CONSTRAINED POLE OPTIMIZATION FOR MODAL REVERBERATION

Esteban Maestre
CAML / CIRMMT
Schulich School of Music
McGill University
Montréal, Canada
esteban@music.mcgill.ca

Jonathan S. Abel, Julius O. Smith
CCRMA
Music Department
Stanford University
Stanford, USA
abel@ccrma.stanford.edu
jos@ccrma.stanford.edu

Gary P. Scavone
CAML / CIRMMT
Schulich School of Music
McGill University
Montréal, Canada
gary@music.mcgill.ca

ABSTRACT

The problem of designing a modal reverberator to match a measured room impulse response is considered. The modal reverberator architecture expresses a room impulse response as a parallel combination of resonant filters, with the pole locations determined by the room resonances and decay rates, and the zeros by the source and listener positions. Our method first estimates the pole positions in a frequency-domain process involving a series of constrained pole position optimizations in overlapping frequency bands. With the pole locations in hand, the zeros are fit to the measured impulse response using least squares. Example optimizations for a medium-sized room show a good match between the measured and modeled room responses.

1. INTRODUCTION

Modal synthesis has long been used in computer music to simulate large resonating acoustic structures [1]. It was arguably first understood by Daniel Bernoulli circa 1733 [2], when he realized that acoustic vibrations could be seen as a superposition of pure harmonic (sinusoidal) vibrations. Constituting a flexible and efficient approach as compared to convolution techniques, modal structures have been recently proposed for implementing reverberation: [3, 4] suggest synthesizing late-field room reverberation using randomly generated modes; [5, 6] describe how to use measurements to design room reverberation and electromechanical reverberators from spectral peak picking, and implement them as a parallel sum of resonators in a structure termed a "modal reverberator."

The modal reverberator relies on a numerically robust parallel structure [7] that provides computational advantages, as the parallel structure can be efficiently computed, and only audible modes need to be implemented. The parallel decomposition leads to precise control over decay rate, equalization, and other reverberation features, as they can be individually adjusted on a mode by mode basis, and are easily slewed over time. Another advantage of the modal reverberator structure is that separate parameters control the spatial and temporal features of the reverberation: The mode frequencies and dampings are properties of the room or resonating object, describing the mode oscillation frequencies and mode decay times. The mode amplitudes are determined by the source and listener positions according to the mode spatial patterns (given room dimensions and boundary conditions). In this way, the poles of the resonant filters are fixed according to the room, and the zeros are derived from the source and listener positions within the room. In this work, we are concerned with designing a modal reverberator—that is, finding the poles and zeros of each resonant mode filter—so that its output approximates a given measured room response.

In the context of adaptive acoustic echo cancellation, [8, 9] proposed methods for estimating the poles and zeros of moderately low-order transfer functions used to represent the low-frequency region of room responses: using multiple impulse responses obtained for different source and listener positions in the same room, a set of "common acoustical poles" are first estimated using a least-squares technique; then, different sets of zeros are estimated and interpolated to model the localization of the source and listener.

In [10], low-frequency modes were identified in the room impulse response spectrogram from local peaks in estimated reverberation time as a function of frequency. This approach can successfully find modes and their associated poles and zeros, though it is only applicable to long-lasting, low-frequency modes.

In [6], the resonant filter parameters were estimated from a measured room impulse by first finding the mode frequencies from peaks in the room impulse response magnitude spectrum. The number of modes was given by the number of spectral peaks above a given threshold relative to the critical-band-smoothed magnitude spectrum. The mode dampings were estimated from the reverberation time in a band of frequencies about the respective mode frequencies. Finally, the mode amplitudes were found via least-squares fit to the measured impulse response, given the estimated mode frequencies and dampings. While this approach produced a good match between the measured and modeled impulse responses, and could be implemented in such a way as to generate audio effects such as pitch shifting and distortion [11, 12], it was thought that further optimization could noticeably improve the result. Consider a pair of modes that are close in frequency. The stronger mode would likely bias the peak frequency of the weaker mode, as its spectral peak would be on the sloping shoulder of the stronger resonance. In addition, while it is common for adjacent modes to have dampings that differ by a fair amount, the damping assigned to adjacent modes using the approach of [6] would be nearly identical.

In this work, we use an approach similar to that of [13, 14, 15, 16] to optimize initial estimates of the mode pole parameters. As a medium-sized room could have upwards of a couple thousand modes to be modeled, our approach optimizes the mode parameters in a set of overlapping frequency bands so that any given band optimization has a manageable number of parameters. Once the mode pole parameters (equivalently, the mode frequencies and dampings) are estimated, a linear least-squares fit to the measured impulse response is used to find the mode filter zeros.

The rest of the paper is organized as follows. Section 2 introduces a modal reverberator parallel structure and outlines the procedure for pole-zero design from an impulse response measurement. Section 3 describes the procedure used for pole initialization, and Section 4 describes the pole optimization algorithm. Finally,

Section 5 presents some preliminary results and Section 6 discusses potential improvements and applications.

2. MODAL REVERBERATOR DESIGN

Given a reverberation impulse response measurement $h(t)$ and an input signal $x(t)$, one can obtain the *reverberated* output signal $y(t)$ from $x(t)$ by $y(t) = h(t) * x(t)$, where '$*$' denotes convolution. The *modal synthesis* approach [1] approximates $h(t)$ by a sum of M parallel components $h_m(t)$, each corresponding to a resonant mode of the reverberant system $h(t)$. In the z-domain, each parallel term can be realized as a recursive second-order system $H_m(z)$. This is expressed as

$$Y(z) = \hat{H}(z)X(z) = \sum_{m=1}^{M} H_m(z)X(z), \tag{1}$$

where $\hat{H}(z)$ is a $2M$-order digital approximation of $h(t)$, and each m-th parallel term $H_m(z)$ is

$$H_m(z) = (g_{0,m} + g_{1,m}z^{-1})R_m(z) \tag{2}$$

with real numerator coefficients $g_{0,m}$ and $g_{1,m}$, and

$$R_m(z) = \frac{1}{1 + a_{1,m}z^{-1} + a_{2,m}z^{-2}} \tag{3}$$

is a resonator with real coefficients defined by a pair of complex-conjugate poles p_m and $p_m^\star$. Denominator coefficients are related to pole angle and radius by $a_{1,m} = -2|p_m|\cos\angle p_m$ and $a_{2,m} = |p_m|^2$, and define the m-th modal frequency f_m and bandwidth β_m via $f_m = f_s\angle p_m/2\pi$ and $\beta_m = -f_s\log|p_m|/\pi$ respectively, where f_s is the sampling frequency in Hz. Numerator coefficients $g_{0,m}$ and $g_{1,m}$ are used to define the complex gain of the m-th mode [17].

