve the use of psychoacoustic models that predict the perceived quality of the synthesized sound field based on the physical properties of the sound field [275]. These physical properties can be measured from acoustic measurements. The most common objective evaluation parameters used in the context of sound field synthesis are Reverberation Time (RT), Early Decay Time (EDT), Clarity, Definition, Sound Strength, Apparent Source Width (ASW), Listener Envelopment, Interaural Cross-Correlation (ICC), Early Support, and Speech Transmission Index (STI), all of which are defined in the ISO3382-1:2009 standard [34].

The perceptual evaluation of acoustics for a specific application can be performed by aggregating the objective evaluation parameters into a single index that represents the perceived quality of the sound field for specific applications. This has been done to perceptually evaluate concert hall acoustics [275,276] by applying a weighting factor to each objective evaluation parameter based on the perceived importance of the parameter for the application.

#### 7.1.2. Subjective evaluation methods

Subjective evaluation methods involve human listeners who evaluate the synthesized sound field based on their perception by listening or interacting with the system and providing feedback through questionnaires [185,196,226,277,278] or interviews [151,270,272,279,280]. The resulting data from the subjective evaluation methods can be processed using statistical tools to analyze the perceived quality of the synthesized sound field [281].

Perceptual evaluation methods can be broadly divided into two categories: *Comparative* and *Descriptive* methods [282]. Comparative methods evaluate aspects of stimuli in comparison to a reference stimulus [182,283]. For example, in preference tests, the listener is asked to compare two stimuli and indicate a preference for one stimulus [185,196,226,270,284–287]. Descriptive methods involve evaluating each stimulus by ranking a set of descriptive attributes based on the perceived intensity of each attribute for every stimulus. The type of evaluation method that is used depends on the research question and the context of the study [282]. For an overview of perceptual evaluation methods used in the context of audio evaluation, the reader is referred to [282].

#### 7.1.3. Sensory evaluation methods

While comparative and descriptive evaluation methods are commonly used in the subjective evaluation of audio, sensory evaluation methods offer a more structured approach to evaluate sensory data. These methods were first introduced in the food industry to evaluate the sensory properties of food and beverages [288,289], and have since been used in the perceptual evaluation of reproduced concert hall

acoustics [290–292], car cabin acoustics [293–295], and small room acoustics [187,296,297].

In sensory evaluation methods, a set of descriptors, also called attributes, should be carefully selected to accurately describe the perceived sensation across a set of stimuli. When experienced or expert assessors are involved in the evaluation, the attributes can be elicited using Consensus Vocabulary Profiling (CVP) where the group of assessors agrees on a common set of attributes that best describe the perceived sensational differences between all stimuli before the evaluation [298]. However, it can be challenging and time-consuming to elicit a common set of attributes that accurately represent the perceived sensation of the stimuli, especially for large groups of assessors. As an alternative, Individual Vocabulary Profiling (IVP) can be used, where the assessors describe the perceived sensations using their own vocabulary [299]. In IVP, the usage of each elicited attribute may not be consistent across assessors, which can make the interpretation of the results challenging. To address this issue, statistical methods, such as Principal Component Analysis (PCA) and Multiple Factor Analysis (MFA), can be used to extract the meaning of the attributes and to perform a correlation analysis between the attributes and the stimuli [299]. Moreover, the relationships between preference ratings, sensory evaluation methods, and acoustic measurements can be visualized using MFA [300], providing valuable insights into the perceived quality of the stimuli. The attribute elicitation phase is followed by the evaluation phase, where the assessors evaluate the stimuli based on the elicited attributes. After the experiment, the data collected from the assessors is analyzed using statistical methods, such as Generalized Procrustes Analysis (GPA), MFA, PCA, and ANCOVA, to evaluate the perceived quality of the stimuli [282].

A widely recognized method for evaluating the sensory properties of food and beverages is Quantitative Descriptive Analysis (QDA), where a panel of trained assessors evaluates the sensory properties of the stimuli using CVP [301]. The assessors undergo training to recognize and describe the sensory properties of the stimuli under evaluation, and the data collected from the assessors are analyzed using statistical methods to evaluate the perceived quality of the stimuli. However, QDA is time-consuming and requires a panel of trained assessors, which can be expensive and difficult to maintain. This limitation has restricted its application in the perceptual evaluation of audio reproduction systems. To address the limitations of QDA, faster sensory evaluation methods that do not require extensive training of the participants such as Free Choice Profiling (FCP) and Flash Profile (FP) have been developed. These methods, which rely on IVP, enable the quick evaluation of sensory properties using untrained or semi-trained assessors [289,302].

During the elicitation phase of FCP and FP, each assessor is asked to identify the descriptors that best describe the perceived sensation of the presented stimuli. These descriptors are individually elicited without the need for a common definition or agreement among the assessors. In FCP, assessors then rate the set of attributes based on the perceived intensity of each attribute for every stimulus. After the experiment, the data is analyzed using GPA to align the data from each individual assessor into a common space, resulting in a consensus profile that represents the perceived quality of the stimuli. In FP, instead of rating the attributes, the assessors are asked to rank the set of attributes based on the perceived intensity of each attribute for every stimulus, and the data collected from the assessors is analyzed using MFA. The advantage of FCP and FP is that they are quick and easy to perform without requiring a panel of trained assessors. Moreover, while the results obtained from FCP and FP may not be as accurate or complete as those obtained from QDA, they provide similar insights into the perceived quality of the stimuli [281,289,302]. According to one comparative study, the FP method was found to display a stronger discrimination capability than QDA and FCP in taste evaluations [303].

Given the efficiency and ease of use of IVP methods, they have been used in the perceptual evaluation of auditorium and concert hall acoustics [290–292], car cabin acoustics [293,294], small room acoustics [297].

The auditory sensory evaluation methods described above are non-interactive, meaning the assessors evaluating the stimuli did not interact with the stimuli during the evaluation. Interactive sensory evaluation, on the other hand, involve the assessors interacting with the auditory stimuli to evaluate the perceived quality of the stimuli. This interaction provides a more realistic evaluation of the stimuli, as the assessors can evaluate the stimuli based on their perception of the stimuli in a real-world scenario. In the context of interactive sensory evaluation for audio reproduction systems, FP has been used to evaluate the perceived quality of virtual acoustic environments [304]. In [305], Dual Multiple MFA was used to evaluate the perceived quality of trumpet performance adjustments in different virtual acoustic environments. Tomasetti et al. [280] compared musicians' preferences for collaborative music performance using binaural synthesis with head tracking over a conventional stereo setup. Similar research was conducted in [306], where musicians' preference for binaural audio with head tracking over stereo was evaluated for digital piano performance.

In [307], the perceptual localization performance of virtual sources was evaluated under real and reproduced acoustic conditions using a 4th order 3D HOA system. Participants were asked to localize virtual acoustic sources in a VR environment by indicating the perceived DOA, apparent source width, and perceived distance of the virtual sources using a VR pointer interface.

### 7.2. Perceptual evaluation in historical worship spaces

The acoustic quality of HWS can be evaluated using objective psychoacoustic-based evaluation parameters, as described in Section 7.1.1. In [308], a synthetic index, an index that is based on a combination of multiple partial indices, was proposed to assess the acoustic properties of sacral buildings. The proposed index was a function of five partial indices, each representing a specific aspect of the acoustic quality of the building. Kosała proposed a new formula for calculating the single number rating of the acoustic quality of church interiors [309] based on a combination of three partial indices using the Singular Value Decomposition (SVD) method. In [310], a double synthetic index was proposed to evaluate the acoustics of churches. The double synthetic index separates the evaluation of the acoustics of churches into two metrics: an index for the perception of music, and an index for the perception of speech [310].

An acoustic quality assessment was carried out by Kosała [100] in the St. Elizabeth of Hungary church in Jaworzno, Poland. The acoustic properties of the church were evaluated using two indices, namely the Global Acoustic Properties Index (GAP) [99] and the Global Index (GI) [311]. These acoustic evaluation methods were also used in the assessment of Orthodox churches in Poland [312].

A widely used perceptual evaluation method for HWS is the measurement of objective acoustic parameters. Common parameters include RT, EDT, clarity, definition, sound strength, sound level pressure, STI, ASW, and ICC. These parameters offer a simple and effective way to evaluate the acoustics of HWS, as they can be directly calculated from RIR measurements. Objective acoustic parameters have been used to perceptually evaluate the acoustics of various HWS, including churches [12,13,17,106,112,115,116,131,137,140,141,143,152,313], cathedrals [16,18,22,107,119,128,145,150,151], mosques [19,98,124,130,134,314], and Buddhist temples [121,315].

Subjective evaluation methods have also been used to evaluate the perceived quality of the acoustics in HWS. In [9], a subjective evaluation of preferred listening conditions in nine Italian Catholic churches was performed. A total of 143 participants were included in the evaluation, where they were asked to state their preference for different binaural auralizations of church acoustics convolved with musical excerpts. The musical excerpts ranged from Gregorian chant to symphonic music. The results of the listening tests were analyzed by

applying Factor Analysis to determine the key acoustic parameters that influenced the participants' preferences [9].

A subjective listening test conducted by Postma and Katz in 2016 [16] compared the perceptual similarity of measured and simulated binaural auralizations of the acoustics of the Saint-Germain-des-Prés abbey church, the Cathédrale Notre-Dame de Paris, and the Théâtre de l'Athénée, France. The participants were asked to compare the auralizations by rating the similarity of eight predefined perceptual attributes for each pair of auralizations [16].

In [316], the perceptual evaluation of Han Chinese Buddhist temples was carried out. The evaluation involved a series of questionnaires which gauged the participants' comfort and preference for the soundscape at four temples. The soundscapes of each temple was measured and compared to the participants' responses to the questionnaires to identify the characteristics of the soundscapes that influenced the participants' comfort and preference [316].

The spatial impression and directional resolution of the reproduced acoustics of the St. Andrew's church, Lyddington, UK, was investigated by Frank and Zotter [14]. A perceptual evaluation was done to compare the performance of acoustic reproduction with first-order spherical harmonics, and the SDM-enhanced spherical harmonic representations of the sound field. Two listening tests were organized. The first listening test was used to determine the size of the perceptual sweet spot by asking participants to localize the reproduced sound at different locations in the reproduction setup. The second listening experiment used a MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA)-like graphical interface where the participants could rate the perceived spatial depth mapping for each auralization [14].

Subjective listening tests have also been carried out by Álvarez-Morales et al. in 2017 [17] to evaluate the visual and acoustic experience of a reproduced VR environment of the Cathedral of Seville, Spain. The evaluation consisted of three tests: a blind acoustic perception test, a visual virtualization test, and a combined visual and acoustic test. The participants were asked to rank the perceived quality of the visual and acoustic experience of the VR environment [17]. For each test, the participants were asked to rank a set of predefined attributes based on the perceived quality of the VR environment.

A perceptual assessment of reverberation in Spanish cathedrals was carried out by Girón et al. in 2020 [317]. A series of binaural listening tests with 42 participants were carried out to evaluate the perceived reverberation in the cathedrals. The participants had to compare the perceived reverberation between sets of two cathedrals. The results indicated that a reverberation time between four and seven seconds was indistinguishable for the listeners [317].

A recent study by López-Mochales et al. [231] investigated the enhancement of feelings of transcendence, tenderness, and expressiveness by music in Christian liturgical spaces. The study involved a series of listening tests of music excerpts, one liturgical piece and one secular piece, that were presented to the listeners in either raw format, or an auralized format with different church acoustics. Using these excerpts, two listening tests were performed: one where the participants had to rate a set of attributes based on the perceived expressiveness and beauty of the music, and one preference test where the participants had to indicate their preference for the music excerpts. The results were analyzed using a factorial ANCOVA analysis to evaluate the effect of the church acoustics on the perceived expressiveness and beauty of the music [231].

The perceptual evaluation of a real-time auralization system reproducing the acoustics of the Chiesa di Sant' Aniceto in Rome, Italy, was carried out by Canfield-Dafilou et al. in 2019 [20]. The evaluation involved two choirs, each choir consisting of three singers, that were asked to perform four verses of a Catholic hymn inside the auralization system. The auralization system reproduced the acoustics of the church in real-time, while varying the reverberation time during the performance. After the performance, the singers were asked to fill in an informal survey to evaluate the perceived quality of the auralization system [20].



(a) (b)

**Fig. 4.** (a) View from the chapel floor looking towards the choir loft. (b) View from the choir loft looking onto the chapel floor. The loudspeaker used during the acoustic measurements is positioned on the right side of the choir loft.

