ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Efficient synthesis of large room impulse responses in the modal domain

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Ducceschi, M., Webb, C.J., Fratoni, G. (2025). Efficient synthesis of large room impulse responses in the modal domain. IEEE [10.1109/I3DA65421.2025.11202053].

Availability:
This version is available at: https://hdl.handle.net/11585/1022876 since: 2025-11-25

Published:

DOI: http://doi.org/10.1109/I3DA65421.2025.11202053

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

15 September 2026

Efficient synthesis of large room impulse responses in the modal domain

Michele Ducceschi
Dept. of Industrial Engineering
University of Bologna
Bologna, Italy
michele.ducceschi@unibo.it

Craig J. Webb
Dept. of Industrial Engineering
University of Bologna
Bologna, Italy
craigjonathan.webb@unibo.it

Giulia Fratoni
Dept. of Industrial Engineering
University of Bologna
Bologna, Italy
giulia.fratoni2@unibo.it

***Abstract*—An efficient methodology for synthesising large-scale room impulse responses (RIRs) is presented, operating entirely in the modal domain. The approach proceeds in two stages: an analysis phase, wherein modal parameters are extracted from measured RIRs using a frequency-domain sub-band implementation of the PolyMAX algorithm; and a synthesis phase, based on time-stepping of a second-order modal equation discretised via a low-dispersion finite difference scheme. The procedure yields a structured parametric model consisting of modal frequencies, decay rates, and shape factors, directly interpretable in physical terms. Sub-band fitting is employed to accommodate high modal densities and broadband content, with stabilisation diagrams used to isolate physically relevant poles. The method is applied to both synthetic and measured responses acquired in three large historical spaces. Results indicate close agreement with the target responses in both frequency and time domains, along with substantial reductions in computational cost. The modal representation further permits statistical post-analysis of decay behaviour and supports real-time synthesis on commodity hardware. The overall framework offers a scalable alternative to traditional numerical solvers for room acoustics modelling.**

I. Introduction

The synthesis of room impulse responses (RIRs) encompasses multiple domains in acoustics and audio signal processing. Applications include virtual acoustics and auralisation systems [1], [2], data-driven modelling and generation of spatial sound fields [3], and the simulation and analysis of perceptual and physiological responses in hearing science [4]. Within these contexts, the task of efficiently generating RIRs—while preserving accuracy—remains central to the success of virtual acoustic systems [5], [6].

The development of methods for RIR analysis and synthesis spans more than a century, beginning with analytical studies such as Sabine's classical formulation of reverberation time [7]. Computational modelling techniques emerged in the 1960s, particularly in the high-frequency regime, where geometrical acoustics offered tractable approximations. Early ray-based strategies include ray tracing schemes [8] and the well-established image source method introduced by Allen and Berkley [9], which continues to provide a baseline model for specular reflections in rectilinear and polyhedral spaces. In parallel, the late 20th century saw the development of wave-based approaches, most notably the application of finite difference time domain (FDTD) methods to low-frequency room simulation [10]. These methods have since scaled with improvements in computing hardware, including GPU-accelerated architectures [11], [12], and now support high-resolution simulations across extended bandwidths.

Present-day wave-based methods include finite difference schemes [13], [14], finite volume discretisations [15], and finite element models [16], [17], each with respective trade-offs regarding numerical dispersion, boundary representation, and computational cost. Hybrid strategies have also emerged, combining geometric and wave-based models to accommodate broadband and multiscale phenomena [18]. More recently, data-driven alternatives using neural networks have appeared, progressing from early parametric estimators [19] to recent systems based on operator learning frameworks capable of reconstructing full-bandwidth 3D fields [20].

The present work addresses a related but distinct problem: extracting structured room models directly from measured RIR data. A natural framework for this task is modal decomposition, wherein the RIR is represented as a sum of exponentially decaying sinusoids. Each sinusoidal term corresponds to a second-order resonant filter; its pole encodes both modal frequency and damping, while the gain reflects source–receiver positioning and excitation strength [21]. This formulation is both physically interpretable and amenable to compact parametric representation. The main challenge lies in the robust and efficient estimation of poles and gains from measurement data.

Several methodologies have been developed for this purpose, some of which avoid the explicit estimation of poles. Notably, Bank's fixed-pole approach fixes resonant frequencies and bandwidths a priori and optimises the gain values using a weighted least-squares fit across frequency bins [22].

General approaches to modal parameter extraction originated in structural dynamics [23], where they were developed to analyse vibrating mechanical systems. These techniques have since been adapted for use in room acoustics, typically in a sub-band setting due to the high modal density of enclosed sound fields. In the time domain, notable examples include sub-band ESPRIT [24], frequency-zoomed ARMA modelling [25], and ESPRIT analysis on a warped frequency axis [26]. These methods rely on constructing Hankel matrices from delayed signal blocks, yielding poles through subspace

analysis and gains via least-squares fitting, a feature analogous to the well-known eigensystem realisation algorithm by Juang and Pappa [27].

Frequency-domain approaches offer an alternative route, as exemplified by the work of Maestre *et al.* [28], [29]. Here, initial pole estimates are obtained from the frequency response, followed by nonlinear constrained optimisation to refine the pole positions. Once finalised, the modal gains are computed analytically via linear least-squares. While capable of yielding high-quality models, this approach is sensitive to initialisation and becomes computationally expensive at higher model orders due to the need for iterative gradient-based refinement.

The approach developed here adopts a frequency-domain strategy based on the PolyMAX algorithm for modal analysis. As with the least-squares complex frequency-domain (LSCF) method, PolyMAX recasts the pole estimation problem as a polynomial, enabling semi-automated pole selection [30], [31]. The fitted model is then converted to state-space form, with modal parameters extracted via eigen-decomposition. The procedure is entirely linear, albeit at the expense of computing zeros of complex polynomials through an eigenvalue procedure, and avoids the convergence issues typical of nonlinear optimisation. More importantly, though, PolyMAX yields stabilisation diagrams—plots of pole location versus model order—which enable the robust identification of physical modes and effective suppression of numerical artefacts. This clear separation of stable and spurious components supports semi-automated pole selection, without reliance on heuristic thresholds or manual post-processing.