2.1. Design problem

The problem of designing $\hat{H}(z)$ from a given measurement $h(t)$ involves a modal decomposition, and it can be posed as

$$\underset{\mathbf{p},\mathbf{g}_0,\mathbf{g}_1}{\text{minimize}} \quad \varepsilon(\hat{H}, H), \tag{4}$$

where $\mathbf{p}$ is a set of M complex poles inside the upper half of the unit circle on the z-plane, $\mathbf{g}_0$ and $\mathbf{g}_1$ are two sets of M real coefficients, and $\varepsilon(\hat{H}, H)$ is an error measure between the model and the measurement. To find good approximations of dense impulse responses, one needs to face decompositions on the order of hundreds or thousands of highly overlapping modes. In this work we solve this non-linear problem in two steps: first, we find a convenient set of M modes via constrained optimization of complex poles $\mathbf{p}$; second, we obtain the modal complex gains by solving a linear problem to find coefficients $\mathbf{g}_0, \mathbf{g}_1$ as described in Section 2.2.

2.2. Estimation of modal gains

Given a target frequency response $H(e^{j\omega})$ and M complex poles $p_1 \cdots p_m \cdots p_M$, it is straightforward to solve for the numerator coefficients by formulating a linear problem. Let vector $\mathbf{h} = [h_1 \cdots h_k \cdots h_K]^T$ contain K samples of $H(e^{j\omega})$ at normalized angular frequencies $0 \le \omega_k < \pi$, i.e., $h_k = H(e^{j\omega_k})$. Likewise, let vector $\mathbf{r}_m^0 = [r_{m,1}^0 \cdots r_{m,k}^0 \cdots r_{m,K}^0]^T$ sample the frequency response of $R_m(z)$ with $r_{m,k}^0 = R_m(e^{j\omega_k})$, and vector $\mathbf{r}_m^1 = [r_{m,1}^1 \cdots r_{m,k}^1 \cdots r_{m,K}^1]^T$ the frequency response of $z^{-1}R_m(z)$ with $r_{m,k}^1 = e^{-j\omega_k}R_m(e^{j\omega_k})$. Next, let $\mathbf{Q}$ be the $K \times 2M$ matrix of basis vectors constructed as $\mathbf{Q} = [\mathbf{r}_1^0 \cdots \mathbf{r}_M^0, \mathbf{r}_1^1 \cdots \mathbf{r}_M^1]$. Finally, let vector $\mathbf{g}$ contain the numerator coefficients arranged as $\mathbf{g} = [g_{0,1} \cdots g_{0,M}, b_{1,1} \cdots b_{1,M}]^T$. Now we can solve the least-squares projection problem

$$\underset{\mathbf{g}}{\text{minimize}} \quad \|\mathbf{Qg} - \mathbf{h}\|^2. \tag{5}$$

2.3. Pole optimization

Recently we proposed methods for modal decomposition of string instrument bridge admittance measurements (e.g., [14, 16]) using constrained pole optimization via quadratic programming as described in [15], with complexities in the order of dozens of modes. In such methods, a quadratic model is used at each step to numerically approximate the gradient of an error function that spans the full frequency band. Because in this work we deal with decompositions of much higher order, using this technique to simultaneously optimize hundreds or thousands of poles is impractical. Therefore, we propose here an extension of the method to allow for several pole optimizations to be carried out separately, for different overlapping frequency bands. Optimized poles are then collected and merged into a single set from which numerator coefficients are estimated by least-squares as described above leading to (5).

3. INITIALIZATION PROCEDURE

A trusted initial set of pole positions is essential for nonlinear, non-convex optimization. Once the order of the system is imposed, pole initialization comprises two main steps: modal frequency estimation and modal bandwidth estimation. Modal frequency estimation is based on spectral peak picking from the frequency response measurement, while modal bandwidths are estimated from analyzing the energy decay profile as obtained from a time-frequency representation of the impulse-response measurement.

To impose a modal density given the model order M, we assume that the decreasing frequency-resolution at higher frequencies in human hearing makes it unnecessary to implement the ever-increasing modal density of reverberant systems. To that end, we devised a peak-picking procedure to favor a uniform modal density over a warped frequency axis approximating a Bark scale [18]

3.1. Estimation of modal frequencies

Estimation of modal frequencies is based on analysis of the log-magnitude spectrum $\Upsilon(f)$ of the impulse response. From $\Upsilon(f)$, all N peaks in $f_{\min} \le f < f_{\max}$, with $0 \le f_{\min} < f_{\max} < f_s/2$, are found by collecting all N local maxima after smoothing of $\Upsilon(f)$. For each peak, a corresponding peak salience descriptor is computed by frequency-warping and integration of $\Upsilon(f)$ around each peak, as described in [16].

Next, a total of B adjacent bands are defined between $f_{\min}$ and $f_{\max}$, each spanning from delimiting frequencies f_l^b to f_r^b and following a Bark scale. This is carried out by uniformly dividing a Bark-warped frequency axis ϖ into B portions spread between warped frequencies $\varpi_{\min}$ and $\varpi_{\max}$, and then mapping delimiting warped frequencies ϖ_l^b and ϖ_r^b to their linear frequency counterparts. To map between linear and warped frequencies, we use the arctangent approximation of the Bark scale as described in [18].