An interactive real-time auralization system for reproducing Cathédrale Notre-Dame de Paris' acoustics was validated using a perceptual analysis [24]. The perceptual analysis was carried out by four singers, all members of a medieval ensemble with experience performing in the Cathédrale Notre-Dame de Paris. The singers were asked to validate the reproduction system by performing excerpts of medieval repertoire in the virtual reproduction space in both binaural and multichannel loudspeaker configurations. The singers' feedback was collected using questionnaires which evaluated the similarity of the virtual acoustics to the real acoustics of the cathedral [24]. It should be noted that the participants could only base their comparison on their memory of the cathedral's acoustics [24], which might affect the reliability of the perceptual evaluation.

The relationship between ancient music practices and the acoustics of the space in which it is performed was investigated using an immersive binaural auralization system that was developed and presented by Mullins and Katz [27]. A subjective analysis in the form of an interactive choir performance was conducted to evaluate the suitability of various acoustical conditions for the performance of ancient music. The choir was asked to perform short excerpts of choral music using the interactive auralization system, and to rate the ease, suitability, and difficulty of performance after each excerpt [27].

## 8. Example case study of the Nassau Chapel

### 8.1. Introduction

In this section, we present a case study of the preservation and auralization of the Nassau Chapel, also called St. George's Chapel, a 16th century Gothic style chapel located in Brussels, Belgium. The chapel was originally founded by Willem van Duvenvoorde in 1344 and was rebuilt in the early 16th century by the Nassau family as part of the Palace of Nassau [318]. The chapel was used as a public place of worship and services were held in the chapel throughout the Ancien Régime. Today, the fully restored Nassau Chapel is incorporated into the buildings of the Royal Library of Belgium where it serves as an exhibition space displaying the manuscript collection of the Burgundian dukes [319].

The aim of the case study is twofold. First, we aim to digitally preserve the acoustics of the Nassau Chapel using room acoustic measurements and SFA to capture the spatial and temporal characteristics of the sound field in the chapel. Second, we aim to use the measured data in a real-time auralization system that serves as an interactive rehearsal space for early music performers to explore the relationship between early music performances and the acoustics of the chapel.

The case study of the Nassau Chapel is divided into five main sections, according to the components of the auralization process presented in Fig. 1. Section 8.2 describes the room acoustic acquisition

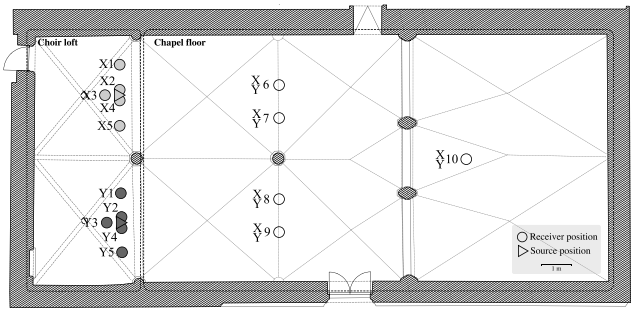



**Fig. 5.** Floor plan of the Nassau Chapel containing the microphone array positions, denoted as circles, and the loudspeaker positions, with the direction of the loudspeaker indicated by the triangle. Two sets of measurements were taken: one with the loudspeaker on the left side of the choir loft (light gray), and the other with the loudspeaker on the right side of the choir loft (dark gray). The microphone array positions were positioned around each loudspeaker position in the choir loft. The microphone positions in the chapel floor are common for both sets of measurements. The measurements are labeled X1 to X10 for measurements made using the left-side loudspeaker and Y1 to Y10 for the right-side loudspeaker.

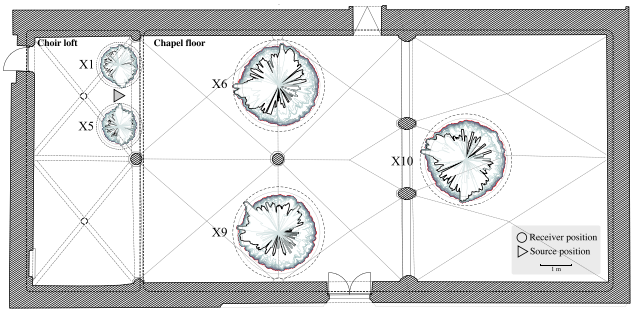



**Fig. 6.** Visualization of the lateral spatio-temporal response at different receiver positions in the Nassau Chapel, with the loudspeaker on the left side of the choir loft. Each line in the spatio-temporal response represents a time instant from 10 ms to 200 ms, with a time-increment of 10 ms. The bold black line represents the cumulative energy at 30 ms, while the outer red line represents the cumulative energy at 200 ms. The dashed circle represents the maximum energy of the spatio-temporal response. Each dotted circle represents −6 dB of the maximum energy. The energy curves are smoothed with a five-degree angular sliding window average.

of the Nassau Chapel, including the measurement setup and procedure used to capture the RIRs at several positions in the chapel. Section 8.3 presents a spatio-temporal analysis of the acoustic measurements made in the chapel using the SDM as a tool to extract the spatial and temporal characteristics of the sound field. In Section 8.4, the reproduction setup that is used to reproduce the sound field is presented. The Nassau Chapel's acoustics are reproduced from the measured data using a loudspeaker array and the NLS method. Section 8.5 addresses the implementation of a real-time auralization system by first reducing the computational complexity of the sound field synthesis using a low-rank approximation of the synthesis filters, followed by the implementation of the acoustic feedback cancelation method discussed in Section 6. Finally, Section 8.6 presents the perceptual evaluation of the acoustics of the Nassau Chapel using objective acoustic metrics.

### 8.2. Room acoustic acquisition

A series of room acoustic measurements was carried out in the Nassau Chapel to capture the sound field in the space during a historical ceremony or service when the choir is presumed to have been positioned on the choir loft, singing towards the chapel floor. The measurements took place during the night of February 27$^{th}$ 2020, during a time when the chapel was undergoing renovations for the installation of a new exhibition. During the measurements, the chapel was empty and in an unfurnished state.

The acoustic measurement setup consisted of a *Genelec 8030C* loudspeaker and a *GRAS 50VI-1 Vector Intensity Probe*, a microphone array consisting of six omnidirectional microphones arranged in a Cartesian grid with a spacing of 25 mm. The sound levels of each microphone were calibrated before the measurements using a *Bruel & Kjaer 4230 Sound Level Calibrator*. A floor plan of the Nassau Chapel containing the loudspeaker and microphone array positions is shown in Fig. 5. Two sets of measurements were taken, one with the loudspeaker on the left side of the choir loft, and the other with the loudspeaker on the right side of the choir loft. For each loudspeaker position in the choir loft, five microphone array positions were chosen around the loudspeaker position to capture the sound field that the choir would have experienced. The acoustic sound field was also captured at the chapel floor during both sets of measurements at five positions. A photograph of the Nassau Chapel with the loudspeaker placed on the right side of the choir loft is shown in Fig. 4.

The room acoustic measurements were carried out by playing back an ESS signal [50] from the loudspeaker and recording the acoustic response of the chapel with the microphone array. The measurement sampling rate was set at 192 kHz, and the ESS signal had a duration of 15 seconds and a frequency range up to 96 kHz. The sweeps were repeated five times with an idle time of ten seconds between each sweep to allow the acoustic response of the chapel to decay to a sufficiently low level. After the measurements took place, the recorded responses were deconvolved to obtain the RIR at each microphone and loudspeaker position.

### 8.3. Sound field analysis

The RIR data obtained from the measurements in the Nassau Chapel were analyzed using the SDM to obtain a spatio-temporal representation of the sound field in the chapel. The analysis was carried out for the measurements taken with the loudspeaker on the left side of the choir loft. The spatio-temporal response at different receiver positions in the chapel is shown in Fig. 6. The spatio-temporal response shows the cumulative energy of the sound field at different time instants from 10 ms to 200 ms, with a time increment of 10 ms. The bold black line represents the cumulative energy at 30 ms, while the outer red line represents the cumulative energy at 200 ms. The dashed circle represents the maximum energy of the spatio-temporal response, while each dotted circle represents −6 dB of the maximum energy. The energy curves are smoothed with a five-degree angular sliding window average. The spatio-temporal response shows the lateral energy distribution of the sound field in the chapel at different time instants, providing insights into the spatial characteristics of the sound field.

It can be seen from the spatio-temporal responses that the cumulative spatial energy distribution at 30 ms is more diffuse at the choir loft compared to the chapel floor. This is due to the superposition of a large number of early reflections at the choir loft due to the proximity of the walls and ceiling. The energy distribution at the chapel floor is sparser, with the energy arriving directly from the loudspeaker, and the early reflections taking longer to arrive due to the increased distance from the walls and ceiling. From this, we can infer that the choir experiences more noticeable early reflections at the choir loft compared to the chapel floor, which can affect the perception of the sound field in the chapel. The perceptual impact of this spatial energy distribution can be further analyzed through perceptual evaluation, which is discussed in Section 8.6.

### 8.4. Sound field synthesis

The measurements and spatial analysis of the Nassau Chapel's acoustics were used to reproduce the acoustics of the chapel in the Alamire Interactive Lab (AIL) at the Library of Voices in Leuven,



**Fig. 7.** The Alamire Interactive Lab at the Library of Voices in Leuven, Belgium. The lab is equipped with a loudspeaker array containing 24 loudspeakers (Martin Audio CDD6) and six microphones (Audio Technica AT4051b) surrounding the loudspeaker array. The loudspeakers and microphones interface with the computer system through a Dante audio-over-Ethernet protocol to enable low-latency audio streaming between the computer system and the loudspeaker array.

Belgium. The AIL is a rehearsal space that is used to study and perform early music, and is equipped with a loudspeaker array containing 24 loudspeakers (Martin Audio CDD6) and six microphones (Audio Technica AT4051b) surrounding the loudspeaker array. The loudspeakers and microphones interface with the computer system through a Dante audio-over-Ethernet protocol to enable low-latency audio streaming between the computer system and the loudspeaker array. The interactive rehearsal space at the AIL is shown in Fig. 7.

The reproduction of the Nassau Chapel's acoustics was carried out using NLS, which maps each image-source from the SDM analysis to the closest loudspeaker in the loudspeaker array based on the DOA of each image-source. This results in a set of synthesis filters that can be convolved with an input signal and played back through the loudspeaker array to reproduce the sound field of the Nassau Chapel. The perceptual evaluation of the reproduced sound field is discussed in Section 8.6.

### 8.5. Real-time auralization

The Nassau Chapel currently serves as an exhibition space in the Royal Library of Belgium and is therefore not accessible for rehearsals or performances. To enable early music performers to interact with the acoustics of the chapel during rehearsals, a real-time auralization system was implemented in the AIL.

Real-time auralization in multichannel playback systems requires a high computational cost to convolve the input signal captured by the microphones with all the synthesis filters, one filter for each loudspeaker in the reproduction setup. Since the Nassau Chapel is a highly reverberant space, the synthesis filters contain many filter taps, resulting in a high computational cost that requires a powerful processor to handle the convolution operations in real-time. When the processor is not able to handle the computational load in real-time, the system introduces latency which can degrade the quality of the auralization and make it unsuitable for real-time applications. To lower the computational cost of convolution with the synthesis filters, a low-rank approximation of the filters was applied to reduce the number of filter taps and therefore the computational complexity of the convolution operation. In [244], the authors applied a low-rank approximation to the synthesis filters, achieving a significant reduction in computational complexity while maintaining key perceptual features of the sound field.

Additionally, real-time auralization using loudspeakers and microphones to reproduce the sound field is prone to acoustic feedback, which occurs when the output signal from the loudspeaker array is picked up by the microphones and fed back into the auralization system. Acoustic feedback causes instability in the system and degrades the quality of the auralization by introducing artifacts such as howling or unwanted echo. To mitigate the effects of acoustic feedback in the AIL, the acoustic feedback cancelation method presented in [30] was implemented for the real-time auralization system.

### 8.6. Perceptual evaluation

The room acoustic measurements were analyzed to extract various room acoustic parameters, including the Reverberation Time ($T_{20}$), EDT, Clarity ($C_{80}$), Definition ($D_{50}$), Center time ($T_s$), and Direct-to-Reverberant Ratio (DRR). The parameters were calculated in accordance with the ISO:3382-1 standard [34]. The objective room acoustic parameters were calculated for each microphone position in the chapel and are presented in Table 4. The results show that the reverberation time in the Nassau Chapel is relatively long, with values ranging from 5.32 to 6.20 s. Although unusual for a chapel of this size, the reverberation time is expected to be higher since the measurements were conducted in an empty chapel due to ongoing renovations, rather than when the chapel is furnished. Moreover, a clear distinction can be observed between the measurements made on the choir loft (1–5), and the measurements made on the chapel floor (6–10). The measurements on the choir loft have a lower reverberation time and center time compared to the measurements on the chapel floor. This is expected since the choir loft is positioned closer to the chapel's ceiling. Moreover, the clarity, definition, energy and DDR are higher on the choir loft compared to the chapel floor, confirming that the sound field is sparser and less reverberant on the choir loft.

