KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

| | |
|---|---|
| **Citation/Reference** | Vairetti G., De Sena E., Catrysse M., Jensen S.H., Moonen M., van Waterschoot T. (2017),<br>**A scalable algorithm for physically motivated and sparse approximation of room impulse responses with orthonormal basis functions**<br>IEEE/ACM Trans. Audio Speech Lang. Process. |
| **Archived version** | Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher. |
| **Published version** | Accepted for publication |
| **Journal homepage** | http://signalprocessingsociety.org/publications-resources/ieeeacm-transactions-audio-speech-and-language-processing/ieeeacm |
| **Author contact** | giacomo.vairetti@esat.kuleuven.be<br>*+ 32 (0)16 321817* |
| **IR** | url in Lirias https://lirias.kuleuven.be/handle/123456789/581178 |

*(article begins on next page)*

# A Scalable Algorithm for Physically Motivated and Sparse Approximation of Room Impulse Responses with Orthonormal Basis Functions†

Giacomo Vairetti*[1], Enzo De Sena[1], *Member, IEEE,*
Michael Catrysse[2], Søren Holdt Jensen[3], *Senior Member, IEEE,*
Marc Moonen[1], *Fellow, IEEE,* and Toon van Waterschoot[1,4], *Member, IEEE*

***Abstract*—Parametric modeling of room acoustics aims at representing room transfer functions (RTFs) by means of digital filters and finds application in many acoustic signal enhancement algorithms. In previous work by other authors, the use of orthonormal basis functions (OBFs) for modeling room acoustics has been proposed. Some advantages of OBF models over all-zero and pole-zero models have been illustrated, mainly focusing on the fact that OBF models typically require less model parameters to provide the same model accuracy. In this paper, it is shown that the orthogonality of the OBF model brings several additional advantages, which can be exploited if a suitable algorithm for identifying the OBF model parameters is applied. Specifically, the orthogonality of OBF models does not only lead to improved model efficiency (as pointed out in previous work), but also leads to improved model scalability and model stability. Its appealing scalability property derives from a previously unexplored interpretation of the OBF model as an approximation to a solution of the inhomogeneous acoustic wave equation. Following this interpretation, a novel identification algorithm is proposed that takes advantage of the OBF model orthogonality to deliver efficient, scalable and stable OBF model estimates, which is not necessarily the case for nonlinear estimation techniques that are normally applied.**

***Index Terms*—Parametric modeling, orthonormal basis function models, room acoustics, matching pursuit.**

## I. Introduction

PARAMETRIC modeling of room acoustics aims at representing room transfer functions (RTFs) by means of rational expressions in the $z$-transform domain, implemented through digital filters, and finds application in a variety of acoustic signal enhancement tasks, e.g. echo cancellation, feedback cancellation, and dereverberation, as well as in auralization systems. The most common parametric models are All-Zero (AZ) models [1], which define a finite impulse response (FIR) filter as a truncation of a sampled room impulse response (RIR). AZ models enable achieving an arbitrary degree of accuracy, but a good approximation of a RIR usually requires a large number of model parameters. Pole-Zero (PZ) models [2], which produce an infinite impulse response (IIR), are used sometimes in order to reduce the number of parameters. PZ models have a more meaningful motivation from a physical point of view, in the sense that the resonant behavior of room acoustic responses can be represented by means of complex-conjugate poles in the transfer function. This is particularly true when a PZ model is implemented using the parallel form of fixed-pole IIR filters [3, p.359]. This parallel filter (PF), proposed in recent years for RTF modeling and audio equalization [4]–[8], consists of second-order all-pole filters, each of which is defined by a pair of complex-conjugate poles. Its transfer function is given by a linear combination of resonances, in analogy with the physical definition of a RTF as an infinite summation of room modes [9]–[11]. However, since RTFs are characterized by a complicated time-frequency evolution and a large number of room resonances, the improvement in modeling efficiency obtained with PZ models compared to AZ models is in some cases only marginal [12]. Moreover, PZ models often suffer from instability and ill-conditioning issues in the estimation of the model parameters, especially for high model orders, which is why AZ models are usually preferred.

In order for models producing an IIR to become a valid alternative to AZ models, models with improved model efficiency and with stable and numerically well-conditioned identification algorithms (and possibly other interesting properties) are sought. Fixed-pole models based on Orthonormal Basis Functions (OBFs) [13]–[16] can be derived directly from an orthogonalization of PF models. OBF models span the same approximation space of PF models for the same set of poles, with the difference that the outputs of each second-order all-pole filter are made orthogonal to each other by a sequence of all-pass filters (i.e. by zero-pole cancellation). The use of single-pole OBF models for acoustic echo cancellation [17], [18], and of multiple-pole OBF models for loudspeaker response equalization and modeling of room and musical instrument responses [19]–[23] have been previously motivated by the possibility of positioning the poles anywhere inside the unit circle, thus providing stability of the filter and giving freedom in the allocation of the frequency resolution. It has been shown that these properties, together with orthogonality, provide a more accurate representation of the RTF for a