From all N initial peaks and corresponding annotated frequencies and saliences, picking of M modal frequencies is carried out via an iterative procedure that starts from B lists of candidate peaks, each containing the initial N_b peaks that lie inside band b. First, we define the following variables: P as the total number of picked peaks, N as the total number of available peaks, and N^b as the number of available peaks in the b-th band. The procedure for peak picking is detailed next:

if $N \leq M$ **then**
  pick all N peaks
else
  $P \leftarrow 0$
  while $N - (M - P) > 0 \ \ \& \ \ M > P$ **do**
    $A \leftarrow \lfloor (N - P)/B \rfloor$
    $C \leftarrow (N - P) \mod B$
    $b \leftarrow 1$
    while $b \leq B$ **do**
      $D \leftarrow A + C \mod b$
      $E \leftarrow \min(D, N^b)$
      pick the highest-salience E peaks in list b;
      remove those E peaks from list b.
      $N_b \leftarrow N^b - E$
      $N \leftarrow N - E$
      $P \leftarrow P + E$
    end while
  end while
end if

Once the M peaks have been selected, parabolic interpolation of $\Upsilon(f)$ around each m-th peak is used for defining the m-th modal frequency f_m.

3.2. Estimation of modal bandwidths

Modal bandwidth estimation starts from computing a spectrogram of the impulse response, leading to a time-frequency log-magnitude representation $\Upsilon(t, f)$ from which we want to estimate the decay rate of all L frequency bins within $f_{\min} \leq f < f_{\max}$. Several methods have been proposed for decay rate estimation in the context of modal reverberation (see [3] and references therein). Since in our case we only aim at providing a trusted initial estimate of the bandwidth of each mode, we employ a simple method based on linear interpolation of a decaying segment of the magnitude envelope $\Upsilon(t, f_l)$ of each l-th frequency bin, as follows.

The decaying segment for the l-th bin is delimited by start time t_s^l and end time t_e^l. First, t_s^l is defined as the time at which $\Upsilon(t, f_l)$ has fallen 10 dB below the envelope maximum. Second, t_e^l is defined as the time at which $\Upsilon(t, f_l)$ has reached a noise floor threshold Υ_{th}. To set Υ_{th}, we first estimate the mean μ_n and standard deviation σ_n of the noise floor magnitude, and then define $\Upsilon_{\text{th}} = \mu_n + 2\sigma_n$.

By linear interpolation of the decay segment of each l-th frequency bin, the decay rate τ_l is obtained from the estimated slope, and the bandwidth β_l is computed as $\beta_l = 1/\pi\tau_l$, leading to the bandwidth profile $\beta(f_l)$. Finally, each modal bandwidth β_m is obtained by interpolation of $\beta(f_l)$ at its corresponding modal frequency f_m.

3.3. Pole positioning

Since optimization is carried out in the z-plane, we use estimated frequencies and bandwidths to define M initial pole positions inside the unit circle on the z-plane by computing their angle ω_m and radius $|p_m|$ as $\omega_m = 2\pi f_m / f_s$ and $|p_m| = e^{-\pi\beta_m/f_s}$. We use $\mathbf{p}^x$ to denote the vector of initial pole positions.

4. POLE OPTIMIZATION ALGORITHM

Pole optimization is carried out by dividing the problem into many optimization subproblems, each focused on one of B frequency bands and dealing with a subset of the initial poles. In each b-th subproblem, we optimize only those poles for which corresponding modal frequencies lay inside the b-th frequency band. To help with mode interaction around the band edges, we configure the bands to be partially overlapping. After optimization, we collect optimized poles only from the non-overlapping regions of the bands.

Prior to segmenting into bands or performing any optimization, we use the set $\mathbf{p}^x$ of initial M poles to solve problem (5). This leads to initial gain coefficient vectors $\mathbf{g}_0^x$ and $\mathbf{g}_1^x$. These, together with initial poles $\mathbf{p}^x$, are used to support band pre-processing and optimization as detailed below.

4.1. Band preprocessing

We first uniformly divide the Bark-warped frequency axis ϖ into B adjacent bands between $\varpi_{\min}$ and $\varpi_{\max}$. Band edges of the b-th band are ϖ_ℓ^b and ϖ_r^b. Next, an overlapping region is added to each side of each band, extending the edges from ϖ_ℓ^b and ϖ_r^b to $\varpi_\mathcal{L}^b$ and $\varpi_\mathcal{R}^b$ respectively. The outer edges are defined as $\varpi_\mathcal{L}^b = \varpi_\ell^b - \delta^b(\varpi_r^b - \varpi_\ell^b)$ and $\varpi_\mathcal{R}^b = \varpi_r^b + \delta^b(\varpi_r^b - \varpi_\ell^b)$, with δ^b being a positive real number. The outer edges define the b-th (extended) optimization frequency band, on which the b-th optimization problem is focused. This is illustrated in Figure 1, where each optimization frequency band is conformed by three subbands: a center subband matching the original, non-overlapping band, and two side subbands.

Once the optimization bands have been configured on the warped frequency axis ϖ, all band edges are mapped back to the linear frequency axis f. Then, for each b-th band, two sets of poles $\mathbf{p}^b$ and $\mathbf{p}^{\neg b}$ are created from the initial set $\mathbf{p}^x$ as follows. Set $\mathbf{p}^b$ includes all U poles whose modal frequencies f_u are in $f_\mathcal{L}^b \leq f_u < f_\mathcal{R}^b$, while set $\mathbf{p}^{\neg b}$ includes the remaining O poles (with $O = M - U$), i.e., those poles whose modal frequency f_o is not in $f_\mathcal{L}^b \leq f_o < f_\mathcal{R}^b$. Moreover, from initial gain vectors $\mathbf{g}_0^x$ and $\mathbf{g}_1^x$ we retrieve the gains corresponding to poles in $\mathbf{p}^b$ and to poles in $\mathbf{p}^{\neg b}$, leading to two pairs of gain vectors $\mathbf{g}_0^b$, $\mathbf{g}_1^b$ and $\mathbf{g}_0^{\neg b}$, $\mathbf{g}_1^{\neg b}$. Finally, with arranged sets of poles and gains, we construct two models $H^b(z)$ and $H^{\neg b}(z)$ of the form of $\hat{H}(z)$ in (1). The first model,

$$H^b(z) = \sum_{u=1}^{U} (g_{0,u}^b + g_{1,u}^b z^{-1}) R_u^b(z) \tag{6}$$

with resonators $R_u^b(z)$ constructed from poles $\mathbf{p}^b$, will be optimized to solve the b-th band subproblem (see Section 4.2). The second model,

$$H^{\neg b}(z) = \sum_{o=1}^{O} (g_{0,o}^{\neg b} + g_{1,o}^{\neg b} z^{-1}) R_o^{\neg b}(z) \tag{7}$$

with resonators $R_o^{\neg b}(z)$ constructed from poles $\mathbf{p}^{\neg b}$, presents fixed coefficients and is used to pre-synthesize a frequency response to be used as a constant offset during optimization of the model (6) above

(see Section 4.3). We include this offset response to account for how (fixed) off-band modes contribute to the frequency response of model (6) during optimization.