The reproduced acoustics of the Nassau Chapel in the AIL were evaluated using the same objective room acoustic parameters to compare the psychoacoustic characteristics of the reproduced sound field with the measured data. The objective parameters were calculated for the auralization of position X7 in the Nassau Chapel and are presented in Table 4. The results show that the objective parameters of the auralization of position X7 are similar to the measured data, with most parameters of the auralization falling within 1 JND of the measured data. This indicates that the auralization system is able to reproduce the acoustics of the Nassau Chapel with the same psychoacoustic characteristics as the measured data. The auralization does show a significantly lower DRR compared to the measured data, indicating that the direct-path component of the sound field is less pronounced in the auralization compared to the measured data. Further investigation is needed to determine the cause of this discrepancy and to improve the auralization system to better match the measured data.

## 9. Conclusions

In this paper, we have presented a comprehensive review regarding the state of research in the field of acoustic preservation of historical Historical Worship Spaces (HWS) through auralization. Acoustic preservation through auralization is a multidisciplinary field that integrates acoustics, signal processing, and psychoacoustics to faithfully recreate the acoustic characteristics of these spaces. Over recent decades, significant advancements have been made, particularly in the development of methods for the acquisition, analysis, synthesis, and evaluation of historical worship space acoustics. The literature reviewed has focused on various stages of the auralization process, including room acoustic measurement, sound field analysis, synthesis techniques, real-time auralization, and perceptual assessment. Furthermore, applications of these methods to historical HWS have been explored in depth.

While established methods such as Higher-Order Ambisonic (HOA), Spherical Harmonic Decomposition (SHD), and Wave Field Synthesis

**Table 4**
Objective acoustic parameters for all measurement positions in the Nassau Chapel, averaged from five consecutive measurements at each position. Objective parameters were also calculated for the Alamire Interactive Lab and the auralization of position X7 in the Nassau Chapel. The objective parameters were calculated according to single number frequency averaging in the 500 Hz to 1 kHz octave band, in accordance with the ISO:3382-1 standard [34].

| Position | $T_{20}$ [s] | EDT [s] | $C_{80}$ [dB] | $D_{50}$ [dB] | $T_s$ [s] | DRR [dB] |
|---|---|---|---|---|---|---|
| X1 | 5.88 | 5.22 | −0.18 | 0.43 | 0.26 | −6.62 |
| Y1 | 5.79 | 5.47 | 0.32 | 0.46 | 0.25 | −6.30 |
| X2 | 5.40 | 4.27 | 5.56 | 0.75 | 0.11 | −0.14 |
| Y2 | 5.32 | 3.76 | 6.02 | 0.77 | 0.10 | 0.32 |
| X3 | 5.83 | 5.25 | 0.15 | 0.44 | 0.26 | −5.12 |
| Y3 | 5.84 | 5.40 | 0.72 | 0.47 | 0.24 | −5.36 |
| X4 | 5.49 | 4.26 | 6.20 | 0.79 | 0.10 | 0.63 |
| Y4 | 5.42 | 3.96 | 6.22 | 0.79 | 0.10 | 0.01 |
| X5 | 5.87 | 5.84 | 0.12 | 0.44 | 0.27 | −6.43 |
| Y5 | 5.84 | 5.41 | 0.08 | 0.45 | 0.25 | −6.47 |
| X6 | 6.07 | 5.85 | −6.74 | 0.12 | 0.42 | −11.53 |
| Y6 | 6.06 | 6.26 | −6.13 | 0.14 | 0.43 | −15.41 |
| X7 | 6.15 | 6.11 | −6.64 | 0.12 | 0.44 | −11.54 |
| Y7 | 6.04 | 6.02 | −5.75 | 0.13 | 0.43 | −13.96 |
| X8 | 6.18 | 6.23 | −7.86 | 0.09 | 0.47 | −15.35 |
| Y8 | 6.07 | 6.06 | −7.43 | 0.10 | 0.45 | −15.38 |
| X9 | 6.17 | 6.14 | −6.11 | 0.15 | 0.43 | −14.84 |
| Y9 | 6.05 | 5.77 | −6.15 | 0.14 | 0.42 | −11.68 |
| X10 | 6.05 | 6.11 | −5.84 | 0.14 | 0.41 | −15.64 |
| Y10 | 6.20 | 6.21 | −6.82 | 0.11 | 0.44 | −11.31 |
| Alamire Interactive Lab | 0.54 | 0.55 | 9.72 | 0.76 | 0.03 | −7.01 |
| Auralization of X7 | 5.98 | 6.42 | −6.87 | 0.12 | 0.48 | −24.32 |

(WFS) have proven effective in concert hall acoustics, their application to historical HWS has been limited. Future research should focus on applying and adapting these techniques to the analysis and auralization of historical HWS. Additionally, the development of interactive auralization tools that allow musicians, researchers, and other stakeholders to experience and interact with the acoustics of historical HWS in real-time is an important avenue for future work. Overcoming challenges related to computational complexity, latency, and acoustic feedback will be crucial for achieving seamless, interactive simulations.

Additionally, the evaluation of the perceptual impact of historical worship space acoustics and auralization systems has primarily relied on objective acoustic parameters to evaluate the psychoacoustic effects on listeners. However, to gain a better understanding of the subjective experience of listeners and how the auralized spaces are perceived, future research should incorporate subjective and sensory evaluation methods.

Several reviewed methods were illustrated in a case study of the Nassau Chapel, where auralization techniques were applied to preserve and valorize the chapel's acoustics through real-time interactive auralization. This case study demonstrated the potential of interactive auralization tools to support preservation efforts, offering musicians the opportunity to rehearse and perform in a virtual recreation of the chapel's original acoustics.

This review paper serves as a reference for researchers, acousticians, architects, and cultural heritage professionals, providing a framework for future preservation projects. An outline of the general process for acquiring, analyzing, synthesizing, and evaluating the acoustics of these spaces has been presented, which can help guide future research in the field.

In conclusion, the preservation and auralization of historical HWS is an important interdisciplinary field of research that combines acoustics, signal processing, and psychoacoustics to safeguard and recreate the original acoustic characteristics of these spaces. The reviewed research offers a solid foundation for further exploration, pointing to key areas for future development.

## Acronyms

**AA** Acoustic Acquisition
**AIL** Alamire Interactive Lab
**ALLRAD** All-Round Ambisonic Decoding
**ART** Acoustic Ray Tracing
**ASW** Apparent Source Width
**CAD** Computer Aided Design
**CVP** Consensus Vocabulary Profiling
**DOA** Direction of Arrival
**DOF** Degrees of Freedom
**DRR** Direct-to-Reverberant Ratio
**EDT** Early Decay Time
**ESS** Exponential Sine Sweep
**FCP** Free Choice Profiling
**FCR** Feedback Canceling Reverberator
**FDN** Feedback Delay Network
**FFT** Fast Fourier Transform
**FIR** Finite Impulse Response
**FOA** First-Order Ambisonics
**FP** Flash Profile
**FVN** Filtered Velvet Noise
**FWHT** Fast Walsh-Hadamard Transform
**GAN** Generative Adversarial Network
**GCC** Generalized Cross-Correlation
**GPA** Generalized Procrustes Analysis
**GPU** Graphics Processing Unit
**HOA** Higher-Order Ambisonic
**HRTF** Head-Related Transfer Function
**HWS** Historical Worship Spaces
**ICC** Interaural Cross-Correlation
**IMU** Inertial Measurement Unit
**IR** Impulse Response
**ISM** Image-Source Method
**IVP** Individual Vocabulary Profiling
**JND** Just Noticeable Difference
**LTI** Linear Time-Invariant
**MDAP** Multiple-Direction Amplitude Panning
**MFA** Multiple Factor Analysis
**MLS** Maximum Length Sequence
**MUSHRA** MUltiple Stimuli with Hidden Reference and Anchor

**NAF** Neural Acoustic Field
**NLS** Nearest Loudspeaker Synthesis
**NUPC** Non-Uniform Partitioned Convolution
**OLA** Overlap-Add
**OLS** Overlap-Save
**PCA** Principal Component Analysis
**PE** Perceptual Evaluation
**QDA** Quantitative Descriptive Analysis
**RIR** Room Impulse Response
**RT** Reverberation Time
**RTA** Real-Time Auralization
**SDM** Spatial Decomposition Method
**SDN** Scattering Delay Network
**SFA** Sound Field Analysis
**SFS** Sound Field Synthesis
**SHD** Spherical Harmonic Decomposition
**SNR** Signal-to-Noise Ratio
**STI** Speech Transmission Index
**SVD** Singular Value Decomposition
**VBAP** Vector Base Amplitude Panning
**VR** Virtual Reality
**WFA** Wave Field Analysis
**WFS** Wave Field Synthesis

## CRediT authorship contribution statement

**Hannes Rosseel:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis. **Toon van Waterschoot:** Writing – review & editing, Supervision.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

The authors express their gratitude to the Alamire Foundation and Ann Kelders for their invaluable information and late night access to the Nassau Chapel. Additionally, special thanks are extended to Randall Ali for assisting with the acoustic measurements conducted in the Nassau Chapel. This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU Leuven internal funds C3/23/056 "HELIXON: Hybrid, efficient, and liquid interpolation of sound in extended reality" and C14/21/075 "A holistic approach to the design of integrated and distributed digital signal processing algorithms for audio and speech communication devices", FWO SBO Project The sound of music - Innovative research and valorization of plainchant through digital technology" (S005319N), FWO Large-scale research infrastructure "The Library of Voices - Unlocking the Alamire Foundation's Music Heritage Resources Collection through Visual and Sound Technology" (I013218N), and SBO Project "New Perspectives on Medieval and Renaissance Courtly Song" (S005525N). The research leading to these results has received funding from the Flemish Government under the AI Research Program and from the European Research Council under the European Union's Horizon 2020 research and innovation program/ERC Consolidator Grant: SONORA (no. 773268). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

## Data availability

Data will be made available on request.