†EDICS: AUD-MAAE (Modeling, Analysis and Synthesis of Acoustic Environments)

*Corresponding author (giacomo.vairetti@esat.kuleuven.be)
[1]KU Leuven, Dept. of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Kasteelpark Arenberg 10, 3001 Leuven, Belgium.
[2]Televic N.V. Leo Bekaertlaan 1, 8870 Izegem, Belgium.
[3]Aalborg University, Dept. of Electronic Systems, Fredrik Bajers Vej 7B, 9220 Aalborg, Denmark.
[4]KU Leuven, Dept. of Electrical Engineering (ESAT-ETC), AdvISe Lab, Kleinhoefstraat 4, 2440 Geel, Belgium.

given number of model parameters, compared to conventional models. Differently from PF models, orthogonality makes the estimation of the parameters that appear linearly in OBF models straightforward and numerically well-conditioned. The poles, on the other hand, appear nonlinearly in the model, which makes their estimation a difficult problem, requiring in principle nonlinear estimation techniques. In [18], the pole parameters of single-pole OBF models were optimized using the Gauss-Newton method. In [19]–[23], multiple poles were estimated with a nonlinear iterative algorithm for FIR-to-IIR filter conversion, called the Brandenstein-Unbehauen (BU) method [24], resembling the Steiglitz-McBride method for PZ modeling [25]. The BU method exploits the orthogonality of OBF filters by minimizing the energy of a target RIR with a sequence of all-pass filters. Although this method is capable of producing accurate pole estimates, it is not exempt from numerical problems for high model orders, in which case the algorithm can converge to a local minimum and even produce unstable poles. Modifications of the BU method have been proposed to overcome this problem, such as through prewarping of the target RIR (warped BU – wBU) [21], [26], in order to approximate a desired frequency resolution, or partitioning of the target RIR in frequency subbands or in time [27]. Furthermore, the BU method and its variants require the model order to be determined before estimation, resulting in a non-scalable algorithm that has to be run every time the number of poles to be estimated changes.

The nonlinear problem of estimating the poles was bypassed in [28] by applying convex optimization to a discrete grid of candidate stable poles. A sparse solution was obtained by selecting basis functions out of a large non-orthogonal dictionary. In [29], a matching-pursuit-based algorithm called OBF-MP was introduced. A similar algorithm was also suggested in [30] for the estimation of the poles of a RIR model described as the linear combination of sampled exponentially decaying sinusoids, but not considering any particular filter implementation of the model (if not the implicit use of FIR filters); however, the choice of this model implies a non-orthogonal dictionary and, consequently, ill-conditioning problems in the estimation of the parameters, which would require the use of computationally more complex versions of the algorithm, such as Orthogonal MP as in [31], or suboptimal iterative procedures [30]. The OBF-MP algorithm [29], instead, exploits the appealing properties of OBF models, i.e. orthogonality, stability and numerical well-conditioning, in order to deliver efficient, scalable and stable OBF model estimates for room acoustic modeling, which can be directly implemented through a stable IIR filter. It is shown in the present work that the scalability property of the algorithm stems from a previously unexplored interpretation of OBF models as an approximation to a solution of the inhomogeneous acoustic wave equation. Indeed, OBF models are physically motivated in the modal region, where the RTF is a linear combination of room resonances, sparse in frequency. The OBF-MP algorithm thus provides a sparse approximation of the most dominant modes in the low-frequency region of the RTF, while approximating the spectral envelope at higher frequencies. In this paper, the OBF-MP algorithm is further investigated and its performance in terms of efficiency and computational complexity is studied for a large set of measured RIRs.

The paper is organized as follows. In Section II, fundamentals of the theory of room acoustics are briefly reviewed, together with an overview of conventional parametric models. In Section III, the OBF models are reviewed in detail, as well as their use in the approximation of a target RIR. The OBF-MP algorithm is described in Section IV and its computational complexity is analyzed. In Section V, the concept of model and filter complexity of different parametric models is introduced. Simulation results are shown in Section VI, comparing the performance in the approximation of a large set of measured RIRs of OBF models estimated using the OBF-MP algorithm with respect to conventional models and OBF models estimated using the BU method. A discussion of the results and future work can be found in Section VII, which also concludes the paper.

## II. Parametric modeling of room acoustics

This section reviews elements of room acoustics and provides an overview of conventional parametric models.