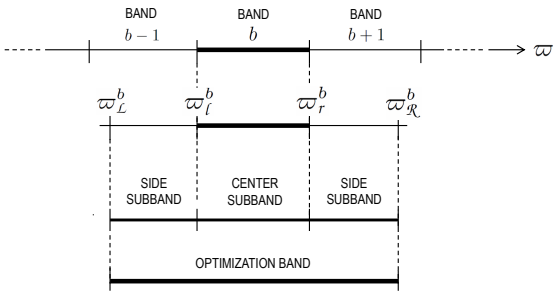

Figure 1: *Optimization preprocessing: partition into overlapping frequency bands on a Bark-warped frequency axis ϖ (depicted is band b).*

4.2. Band optimization

We parametrize the initial set of U modes inside the b-th band by representing each u-th mode as a z-plane complex pole pair p_u^b in terms of two parameters: an angle parameter $w_u^b = |\angle p_u^b|$ and a radius parameter $s_u^b = -\log(1 - |p_u^b|)$. This leads to two parameter sets: a set $\boldsymbol{w}^b = \{w_1^b \cdots w_U^b\}$ of angle parameter values, and a set $\mathbf{s}^b = \{s_1^b \cdots s_U^b\}$ of radius parameter values. With this parametrization, we state the b-th band optimization problem as

$$\begin{aligned} \underset{\boldsymbol{w}^b, \mathbf{s}^b}{\text{minimize}} \quad & \varepsilon(\hat{H}^b, H^b) \\ \text{subject to} \quad & \mathbf{C}^b, \end{aligned} \tag{8}$$

where $\mathbf{C}^b$ is a set of linear constraints specific to band b and $\varepsilon(\hat{H}^b, H^b)$ is an error function, also specific to band b, computed as described in Section 4.3. Note that numerator coefficients have been left out as they are not exposed as variables in the optimization (see [15]). Constraints $\mathbf{C}^b$ are used to restrict the position and arrangement of poles inside the b-th unit circle sector used to represent the b-th band on the z-plane. We have schematically represented the optimization process in Figure 2. We map band edge frequencies f_L^b and f_R^b to sector edge angles ω_L^b and ω_R^b via $\omega_L^b = 2\pi f_L^b / f_s$ and $\omega_R^b = 2\pi f_R^b / f_s$ respectively.

A key step before constraint definition is to sort the pole parameter sets so that linear constraints can be defined in a straightforward manner to ensure that the arrangement of poles in the b-th unit circle sector is preserved during optimization, therefore reducing the number of crossings over local minima (see [15]). Elements in sets $\boldsymbol{w}^b$ and $\mathbf{s}^b$ are jointly sorted as pairs (each pair corresponding to a complex-conjugate pole) by ascending angle parameter w_u.

From ordered sets $\boldsymbol{w}^b$ and $\mathbf{s}^b$, linear constraints $\mathbf{C}^b$ are defined as follows. First, stability is enforced by $0 < s_u^b$. Then, poles are constrained to stay in the b-th sector via $\omega_L^b \leq w_u^b < \omega_R^b$. Next, to aid convergence we constrain the pole sequence order in set $\boldsymbol{w}^b$ to be respected. This is expressed by $w_{u-1}^b < w_u^b < w_{u+1}^b$. Moreover, assuming that initialization provides an already trusted first solution, we can bound the search to a region around the initial pole positions. This can be expressed via the additional inequalities $w^- < w_u^b < w^+$ and $s^- < s_u^b < s^+$, where '$-$' and '$+$' superscripts are used to indicate lower and upper bounds, respectively.

We solve this problem by means of sequential quadratic programming [19]. At each step, the error surface is quadratically approximated by successive evaluations of the band approximation error function described in Section 4.3.

4.3. Band error computation

At each i-th step of the optimization, given a set of poles $\mathbf{p}^b|_i$ defined from current values in parameter sets $\boldsymbol{w}^b$ and $\mathbf{s}^b$, the error $\varepsilon(\hat{H}^b|_i, H^b)$ is computed by solving a linear problem similar to 5, but restricted to the b-th frequency band.

First, let vector $\mathbf{h} = [h_1 \cdots h_k \cdots h_K]^T$ contain K samples of the measurement $H(e^{j\omega})$ at normalized angular frequencies $\omega_L^b \leq \omega_k < \omega_R^b$, i.e., $h_k = H(e^{j\omega_k})$. Similarly, let vector $\mathbf{v} = [v_1 \cdots v_k \cdots v_K]^T$ contain K samples of the frequency response of model $\hat{H}^{\neg b}(z)$, i.e., $v_k = \hat{H}^{\neg b}(e^{j\omega_k})$. Then, we obtain U frequency response vectors $\mathbf{r}_u^0|_i = [r_{u,1}^0|_i \cdots r_{u,k}^0|_i \cdots r_{u,K}^0|_i]^T$ by using $\mathbf{p}^b|_i$ to evaluate each $R_u^b(z)$ in the same frequency range, i.e., $r_{u,k}^0|_i = R_u^b(e^{j\omega_k})|_i$. Likewise, we obtain U vectors $\mathbf{r}_u^1|_i = [r_{u,1}^1|_i \cdots r_{u,k}^1|_i \cdots r_{u,K}^1|_i]^T$ by evaluating each $z^{-1}R_u^b(z)$, i.e., $r_{u,k}^1|_i = e^{-j\omega_k} R_u^b(e^{j\omega_k})|_i$. Next, all $2U$ vectors $\mathbf{r}_u^0|_i$ and $\mathbf{r}_u^1|_i$ are arranged to form a basis matrix $\mathbf{Q}^b|_i$ of size $K \times 2U$ as $\mathbf{Q}^b|_i = [\mathbf{r}_1^0|_i \cdots \mathbf{r}_U^0|_i, \mathbf{r}_1^1|_i \cdots \mathbf{r}_U^1|_i]$. Finally, let vector $\mathbf{g}^b$ contain the numerator coefficients of model $H^b(z)$ arranged as $\mathbf{g}^b = [g_{0,1}^b \cdots g_{0,U}^b, g_{1,1}^b \cdots g_{1,U}^b]^T$. With all these, we solve the least-squares problem