The paper is organised as follows: Sec. II presents the overall pole-residue identification theory. Practical implementation details are given in Sec. III. Examples are discussed in Sec. IV, including a simple synthetic resonator used as a test, and three large central-plan halls. Results are then discussed, and a basic statistical analysis of the modal decay parameters is given.

II. Theory

For conciseness, the following discussion assumes a single-input single-output (SISO) system. Generalisations to the multi-input - multi-output (MIMO) case are possible, and the reader is referred to Peeters *et. al* [31] for details. Note that, in the SISO case, there is no distinction between the PolyMAX algorithm and standard LSCF.

In a SISO system, a frequency response function in the z domain $H(z)$ can be modelled as a rational function of the complex variable z:

$$H(z) \;=\; \frac{B(z)}{A(z)} \;=\; \frac{b_0 + b_1 z + \cdots + b_p z^p}{a_0 + a_1 z + \cdots + a_p z^p}\,, \tag{1}$$

where $A(z)$ and $B(z)$ are polynomials in z. The frequency response is obtained by replacing $z \to e^{j\omega T}$ for continuous radian frequency ω and constant time interval T. The denominator $A(z)$ of degree p captures the poles of the system (the resonant modal frequencies and damping). In contrast, the numerator $B(z)$, again of order p, captures the zeros and overall gain (mode shape contributions for the given input-output pair). Before proceeding, it is worth noting that the constant T is not intended here as the sampling interval of a measured RIR; rather, it is an artificial parameter used to scale the exponential. An appropriate choice of T is essential for the well-posedness of the optimisation routine. A practical choice for T will be illustrated later.

Given discrete frequency response data from a measured RIR, the transfer function above is evaluated on a number N of discrete frequency bins, such that:

$$\hat{H}_k := \hat{H}\left(e^{j\omega_k T}\right), \quad k = 0, 1, ..., N-1, \tag{2}$$

where the "hat" notation denotes a measured response. The coefficients of $A(z)$ and $B(z)$ can be determined by solving a linear least-squares problem. Specifically, approximating (1) with the measured response gives:

$$\epsilon_k := B_k - A_k\,\hat{H}_k \approx 0, \tag{3}$$

for each frequency sample ω_k in the dataset. In matrix form:

$$\boldsymbol{\epsilon} = [\mathbf{X}, \mathbf{Y}] \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix}, \tag{4}$$

where $\mathbf{X}, \mathbf{Y} \in \mathbb{C}^{N\times(p+1)}$, with:

$$(\mathbf{X})_{k,:} \;=\; \left[1, e^{j\omega_k T}, ..., e^{jp\omega_k T}\right], \tag{5}$$

$$(\mathbf{Y})_{k,:} \;=\; -\hat{H}_k \cdot (\mathbf{X})_{k,:}, \tag{6}$$

$$\boldsymbol{\beta} \;=\; [b_0, b_1, ..., b_p]^\intercal, \tag{7}$$

$$\boldsymbol{\alpha} \;=\; [a_0, a_1, ..., a_p]^\intercal. \tag{8}$$

A. Least-square minimisation

A linear-in-parameters least-squares minimisation is readily available for the norm of $\boldsymbol{\epsilon}$. The optimisation problem is defined as finding:

$$\min_{\boldsymbol{\beta},\boldsymbol{\alpha}\in\mathbb{R}^{p+1}} \frac{1}{2}\|\boldsymbol{\epsilon}\|_2^2\,. \tag{9}$$

First, note that:

$$\|\boldsymbol{\epsilon}\|_2^2 = [\boldsymbol{\beta}^\intercal, \boldsymbol{\alpha}^\intercal] \begin{bmatrix} \mathbf{R} & \mathbf{S} \\ \mathbf{S}^\intercal & \mathbf{T} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix}, \tag{10}$$

with $\mathbf{R} := \mathrm{Re}(\mathbf{X}^\dagger\mathbf{X})$, $\mathbf{S} := \mathrm{Re}(\mathbf{X}^\dagger\mathbf{Y})$, $\mathbf{T} := \mathrm{Re}(\mathbf{Y}^\dagger\mathbf{Y})$. The symbol † denotes the Hermitian transpose and, hence, all three matrices $\in \mathbb{R}^{(p+1)\times(p+1)}$. Thus, taking gradients of the cost function (9) and setting them to zero yields the so-called *normal equations*:

$$\mathbf{R}\boldsymbol{\beta} \;=\; -\mathbf{S}\boldsymbol{\alpha}, \tag{11a}$$

$$\mathbf{T}\boldsymbol{\alpha} \;=\; -\mathbf{S}^\intercal\boldsymbol{\beta}. \tag{11b}$$

Using (11a) in (11b) to eliminate $\boldsymbol{\beta}$, one gets:

$$\left(\mathbf{T} - \mathbf{S}^\intercal\mathbf{R}^{-1}\mathbf{S}\right)\boldsymbol{\alpha} := \mathbf{M}\boldsymbol{\alpha} = 0. \tag{12}$$