## References

[1] S. Girón, L. Álvarez Morales, T. Zamarreño, Church acoustics: A state-of-the-art review after several decades of research, J. Sound Vib. 411 (2017) 378–408.
[2] B.F.G. Katz, D. Murphy, A. Farina, Exploring cultural heritage through acoustic digital reconstructions, Phys. Today 73 (12) (2020) 32–37.
[3] B.F.G. Katz, D. Murphy, A. Farina, The past has ears (PHE): XR explorations of acoustic spaces as cultural heritage, in: Augmented Reality, Virtual Reality, and Computer Graphics, Springer International Publishing, Cham, 2020, pp. 91–98.
[4] Unesco, Basic texts of the 2003 convention for the safeguarding of the intangible cultural heritage, 2022.
[5] D. Zhang, Y. Shan, X. Chen, Z. Wang, Soundscape in religious historical buildings: a review, Herit. Sci. 12 (1) (2024) 45.
[6] M. Vorländer, Auralization, Springer, Berlin, Heidelberg, 2020.
[7] C. Weitze, C. Christensen, J. Rindel, Comparison between in-situ recordings and auralizations for mosques and Byzantine churches, in: Proc. of Joint Baltic-Nordic Acoustical Meeting, 2003, pp. 53–57.
[8] P. Koutsivitis, C. Tsakostas, A. Floros, D. Troupis, N.-A. Tatlas, Reproduction of audio-visual interactive events in virtual ancient Greek spaces, in: Forum Acusticum 2005, 2005, pp. 2481–2486.
[9] F. Martellotta, Subjective study of preferred listening conditions in Italian catholic churches, J. Sound Vib. 317 (1) (2008) 378–399.
[10] B. Fazenda, I. Drumm, Recreating the sound of Stonehenge, Acta Acust. United Acust. 99 (1) (2013) 110–117.
[11] A. Pedrero, A. Díaz-Chyla, C. Díaz, S. Pelzer, M. Vorländer, Virtual restoration of the sound of the Hispanic rite, in: Proc. of Forum Acusticum 2014, 2014, pp. 1–5.
[12] B.N.J. Postma, B.F.G. Katz, Creation and calibration method of acoustical models for historic virtual reality auralizations, Virtual Real. 19 (3) (2015) 161–180.
[13] L. Elicio, F. Martellotta, Acoustics as a cultural heritage: The case of Orthodox churches and of the Russian church in Bari, J. Cult. Herit. 16 (6) (2015) 912–917.
[14] M. Frank, F. Zotter, Spatial impression and directional resolution in the reproduction of reverberation, in: Proc. 42. Jahrestagung Für Akustik (DAGA), vol. 2016, 2016, pp. 1304–1307.
[15] B.N.J. Postma, D. Poirier-Quinot, J. Meyer, B.F.G. Katz, Virtual reality performance auralization in a calibrated model of Notre-Dame Cathedral, in: Proc. EuroRegio 2016, 2016, pp. 1–10.
[16] B.N.J. Postma, B.F.G. Katz, Perceptive and objective evaluation of calibrated room acoustic simulation auralizations, J. Acoust. Soc. Amer. 140 (6) (2016) 4326–4337.
[17] L. Álvarez Morales, J.F. Molina-Rozalem, S. Girón, A. Alonso, B. Pedro, Álvarez Corbacho Ángel, Virtual reality in church acoustics: Visual and acoustic experience in the cathedral of seville, in: Proc. 24th Int. Congress on Sound and Vibration, 2017, pp. 1–8.
[18] F. Martellotta, L. Álvarez Morales, S. Girón, T. Zamarreño, An investigation of multi-rate sound decay under strongly non-diffuse conditions: The crypt of the Cathedral of Cadiz, J. Sound Vib. 421 (2018) 261–274.
[19] R. Suárez, A. Alonso, J.J. Sendra, Virtual acoustic environment reconstruction of the hypostyle mosque of Cordoba, Appl. Acoust. 140 (2018) 214–224.
[20] E.K. Canfield-Dafilou, E.F. Callery, J.S. Abel, J. Berger, A method for studying interactions between music performance and rooms with real-time virtual acoustics, in: Proc. AES 146th Convention, 2019, pp. 1–10.
[21] B.F. Katz, A. Weber, An acoustic survey of the Cathédrale Notre-Dame de Paris before and after the fire of 2019, Acoustics 2 (4) (2020) 791–802.
[22] L. Álvarez-Morales, M. Lopez, Á. Álvarez-Corbacho, The acoustic environment of York Minster's Chapter House, Acoustics 2 (1) (2020) 13–36.
[23] H. Autio, M. Barbagallo, C. Ask, D. Bard Hagberg, E. Lindqvist Sandgren, K. Strinnholm Lagergren, Historically based room acoustic analysis and auralization of a church in the 1470s, Appl. Sci. 11 (4) (2021).
[24] N. Eley, S. Mullins, P. Stitt, B.F. Katz, Virtual Notre-Dame: Preliminary results of real-time auralization with choir members, in: Proc. 2021 Immersive and 3D Audio: From Architecture To Automotive (I3DA), 2021, pp. 1–6.
[25] O. Warusfel, S. Emerit, Assessing the acoustics of an ancient Egyptian temple, in: Proc. 2021 Immersive and 3D Audio: From Architecture To Automotive (I3DA), 2021, pp. 1–6.
[26] J. Berger, N. Farzaneh, E. Murakami, L. Valentin, Exploring the past with virtual acoustics and virtual reality, in: Proc. 2023 Immersive and 3D Audio: From Architecture To Automotive (I3DA), 2023, pp. 1–8.
[27] S.S. Mullins, B.F.G. Katz, Immersive auralisations for choral ensembles, in: Proc. 11th Int. Conf. Auditorium Acoustics, vol. 45, Institute of Acoustics, Athens, Greece, 2023, pp. 1–8.
[28] D. Schröder, M. Vorländer, RAVEN: A real-time framework for the auralization of interactive virtual environments, in: Proc. of the Forum Acusticum 2011, 2011, pp. 1541–1546.
[29] D. Murphy, F. Stevens, [Openair], 2024, accessed: 2024-10-30. URL https://www.openair.hosted.york.ac.uk/.
[30] J.S. Abel, E.F. Callery, E.K. Canfield-Dafilou, A feedback canceling reverberator, in: Proc. 21st Int. Conf. Digital Audio Effects (DAFx '18), 2018, pp. 100–106.

[31] M. Vorländer, M. Kob, Practical aspects of MLS measurements in building acoustics, Appl. Acoust. 52 (3) (1997) 239–258.

[32] K. Prawda, S. Schlecht, V. Välimäki, Time variance in measured room impulse responses, in: Proc. Forum Acusticum 2023, Torino, Italy, 2023, pp. 1639–1646.

[33] S. Tervo, J. Pätynen, A. Kuusinen, T. Lokki, Spatial decomposition method for room impulse responses, J. Audio Eng. Soc. 61 (1/2) (2013) 17–28.

[34] International Organization for Standardization, Acoustics - measurement of room acoustic parameters - part 1: Performance spaces, International Standard 3382-1, ISO, Geneva, Switzerland, 2009.

[35] International Organization for Standardization, Acoustics - measurement of room acoustic parameters - part 2: Reverberation time in ordinary rooms, International Standard 3382-2, Iso, Geneva, Switzerland, 2008.

[36] H. Kuttruff, Room Acoustics, Sixth ed., CRC Press, 2017.

[37] D. Griesinger, Beyond MLS-occupied hall measurement with FFT techniques, in: Preprints AES 101st Convention, Los Angeles, CA, 1996, pp. 1–23.

[38] M. Horvat, K. Jambrošić, H. Domitrović, A comparison of impulse-like sources to be used in reverberation time measurements, J. Acoust. Soc. Amer. 123 (2008) 3501.

[39] J. Pätynen, B.F. Katz, T. Lokki, Investigations on the balloon as an impulse source, J. Acoust. Soc. Amer. 129 (1) (2011) El27–el33.

[40] M. Lamothe, J. Bradley, Acoustical characteristics of guns as impulse sources, Can. Acoust. 13 (2) (1985) 16–24.

[41] M.R. Schroeder, Integrated-impulse method measuring sound decay without using impulses, J. Acoust. Soc. Amer. 66 (2) (1979) 497–500.

[42] J. Borish, J.B. Angell, An efficient algorithm for measuring the impulse response using pseudorandom noise, J. Audio Eng. Soc. 31 (1983) 478–488.

[43] D.D. Rife, J. Vanderkooy, Transfer-function measurement with maximum-length sequences, J. Audio Eng. Soc. 37 (1989) 419–444.

[44] C. Dunn, M. Hawksford, Distortion immunity of MLS-derived impulse response measurements, J. Audio Eng. Soc. 41 (1993) 314–335.

[45] J. Vanderkooy, Aspects of MLS measuring systems, J. Audio Eng. Soc. 42 (1994) 219–231.

[46] A. Novak, F. Rund, P. Honzik, Impulse response measurements using mls technique on nonsynchronous devices, J. Audio Eng. Soc. 64 (2016) 978–987.

[47] F. Jacobsen, P.M. Juhl, Fundamentals of General Linear Acoustics, John Wiley & Sons, 2013, pp. 260–262, Ch. The Method Based on 'Maximum Length Sequences' (MLS).

[48] J.L. Nielsen, Improvement of signal-to-noise ratio in long-term MLS measurements with high-level nonstationary disturbances, J. Audio Eng. Soc. 45 (1997) 1063–1066.

[49] P. Svensson, J.L. Nielsen, Errors in MLS measurements caused by time variance in acoustic systems, J. Audio Eng. Soc. 47 (1999) 907–927.

[50] A. Farina, Simultaneous measurement of impulse response and distortion with a swept-sine technique, in: Preprints Audio Eng. Soc. Convention, vol. 108, 2000.

[51] S. Müller, P. Massarani, Transfer-function measurement with sweeps, J. Audio Eng. Soc. 49 (2001).

[52] G.-B. Stan, J.-J. Embrechts, D. Archambeau, Comparison of different impulse response measurement techniques, J. Audio Eng. Soc. 50 (2002) 249–262.

[53] P. Guidorzi, L. Barbaresi, D. D'Orazio, M. Garai, Impulse responses measured with mls or swept-sine signals applied to architectural acoustics: An in-depth analysis of the two methods and some case studies of measurements inside theaters, Energy Procedia 78 (2015) 1611–1616.

[54] A.J. Berkhout, D. de Vries, M.M. Boone, A new method to acquire impulse responses in concert halls, J. Acoust. Soc. Amer. 68 (1) (1980) 179–183.

[55] A. Farina, Advancements in impulse response measurements by sine sweeps, in: Preprints AES 122th Convention, 2007, pp. 1–21.

[56] K. Prawda, S.J. Schlecht, V. Välimäki, Non-stationary noise removal from repeated sweep measurements, JASA Express Lett. 4 (8) (2024) 081601.

[57] E.K. Canfield-Dafilou, J.S. Abel, An allpass chirp for constant signal-to-noise ratio impulse response measurement, in: Proc. AES 144th Convention, 2018, pp. 1–7.

[58] F. Martellotta, E. Cirillo, A. Carbonari, P. Ricciardi, Guidelines for acoustical measurements in churches, Appl. Acoust. 70 (2) (2009) 378–388.

[59] S. Yannick, V. Anneleen, M. Sanne, R. Monika, D.J. Krista, An archaeoacoustic reconstruction of the great hall of the palace of Binche (1545-1554), in: Proc. 2021 Immersive and 3D Audio: From Architecture To Automotive (I3DA), 2021, pp. 1–8.

[60] Y. Cui, Q. Li, B. Yang, W. Xiao, C. Chen, Z. Dong, Automatic 3-d reconstruction of indoor environment with mobile laser scanning point clouds, IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 12 (8) (2019) 3117–3130.

[61] S. Montelpare, M. Pierantozzi, S. Di Loreto, A. Ricciutelli, M. ferrara, Geometrical acoustics in cultural heritage conservation and promotion: Digitalization of the acoustic characteristics, in: Advanced Research in Technologies, Information, Innovation and Sustainability, Springer Nature, Switzerland, 2024, pp. 292–303.

[62] M. Vorländer, Computer simulations in room acoustics: Concepts and uncertainties, J. Acoust. Soc. Amer. 133 (3) (2013) 1203–1213.

[63] J. Saarelma, Finite-difference time-domain method for room acoustic simulation - auralization and applications, (Ph.D. thesis), Aalto University, Finland, 2019.

[64] L. Savioja, U.P. Svensson, Overview of geometrical room acoustic modeling techniques, J. Acoust. Soc. Amer. 138 (2) (2015) 708–730.

[65] L. Savioja, Real-time 3D finite-difference time-domain simulation of low- and mid-frequency room acoustics, in: Proc. 13th Int. Conf. Digital Audio Effects (DAFx '10), 2010, pp. 1–8.

[66] L. Savioja, D. Manocha, M. Lin, Use of GPUs in room acoustic modeling and auralization, in: Proc. Int. Symposium Room Acoustics (ISRA 2010), 2010, pp. 1–7.

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

[68] J. Borish, Extension of the image model to arbitrary polyhedra, J. Acoust. Soc. Amer. 75 (6) (1984) 1827–1836.

[69] E. De Sena, N. Antonello, M. Moonen, T. van Waterschoot, On the modeling of rectangular geometries in room acoustic simulations, IEEE/ ACM Trans. Audio Speech Lang. Process. 23 (4) (2015) 774–786.

[70] N. Declercq, J. Degrieck, R. Briers, O. Leroy, A theoretical study of special acoustic effects caused by the staircase of the EI Castillo pyramid at the Maya ruins of Chichen-Itza in Mexico, J. Acoust. Soc. Amer. 116 (2005) 3328–3335.

[71] F.A. Bilsen, Repetition pitch glide from the step pyramid at Chichen Itza, J. Acoust. Soc. Amer. 120 (2) (2006) 594–596.

[72] D. Schröder, T. Lentz, Real-time processing of image sources using binary space partitioning, J. Audio Eng. Soc. 54 (2006) 604–619.

[73] M. Vorländer, Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm, J. Acoust. Soc. Amer. 86 (1) (1989) 172–178.

[74] G. Naylor, ODEON – another hybrid room acoustical model, Appl. Acoust. 38 (2) (1993) 131–143.

[75] A. Krokstad, S. Strøm, S. Sorsdal, Calculating the acoustical room response by the use of a ray tracing technique, J. Sound Vib. 8 (1) (1968) 118–125.

[76] A. Krokstad, U.P. Svensson, S. Strøm, Acoustics, Information, and Communication: Memorial Volume in Honor of Manfred R. Schroeder, Springer International Publishing, Cham, 2015, pp. 15–31, Ch. The Early History of Ray Tracing in Acoustics.

[77] H. Autio, N.-G. Vardaxis, D. Bard Hagberg, An iterative ray tracing algorithm to increase simulation speed while maintaining overall precision, Acoustics 5 (1) (2023) 320–342.

[78] J.K. Haviland, B.D. Thanedar, Monte Carlo applications to acoustical field solutions, J. Acoust. Soc. Amer. 54 (6) (1973) 1442–1448.