$$\underset{\mathbf{g}^b}{\text{minimize}} \quad \|\mathbf{Q}^b|_i \mathbf{g}^b + \mathbf{v} - \mathbf{h}\|^2 \tag{9}$$

and use obtained vector $\mathbf{g}^b$ to compute the i-th step error as

$$\varepsilon(H^b|_i, \hat{H}^b) = \|\mathbf{Q}^b|_i \mathbf{g}^b + \mathbf{v} - \mathbf{h}\|^2. \tag{10}$$

4.4. Pole collection and problem solution

Once the poles of the B overlapping frequency bands have been optimized, the final set poles $\mathbf{p}$ is constructed by collecting all optimized poles inside each of the B center subsectors, i.e., all poles whose angle parameter w satisfy $\omega_\ell^b \leq w < \omega_r^b, \forall b \in \{1 \dots B\}$. Using collected poles, we solve problem (5) of Section 2.2.

5. EXAMPLE RESULTS

We carried out a set of test examples to model an impulse response measurement taken from a medium-sized room at a sampling frequency of 48000 Hz. In our models, we explored different orders $M = 800, 1200, 1600, 1800$, with modes in the region between 30 Hz and 20 kHz. In all cases, we chose to use a sufficiently large number of bands $B = 200$, with a constant overlapping $\delta^b = 1.0$. Example synthesized impulse responses are available online[1].

In Figure 3 we display the spectrogram of the target impulse response plus three example models of order $M = 800, 1200, 1800$. While we observe an overall good match both in frequency and time, it is clear how the response of lower order models (e.g., $M = 800$) present a more sparse modal structure, especially in the high end.

[1] http://ccrma.stanford.edu/∼esteban/modrev/dafx2017

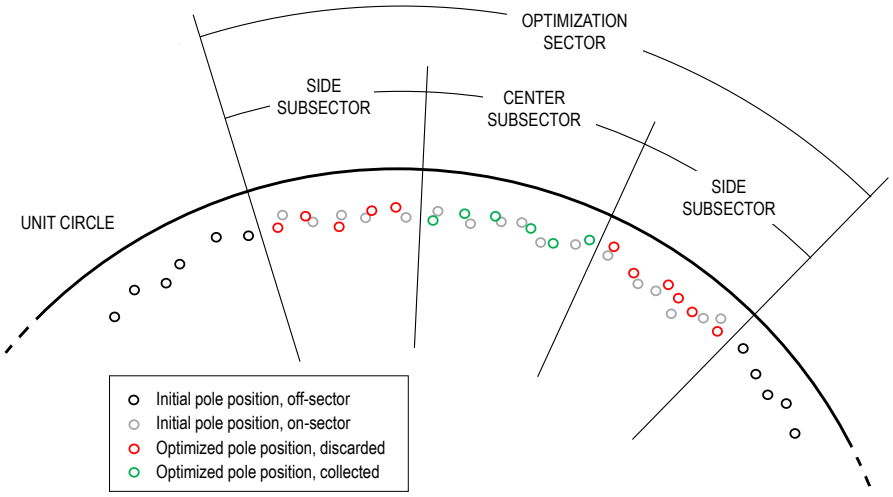

Figure 2: *Schematic representation of pole optimization inside a frequency band, which is mapped to a sector of the unit circle on the z-plane.*

As it can be perceived from listening to the modeled responses, this leads to a metallic character in the sound.

To get an idea of the how the modeling error is reduced during optimization, we compare the error $H(\omega) - \hat{H}(\omega)$ obtained before and after optimization. This is featured in Figure 4 for $M = 1600$, where it is possible to observe the coarse envelope of the error to decrease by 5 to 10 dB in all frequency regions.

For a detail of how optimization improves the modeling accuracy, in Figure 5 appear the magnitude responses of the target measurement (middle), initial (bottom) and optimized models (top) of $M = 1600$ for three different frequency regions. With regard to the time domain, initial and optimized impulse response models are compared to the target measurement in Figure 6, where it is possible to observe how optimization helps to significantly reduce the pre-onset ripple of the model.

6. CONCLUSION

We have presented a frequency-domain pole-zero optimization technique to fit the coefficients of a modal reverberator, applied to model a mid-sized room impulse-response measurement. Our method, which includes a pole initialization procedure to favor a constant density of modes over a Bark-warped frequency axis, is based on constrained optimization of pole positions within a number of overlapping frequency bands. Once the pole locations are estimated, the zeros are fit to the measured impulse response using linear least squares. Our initial explorations with example models of a medium-sized room display a good agreement between the measured and modeled room responses, demonstrating how our pole optimization technique can be of practical use in modeling problems that require thousands of modes to accurately simulate dense impulse responses.

Our initial results show a promising path for improving the accuracy of efficient modal reverberators. At the same time, optimization could lead to a reduction of the computational cost given a required accuracy. Besides carrying out a more exhaustive exploration of model orders and parameter values (e.g. frequency-dependent overlapping factor), gathering data from subjective tests could provide a good compromise between perceptual quality and computational cost. In terms of constraint definition, upper and lower bounds are still set by hand—further exploration could lead to methods for designing bounds for pole radii by attending to a statistical analysis of observed modal decays and therefore avoid an excess in pole selectivity. A potential development aims at making the whole process to be iterative: use the optimized solution as an initial point, perform again the optimization, and repeat until no improvement is obtained; this could perhaps have a positive effect in the vicinity of band edges, though its significance would need to be tested.