### A. Fundamentals of room acoustics

The RTF between an omnidirectional point source $s(\boldsymbol{r}, t) = s(t)\delta(\boldsymbol{r} - \boldsymbol{r}_s)$ at position $\boldsymbol{r}_s = (x_s, y_s, z_s)$ (with $s(t)$ a given source function and $\delta(\cdot)$ the Kronecker delta function) and a receiver at position $\boldsymbol{r} = (x, y, z)$, can be seen as a linear superposition of room modes, mutually orthogonal in the space dimension, with the mode amplitudes depending on $\boldsymbol{r}$ and $\boldsymbol{r}_s$, and on the strength of the source. This is described by the Green's Function (GF) of the inhomogeneous acoustic wave equation, which, neglecting higher-order terms such as the variability of the temperature and of the density of the medium [9], [10], is given by

$$P(\boldsymbol{r}, \boldsymbol{r}_s, \omega) = G(\omega) \sum_{i=1}^{\infty} \frac{\psi_i(\boldsymbol{r})\psi_i(\boldsymbol{r}_s)\ j\omega}{\omega^2 - \omega_i^2 - 2j\zeta_i\omega_i + \zeta_i^2}, \quad (1)$$

with $P(\boldsymbol{r}, \boldsymbol{r}_s, \omega)$ the sound pressure in a room at the driving frequency $\omega$ for given receiver and source positions $\boldsymbol{r}$ and $\boldsymbol{r}_s$, and $G(\omega)$ a frequency-dependent gain constant. The eigenfrequencies $\omega_i$, also called resonance frequencies [9], are the values of $\omega$ for which the acoustic wave equation has non-zero solutions satisfying the boundary conditions. The eigenfunction $\psi_i$ corresponding to eigenfrequency $\omega_i$ defines a three-dimensional standing wave, called a room mode. A given room mode is dominant when the driving frequency $\omega$ is close to its resonance frequency $\omega_i$, while it has no contribution to the sound field when the source or the receiver is placed on one of its nodal surfaces, i.e. where either $\psi_i(\boldsymbol{r}_s)$ or $\psi_i(\boldsymbol{r}_0)$ is zero. The damping constant $\zeta_i$ accounts for frequency-dependent energy losses at the walls and determines the half-bandwidth at -3 dB of the room resonance, which is $B = \zeta_i/\pi$ (in Hz) [9], [10].

The inverse Fourier Transform (FT) of (1) gives the RIR, which, for $t \geq 0$, is a sum of exponentially decaying sinusoids,

$$h(\boldsymbol{r}, \boldsymbol{r}_s, t) = \sum_{i=1}^{\infty} c_i e^{-\zeta_i t} \cos(\omega_i t + \phi_i), \quad (2)$$

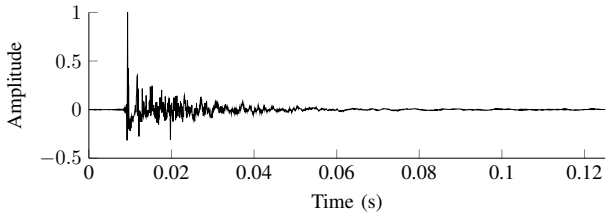


Fig. 1. RIR measured in the Speech Lab at KU Leuven [33].

where the $i^{\text{th}}$ sinusoid has amplitude $c_i$ and phase $\phi_i$, resonance frequency $\omega_i$ and a decay determined by the damping constant $\zeta_i$. The GF describes the sound field for any possible position of the source and the receiver inside any kind of room. However, a closed-form analytical expression for $\psi_i$ exists only for simple room shapes and for simple boundary conditions.

The problem of modeling a RIR presents many challenges, mainly because of its complicated time-frequency structure. The RIR measured in a reverberant room has typically a very long duration and presents a complicated pattern of the arrival of reflections. An example of a typical RIR is shown in Fig. 1. Furthermore, the modal density increases with the square of the frequency $\omega$, i.e. the approximate number of eigenfrequencies per Hz is given by

$$n_{\omega_i}(\omega) \approx \frac{V}{c^3\pi}\omega^2\,, \tag{3}$$

with $V$ the volume of the room and $c$ the sound velocity. The expression (3) is derived for rectangular rooms, but is asymptotically valid for rooms of any shape [9], [10]. It follows that the modes are well separated only at low frequencies, while they tend to overlap at higher frequencies. The so-called 'Schroeder frequency' [32] gives an indication as to where the transition between these two regions occurs:

$$f_{\text{Sch}} \approx 2000\sqrt{\frac{T}{V}}\,, \tag{4}$$

where $T$ is the reverberation time, defined as the time it takes for the RIR to decay to 60 dB below its starting level, which depends on the damping characteristics of the walls [9]. This expression shows that the overlap is strong already at low frequencies especially for large halls and for rooms with highly absorptive surfaces, for which resonances have larger bandwidth. A consequence of the overlap is that in diffuse field conditions, i.e. above the Schroeder frequency $f_{\text{Sch}}$, the number of magnitude peaks in the RTF in a given range is much lower than the theoretical number of modes [9]. The idea of modeling a RTF using OBF models is then to use a finite number of resonant responses, as opposed to the infinite summation in equations (1) and (2), to model accurately low-frequency well-separated dominant room modes and to approximate the spectral envelope of overlapping modes at higher frequencies.