[79] J. Segura, A. Giménez, J. Romero, S. Cerdá, A comparison of different techniques for simulating and measuring acoustic parameters in a place of worship: Sant Jaume Basílica in Valencia, Spain, Acta Acust. United Acust. 97 (1) (2011) 155–170.

[80] J.-P. Vian, J. Martin, Binaural room acoustics simulation: Practical uses and applications, Appl. Acoust. 36 (3) (1992) 293–305.

[81] D. van Maercke, J. Martin, The prediction of echograms and impulse responses within the Epidaure software, Appl. Acoust. 38 (2) (1993) 93–114.

[82] CATT-Acoustic, [Catt-acoustic] software, 2024, accessed: 2024-09-12. URL https://www.catt.se.

[83] A. Farina, P. Galaverna, G. Truffelli, [Ramsete]: Room acoustics modeling on PC, 2025, accessed: 2025-02-23. URL https://www.ramsete.com/.

[84] AFMG Technologies GmbH, [Ease] 5, 2024, accessed: 2024-09-12. URL https://www.afmg.eu/en/ease.

[85] Treble Technologies, [Treble], 2024, accessed: 2024-09-12. URL https://www.treble.tech/.

[86] S.V. Amengual Garí, P.W. Robinson, P.T. Calamia, Room acoustic characterization for binaural rendering: From spatial room impulse responses to deep learning, in: Proc. 24th Int. Congress on Acoustics (ICA 2022), 2022, pp. 1–10.

[87] L. Kelley, D.D. Carlo, A.A. Nugraha, M. Fontaine, Y. Bando, K. Yoshii, RIR-in-a-box: Estimating room acoustics from 3D mesh data through shoebox approximation, in: Proc. Interspeech 2024, 2024, pp. 3255–3259.

[88] J. Lin, G. Götz, S.J. Schlecht, Deep room impulse response completion, 2024, unpublished, arXiv:2402.00859.

[89] T. Kim, Y. Kwon, S.-E. Yoon, Real-time 3-D mapping with estimating acoustic materials, in: Proc. IEEE/SICE Int. Symposium on System Integration (SII 2020), 2020, pp. 646–651.

[90] Z. Tang, N.J. Bryan, D. Li, T.R. Langlois, D. Manocha, Scene-aware audio rendering via deep acoustic analysis, IEEE Trans. Vis. Comput. Graphics 26 (5) (2020) 1991–2001.

[91] A. Ratnarajah, Z. Tang, R. Aralikatti, D. Manocha, MESH2IR: Neural acoustic impulse response generator for complex 3D scenes, in: Proc. 30th ACM Int. Conf. Multimedia, 2022, pp. 924–933.

[92] A. Luo, Y. Du, M.J. Tarr, J.B. Tenenbaum, A. Torralba, C. Gan, Learning neural acoustic fields, in: Proc. 36th Conf. Neural Information Processing Systems (NeurIPS 2022), 2022, pp. 1–13.

[93] J. Navarro, J.J. Sendra, S. Muñoz, The western latin church as a place for music and preaching: An acoustic assessment, Appl. Acoust. 70 (6) (2009) 781–789.

[94] P. Małecki, J. Wiciak, D. Nowak, Acoustics of Orthodox Churches in Poland, Arch. Acoust. 42 (4) (2017) 579–590.

[95] A.C. Raes, G.G. Sacerdote, Measurements of the acoustical properties of two Roman basilicas, J. Acoust. Soc. Amer. 25 (5) (1953) 954–961.

[96] J. Anderson, M. Bratos-Anderson, Acoustic coupling effects in St Paul's Cathedral, London, J. Sound Vib. 236 (2) (2000) 209–225.
[97] R.S. Shankland, H.K. Shankland, Acoustics of St. Peter's and patriarchal basilicas in Rome, J. Acoust. Soc. Amer. 50 (2a) (1971) 389–396.
[98] Z. Sü Gül, M. Çalışkan, Impact of design decisions on acoustical comfort parameters: Case study of Doğramacızade Ali Paşa Mosque, Appl. Acoust. 74 (6) (2013) 834–844.
[99] K. Kosał a, Z.W. Engel, Assessing the acoustic properties of Roman Catholic churches: A new approach, Appl. Acoust. 74 (10) (2013) 1144–1152.
[100] K. Kosał a, Calculation models for acoustic analysis of St. Elizabeth of Hungary church in Jaworzno Szczakowa, Arch. Acoust. 41 (3) (2016) 485–498.
[101] A.A. Abdou, Measurement of acoustical characteristics of mosques in Saudi Arabia, J. Acoust. Soc. Amer. 113 (3) (2003) 1505–1517.
[102] E. Cirillo, F. Martellotta, Acoustics of Apulian-Romanesque churches: An experimental survey, Build. Acoust. 9 (4) (2002) 271–288.
[103] E. Cirillo, F. Martellotta, Acoustics of Apulian-Romanesque churches: Correlations between architectural and acoustic parameters, Build. Acoust. 10 (1) (2003) 55–76.
[104] M. Galindo, S. Girón, T. Zamarreño, Correlations of the acoustic fields of Mudejar-Gothic Churches, Build. Acoust. 15 (3) (2008) 175–195.
[105] M. Galindo, T. Zamarreño, S. Girón, Acoustic simulations of Mudejar-Gothic churches, J. Acoust. Soc. Amer. 126 (3) (2009) 1207–1218.
[106] V. Desarnaulds, A.P. Carvalho, G. Monay, Church acoustics and the influence of occupancy, Build. Acoust. 9 (1) (2002) 29–47.
[107] G. Iannace, Acoustic correction of monumental churches with ceramic material: The case of the Cathedral of Benevento (Italy), J. Low Freq. Noise, Vib. Act. Control. 35 (3) (2016) 230–239.
[108] D. De Vries, S. Brix, E.M. Hulsebos, Measurements of church impulse responses using a circular microphone array for natural spatial reproduction of a choir concert recording, J. Audio Eng. Soc. (2002).
[109] L. Tronchin, D. Knight, The acoustical survival of san vitale, ravenna, Italy through two millennia, in: Proc. Auditorium Acoustics 2008, Vol. 30, Institute of Acoustics, Oslo, 2008, pp. 301–308.
[110] F. Martellotta, Identifying acoustical coupling by measurements and prediction-models for St. Peter's Basilica in Rome, J. Acoust. Soc. Amer. 126 (3) (2009) 1175–1186.
[111] Y. Soeta, K. Ito, R. Shimokura, S. i. Sato, T. Ohsawa, Y. Ando, Effects of sound source location and direction on acoustic parameters in Japanese churches, J. Acoust. Soc. Amer. 131 (2) (2012) 1206–1220.
[112] A. Alonso, F. Martellotta, Room acoustic modelling of textile materials hung freely in space: from the reverberation chamber to ancient churches, J. Build. Perform. Simul. 9 (5) (2016) 469–486.
[113] F. Martellotta, Understanding the acoustics of Papal Basilicas in rome by means of a coupled-volumes approach, J. Sound Vib. 382 (2016) 413–427.
[114] D. D'Orazio, G. Fratoni, E. Rossi, M. Garai, Understanding the acoustics of St. John's Baptistery in Pisa through a virtual approach, J. Build. Perform. Simul. 13 (3) (2020) 320–333.
[115] L. Tronchin, A. Bevilacqua, Evaluation of acoustic similarities in two Italian churches honored to S. Dominic, Appl. Sci. 10 (20) (2020).
[116] E. Alberdi, M. Galindo, A. León-Rodríguez, J. León, Acoustic behaviour of polychoirs in the Baroque church of Santa María Magdalena, Seville, Appl. Acoust. 175 (2021) 107814.
[117] A. Ricciutelli, V. Lori, C. Lops, F. Serpilli, S. Montelpare, Churches acoustics as intangible cultural heritage: Experimental and numerical characterization of the temple of san biagio, Int. J. Archit. Herit. 18 (7) (2024) 1117–1130.
[118] L. Álvarez, A. Alonso, T. Zamarreño, S. Girón, M. Galindo, Virtual acoustics of the cathedral of malaga (Spain), in: Proc. of the Forum Acusticum 2011, 2011, pp. 1–6.
[119] A. Alonso, J.J. Sendra, R. Suárez, T. Zamarreño, Acoustic evaluation of the cathedral of Seville as a concert hall and proposals for improving the acoustic quality perceived by listeners, J. Build. Perform. Simul. 7 (5) (2014) 360–378.
[120] A. Alonso, R. Suárez, J.J. Sendra, Virtual reconstruction of indoor acoustics in cathedrals: The case of the Cathedral of Granada, Build. Simul. 10 (4) (2017) 431–446.
[121] Y. Soeta, R. Shimokura, Y.H. Kim, T. Ohsawa, K. Ito, Measurement of acoustic characteristics of Japanese Buddhist temples in relation to sound source location and direction, J. Acoust. Soc. Amer. 133 (5) (2013) 2699–2710.
[122] P. Fausti, R. Pompoli, N. Prodi, Comparing the acoustics of mosques and Byzantine churches, in: Proc. of the 19th International. Symposium CIPA, 2003, pp. 1–5.
[123] Z. Sü Gül, Exploration of room acoustics coupling in Hagia Sophia of Istanbul for its different states, J. Acoust. Soc. Amer. 149 (1) (2021) 320–339.
[124] K. Kitapci, G. Çelik Başok, The acoustic characterization of worship ambiance and speech intelligibility in wooden hypostyle structures: The case of the Aslanhane Mosque, Acoust. Aust. 49 (3) (2021) 425–440.
[125] T.J. Cox, B.M. Fazenda, S.E. Greaney, Using scale modelling to assess the prehistoric acoustics of Stonehenge, J. Archaeol. Sci. 122 (2020) 105218.
[126] E.G. Tzekakis, Reverberation time of the Rotunda of Thessaloniki, J. Acoust. Soc. Amer. 57 (5) (1975) 1207–1209.
[127] A. Pyshin, L. Solovejtchik, Y. Tikhomirov, S. Yavorsky, Acoustics of the Smolny Cathedral in St. Petersburg, Build. Acoust. 8 (1) (2001) 75–82.
[128] A. Pedrero, R. Ruiz, A. Díaz-Chyla, C. Díaz, Acoustical study of Toledo Cathedral according to its liturgical uses, Appl. Acoust. 85 (2014) 23–33.
[129] A. Elkhateeb, S. Eldakdoky, The acoustics of Mamluk masjids: A case study of Iwan-type masjids in Cairo, Appl. Acoust. 178 (2021).
[130] S. Sukaj, U. Berardi, G. Ciaburro, G. Iannace, A. Trematerra, A. Bevilacqua, Two Albanian mosques: the acoustics discovery inside prayer rooms, in: Proc. 2021 Immersive and 3D Audio: From Architecture To Automotive (I3DA), 2021, pp. 1–6.
[131] S. Sukaj, A. Bevilacqua, G. Iannace, I. Lombardi, R. Parente, A. Trematerra, Byzantine churches in Albania: How geometry and architectural composition influence the acoustics, Buildings 12 (3) (2022).
[132] W. Orfali, W. Ahnert, Acoustical simulation and auralization in mosques, in: Preprints AES 118th Convention, 2005, pp. 1–10.
[133] Z. Sü Gül, N. Xiang, M. Çalışkan, Investigations on sound energy decays and flows in a monumental mosque, J. Acoust. Soc. Amer. 140 (1) (2016) 344–355.
[134] M. Kavraz, The acoustic characteristics of the çarşı mosque in trabzon, turkey, Indoor Built Environ. 25 (1) (2016) 128–136.
[135] A. Bellia, Experiencing sounds in ancient spaces: An overview, Rev. Do Mus. de Arqueol. E Etnol. 41 (2023) 17–31.
[136] R. Suárez, J.J. Sendra, A. Alonso, Acoustics, liturgy and architecture in the Early Christian Church. from the domus ecclesiae to the basilica, Acta Acust. United Acust. 99 (2) (2013) 292–301.
[137] U. Berardi, Simulation of acoustical parameters in rectangular churches, J. Build. Perform. Simul. 7 (1) (2013).
[138] L. Alvarez-Morales, F. Martellotta, A geometrical acoustic simulation of the effect of occupancy and source position in historical churches, Appl. Acoust. 91 (2015) 47–58.
[139] R. Suárez, A. Alonso, J.J. Sendra, Archaeoacoustics of intangible cultural heritage: The sound of the Maior Ecclesia of Cluny, J. Cult. Herit. 19 (2016) 567–572.
[140] S. Zhao, E. Cheng, X. Qiu, P. Alambeigi, J. Burry, M.C. Burry, A preliminary investigation on the sound field properties in the Sagrada Família Basilica, in: Proc. of Acoustics 2016, Brisbane, Australia, 2016, pp. 1–10.
[141] P.R. Sender, S.J., G.A., Virtual acoustic reconstruction of a lost church: application to an Order of Saint Jerome monastery in Alzira, Spain, J. Build. Perform. Simul. 11 (3) (2018) 369–390.
[142] D. Hudokova, V. Chmelik, D. Urban, L. Zelem, M. Rychtarikova, Impact of changes in liturgy on speech intelligibility in catholic church. Investigation through a room acoustic prediction software, in: Proc. 2021 Immersive and 3D Audio: From Architecture To Automotive (I3DA), Bologna, Italy, 2021, pp. 1–7.
[143] E. Alberdi, M. Galindo, Á.L. León-Rodríguez, Evolutionary analysis of the acoustics of the Baroque church of San Luis de los Franceses (Seville), Appl. Sci. 11 (4) (2021).
[144] J.B. Pantea Alambeigi, M. Burry, Auralising the soundscape of sagrada família basilica: a virtual journey through sound, Archit. Sci. Rev. (2024) 1–11.
[145] R. Suárez, A. Alonso, J.J. Sendra, Intangible cultural heritage: The sound of the Romanesque cathedral of Santiago de Compostela, J. Cult. Herit. 16 (2) (2015) 239–243.
[146] F. Martellotta, L. Pon, On-site acoustical characterization of Baroque tapestries: The Barberini collection at St. John the Divine Cathedral, J. Acoust. Soc. Amer. 144 (3) (2018) 1615–1626.
[147] S. Mullins, E.K. Canfield-Dafilou, B.F.G. Katz, The development of the early acoustics of the chancel in Notre-Dame de Paris: 1160–1230, in: Proc. 2nd Symposium: The Acoustics of Ancient Theatres, Verona, Italy, 2022, pp. 1–4.
[148] E.K. Canfield-Dafilou, S. Mullins, B.F.G. Katz, Opening the lateral chapels and the acoustics of Notre-Dame de Paris: 1225–1320, in: Proc. 2nd Symposium: The Acoustics of Ancient Theatres, Verona, Italy, 2022, pp. 1–4.
[149] E.K. Canfield-Dafilou, S.S. Mullins, B.F.G. Katz, Can you hear the paintings? the effect of votive offerings on the acoustics of Notre-Dame, in: Proc. Forum Acusticum 2023, 2023, pp. 2537–2544.
[150] E.K. Canfield-Dafilou, N. Buchs, B. Caseau Chevallier, The voices of children in Notre-Dame de Paris during the late middle ages and the modern period, J. Cult. Herit. 65 (2024) 160–168.
[151] J. De Muynke, M. Baltazar, M. Monferran, M. Voisenat, B.F.G. Katz, Ears of the past, an inquiry into the sonic memory of the acoustics of Notre-Dame before the fire of 2019, J. Cult. Herit. 65 (2024) 169–176.
[152] B. Postma, A. Tallon, B. Katz, Calibrated auralization simulation of the abbey of Saint-Germain-des-Prés for historical study, in: Proc. of the Institute of Acoustics, vol. 37, 2015, pp. 1–8.
[153] L. Álvarez Morales, T. Zamarreño, S. Girón, M. Galindo, A methodology for the study of the acoustic environment of Catholic cathedrals: Application to the Cathedral of Malaga, Build. Environ. 72 (2014) 102–115.
[154] G. Navas-Reascos, L.M. Alonso-Valerdi, D.I. Ibarra-Zarate, Archaeoacoustics around the world: A literature review (2016–2022), Appl. Sci. 13 (4) (2023).
[155] A.J. Berkhout, D. de Vries, J.J. Sonke, Array technology for acoustic wave field analysis in enclosures, J. Acoust. Soc. Amer. 102 (5) (1997) 2757–2770.