This equation is never solved directly. To avoid computing the trivial solution $\boldsymbol{\alpha} = \mathbf{0}$, the last element of $\boldsymbol{\alpha}$ is set to one (i.e. $a_p = 1$). Then, the following sub-problem is extracted:

$$\mathbf{A}\,\boldsymbol{\alpha}_{0:p-1} = \mathbf{b} \tag{13}$$

with $\mathbf{A} := \mathbf{M}_{0:p-1,0:p-1}$, $\mathbf{b} := \mathbf{M}_{0:p-1,p}$. This yields the least-squares denominator coefficients:

$$\boldsymbol{\alpha} = [a_0, a_1, ..., 1]^\intercal. \tag{14}$$

B. Pole Extraction

Given the denominator coefficients of the transfer function, one may proceed with the extraction of poles by solving the characteristic equation $A(z) = 0$. Rather than computing the roots directly—an approach that can be numerically unstable for higher-order systems—it is preferable to recast the problem in terms of an eigenvalue computation:

$$\mathbf{CV} = \mathbf{V\Lambda}, \tag{15}$$

where $\mathbf{C}$ denotes the companion matrix, given explicitly as:

$$\mathbf{C} := \begin{bmatrix} 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ -a_0 & -a_1 & \cdots & -a_{p-2} & -a_{p-1} \end{bmatrix}, \tag{16}$$

and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues. The eigenvalues of $\mathbf{C}$ are of the form $e^{-\lambda_i T}$, where

$$\lambda_i := -\sigma_i + j\sqrt{\Omega_i^2 - \sigma_i^2}, \tag{17}$$

with $\sigma_i := 3\log(10)/\tau_{60}$ the decay rate associated with mode i, defined in terms of the 60 dB decay time τ_{60}, and Ω_i the corresponding undamped modal frequency.

C. Modal Shape Recovery via Pole–Residue Least-Squares Fitting

Upon determination of the modal poles $\{\lambda_i\}$, the corresponding residues may be obtained via a least-squares fitting procedure. In a general MIMO framework, one distinguishes between input participation and output modal shape vectors. In the SISO case under consideration here, both reduce to a single complex-valued scalar, the residue r_i, which encodes the amplitude and phase of mode i within the system's transfer function.

The frequency response function (FRF) is expressed in pole–residue form as:

$$\begin{aligned} H(j\omega) &= \sum_{i=1}^{M} H_i(j\omega), \text{ where} \\ H_i(j\omega) &= \frac{r_i}{j\omega - \lambda_i} + \frac{r_i^*}{j\omega - \lambda_i^*}, \end{aligned} \tag{18}$$

where λ_i and r_i denote the complex-valued poles and residues, and * indicates complex conjugation.

Given fixed pole locations $\{\lambda_i\}$, the representation (18) is linear in the unknowns $\{r_i\}$, and may thus be estimated via linear least-squares fitting against sampled frequency-domain data. This constitutes the second stage of the PolyMAX algorithm, wherein residue estimation follows the selection of stable poles through iterative refinement.

D. Time-Domain Resynthesis

From the pole–residue model (18), one obtains a corresponding time-domain formulation for the modal coordinate q_i driven by an external input $f(t)$:

$$\ddot{q}_i + 2\sigma_i \dot{q}_i + \Omega_i^2 q_i = c_i f(t) + d_i \dot{f}(t), \tag{19}$$

with coefficients defined as:

$$\begin{aligned} c_i &= 2\sigma_i \operatorname{Re}(r_i) - 2\sqrt{\Omega_i^2 - \sigma_i^2}\operatorname{Im}(r_i), \\ d_i &= 2\operatorname{Re}(r_i). \end{aligned}$$

Here, c_i corresponds to the real-valued modal shape for mode i. For numerical simulation, the continuous-time equation (19) may be discretised using the time sequence $q_i^n \approx q_i(nT_s)$, where T_s is the sample interval (not to be confused with the time scaling constant T used previously). A finite-difference scheme with exponential damping yields:

$$\begin{aligned} q_i^{n+1} = {} & 2q_i^n \cos\left(\sqrt{\Omega_i^2 - \sigma_i^2}\, T_s\right) e^{-\sigma T_s} \\ & - e^{-2\sigma T_s} q_i^{n-1} \\ & + T_s e^{-\sigma T_s}\left(T_s c_i f^n + d_i(f^n - f^{n-1})\right), \end{aligned}$$

which is unconditionally stable and exhibits low numerical dispersion, as discussed in [32]–[34].

Each modal coordinate q_i^n may be updated independently, allowing for efficient parallel implementation. The final system response is obtained as:

$$y^n = \sum_{i=1}^{M} q_i^n. \tag{20}$$

III. Implementation

Due to the extremely high modal density of measured room impulse responses, full-band optimisation is generally infeasible. Instead, a sub-band fitting approach is adopted.

A. Sub-Band Optimisation

Let a sub-band be defined by the lower and upper frequency bounds ω_l and ω_r, respectively. The fitting performance within this band is determined primarily by two design parameters: the time-scaling constant T in (1), and the model order p. If T is too small, the basis functions vary too slowly across the data window, and the matrix $\mathbf{A}$ in (13) becomes ill-conditioned. Conversely, overly large values of T lead to excessive variation and instability. Following standard practice in structural dynamics, a robust choice is $T = \pi/(\omega_r - \omega_l)$ [35], which is adopted throughout this work.

The model order p is not selected directly. Instead, a stabilisation diagram is constructed by sweeping p over a specified range. At each step, poles are retained only if their frequency and decay rate remain within specified thresholds relative to the previous iteration. Clusters of stable poles are then formed using a distance-based metric. See Fig. 3 for an example of such a stabilisation diagram.

Residues are computed using the procedure described in Section II-C. The output of the sub-band b fit consists of M_b poles and residues.

B. Global Optimisation

To cover a broader frequency range, the optimisation domain is partitioned into sub-bands. Various strategies may be employed, including linear, logarithmic, or perceptually motivated subdivisions such as the Bark scale [29], [36]. The optimisation is carried out independently in each band, and the pole–residue pairs are aggregated across bands.

To minimise edge artefacts and spurious contributions from neighbouring bands, it is often beneficial to optimise over a band with additional overlap and retain only the central band's poles and residues post-fit [29]. Once all N_B bands are optimised, the total number of identified modes is $M = \sum_{b=1}^{N_B} M_b$, yielding the $4M$ real coefficients Ω_i, σ_i, c_i, and d_i needed for direct synthesis via (19).