### B. Conventional parametric models for room acoustics

Parametric modeling of room acoustics aims at approximating the GF in (1) by a rational function in the $z$-domain,

$$H(\ddot{\boldsymbol{r}}, z) = \frac{B(\ddot{\boldsymbol{r}}, z)}{A(\ddot{\boldsymbol{r}}, z)} = \frac{\sum_{i=0}^{Q} b_i(\ddot{\boldsymbol{r}}) z^{-i}}{1 + \sum_{i=1}^{P} a_i(\ddot{\boldsymbol{r}}) z^{-i}}, \tag{5}$$

where $\ddot{\boldsymbol{r}} = (\boldsymbol{r}_s, \boldsymbol{r})$ denotes a particular source-receiver position pair. Common assumptions to be made are stability, causality, linearity, and time-invariance of the acoustic system. The expression in (5) can be rewritten in a pole-zero form by factorizing the numerator and denominator polynomials, yielding

$$H(\ddot{\boldsymbol{r}}, z) = \frac{B(\ddot{\boldsymbol{r}}, z)}{A(\ddot{\boldsymbol{r}}, z)} = b_0(\ddot{\boldsymbol{r}}) \frac{\prod_{i=1}^{Q} \{1 - q_i(\ddot{\boldsymbol{r}}) z^{-1}\}}{\prod_{i=1}^{P} \{1 - p_i(\ddot{\boldsymbol{r}}) z^{-1}\}}\,. \tag{6}$$

The zeros $q_i$ represent anti-resonances and time delays in the RIR, while poles $p_i$ are associated with room resonances.

AZ models [1], for which $A(\ddot{\boldsymbol{r}}, z) = 1$, can achieve an arbitrary degree of accuracy by using a high-order FIR filter. The main problem is that the number of parameters of the filter necessary to model the resonant behavior of the system often has to be quite large, depending on the sampling frequency $f_s$ and the reverberation time $T$. Furthermore, the RIR strongly depends on the source and receiver position, so that the parameter values obtained for approximating a RIR at a given source-receiver position $\ddot{\boldsymbol{r}}_1 = (\boldsymbol{r}_{s_1}, \boldsymbol{r}_1)$ are in general significantly different from those for a RIR at another position $\ddot{\boldsymbol{r}}_2 = (\boldsymbol{r}_{s_2}, \boldsymbol{r}_2)$.

Models producing an IIR are used in an attempt to reduce the number of parameters needed to approximate a target RIR [34]. PZ models [2] uses both zeros and poles, so that both room resonances and time delays can be modeled, as well as the non-minimum-phase components of the RTF. However, since both $A(\ddot{\boldsymbol{r}}, z)$ and $B(\ddot{\boldsymbol{r}}, z)$ in (6) are non-constant polynomials in $z^{-1}$, no closed-form solution exists to the model parameter estimation problem and nonlinear optimization methods are required. These methods usually start from the estimation of an all-pole model and then iteratively compute optimal parameter values in the Least Squares (LS) sense. The most popular one is the so-called Steiglitz-McBride (STMCB) method [25], which, however, is not guaranteed to converge and may become unstable, especially for high model orders. Another difficulty lies in determining the optimal values for $Q$ and $P$ in (5) or (6), i.e. the order of the numerator and denominator polynomial, respectively.

PF models, which use the parallel form of fixed-pole IIR filters [3, p.359] consisting of a parallel of second-order all-pole filters, result from a partial fraction expansion (PFE) of the transfer function in (5), which, for $Q < P$, can be written as

$$H(\ddot{\boldsymbol{r}}, z) = \sum_{i=1}^{P} \frac{R_i(\ddot{\boldsymbol{r}})}{(1 - p_i(\ddot{\boldsymbol{r}}) z^{-1})}, \tag{7}$$

where $R_i$ are the residues of the poles $p_i$. If $Q \geq P$, an FIR filter of order $Q - P + 1$ should be added to the right-hand side of the equation [3, pp.112-114], [7], [35]. When the coefficients of $A(z)$ and $B(z)$ in (5) are real, complex poles

will occur in conjugate pairs, so that for each one-pole filter defined by $(R_i, p_i)$ there will be a one-pole filter defined by $(R_i^*, p_i^*)$. These two terms can be added together to form a real second-order section, so that (7) becomes

$$H(\ddot{\boldsymbol{r}}, z) = \sum_{i=1}^{P/2} \left\{ \frac{R_i(\ddot{\boldsymbol{r}})}{1 - p_i(\ddot{\boldsymbol{r}}) z^{-1}} + \frac{R_i^*(\ddot{\boldsymbol{r}})}{1 - p_i^*(\ddot{\boldsymbol{r}}) z^{-1}} \right\}, \tag{8}$$

whose impulse response, with $n = t f_s$ the discrete time variable, is given by

$$h(\ddot{\boldsymbol{r}}, n) = \sum_{i=1}^{P/2} \left\{ R_i(\ddot{\boldsymbol{r}}) \left[p_i(\ddot{\boldsymbol{r}})\right]^n + R_i^*(\ddot{\boldsymbol{r}}) \left[p_i^*(\ddot{\boldsymbol{r}})\right]^n \right\}, \tag{9}$$

which is a finite sum of pairs of geometric series, each for a pair of complex-conjugate poles. After some elaborations, this can be shown to be equivalent to

$$h(\ddot{\boldsymbol{r}}, n) = \sum_{i=1}^{P/2} 2|R_i(\ddot{\boldsymbol{r}})| \rho_i^n \cos(\sigma_i n + \angle R_i(\ddot{\boldsymbol{r}})), \tag{10}$$