In conclusion, we note that *modal synthesis* supports a *dynamic* representation of a space, at a fixed cost, so that one may simulate walking through the space or listening to a moving source by simply changing the coefficients of the fixed modes being summed. The modal approach may thus be considered competitive with three-dimensional physical models such as Scattering Delay Networks [20, 21, 22] in terms of psychoacoustic accuracy per unit of computational expense. In that direction, an imminent extension of our method deals with simultaneously using N impulse response measurements of the same space (a set of measurements taken for different combinations of source and receiver locations, as previously proposed in [8] for low-order equalization models of the low frequency region) to optimize a common set of poles: at each step, poles in each frequency band are optimized via minimization of a global error function that simultaneously accounts for the approximation error of N models (one per impulse response) constructed from the the same set of poles. A common set of poles suffices for each fixed spatial geometry, or coupled geometries.

7. REFERENCES

[1] J. M. Adrien, "The missing link: Modal synthesis," in *Representations of Musical Signals*, G. De Poli, A. Piccialli, and C. Roads, Eds., pp. 269–267. MIT Press, 1991.

[2] O. Darrigol, "The acoustic origins of harmonic analysis," *Archive for History of the Exact Sciences*, vol. 61, no. 4, 2007.

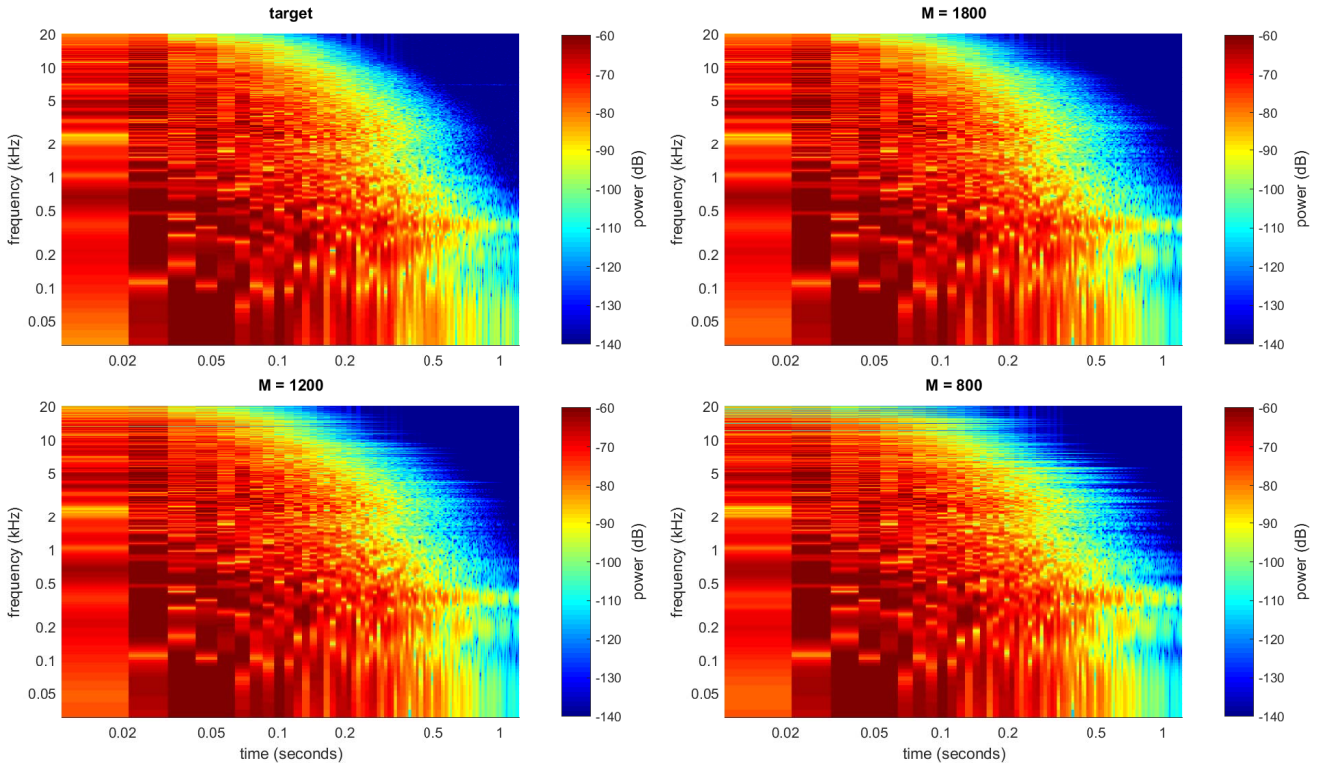

Figure 3: *Spectrogram of the target impulse response plus three example models with* $M = 800, 1200, 1800$.

[3] M. Karjalainen, P. Antsalo, A. Mäkivirta, T. Peltonen, and V. Välimäki, "Estimation of modal decay parameters from noisy response measurements," *Journal of the Audio Engineering Society*, vol. 50:11, pp. 867–878, 2002.

[4] M. Karjalainen and H. Jarvelainen, "More about this reverberation science: Perceptually good late reverberation," in *Audio Engineering Society Convention 111*, New York, U.S.A, 2001.

[5] J. S. Abel, S. A. Coffin, and K. S. Spratt, "A modal architecture for artificial reverberation," *The Journal of the Acoustical Society of America*, vol. 134:5, pp. 4220, 2013.

[6] J. S. Abel, S. A. Coffin, and K. S. Spratt, "A modal architecture for artificial reverberation with application to room acoustics modeling," in *Proc. of the Audio Engineering Society 137th Convention*, 2014.

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

[8] Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *IEEE Transactions on Speech and Audio Processing*, vol. 2:2, pp. 320–328, 1994.

[9] Y. Haneda, Y. Kaneda, and N. Kitawaki, "Common-acoustical-pole and residue model and its application to spatial interpolation and extrapolation of a room transfer function," *IEEE Transactions on Speech and Audio Processing*, vol. 7:6, pp. 709–717, 1999.