IV. Examples

As a first case study, a synthetic resonator is constructed in MATLAB, comprising 100 modes with randomly generated modal frequencies, decay times, and real-valued modal shapes such that $d_i = 0$ in (19). Frequencies are drawn within the range $[20, 520]$ Hz.

The second set of test cases is a sample of large RIRs acquired in three different centrally planned halls through acoustic surveys, pictures of which are visible in Fig. 1. The halls have been the subject of previous studies [37].

The halls: The first space is the Odeo Cornaro in Padua (Italy), which is a 16th-century Renaissance music hall that continues until today to serve its original function, hosting musical and cultural events [38], [39], whose central hall is visible in Fig. 1(a). The structure was built between 1524 and 1534 as part of a vital villa and has remained largely intact over the centuries. The architecture includes a sequence of richly frescoed chambers used for musical and intellectual gatherings. The central musical hall is particularly interesting, as it is an octagonal marble room (220 m^3) that exemplifies Renaissance proportion and central-plan harmony. Architectural historians highlight its exceptional acoustic response, attributing sound sustain during vocal performances to the concave geometry of the ceiling and the four large lateral niches.

The second environment is the Rotunda of Madonna del Monte in Bologna (Italy), a 12th-century circular building remarkably preserved in its original form [40], as seen in Fig. 1(b). Originally part of a monastic complex, this Romanesque central-plan sacred space was used by monks for private prayer and chanting. The environment (715 m^3) features a dome with a central oculus, reminiscent of ancient sepulchral spaces, such as the Pantheon in Rome. The Rotunda's design reflects theological and symbolic intentions, contributing to its unique acoustic profile. The surviving 19 niches, although never filled with the intended sculptures, remain key elements enhancing sound diffusion, along with the rough masonry on the inner walls.

Built between the 12th and 14th centuries, the third hall is the St. John's Baptistery, close to the renowned Leaning Tower in Pisa, Italy [41]. Part of the hall is visible in Fig. 1(c).

Fig. 1. Pictures of the central halls of (a): Odeo Cornaro; (b): Rotunda of Madonna del Monte; (c): the Pisa Baptistery. Panels (a) and (b) reproduced with permission by the authors [39], [40]. Picture in Panel (c) by Mstyslav Chernov, Wikipedia, reproduced without changes, licensed under https://creativecommons.org/licenses/by-sa/3.0/.

Initially designed for baptism ceremonies, the Baptistery's architecture follows theological and traditional symbolism. It features a circular plan and a high conical dome, achieving an impressive overall volume of 23,000 m^3. The ground floor features a central zone, marked by an octagonal font, altar, and pulpit, as well as an external circular ambulatory. The space includes marble linings, columns with decorated capitals, and floor-level tombs. On the first level, the *matronaeum*, a more austere gallery intended to host women, overlooks the central space. Although direct records of musical use are rare, the environment likely supported Gregorian chant and later polyphony, as suggested by comparative acoustic studies in similar contexts (e.g., St. Vitale in Ravenna and Notre-Dame in Paris). More details about the three spaces are provided by Ref. [37].

Measurements: A series of acoustic surveys allowed the collection of measured RIRs in each environment in compliance with ISO 3382-1 [42]. During the measurements, the halls were furnished and unoccupied, except for the presence of two operators. RIRs were acquired using a dodecahedron and a subwoofer as sound sources, along with half-inch free-field microphones as monaural receivers. Both sound sources were calibrated in a reverberation room following ISO 3741 [43]. The setup also included a MOTU soundcard, a laptop launching the exponential sine sweep (ESS) signals (512 k at 48 kHz), and Dirac 6.0 commercial software to acquire the measured RIRs and derive the key acoustic indicators in the postprocessing step.

Sound sources and receivers were placed at 1.5 m and 1.2 m above the floor level, respectively. The spatial configuration

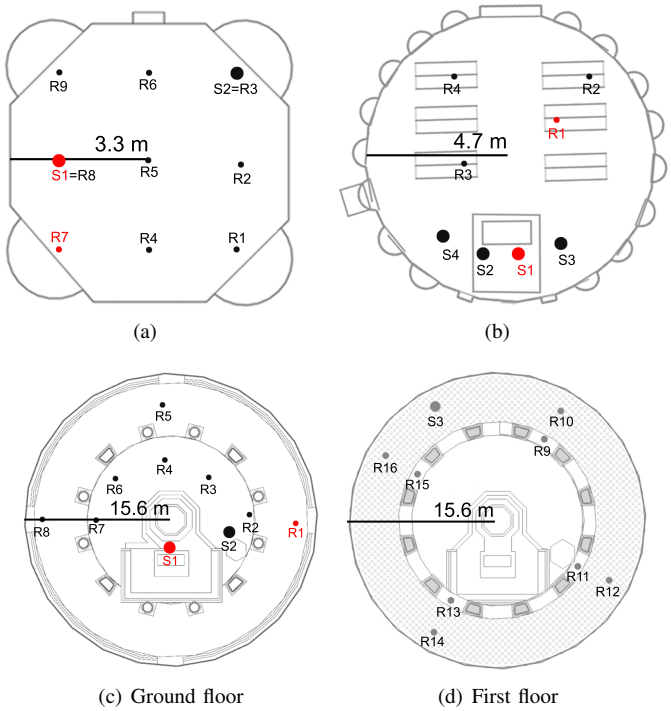

Fig. 2. Sources (S) and receivers (R) layout during the ISO 3382-1 measurements in Odeo Cornaro (a), Rotunda of Madonna del Monte (b), St John's Baptistery ground floor (c) and first floor (d). The main dimension of each hall is provided (images inspired by Ref. [37]). Red labels in each panel indicate the source and receiver locations for the test cases in this section.