with $\rho_i$ and $\sigma_i$ respectively the radius and the angle of the pole $p_i = \rho_i e^{j\sigma_i}$, which is a finite linear combination of exponentially decaying sinusoids sampled in time, with amplitude and phase determined by the residues $R_i$. It is evident by comparing the expressions in (10) and (2) that a RIR can be approximated by a PF model using poles with radius and angle defined by the damping constants $\zeta_i$ and the resonance frequencies $\omega_i$ as

$$\begin{aligned} \rho_i &= e^{-\zeta_i / f_\mathrm{s}}, \\ \sigma_i &= \omega_i / f_\mathrm{s}\,. \end{aligned} \tag{11}$$

Notice that, when $\rho_i$ is small and $\sigma_i$ is close to either 0 or $\pi$, the resonance generated by $p_i$ is influenced by the resonance generated by $p_i^*$, so that their magnitude peaks have frequency slightly different from $\pm\omega_i$ [35], [36]. The PF model as an approximation of the GF was first discussed in [11] in relation to the modeling of a RTF by using common acoustical poles and their residues (CAPR). It has been shown that the GF in (1) is a PFE for the resonance frequencies, which can be approximated by a PF model, assuming $\zeta_i \ll \omega_i$. It is also shown that the residues $R_i(\ddot{\boldsymbol{r}})$ are related to the eigenfunctions $\psi_i$ of the GF, thus expressing the variation of the RTF at different source and receiver positions.

The transfer function of the PF in (8) can be rearranged as

$$H(\ddot{\boldsymbol{r}}, z) = \sum_{i=1}^{P/2} \left[ \frac{d_{i,0}(\ddot{\boldsymbol{r}}) + d_{i,1}(\ddot{\boldsymbol{r}}) z^{-1}}{(1 - p_i(\ddot{\boldsymbol{r}}) z^{-1})(1 - p_i^*(\ddot{\boldsymbol{r}}) z^{-1})} \right],$$

$$d_{i,0}(\ddot{\boldsymbol{r}}) = \mathrm{Re}\{R_i(\ddot{\boldsymbol{r}})\} = |R_i(\ddot{\boldsymbol{r}})| \cos(\angle R_i(\ddot{\boldsymbol{r}})), \tag{12}$$

$$d_{i,1}(\ddot{\boldsymbol{r}}) = \mathrm{Re}\{R_i(\ddot{\boldsymbol{r}}) p_i^*(\ddot{\boldsymbol{r}})\} = |R_i(\ddot{\boldsymbol{r}})\, p_i| \cos(\sigma_i - \angle R_i(\ddot{\boldsymbol{r}})),$$

and implemented as a parallel of second-order filters, shown in Fig. 2, which is linear in the parameters $\{d_{i,0}, d_{i,1}\}$, but nonlinear in the poles $\{p_i, p_i^*\}$. Each second-order section models a room resonance, with resonance frequency and bandwidth determined by the position of $\{p_i, p_i^*\}$, within the unit circle in order to ensure stability. Particular attention should be given to repeated poles, which produce polynomial amplitude envelopes on the decaying exponentials [35], the order of which is determined by the multiplicity of the repeated pole. It should be noticed that, in the presence of repeated poles, the model structure in Fig. 2 has to be modified accordingly.

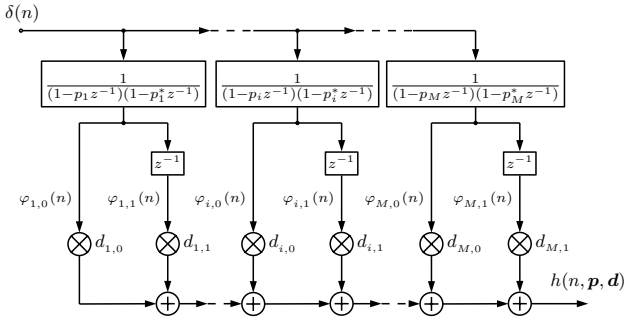


Fig. 2. The PF model structure (with $M = P/2$). The impulse responses to the second-order IIR filters, denoted by $\varphi_{i,0}$ and $\varphi_{i,1}$ for $i = 1, \ldots, M$, are used as basis functions in a linear-in-the-parameters model structure.

## III. Orthonormal Basis Function models

Parametric models based on OBFs can be derived from an orthogonalization of PF models. The orthogonality of the basis functions, together with the linearity in the parameters, introduces some desirable properties which bring a number of advantages in terms of efficiency and numerical stability in the modeling of RIRs. In this section, OBF models are also described as a generalization of other parametric models. Furthermore, their properties are described along with their application in the approximation of a target RIR.