[156] J. Baan, D. de Vries, Array technology for discrimination between specular and non-specular reflections in enclosed spaces, in: Preprints AES 104th Convention, 1998, pp. 1–19.
[157] D. de Vries, M. Boone, Wave field synthesis and analysis using array technology, in: Proc. 1999 IEEE Workshop Appls. Signal Process. Audio Acoust., WASPAA'99, 1999, pp. 15–18.
[158] M. Kuster, D. de Vries, E.M. Hulsebos, A. Gisolf, Acoustic imaging in enclosed spaces: Analysis of room geometry modifications on the impulse response, J. Acoust. Soc. Amer. 116 (4) (2004) 2126–2137.
[159] A. Berkhout, D. de Vries, M.C. Brink, Array technology for bending wave field analysis in constructions, J. Sound Vib. 300 (1) (2007) 25–42.
[160] T.D. Abhayapala, A. Gupta, Spherical harmonic analysis of wavefields using multiple circular sensor arrays, IEEE Trans. Audio Speech Lang. Process. 18 (6) (2010) 1655–1666.
[161] H. Chen, T.D. Abhayapala, W. Zhang, 3D sound field analysis using circular higher-order microphone array, in: Proc. 23rd European Signal Processing Conference (EUSIPCO 2015), 2015, pp. 1153–1157.
[162] H. Sun, S. Yan, U.P. Svensson, H.-F. Sun, Spherical harmonics based optimal minimum sidelobe beamforming for spherical sensor arrays, in: Proc. Int. ITG Workshop Smart Antennas (WSA 2010), 2010, pp. 286–291.
[163] B. Rafaely, Y. Peled, M. Agmon, D. Khaykin, E. Fisher, Speech Processing in Modern Communication: Challenges and Perspectives, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 281–305, Ch. Spherical Microphone Array Beamforming.
[164] Q. Huang, R. Hu, Y. Fang, Real-valued MVDR beamforming using spherical arrays with frequency invariant characteristic, Digit. Signal Process. 48 (2016) 239–245.
[165] L. Kumar, R.M. Hegde, Near-field acoustic source localization and beamforming in spherical harmonics domain, IEEE Trans. Signal Process. 64 (13) (2016) 3351–3361.
[166] S. Tervo, A. Politis, Direction of arrival estimation of reflections from room impulse responses using a spherical microphone array, IEEE/ ACM Trans. Audio Speech Lang. Process. 23 (10) (2015) 1539–1551.
[167] W. Gao, H. Zhao, W. Xu, Direction of arrival estimation based on spherical harmonics decomposition, in: Proc. OCEANS 2016 MTS/IEEE Monterey, 2016, pp. 1–5.
[168] M. Park, B. Rafaely, Sound-field analysis by plane-wave decomposition using spherical microphone array, J. Acoust. Soc. Amer. 118 (5) (2005) 3094–3103.
[169] D.N. Zotkin, R. Duraiswami, N.A. Gumerov, Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays, IEEE Trans. Audio Speech Lang. Process. 18 (1) (2010) 2–16.
[170] M. Imran, J.Y. Jeon, A. Hussain, Plane wave decomposition and beamforming for directional spatial sound localization, in: Proc. 2016 13th Int. Bhurban Conf. Applied Sciences and Technology, IBCAST, 2016, pp. 528–534.
[171] M.A. Gerzon, Periphony: With-height sound reproduction, J. Audio Eng. Soc. 21 (1973) 2–10.
[172] M.A. Gerzon, Ambisonics in multichannel broadcasting and video, J. Audio Eng. Soc. 33 (1985) 859–871.
[173] C. Knapp, G. Carter, The generalized correlation method for estimation of time delay, IEEE Trans. Audio Speech Lang. Process. 24 (4) (1976) 320–327.
[174] D.P. Jarrett, E.A.P. Habets, P.A. Naylor, 3D source localization in the spherical harmonic domain using a pseudointensity vector, in: Proc. 18th European Signal Processing Conference (EUSIPCO 2010), 2010, pp. 442–446.
[175] M. Zaunschirm, M. Frank, F. Zotter, BRIR synthesis using first-order microphone arrays, in: Preprints AES 144th Convention, 2018, pp. 1–10.
[176] S.V. Amengual Garí, J.M. Arend, P.T. Calamia, P.W. Robinson, Optimizations of the spatial decomposition method for binaural reproduction, J. Audio Eng. Soc. 68 (2020) 959–976.
[177] X. Lai, H. Torp, Interpolation methods for time-delay estimation using cross-correlation method for blood velocity measurement, IEEE Trans. Ultrason. Ferroelect. Freq. Control. 46 (2) (1999) 277–290.
[178] L. Zhang, X. Wu, On cross correlation based-discrete time delay estimation, in: Proc. 2005 IEEE Int. Conf. Acoust. Speech, Signal Process., ICASSP'05, vol. 4, 2005, iv/981–iv/984.
[179] H. Rosseel, T. van Waterschoot, Improved acoustic source localization by time delay estimation with subsample accuracy, in: Proc. 2021 Immersive and 3D Audio: From Architecture To Automotive (I3DA), 2021, pp. 1–8.
[180] J. Pätynen, S. Tervo, T. Lokki, Spatial analysis of concert hall impulse responses, in: Proc. Int. Symposium Room Acoustics (ISRA 2013), 2013, pp. 1–8.
[181] S. Tervo, J. Pätynen, T. Lokki, Spatio-temporal energy measurements in renowned concert halls with a loudspeaker orchestra, Proc. Meet. Acoust. 19 (1) (2013) 015019.
[182] S.V. Amengual Garí, P. Jukka, L. Tapio, Physical and perceptual comparison of real and focused sound sources in a concert hall, J. Audio Eng. Soc. 64 (2016) 1014–1025.
[183] H. Möller, J. Pätynen, Spatial acoustic measurements in conert halls with a reduced virtual orchestra, in: Proc. Forum Acusticum 2023, 2023, pp. 1–4.
[184] S. Tervo, J. Saarelma, J. Pätynen, I. Huhtakallio, P. Laukkanen, Spatial analysis of the acoustics of rock clubs and nightclubs, in: Proc. 9th Int. Conf. Auditorium Acoustics, Paris, France, October 29-31, Institute of Acoustics, 2015, pp. 551–558.
[185] S. Tervo, P. Laukkanen, J. Pätynen, T. Lokki, Preferences of critical listening environments among sound engineers, J. Audio Eng. Soc. 62 (5) (2014) 300–314.
[186] S. Tervo, J. Pätynen, N. Kaplanis, M. Lydolf, S. Bech, T. Lokki, Spatial analysis and synthesis of car audio system and car cabin acoustics with a compact microphone array, J. Audio Eng. Soc. 63 (2015) 914–925.
[187] W. Lachenmayr, N. Meyer-Kahlen, O. Colella Gomes, A. Kuusinen, T. Lokki, Chamber music hall acoustics: Measurements and perceptual differences, J. Acoust. Soc. Amer. 154 (1) (2023) 388–400.
[188] N. Meyer-Kahlen, S. Amengual, T. Lokki, What the spatial decomposition method can and cannot do, in: Proc. ICA 2022, Acoustical Society of Korea (ASK), 2022, pp. 1–8.
[189] B. Alary, P. Massé, V. Välimäki, M. Noisternig, Assessing the anisotropic features of spatial impulse responses, in: Proc. EAA Spatial Audio Signal Processing Symposium, Paris, France, 2019, pp. 43–48.
[190] S.V. Amengual Garí, W. Lachenmayr, E. Mommertz, Spatial analysis and auralization of room acoustics using a tetrahedral microphone, J. Acoust. Soc. Amer. 141 (4) (2017) El369–el374.
[191] Iris, [Iris], 2024, accessed: 2024-10-20. URL https://www.iris.co.nz.
[192] E. De Sena, H. Hacıotahabiboğlu, Z. Cvetković, J.O. Smith, Efficient synthesis of room acoustics via scattering delay networks, IEEE/ ACM Trans. Audio Speech Lang. Process. 23 (9) (2015) 1478–1492.
[193] V. Pulkki, Virtual sound source positioning using vector base amplitude panning, J. Audio Eng. Soc. 45 (1997) 456–466.
[194] P. Ville, Localization of amplitude-panned virtual sources II: Two- and three-dimensional panning, J. Audio Eng. Soc. 49 (2001) 753–767.
[195] V. Pulkki, Uniform spreading of amplitude panned virtual sources, in: Proc. 1999 IEEE Workshop Appls. Signal Process. Audio Acoust., WASPAA'99, 1999, pp. 187–190.
[196] J. Pätynen, S. Tervo, T. Lokki, Amplitude panning decreases spectral brightness with concert hall auralizations, in: Proc. AES 55th Int. Conf. Spatial Audio, 2014, pp. 1–8.
[197] A.J. Berkhout, A holographic approach to acoustic control, J. Audio Eng. Soc. 36 (1988) 977–995.
[198] D. de Vries, E.M. Hulsebos, Auralization of room acoustics by wave field synthesis based on array measurements of impulse responses, in: Proc. 12th European Signal Processing Conference (EUSIPCO 2004), 2004, pp. 1377–1380.
[199] A.J. Berkhout, D. de Vries, P. Vogel, Acoustic control by Wave Field Synthesis, J. Acoust. Soc. Amer. 93 (5) (1993) 2764–2778.
[200] E. Verheijen, Sound reproduction by wave field synthesis, (Ph.D. thesis), Technische Universiteit Delft, 1997.
[201] S. Spors, R. Rabenstein, Spatial aliasing artifacts produced by linear and circular loudspeaker arrays used for wave field synthesis, in: Preprints AES 120th Convention, 2006, pp. 1–14.
[202] J. Ahrens, S. Spors, H. Wierstorf, Comparison of higher order ambisonics and wave field synthesis with respect to spatial discretization artifacts in time domain, in: Proc. AES 40th Int. Conf. Spatial Audio: Sense the Sound of Space, 2010, pp. 1–8.
[203] F. Gergely, F. Péter, Spatial aliasing and loudspeaker directivity in unified wave field synthesis theory, in: Proc. 44. Jahrestagung Für Akustik (DAGA 2018), 2018, pp. 1314–1317.
[204] F. Winter, F. Schultz, G. Firtha, S. Spors, A geometric model for prediction of spatial aliasing in 2.5d sound field synthesis, IEEE/ ACM Trans. Audio Speech Lang. Process. 27 (6) (2019) 1031–1046.
[205] E. Hulsebos, D. de Vries, E. Bourdillat, Improved microphone array configurations for auralization of sound fields by wave-field synthesis, J. Audio Eng. Soc. 50 (2002) 779–790.
[206] D. de Vries, L. Hörchens, P. Grond, Extraction of 3D information from circular array measurements for auralization with wave field synthesis, EURASIP J. Audio, Speech, Music. Process. 2007 (1) (2007) 190.
[207] S.V. Amengual Garí, The spatial decomposition method meets wave field synthesis: A feasibility study, in: Proc. 50. Jahrestagung Für Akustik (DAGA 2024), 2024, pp. 1630–1633.
[208] M.A. Gerzon, Multi-system ambisonic decoder: Part 1, Wirel. World 83 (1499) (1977) 43–47.
[209] M.A. Gerzon, Multi-system ambisonic decoder: Part 2, Wirel. World 83 (1500) (1977) 69–73.
[210] M.A. Gerzon, Ambisonic decoders for HDTV, in: Preprints AES 92nd Convention, 1992, pp. 1–42.
[211] F. Zotter, M. Frank, All-round ambisonic panning and decoding, J. Audio Eng. Soc. 60 (2012) 807–820.
[212] N. Rozenn, M. Emerit, 3D-sound reproduction over an extensive listening area: a hybrid method derived from holophony and ambisonic, in: Proc. 16th AES Int. Conf. Spatial Sound Reproduction, 1999, pp. 436–453.
[213] J. Daniel, R. Nicol, S. Moreau, Further investigations of high-order ambisonics and wavefield synthesis for holophonic sound imaging, in: Preprints AES 114th Convention, 2003, pp. 1–18.
[214] E.M. Wenzel, M. Arruda, D.J. Kistler, F.L. Wightman, Localization using nonindividualized head-related transfer functions, J. Acoust. Soc. Amer. 94 (1) (1993) 111–123.