of each environment's internal volumes guided the placement of sources and receivers throughout the spaces. The central octagonal space of the first example hosted two sound sources and nine receiver points. In the second hall, sources were placed behind the altar, i.e., the most likely location occupied by speakers and singers, with four receivers spread among the wooden pews. The third space featured two sound sources on the ground floor — at the altar and the pulpit — in accordance with their liturgical relevance, and a third source within the *matroneum* on the first floor. Receiver points were spread among the ambulatory, the baptismal font, and the *matroneum*. Figure 2 shows the layout of sources and receivers during the acoustic measurements, highlighting the source-receiver pairs corresponding to the RIRs assessed in this study. Table

TABLE I
TOTAL VOLUME (V), HEIGHT (H), AND MEAN REVERBERATION TIME VALUE AT 500 - 1000 HZ ($\tau_{30,\text{avg}}$) OF THE THREE TEST CASES.

Hall	V (m^3)	H (m)	$\tau_{30,\text{avg}}$ (s)
Odeo Cornaro (Padua)	220	5.5	2.81
Rotunda of Madonna del Monte (Bologna)	715	12.2	2.78
St. John's Baptistery (Pisa)	23,000	40.0	13.01

I reports the mean values of measured reverberation time at mid-frequencies along with the main geometrical features of each hall (volume and height).

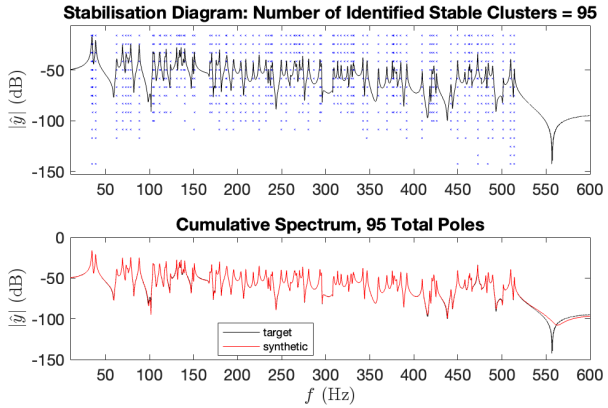

Fig. 3. Stabilisation diagram and globally optimised modal spectrum for the simple resonator described in Section IV-A, yielding 100 modes.

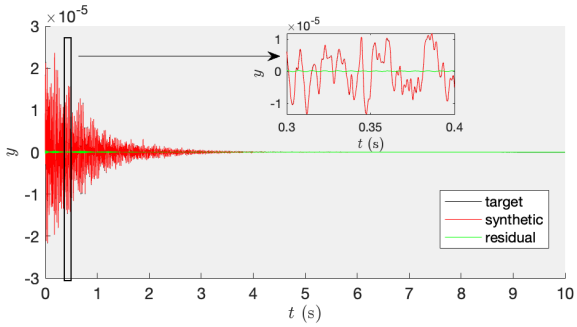

Fig. 4. Impulse response comparison for the system of Section IV-A. The original and synthesised responses are overlaid. Frequency-domain comparisons appear in Fig. 3.

A. Simple Resonator

As a preliminary test case, the PolyMAX algorithm is applied to a simple resonator system over a single frequency band extending from 10 Hz to 600 Hz. The model order p is swept over the range $100 \leq p \leq 400$ in increments of 20. At each iteration, modal stability is assessed, and a threshold of 10% is imposed on both modal frequency and decay time variation to ensure consistent identification.

The resulting stabilisation diagram, shown in Fig. 3, indicates the presence of 95 stable poles across the sweep range. A comparison between the target system response and that obtained from the identified model is also presented. In the time domain, the impulse response reconstruction is shown in Fig. 4, together with the residual error.

An assessment of identification accuracy in terms of modal parameters is given in Fig. 5. Errors are computed relative to the ground truth. While five modes are not explicitly recovered, their spectral locations closely coincide with those of identified components, suggesting masking effects. Frequency errors remain generally below 0.1%, while decay rate and mode shape deviations are typically within 2%, with only minor localised exceptions.

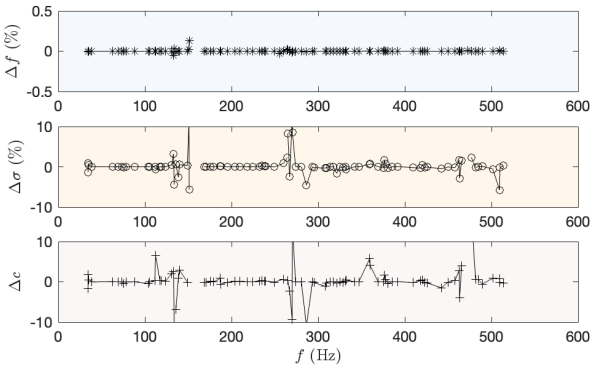

Fig. 5. Relative errors in identified modal parameters: frequency, decay constant, and mode shape, for the case of Figs. 3 and 4.

B. Measured Room Impulse Responses

Application to the hall impulse responses is now considered. The PolyMAX procedure is employed using a sub-band decomposition strategy, as introduced in Section III. The frequency band of interest spans $[20, 12 \cdot 10^3]$ Hz for each of the three examined spaces.

Optimisation Parameters: For the Odeo Cornaro, the smallest of the three spaces, the band is partitioned into 400 uniform sub-bands with 3-band overlaps, resulting in effective band widths of approximately 30 Hz and 90 Hz of overlap per sub-band. For each band, model orders p_{min} and p_{max} are selected within the bounds $50 \leq p_{min} \leq 200$ and $100 \leq p_{max} \leq 500$, depending on frequency. Linear interpolation is used across bands. The model order step is fixed at 50. Poles with decay times exceeding 15 s are discarded. Stability thresholds are set at 10% (frequency) and 30% (decay time).