### A. Construction of OBF models

OBF models are derived with a Gram-Schmidt orthonormalization procedure applied to one- and two-pole filters [13]–[15]. Starting from a normalized first-order IIR filter with pole $p_1$ and transfer function

$$\Psi_1(z, p_1) = \frac{A_1}{1 - p_1 z^{-1}}, \tag{13}$$

where $A_1 = \sqrt{1 - |p_1|^2}$ is a normalization factor, a second-order filter with poles $[p_1, p_2]$ and transfer function orthogonal to (13) can be obtained as

$$\Psi_2(z, [p_1, p_2]^T) = \frac{A_2(z^{-1} - p_1^*)}{(1 - p_1 z^{-1})(1 - p_2 z^{-1})}, \tag{14}$$

with $A_2 = \sqrt{1 - |p_2|^2}$ and with * indicating complex conjugation. The orthogonality of $\Psi_1$ and $\Psi_2$ is provided by the zero in $z = 1/p_1^*$ and can be investigated via the inner product on the Hardy space on the unit circle $\mathcal{H}_2(\mathbb{T})$ (with $\mathbb{T} \triangleq \{z : |z| = 1\}$) as (see [15])

$$\langle \Psi_1, \Psi_2 \rangle = \frac{1}{2\pi j} \oint_{\mathbb{T}} \Psi_1(z) \Psi_2^*(z) \frac{dz}{z} = 0. \tag{15}$$

The transfer function in (14) can be seen as the product of a normalized first-order IIR filter defined by $p_2$ and a first-order all-pass filter defined by $p_1$. By repeating the procedure for a set of poles $\boldsymbol{p}_i = \{p_1, \ldots, p_i\}$, the $i^{\text{th}}$ transfer function will

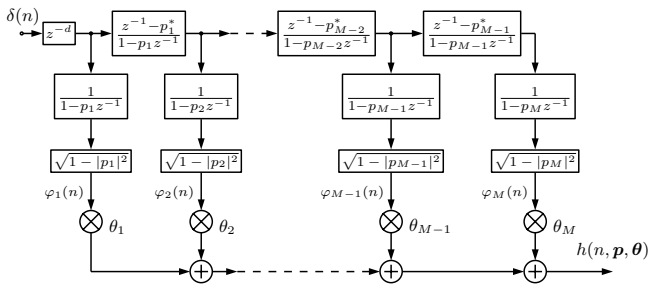


Fig. 3. The Takenaka-Malmquist OBF model structure for $M$ real poles.

consist of a normalized first-order IIR filter defined by $p_i$ and a sequence of first-order all-pass filters defined by the pole set $\boldsymbol{p}_{i-1} = \{p_1, \ldots, p_{i-1}\}$,

$$\Psi_i(z, \boldsymbol{p}_i) = \left( \frac{\sqrt{1-|p_i|^2}}{1-p_i z^{-1}} \right) \prod_{l=1}^{i-1} \left( \frac{z^{-1}-p_l^*}{1-p_l z^{-1}} \right), \qquad (16)$$

which is also known as the Takenaka-Malmquist function [13]. The corresponding model structure is shown in Fig. 3, where the model output $h(n, \boldsymbol{p}, \boldsymbol{\theta})$ is a linear combination of the responses of the basis functions, weighted by the linear parameters $\theta_i$.

An OBF model based on the functions in (16) can be seen as a generalization of other well-known models. If all the poles are identical and real, the Laguerre model [37] is obtained, which is in turn a normalized version of a so-called warped FIR filter model [26], with the value of the warping parameter the repeated real pole. If the pole is placed in the origin, then the Laguerre filter simplifies to an AZ model.

When the pole set $\boldsymbol{p}_i$ contains complex poles, the basis functions in (16) are generally complex-valued and are thus not useful for the identification of real systems. As for PF models, two real-valued basis functions can be obtained by combining pairs of complex-conjugate poles, and by orthogonalizing each pair of basis functions with respect to each other (plus a normalization factor). Different realizations of an OBF model can be obtained for particular choices of these normalization factors, as explained in [15]. A combination of a Takanaka-Malmquist model and the so-called Kautz model can be used, as suggested in [20], modeling real and complex poles, respectively. This model structure, henceforth called mixed-Kautz model, is shown in Fig. 4 for $\dot{m}$ real poles and $\ddot{m}$ pairs of complex-conjugate poles. The basis functions of a mixed-Kautz model are defined for a real pole $p_i$ as

$$\Psi_i(z, \boldsymbol{p}_i) = \left( \frac{A_i}{1-p_i z^{-1}} \right) \prod_{l=1}^{i-1} \left( \frac{z^{-1}-p_l^*}{1-p_l z^{-1}} \right), \qquad (17)$$

or for a complex-conjugate pole pair $\{p_{i-1}, p_i\} = \{p_i, p_i^*\}$ as

$$\begin{aligned} \Psi_i'(z, \boldsymbol{p}_i) &= \frac{c_i'(z^{-1}+1)}{(1-p_i z^{-1})(1-p_i^* z^{-1})} \prod_{l=1}^{i-2} \frac{(z^{-1}-p_l^*)}{(1-p_l z^{-1})}, \\ \Psi_i''(z, \boldsymbol{p}_i) &= \frac{c_i''(z^{-1}-1)}{(1-p_i z^{-1})(1-p_i^* z^{-1})} \prod_{l=1}^{i-2} \frac{(z^{-1}-p_l^*)}{(1-p_l z^{-1})}. \end{aligned} \qquad (18)$$