[215] V. Algazi, R. Duda, D. Thompson, C. Avendano, The CIPIC HRTF database, in: Proc. 2001 IEEE Workshop Appls. Signal Process. Audio Acoust., WASPAA'01, 2001, pp. 99–102.
[216] Acoustics Research Institute, [Ari hrtf] database, 2024, accessed: 2024-09-13. URL https://www.oeaw.ac.at/en/ari/das-institut/software/hrtf-database.
[217] K. Iida, Head-Related Transfer Function and Acoustic Virtual Reality, Springer Singapore, Singapore, 2019, pp. 171–177, Ch. Comparison of HRTF Databases.
[218] M.-V. Laitinen, T. Pihlajamäki, S. Lösler, V. Pulkki, Influence of resolution of head tracking in synthesis of binaural audio, in: Proc. AES 132nd Convention, 2012, pp. 1–8.
[219] S.V. Amengual Garí, W.O. Brimijoin, H.G. Hassager, P.W. Robinson, Flexible binaural resynthesis of room impulse responses for augmented reality research, in: Proc. EAA Spatial Audio Signal Processing Symposium, Paris, France, 2019, pp. 161–166.
[220] E. Hendrickx, P. Stitt, J.-C. Messonnier, J.-M. Lyzwa, B.F. Katz, C. de Boishéraud, Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis, J. Acoust. Soc. Amer. 141 (3) (2017) 2011.
[221] M. Otani, H. Shigetani, M. Mitsuishi, R. Matsuda, Binaural ambisonics: Its optimization and applications for auralization, Acoust. Sci. Technol. 41 (2020) 142–150.
[222] P. Franček, K. Jambrošić, M. Horvat, V. Planinec, The performance of inertial measurement unit sensors on various hardware platforms for binaural head-tracking applications, Sensors 23 (2) (2023).
[223] W. Hess, Head-tracking techniques for virtual acoustics applications, in: Preprints AES 133rd Convention, 2012, pp. 1–15.
[224] F.L. Wightman, D.J. Kistler, Resolution of front–back ambiguity in spatial hearing by listener and source movement, J. Acoust. Soc. Amer. 105 (5) (1999) 2841–2853.
[225] O. Puomio, J. Pätynen, T. Lokki, Optimization of virtual loudspeakers for spatial room acoustics reproduction with headphones, Appl. Sci. 7 (12) (2017).
[226] J. Ahrens, Perceptual evaluation of binaural auralization of data obtained from the spatial decomposition method, in: Proc. 2019 IEEE Workshop Appls. Signal Process. Audio Acoust., WASPAA'19, 2019, pp. 65–69.
[227] H. Wierstorf, A. Raake, S. Spors, The Technology of Binaural Listening, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 255–278, Ch. Binaural Assessment of Multichannel Reproduction.
[228] H. Wierstorf, C. Hohnerlein, S. Spors, A. Raake, Coloration in wave field synthesis, in: Proc. AES 55th Int. Conf. Spatial Audio, 2014, pp. 1–8.
[229] M. Noisternig, A. Sontacchi, T. Musil, R. Holdrich, A 3D ambisonic based binaural sound reproduction system, in: Proc. AES 24th Int. Conf. Multichannel Audio, 2003, pp. 1–5.
[230] F. Zotter, M. Frank, Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality, in: Vol. 19 of Springer Topics in Signal Processing, Springer International Publishing, Cham, 2019.
[231] S. López-Mochales, R. Jiménez-Pasalodos, J. Valenzuela, C. Gutiérrez-Cajaraville, M. Díaz-Andreu, C. Escera, Experimental enhancement of feelings of transcendence, tenderness, and expressiveness by music in christian liturgical spaces, Front. Psychol. 13 (2022).
[232] T. van Waterschoot, M. Moonen, Fifty years of acoustic feedback control: State of the art and future challenges, Proc. the IEEE 99 (2) (2011) 288–327.
[233] J.W. Cooley, J.W. Tukey, An algorithm for the machine calculation of complex fourier series, Math. Comp. 19 (90) (1965) 297–301.
[234] T.G. Stockham, High-speed convolution and correlation, in: Proc. of the April 26-28, 1966, in: Spring Joint Computer Conference, Association for Computing Machinery, 1966, pp. 229–233.
[235] A. Torger, A. Farina, Real-time partitioned convolution for Ambiophonics surround sound, in: Proc. 2001 IEEE Workshop Appls. Signal Process. Audio Acoust., WASPAA'01, IEEE, 2001, pp. 195–198.
[236] F. Wefers, Partitioned convolution algorithms for real-time auralization, (Ph.D. thesis), RWTH Aachen University, Institute of Technical Acoustics, Aachen, Germany, 2014.
[237] F. Wefers, J. Berg, High-performance real-time FIR-filtering using fast convolution on graphics hardware, in: Proc. 13th Int. Conf. Digital Audio Effects, (DAFx '10), 2010, pp. 1–8.
[238] T. Wendt, S. van de Par, S.D. Ewert, A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation, J. Audio Eng. Soc. 62 (2014) 748–766.
[239] J. Cadavid, M.B. Møller, S. Bech, T. van Waterschoot, J. Østergaard, Performance of low frequency sound zones based on truncated room impulse responses, in: Proc. 17th Int. Audio Mostly Conference, Am '22, Association for Computing Machinery, New York, NY, USA, 2022, pp. 239–245.
[240] M. Jälmby, F. Elvander, T. van Waterschoot, Low-rank tensor modeling of room impulse responses, in: Proc. 2021 29th European Signal Processing Conference (EUSIPCO), 2021, pp. 111–115.
[241] M. Jälmby, F. Elvander, T. van Waterschoot, Fast low-latency convolution by low-rank tensor approximation, in: Proc. 2023 IEEE Int. Conf. Acoust., Speech, Signal Process., ICASSP'23, 2023, pp. 1–5.
[242] M. Jälmby, F. Elvander, T. van Waterschoot, Compression of room impulse responses for compact storage and fast low-latency convolution, EURASIP J. Audio, Speech, Music. Process. 45 (1) (2024) 1–23.
[243] M. Jälmby, F. Elvander, T. van Waterschoot, Multi-channel low-rank convolution of jointly compressed room impulse responses, IEEE Open J. Signal Process. (2024) 1–9.
[244] H. Rosseel, M. Jälmby, T. van Waterschoot, Low-rank multi-channel fast convolution for real-time auralization of church acoustics, in: Proc. 32nd European Signal Processing Conference (EUSIPCO 2024), 2024, pp. 1017–1021.
[245] M. Karjalainen, H. Järveläinen, Reverberation modeling using velvet noise, in: Proc. AES 30th Int. Conf. Intelligent Audio Environments, 2007, pp. 1–8.
[246] V. Välimäki, B. Holm-Rasmussen, B. Alary, H.-M. Lehtonen, Late reverberation synthesis using filtered velvet noise, Appl. Sci. 7 (5) (2017).
[247] S. Lee, H.-S. Choi, K. Lee, Differentiable artificial reverberation, IEEE Trans. Audio Speech Lang. Process. 30 (2022) 2541–2556.
[248] J.P. Stautner, M. Puckette, Designing multi-channel reverberators, Comput. Music J. 6 (1982) 52.
[249] J.-M. Jot, C. Antoine, Digital delay networks for designing artificial reverberators, in: Preprints AES 90th Convention, 1991, pp. 1–16.
[250] V. Välimäki, J.D. Parker, L. Savioja, J.O. Smith, J.S. Abel, Fifty years of artificial reverberation, IEEE Trans. Audio Speech Lang. Process. 20 (5) (2012) 1421–1448.
[251] H. Bai, G. Richard, L. Daudet, Late reverberation synthesis: From radiance transfer to feedback delay networks, IEEE/ ACM Trans. Audio Speech Lang. Process. 23 (12) (2015) 2260–2271.
[252] J. Shen, R. Duraiswami, Data-driven feedback delay network construction for real-time virtual room acoustics, in: Proc. 15th Int. Audio Mostly Conference, AM '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 46–52.
[253] S. Lyster, C. Erkut, A differentiable neural network approach to parameter estimation of reverberation, in: R. Michon, L. Pottier, Y. Orlarey (Eds.), Proc. of the 19th Sound and Music Computing Conference, Sound and Music Computing Network, 2022, pp. 358–364.
[254] R. Bona, D. Fantini, G. Presti, M. Tiraboschi, J.I. Engel Alonso-Martinez, F. Avanzini, Automatic parameters tuning of late reverberation algorithms for audio augmented reality, in: Proc. 17th Int. Audio Mostly Conference, AM '22, Association for Computing Machinery, New York, NY, USA, 2022, pp. 36–43.
[255] I. Ibnyahya, J.D. Reiss, A method for matching room impulse responses with feedback delay networks, in: Preprints AES 153rd Convention, New York, NY, USA, 2022, Express Paper.
[256] G. Dal Santo, K. Prawda, S.J. Schlecht, V. Välimäki, Differentiable feedback delay network for colorless reverberation, in: Proc. 26th Int. Conf. Digital Audio Effects (DAFx '23), 2023, pp. 1–9.
[257] T. Carpentier, M. Noisternig, O. Warusfel, Hybrid reverberation processor with perceptual control, in: Proc. 17th Int. Conf. Digital Audio Effects (DAFx '14), 2014, pp. 1–8.
[258] M.R. Schroeder, Improvement of Acoustic-Feedback Stability by Frequency Shifting, J. Acoust. Soc. Amer. 36 (9) (1964) 1718–1724.
[259] J.L. Nielsen, U.P. Svensson, Performance of some linear time-varying systems in control of acoustic feedback, J. Acoust. Soc. Amer. 106 (1) (1999) 240–254.
[260] M.A. Poletti, The stability of multichannel sound systems with frequency shifting, J. Acoust. Soc. Amer. 116 (2) (2004) 853–871.
[261] M. Er, T. Ooi, L. Li, C. Liew, A dsp-based acoustic feedback canceller for public address systems, Microprocess. Microsyst. 18 (1) (1994) 39–47.
[262] P. Gil-Cacho, T. van Waterschoot, M. Moonen, S.H. Jensen, Regularized adaptive notch filters for acoustic howling suppression, in: Proc. 17th European Signal Processing Conference (EUSIPCO 2009), 2009, pp. 2574–2578.
[263] T. van Waterschoot, G. Rombouts, M. Moonen, On the performance of decorrelation by prefiltering for adaptive feedback cancellation in public address systems, in: Proc. of the 2004 IEEE Benelux Signal Processing Symposium (SPS 2004), 2004, pp. 167–170.
[264] A. Spriet, G. Rombouts, M. Moonen, J. Wouters, Adaptive feedback cancellation in hearing aids, J. Franklin Inst. 343 (6) (2006) 545–573.
[265] T. van Waterschoot, M. Moonen, Adaptive feedback cancellation for audio applications, Signal Process. 89 (11) (2009) 2185–2201.
[266] V. Baran, K. Yin-Ying Zhang, C. Rodriguez Escobar, A. Aydin, R. King, W. Woszczyk, The optimization of microphone techniques for capturing virtual acoustic environments, in: Preprints AES 155th Convention, 2023, pp. 1–8.
[267] G. Grazioli, V. Baran, K.Y.-Y. Zhang, A. Aydin, W. Woszczyk, Preliminary evaluation of a sound system employing a cancelling auralizer for producing virtual rooms in a modern recording studio, in: Proc. 2023 Immersive and 3D Audio: From Architecture To Automotive (I3DA), 2023, pp. 1–7.
[268] E. Callery, J.S. Abel, E.K. Canfield-Dafilou, J. Perišić, Convolution, virtual acoustics, and interactions with impossible worlds, in: Proc. 2023 Immersive and 3D Audio: From Architecture To Automotive (I3DA), 2023, pp. 1–11.
[269] K. Schiltz, Church and chamber: The influence of acoustics on musical composition and performance, Early Music. 31 (1) (2003) 65–78.
[270] K. Ueno, H. Tachibana, Experimental study on the evaluation of stage acoustics by musicians using a 6-channel sound simulation system, Acoust. Sci. Technol. 24 (2003) 130–138.