[10] A. Mäkivirta, P. Antsalo, M. Karjalainen, and V. Välimäki, "Low-frequency modal equalization of loudspeaker-room responses," in *Audio Engineering Society 111th Convention*, 2001.

[11] J. S. Abel and K. J. Werner, "Distortion and pitch processing using a modal reverberator," in *Proc. of the International Conference on Digital Audio Effects*, 2015.

[12] K. J. Werner and J. S. Abel, "Modal processor effects inspired by Hammond tonewheel organs," *Applied Sciences*, vol. 6(7), pp. 185, 2016.

[13] E. Maestre, G. P. Scavone, and J. O. Smith, "Modeling of a violin input admittance by direct positioning of second-order resonators," *The Journal of the Acoustical Society of America*, vol. 130, pp. 2364, 2011.

[14] E. Maestre, G. P. Scavone, and J. O. Smith, "Digital modeling of bridge driving-point admittances from measurements on violin-family instruments," in *Proc. of the Stockholm Music Acoustics Conference*, 2013.

[15] E. Maestre, G. P. Scavone, and J. O. Smith, "Design of recursive digital filters in parallel form by linearly constrained pole optimization," *IEEE Signal Processing Letters*, vol. 23:11, pp. 1547–1550, 2016.

[16] E. Maestre, G. P. Scavone, and J. O. Smith, "Joint modeling of bridge admittance and body radiativity for efficient synthesis of string instrument sound by digital waveguides," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 25:5, pp. 1128–1139, 2017.

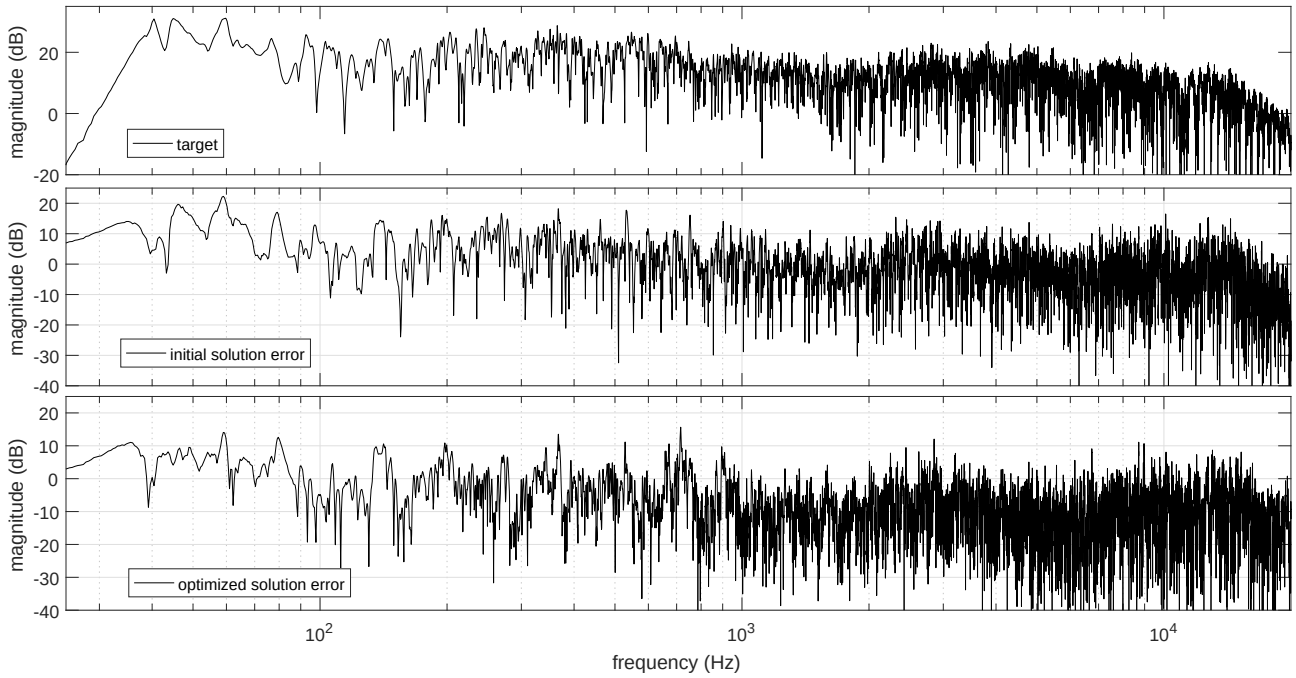

Figure 4: *Comparison of the approximation error as as function of frequency, for* $M = 1600$.

[17] J.O. Smith, *Physical Audio Signal Processing*, W3K Publishing, 2004, online book: `http://ccrma.stanford.edu/~jos/pasp/`.

[18] J. O. Smith and J. S. Abel, "Bark and ERB bilinear transforms," *IEEE Transactions on Speech and Audio Processing*, vol. 7:6, pp. 697–708, 1999.

[19] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, 2006.

[20] E. De Sena, H. Hacihabiboglu, Z. Cvetkovic, and J.O. Smith, "Efficient synthesis of room acoustics via scattering delay networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 9, pp. 1478–1492, 2015.

[21] H. Hacihabiboglu, E. De Sena, Z. Cvetkovic, J. Johnston, and J. O. Smith, "Perceptual spatial audio recording, simulation, and rendering," *IEEE Signal Processing Magazine*, pp. 36–54, 2017.

[22] A. Meacham, L. Savioja, S. R. Martín, and J. O. Smith, "Digital waveguide network reverberation in non-convex rectilinear spaces," in *Audio Engineering Society 141st Convention*, 2016.