For the Rotunda, the band decomposition uses 600 sub-bands with a 4-band overlap. Each band spans approximately 20 Hz, with a total overlap of 80 Hz. Identification parameters are held consistent with those of the Odeo Cornaro.

For the Baptistery, which constitutes a considerably larger volume, an analogous procedure is employed with appropriate scaling of model order and band configuration. Here, 1500 equally-spaced bands are considered with a 4-band overlap, resulting in a 8 Hz band length and a 32 Hz band overlap. The polynomial order p_{min} is linearly interpolated across bands between 300 and 200, and p_{max} between 400 and 300. The same stabilisation thresholds are employed as previously.

Results: The identified frequency responses for each space are plotted in Fig. 6, overlaid on their respective measured targets. The residual response is also shown. These results indicate high accuracy of the reconstruction achieved by the PolyMAX algorithm.

Time-domain reconstruction for the Odeo Cornaro is displayed in Fig. 7. The synthesised impulse closely matches the measured response. However, post-response ringing—due to poorly stabilised spurious modes—is observable. Tighter decay time thresholds alleviate this issue but may reduce identification efficiency. An alternative remedy is to apply a moving average smoothing to the decay times, followed by least-squares refinement of the modal residues. This procedure reduces artefactual ringing but introduces a slight phase shift in the early decay.

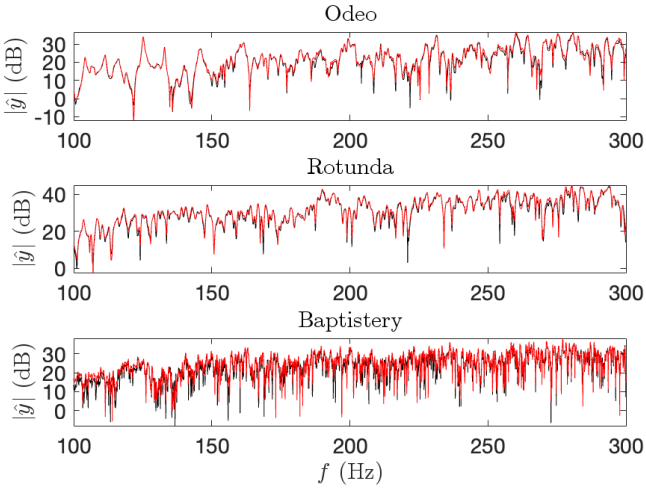

Fig. 6. Low-range target (black) and identified (red) frequency responses for the three analysed halls.

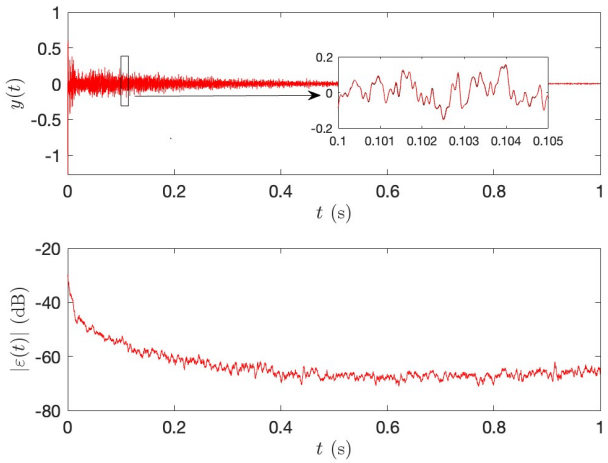

Fig. 7. Impulse responses for the Odeo Cornaro. (a): Target (black), synthesised with decay averaging (red). (b): Residual signal in decibels.

V. Discussion

Modal decomposition provides a structured approach to representing room acoustics. Full-band direct extraction of modal parameters remains computationally challenging, particularly for large environments. The sub-band PolyMAX approach presented here allows for tractable identification over broad frequency ranges. Once modal data are acquired, further statistical analyses become possible.

Decay time distributions, derived from the modal decay constants, are plotted in Fig. 8. Here, a running average over five consecutive modes is shown alongside standard deviation bands. Dashed red lines indicate the 500–1000 Hz octave band boundaries. These distributions may be compared

TABLE II
IDENTIFIED MODAL COUNT (M) AND MEAN MODAL REVERBERATION TIME τ_{60} (500–1000 HZ) FOR THE THREE TEST CASES.

Hall	M	τ_{60} (s)
Odeo Cornaro (Padua)	9072	2.8
Rotunda of Madonna del Monte (Bologna)	11475	2.7
St. John's Baptistery (Pisa)	52470	12.5

to conventional reverberation times obtained from Schroeder integration.

A summary of the average modal reverberation time τ_{60}, computed from the identified modes in the 500–1000 Hz band, is given in Table II. These values closely align with those obtained via classical techniques. The additional capability to analyse statistical variation—unavailable in the Schroeder-based approach—is a key advantage.

Lastly, the numerical integration scheme employed for time-domain synthesis (see Section II-D) supports a highly efficient implementation, suitable for real-time rendering. This is made possible by the low dispersion of the numerical scheme and the ability to run modal banks at reduced update rates. Benchmarks show that up to 20,000 modes may be simulated in real-time using under 15% of a single CPU core on an M2-based machine, when deployed as an audio VST plugin.

A summary of the results, including audio examples, is available on the companion webpage[1]

VI. CONCLUSIONS

A frequency-domain modal decomposition framework has been described for the efficient characterisation and synthesis of large-scale room impulse responses. The core identification procedure utilises the PolyMAX algorithm in a sub-band setting, allowing for the tractable estimation of modal parameters under conditions of high modal overlap and extended bandwidth. The use of stabilisation diagrams across model orders provides a systematic means for distinguishing physically consistent modes from spurious numerical artefacts, allowing for the construction of stable, low-order models without reliance on nonlinear optimisation.