with $A_i = \sqrt{1-|p_i|^2}$, and normalization factors $c_i' = |1-p_i|A_i/\sqrt{2}$ and $c_i'' = |1+p_i|A_i/\sqrt{2}$. Notice that the pair of basis functions in (18) are built as a product of a sequence of $i-2$ first-order all-pass filters given by the poles in $\boldsymbol{p}_{i-2}$, a second-order all-pole filter defined by $\{p_i, p_i^*\}$ and a normalization term, so that the model structure for complex-conjugate poles is given by a parallel of orthonormalized second-order IIR filters. However, real poles may not be of much interest in the approximation of measured RTFs; even though positive real poles would be useful for modeling the cavity mode of a room response, a measured RTF has a band-pass characteristic, with a cut-off at low frequencies determined by the response of the high-pass filter of the loudspeaker, and a cut-off at high frequencies given by the low-pass behavior of the loudspeaker or the anti-aliasing filter. For this reason, only complex-conjugate poles can be considered, thus resulting in the use of a Kautz model.

### *B. Properties of OBF models*

The orthogonality of the basis functions provides some desirable properties. First, the OBFs form a complete set in $\mathcal{H}_2(\mathbb{T})$, under the assumption that $\sum_{i=0}^{\infty}(1-|p_i|) = \infty$ [15]. Thus, by decomposing a target RIR in terms of an orthogonal expansion, the approximation error can be made arbitrarily small by choosing a large enough number of poles.

Second, orthogonality provides flexibility, which results from the fact that poles can be arbitrarily positioned inside the unit circle (for the sake of stability), and that frequency resolution can be allocated unevenly in different regions of the spectrum without numerical conditioning problems, regardless of the model order. This is not the case, for example, for PZ models, where problems of ill-conditioning and instability can arise for high model orders.

Third, OBF models are linear in the parameters $\theta_i$, which means that linear regression can be applied in order to estimate their optimal values. Moreover, due to the orthogonality of the basis functions, it is not necessary to carry out a matrix inversion, which is often a source of numerical problems. Another consequence of orthogonality is the fact that the parameters $\theta_i$ for each IIR filter are independent from the ones for others filters in the structure, so that a model of lower order can be obtained from a model of higher order only by truncation, and similarly additional poles can be included without recomputing the values of the $\theta_i$'s corresponding to the poles already used. An additional advantage of OBF models over PF models is that the same pole can be included more than once (e.g. to model modes with a double decay) without the need to modify the structure. These properties are exploited in the scalable algorithm described in Section IV.

### *C. Approximation of a RIR with an OBF model*

The approximation of a target RIR $h(n)$ using an OBF model consists in estimating the parameters in the pole set $\boldsymbol{p} = \{p_i\}$ and in the set of parameters $\boldsymbol{\theta} = \{\theta_i\}$, with $i = 1, \ldots, M$ (cfr. Fig. 4 where $M = \dot{m} + 2\ddot{m}$), that minimize the distance between a target RIR $h(n)$ and the model response $h(n, \boldsymbol{p}, \boldsymbol{\theta})$ for $n = 1, \ldots, N$. For a fixed set of poles $\boldsymbol{p}$, the

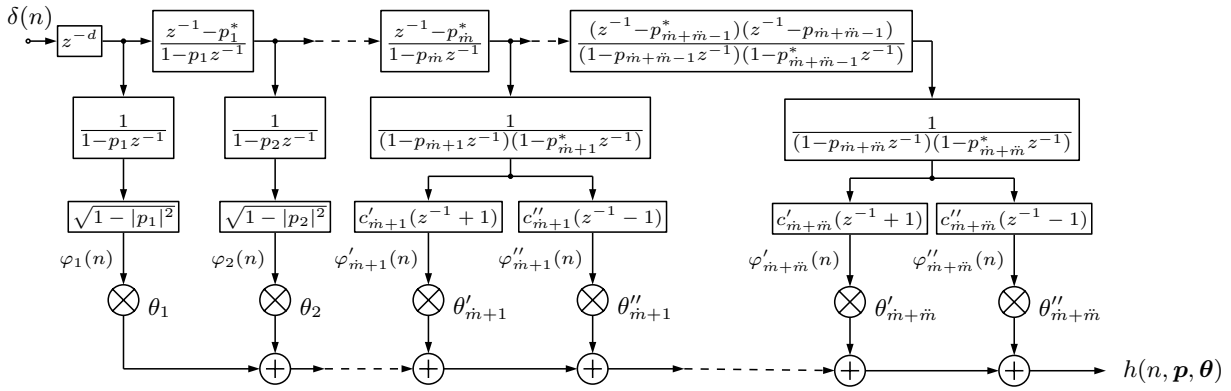


Fig. 4. The mixed-Kautz model structure for $\dot{m}$ real poles and $\ddot{m}$ pairs of complex-conjugate poles. For convenience, the basis functions corresponding to real poles defined in (17) are followed by the basis functions corresponding to complex-conjugate pole pairs defined in (18).