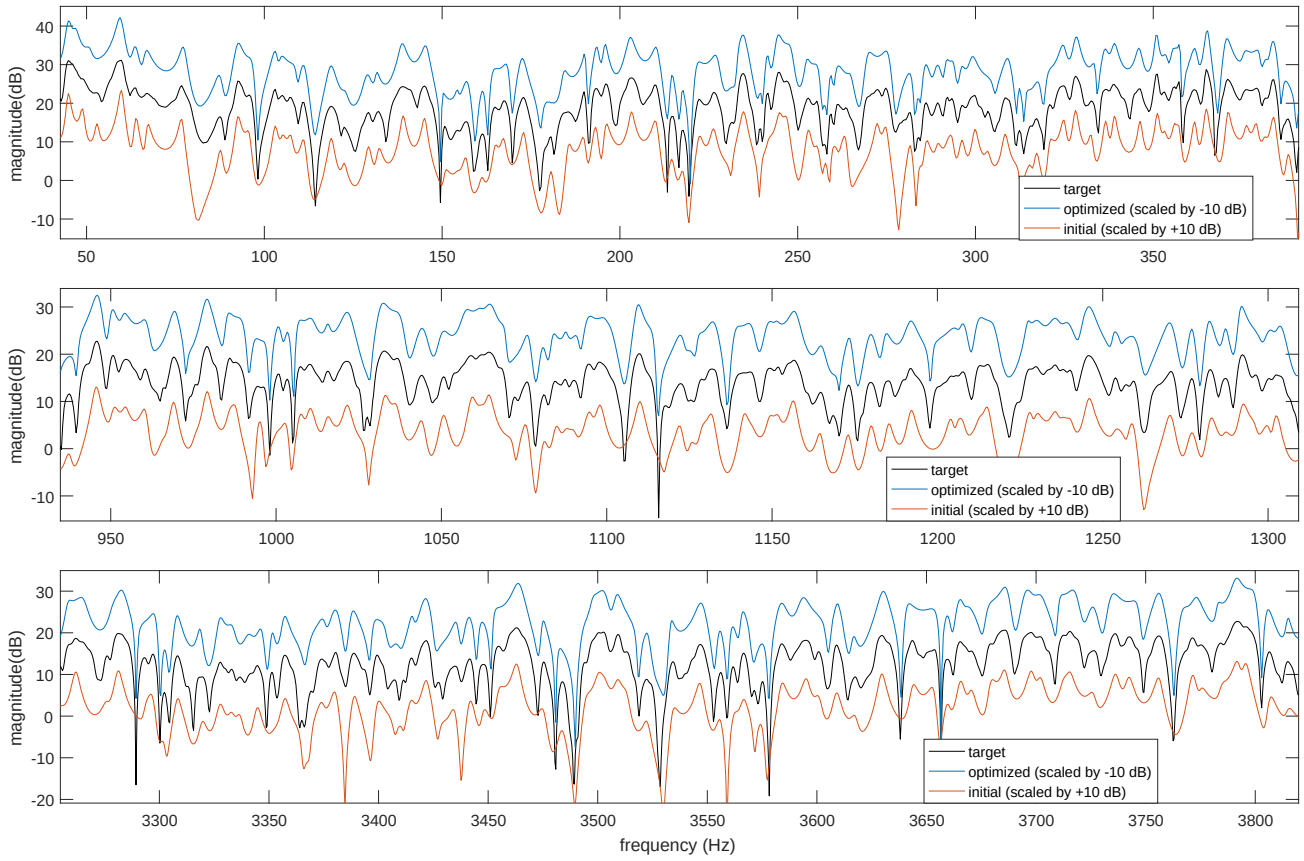

Figure 5: *Magnitude response of an example model with* $M = 1600$*, displayed for three different frequency regions.*

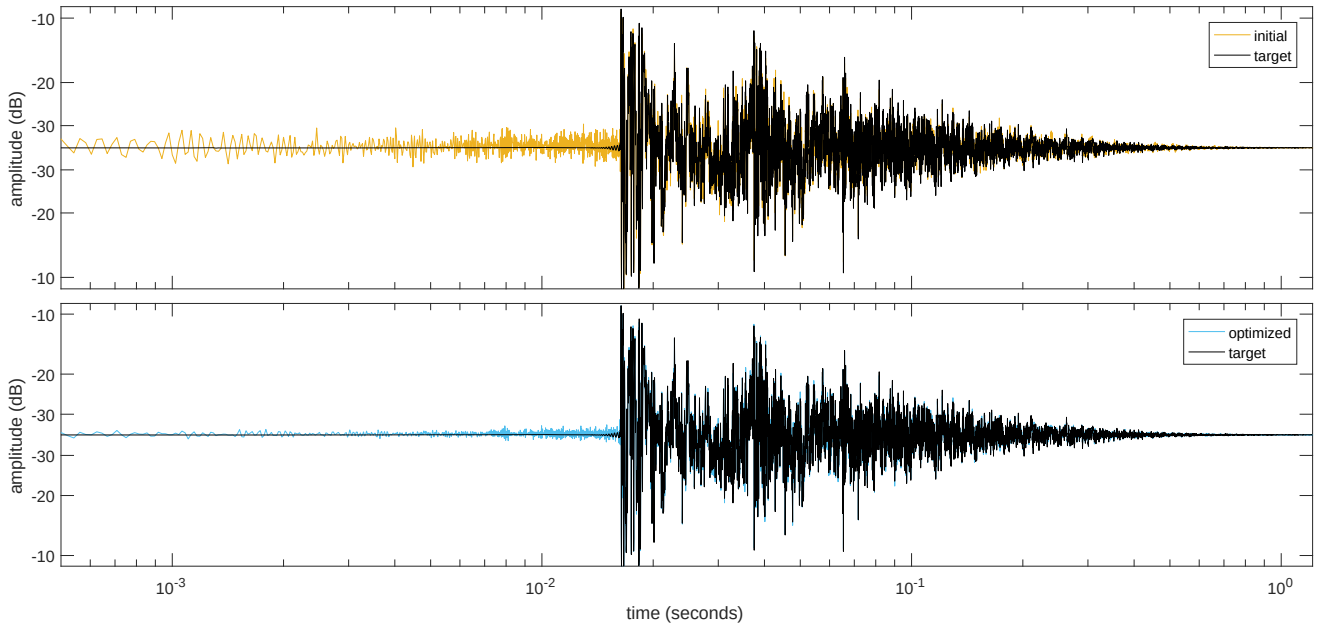

Figure 6: *Initial (top) and optimized (bottom) impulse response example models, for* $M = 1600$*.*