Time-domain synthesis proceeds via a second-order oscillator formulation, discretised using a stable and dispersion-controlled scheme. This enables the real-time resynthesis of RIRs with high fidelity, utilising only a fraction of the available computational resources. A variable-rate update strategy may be adopted across modal banks, without appreciable degradation in output quality, owing to the intrinsic properties of the integration method.

The proposed method has been applied to both synthetic and measured responses, including data collected from three distinct architectural spaces of historical and acoustic interest. Results indicate a close match to measured data in both frequency and time domains. Modal reverberation times derived from the decomposition align closely with those computed via standard techniques, while also permitting statistical analysis over the ensemble of modes—a feature not available through traditional approaches.

The method admits direct extension to multi-channel configurations, and may serve as a basis for parametric spatial modelling, artificial reverberation design, or perceptually informed sound field rendering. Integration into real-time audio engines is straightforward, with performance metrics indicating the capability of simulating modal banks comprising tens of thousands of oscillators using a small computational footprint on modern hardware.

[1]https://mdphys.org/I3DA_2025.html

ACKNOWLEDGEMENTS

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, Grant agreement No. 950084 NEMUS.

REFERENCES

[1] L. Savioja and N. Xiang, "Simulation-based auralization of room acoustics," *Acoust. Today*, vol. 16, no. 4, pp. 49–57, 2020.
[2] S. V. Amengual Garí, D. Eddy, M. Kob, and T. Lokki, "Real-time auralization of room acoustics for the study of live music performance," in *Proceedings of DAGA*, 2016.
[3] P. Masztalski, M. Matuszewski, K. Piaskowski, and M. Romaniuk, "Storir: Stochastic room impulse response generation for audio data augmentation," *arXiv preprint arXiv:2008.07231*, 2020.
[4] A. Ratnarajah, Z. Tang, and D. Manocha, "Room impulse response generator for far-field speech recognition," in *Proc. Interspeech*, 2021.
[5] C.-H. Jeong, "Design, simulation, and virtual prototyping of room acoustics: Challenges and opportunities," in *Proc. Int. Congr. Acoust.*, 2022.
[6] M. Vorländer, *Auralization*. Springer, 2020.
[7] W. C. Sabine, *Collected Papers on Acoustics*. Harvard University Press, 1900.
[8] A. Krokstad, S. Strøm, and S. Sørsdal, "Calculating the acoustical room response by the use of a ray tracing technique," *J. Sound Vib.*, vol. 8, no. 1, pp. 118–125, 1968.
[9] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.
[10] D. Botteldooren, "Finite-difference time-domain simulation of low-frequency room acoustic problems," *J. Acoust. Soc. Am.*, vol. 98, no. 6, pp. 3302–3308, 1995.
[11] L. Savioja, D. Manocha, and M. Lin, "Use of gpus in room acoustic modeling and auralization," in *Proc. ISRA*, 2010, p. 3.
[12] C. J. Webb and S. Bilbao, "Computing room acoustics with cuda-3d fdtd schemes with boundary losses and viscosity," in *2011 IEEE Proc. ICASSP*. IEEE, 2011, pp. 317–320.
[13] K. Kowalczyk and M. van Walstijn, "Room acoustics simulation using 3-d compact explicit fdtd schemes," *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, no. 1, pp. 34–46, 2011.
[14] S. Bilbao and B. Hamilton, "Modeling of complex geometries and boundary conditions in finite difference/finite volume time domain room acoustics simulation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 7, pp. 1524–1533, 2013.
[15] S. Bilbao, B. Hamilton, J. Botts, and L. Savioja, "Finite volume time domain room acoustics simulation under general impedance boundary conditions," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 24, no. 1, pp. 161–173, 2016.
[16] H. Wang, I. Sihar, R. Pagán Muñoz, and M. Hornikx, "Room acoustics modelling in the time-domain with the nodal discontinuous galerkin method," *J. Acoust. Soc. Am.*, vol. 145, no. 4, pp. 2650–2663, 2019.
[17] F. Pind, C.-H. Jeong, A. P. Engsig-Karup, J. S. Hesthaven, and J. Strømann-Andersen, "Time-domain room acoustic simulations with extended-reacting porous absorbers using the discontinuous galerkin method," *J. Acoust. Soc. Am.*, vol. 148, no. 5, pp. 2851–2863, 2020.

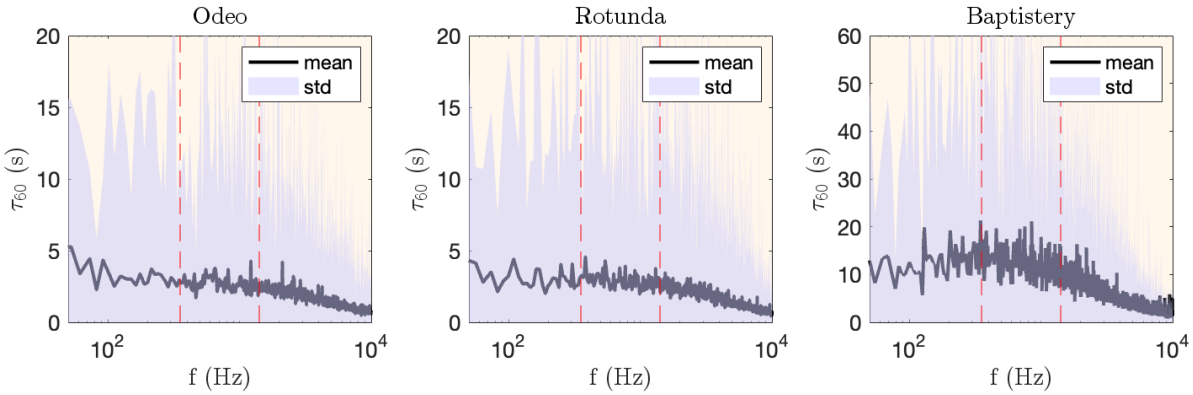

Fig. 8. Statistical analysis of modal decay times. A running average over five consecutive modes is plotted, along with standard deviation bands. Dashed red lines indicate the 500–1000 Hz octave band limits.