[271] A. Bassuet, Acoustics of early music spaces from the 11th to 18th century: Rediscovery of the acoustical excellence of medium-sized rooms and new perspectives for modern concert hall design, J. Acoust. Soc. Am. 115 (5) (2004) 2582.
[272] K. Ueno, K. Kato, K. Kawai, Effect of room acoustics on musicians' performance, Part I: Experimental investigation with a conceptual model, Acta Acust United Ac 96 (2010) 505–515.
[273] Cycling '74, [What is max]?, 2024, accessed: 2024-10-20. URL https://cycling74.com/products/max.
[274] M.S. Puckette, Pure data: another integrated computer music environment, in: Proc. Second Intercollege Computer Music Concerts, 1997, pp. 37–41.
[275] L. Beranek, Concert Halls and Opera Houses: Music, Acoustics, and Architecture, Springer, New York, NY, 2004.
[276] Y. Ando, Architectural Acoustics : Blending Sound Sources, Sound Fields, and Listeners, Springer, New York, NY, 1998.
[277] M. Barron, Subjective study of british symphony concert halls, Acta Acust. United Acust. 66 (1) (1988) 1–14.
[278] P. Hüttenmeister, W.L. Martens, Perceived cathedral ceiling height in a multi-channel virtual acoustic rendering for gregorian chant, in: Proc. of ACOUSTICS, 1/2, 2016, pp. 587–595.
[279] S. Amengual Gari, M. Kob, T. Lokki, et al., Investigations on stage acoustic preferences of solo trumpet players using virtual acoustics, in: Proc. Sound and Music Computing Conference, 2017, pp. 167–174.
[280] M. Tomasetti, L. Turchet, Playing with others using headphones: musicians prefer binaural audio with head tracking over stereo, IEEE Trans. Human–Mach. Syst. 53 (3) (2023) 501–511.
[281] C. Mendonça, S. Delikaris-Manias, Statistical tests with MUSHRA data, in: Preprints AES 144th Convention, United States, 2018, pp. 859–868.
[282] S. Bech, N. Zacharov, Perceptual Audio Evaluation - Theory, Method and Application, John Wiley & Sons, Ltd, 2006.
[283] C. Philip, E. Nicolas, V. Satvik, R. Frederic, Exploring perceptual annoyance and colouration assessment in active acoustic environments, in: Proc. Int. Conf. Acoustics & Sound Reinforcement, 2024, pp. 1–10.
[284] M.R. Schroeder, D. Gottlob, K.F. Siebrasse, Comparative study of European concert halls: correlation of subjective preference with geometric and acoustic parameters, J. Acoust. Soc. Amer. 56 (4) (1974) 1195–1201.
[285] Y. Ando, M. Imamura, Subjective preference tests for sound fields in concert halls simulated by the aid of a computer, J. Sound Vib. 65 (2) (1979) 229–239.
[286] Y. Ando, Springer Handbook of Acoustics, Springer, New York, New York, NY, 2014, pp. 367–402, Ch. Concert Hall Acoustics Based on Subjective Preference Theory.
[287] T. Lokki, J. Pätynen, A. Kuusinen, S. Tervo, Concert hall acoustics: Repertoire, listening position, and individual taste of the listeners influence the qualitative attributes and preferences, J. Acoust. Soc. Amer. 140 (1) (2016) 551–562.
[288] A.A. Williams, G.M. Arnold, A comparison of the aromas of six coffees characterised by conventional profiling, free-choice profiling and similarity scaling methods, J. Sci. Food Agric. 36 (3) (1985) 204–214.
[289] H.T. Lawless, H. Heymann, Sensory Evaluation of Food: Principles and Practices, Springer Science & Business Media, 1999.
[290] T. Lokki, H. Vertanen, A. Kuusinen, J. Pätynen, S. Tervo, Auditorium acoustics assessment with sensory evaluation methods, in: Proc. Int. Symposium Room Acoustic (ISRA 2010), 2010, pp. 29–31.
[291] T. Lokki, J. Pätynen, A. Kuusinen, H. Vertanen, S. Tervo, Concert hall acoustics assessment with individually elicited attributes, J. Acoust. Soc. Am. 130 (2) (2011) 835–849.
[292] T. Lokki, J. Pätynen, A. Kuusinen, S. Tervo, Disentangling preference ratings of concert hall acoustics using subjective sensory profiles, J. Acoust. Soc. Amer. 132 (5) (2012) 3148–3161.
[293] N. Kaplanis, S. Bech, S. Tervo, J. Pätynen, T. Lokki, T. van Waterschoot, S.H. Jensen, Perceptual aspects of reproduced sound in car cabin acoustics, J. Acoust. Soc. Amer. 141 (3) (2017) 1459–1469.
[294] N. Kaplanis, S. Bech, S. Tervo, J. Pätynen, T. Lokki, T. van Waterschoot, S.H. Jensen, A rapid sensory analysis method for perceptual assessment of automotive audio, J. Audio Eng. Soc. 65 (1/2) (2017) 130–146.
[295] J. Camilleri, N. Kaplanis, E. De Sena, Evaluation of car cabin acoustics using auralisation over headphones, in: Proc. 2019 AES Int. Conf. Immersive and Interactive Audio, 2019, pp. 1–10.
[296] N. Kaplanis, S. Bech, S. Jensen, T. van Waterschoot, Perception of reverberation in small rooms: A literature study, in: Proc. AES 55th Int. Conf. Spatial Audio, Helsinki, Finland, 2014, pp. 1–14.
[297] N. Kaplanis, S. Bech, T. Lokki, T. van Waterschoot, S. Holdt Jensen, Perception and preference of reverberation in small listening rooms for multi-loudspeaker reproduction, J. Acoust. Soc. Am. 146 (5) (2019) 3562–3576.
[298] A. Lindau, V. Erbes, S. Lepa, H.-J. Maempel, F. Brinkman, S. Weinzierl, A spatial audio quality inventory (SAQI), Acta Acust. United Acust. 100 (5) (2014) 984–994.
[299] G. Lorho, Individual vocabulary profiling of spatial enhancement systems for stereo headphone reproduction, in: Preprints AES 119th Convention, 2005, pp. 1–19.
[300] A. Kuusinen, J. Pätynen, S. Tervo, T. Lokki, Relationships between preference ratings, sensory profiles, and acoustical measurements in concert halls, J. Acoust. Soc. Amer. 135 (1) (2014) 239–250.
[301] H. Stone, J. Sidel, S. Oliver, A. Woolsey, R.C. Singleton, Sensory Evaluation By Quantitative Descriptive Analysis, John Wiley & Sons, Ltd, 2004, pp. 23–34, Ch. 1.3.
[302] V. Dairou, J.-M. Sieffermann, A comparison of 14 jams characterized by conventional profile and a quick original method, the flash profile, J. Food Sci. 67 (2002) 826–834.
[303] J. Liu, W.L. Bredie, E. Sherman, J.F. Harbertson, H. Heymann, Comparison of rapid descriptive sensory methodologies: Free-choice profiling, flash profile and modified flash profile, Food Res. Int. 106 (2018) 892–900.
[304] H. Rosseel, T. van Waterschoot, Interactive perceptual evaluation of auralized acoustics in rehearsal spaces for vocal music, in: Proc. Forum Acusticum 2023, 2023, pp. 4123–4130.
[305] S.V. Amengual Garí, M. Kob, T. Lokki, Analysis of trumpet performance adjustments due to room acoustics, in: Proc. International Symposium on Room Acoustics, Netherlands, 2019, pp. 65–73.
[306] P. Chen, B. Boren, Enhancing realism for digital piano players: A perceptual evaluation of head-tracked binaural audio, in: Proc. AES 157th Convention, 2024, pp. 1–12.
[307] S. Fargeot, A. Vidal, M. Aramaki, R. Kronland-Martinet, Perceptual evaluation of an ambisonic auralization system of measured 3d acoustics, Acta Acust. 7 (2023) 56.
[308] Z. Engel, K. Kosał a, Index method of the acoustic quality assessment of sacral buildings, Arch. Acoust. 32 (3) (2007) 455–474.
[309] K. Kosała, A single number index to assess selected acoustic parameters in churches with redundant information, Arch. Acoust. 36 (3) (2011).
[310] U. Berardi, A double synthetic index to evaluate the acoustics of churches, Arch. Acoust. 37 (4) (2012) 521–528.
[311] K. Kosał a, The comparative analysis of acoustic properties of Roman Catholic churches using the index method, Acta Phys. Pol. A 125 (4-a) (2014).
[312] K. Kosał a, P. Mał ecki, Index assessment of the acoustics of Orthodox churches in Poland, Appl. Acoust. 130 (2018) 140–148.
[313] P. Brezina, Measurement of intelligibility and clarity of the speech in Romanesque churches, J. Cult. Herit. 16 (3) (2015) 386–390.
[314] N.R. Syamsiyah, A. Dharoko, S.S. Utami, Sound preservation at the Grand Mosque of Yogyakarta in Indonesia: The acoustic performance of the traditional architecture, AIP Conf. Proc. 1977 (1) (2018) 040032-1–040032-7.
[315] D. Zhang, R. Lai, M. Zhang, J. Kang, Effects of spatial elements and sound sources on sound field in Main Hall of Chinese Buddhist temple, J. Acoust. Soc. Amer. 147 (3) (2020) 1516–1530.
[316] D. Zhang, M. Zhang, D. Liu, J. Kang, Soundscape evaluation in Han Chinese Buddhist temples, Appl. Acoust. 111 (2016) 188–197.
[317] S. Girón, M. Galindo, T. Gómez-Gómez, Assessment of the subjective perception of reverberation in Spanish cathedrals, Build. Environ. 171 (2020) 106656.
[318] B. Vannieuwenhuyze, S. Meijering, Het Brusselse hof van Nassau: De oprichting van een laatmiddeleeuwse stadsresidentie, Rev. Belg. de Philol. et D'Hist. 88 (2) (2010) 349–376.
[319] Royal Library of Belgium, [Kbr - royal library of Belgium], 2024, accessed: 2024-10-30. URL https://www.kbr.be/en/.