[18] W. Wittebol, H. Wang, M. Hornikx, and P. Calamia, "A hybrid room acoustic modeling approach combining image source, acoustic diffusion equation, and time-domain discontinuous galerkin methods," *Appl. Acoust.*, vol. 223, p. 110068, 2024.

[19] J. Nannariello, F. A. DiBella, and H. Saluja, "The use of neural network analysis to predict the acoustic performance of large rooms – part i: Predictions of g utilizing numerical simulations," *Appl. Acoust.*, vol. 62, no. 8, pp. 917–950, 2001.

[20] C.-H. Jeong, Y. Yang, and G. E. Karniadakis, "Sound propagation in realistic interactive 3d scenes with deep neural operators," *Proc. Natl. Acad. Sci. USA*, vol. 121, no. 2, p. e2308837120, 2024.

[21] J. S. Abel, S. Coffin, and K. Spratt, "A modal architecture for artificial reverberation with application to room acoustics modeling," in *AES Conv.*, vol. 137, 2014.

[22] B. Bank, "Perceptually motivated audio equalization using fixed-pole parallel second-order filters," *IEEE Signal Process. Lett.*, vol. 15, pp. 477–480, 2008.

[23] P. Avitabile, *Modal testing: A practitioner's guide*. Hoboken, New Jersey: John Wiley & Sons, 2017.

[24] C. Kereliuk, W. Herman, R. Wedelich, and D. J. Gillespie, "Modal analysis of room impulse responses using subband esprit," in *Proc. DAFx*, 2018.

[25] M. Karjalainen, P. A. Esquef, P. Antsalo, A. Mäkivirta, and V. Välimäki, "Frequency-zooming arma modeling of resonant and reverberant systems," *J. Audio Eng. Soc.*, vol. 50, no. 12, pp. 1012–1029, 2002.

[26] O. Das and J. S. Abel, "Modal estimation on a warped frequency axis for linear system modeling," *arXiv preprint arXiv:2202.11192*, 2022.

[27] J.-N. Juang and R. S. Pappa, "An eigensystem realization algorithm for modal parameter identification and model reduction," *J. Guid. Control Dyn.*, vol. 8, no. 5, pp. 620–627, 1985.

[28] E. Maestre, G. Scavone, and J. O. Smith, "Design of recursive digital filters in parallel form by linearly constrained pole optimization," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1547–1550, 2016.

[29] E. Maestre, J. S. Abel, J. O. Smith, and G. Scavone, "Constrained pole optimization for modal reverberation," in *Proceedings of the 20th Proc. DAFx (DAFx-17)*, Edinburgh, UK, 2017.

[30] P. Guillaume, P. Verboven, S. Vanlanduit, H. Van Der Auweraer, and B. Peeters, "A poly-reference implementation of the least-squares complex frequency-domain estimator," in *Proc. of IMAC*, vol. 21, no. 01, 2003.

[31] B. Peeters, H. Van der Auweraer, P. Guillaume, and J. Leuridan, "The polymax frequency-domain method: a new standard for modal parameter estimation?" *Shock Vib.*, vol. 11, no. 3-4, pp. 395–409, 2004.

[32] M. van Walstijn, J. Bridges, and S. Mehes, "A real-time synthesis oriented tanpura model," in *Proc. DAFx*, 2016, pp. 175–182.

[33] M. van Walstijn, V. Chatziioannou, and A. Bhanuprakash, "Implicit and explicit schemes for energy-stable simulation of string vibrations with collisions: Refinement, analysis, and comparison," *J Sound Vib*, vol. 569, p. 117968, 2024.

[34] J. L. Cieśliński, "On the exact discretization of the classical harmonic oscillator equation," *J. Differ. Equ. Appl.*, vol. 17, no. 11, pp. 1673–1694, 2011.

[35] B. Peeters, G. Lowet, H. Van der Auweraer, and J. Leuridan, "A new procedure for modal parameter estimation," *Sound Vib.*, vol. 38, no. 1, pp. 24–29, 2004.

[36] J. Smith and J. Abel, "Bark and erb bilinear transforms," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 6, pp. 697–708, 2002.

[37] G. Fratoni, M. Garai, and D. D'Orazio, "Assessment of modal density and free path distribution in central-planned halls," *J. Acoust. Soc. Am.*, vol. 154, no. 6, pp. 3604–3614, 2023.

[38] G. Fratoni, D. D'Orazio, M. Ducceschi, and M. Garai, "Acoustic analysis of a well-preserved Renaissance music space: the Odeo Cornaro in Padua," *Acta Acust.*, vol. 8, p. 25, 2024.

[39] G. Fratoni, D. D'Orazio, M. Ducceschi, and M. Garai, "The coupled rooms of Odeo Cornaro (1534) as support for Renaissance musicians and soloists," in *Proc. Int. Congr. Acoust. ICA 2022-Proceedings*, 2022, pp. 617–624.

[40] G. Fratoni, B. Hamilton, and D. D'Orazio, "Rediscovering the acoustics of a XII-Century Rotunda through FDTD simulation," in *2021 Proc. I3DA*. IEEE, 2021, pp. 1–8.

[41] D. D'Orazio, G. Fratoni, E. Rossi, and M. Garai, "Understanding the acoustics of St. John's Baptistery in Pisa through a virtual approach," *J. Build. Perform. Simul.*, vol. 13, no. 3, pp. 320–333, 2020.

[42] *ISO 3382-1:2009 Acoustics — Measurement of room acoustic parameters*.

[43] *ISO 3741:2010 Acoustics — Determination of sound power levels and sound energy levels of noise sources using sound pressure — Precision methods for reverberation test rooms*.