problem of estimating $\boldsymbol{\theta}$ is linear and can be solved in closed form. The response $h(n, \boldsymbol{p}, \boldsymbol{\theta})$ of an OBF model for an impulse input signal $\delta(n)$ is the linear combination of the responses $\varphi_i(n, \boldsymbol{p}_i)$ of the $M$ basis functions $\Psi_i(z, \boldsymbol{p}_i)$ (see e.g. Fig. 3 or Fig. 4),

$$\begin{aligned} h(n, \boldsymbol{p}, \boldsymbol{\theta}) &= \sum_{i=1}^{M} \theta_i \Psi_i(z, \boldsymbol{p}_i) \delta(n) \\ &= \sum_{i=1}^{M} \theta_i \varphi_i(n, \boldsymbol{p}_i) = \boldsymbol{\varphi}(n, \boldsymbol{p})^T \boldsymbol{\theta}, \end{aligned} \tag{19}$$

where $\boldsymbol{\varphi}(n, \boldsymbol{p})$ is a vector containing the responses $\varphi_i(n, \boldsymbol{p}_i)$ at time $n$. By stacking all the vectors $\boldsymbol{\varphi}(n, \boldsymbol{p})$ for $n = 1, \ldots, N$ in a matrix $\boldsymbol{\Phi}(\boldsymbol{p})$ of size $N \times M$, the optimal values for $\boldsymbol{\theta}$ for a given input-output set $\{\boldsymbol{\delta}, \boldsymbol{h}\} = \{\delta(n), h(n)\}_{n=1}^{N}$ can be estimated in LS sense as

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\Phi}(\boldsymbol{p})^T \boldsymbol{h}. \tag{20}$$

Note that the LS estimation does not require any matrix inversion, given that the orthonormality of the basis functions implies $\boldsymbol{\Phi}(\boldsymbol{p})^T \boldsymbol{\Phi}(\boldsymbol{p}) = \boldsymbol{I}_M$. It can be seen from (20) that the optimal estimate for $\boldsymbol{\theta}$ corresponds to the correlation of the basis functions in $\boldsymbol{\Phi}(\boldsymbol{p})$ with the target RIR vector $\boldsymbol{h}$, so that $\hat{\boldsymbol{\theta}}$ gives the degree of similarity between each basis function and the target RIR.

The problem of estimating the optimal pole set $\hat{\boldsymbol{p}}$ can be then regarded as finding the poles that generate basis functions that are maximally correlated with the target RIR, so that the approximation error between the model response and the target RIR is minimized. However, no closed-form solution to the pole estimation problem is available. The state-of-the-art approach for the multiple-poles case is the BU method [19]–[23], an iterative nonlinear method based on FIR-to-IIR filter conversion [24]. Frequency prewarping of the target response has been proposed for audio applications in order to match a particular frequency-scale mapping, such as the Bark scale [38]. The BU method exploits the orthogonality of OBF models and provides accurate estimates for the pole parameters. However, the model order has to be predetermined, and stability problems can arise from numerical issues at high model orders.

## IV. The OBF-MP algorithm

The problem of sparse linear approximation of a signal consists in finding a compact representation by a combination of functions taken from an overcomplete basis. These functions are usually called predictors or atoms, which altogether form a basis, sometimes called dictionary. The most popular methods for sparse approximation can be divided in two main categories [39]. In the first one, convex optimization techniques are used to minimize a functional, such as the $\ell_1$ norm in the Least Absolute Shrinkage and Selection Operator (LASSO) [40]. The second category includes iterative greedy algorithms, such as Orthogonal Matching Pursuit (OMP) [41]–[43].

Our approach aims to find a sparse approximation of a target RIR as a linear combination of a finite number of OBFs. A RIR cannot be considered a sparse time-frequency signal itself, with a certain degree of sparsity only in the modal region. By first modeling dominant low-frequency modes and the spectral envelope at higher frequencies, the proposed algorithm is able to provide a sparse approximation of a RIR using a finite-order OBF model. In this section, an OMP-based greedy algorithm, which is termed here OBF-MP [29], is used to iteratively select poles from a large set of candidate poles distributed over the unit disc, thus bypassing the inherent nonlinear problem. At each iteration of the OMP algorithm, the predictor that has the highest correlation with the current residual response is selected. The problem in the OBF-MP algorithm, given that OBFs are defined by previous poles in the structure, consists in defining a dictionary of candidate predictors, where the dictionary has to be updated at each iteration using the predictor selected at the previous iteration. The advantage of OBF-MP over the conventional OMP algorithm is that the orthogonal projection of the current residual response onto the set of predictors selected at the previous iterations is not necessary. The predictors, in fact, are already orthogonal to each other by construction. This ensures that the algorithm does not contain any matrix inversion, thus avoiding ill-conditioning problems. Moreover, since the candidate predictors are orthogonal to the predictors selected at previous iterations, computing the correlation with the current residual response is equivalent to computing the correlation with the target RIR.

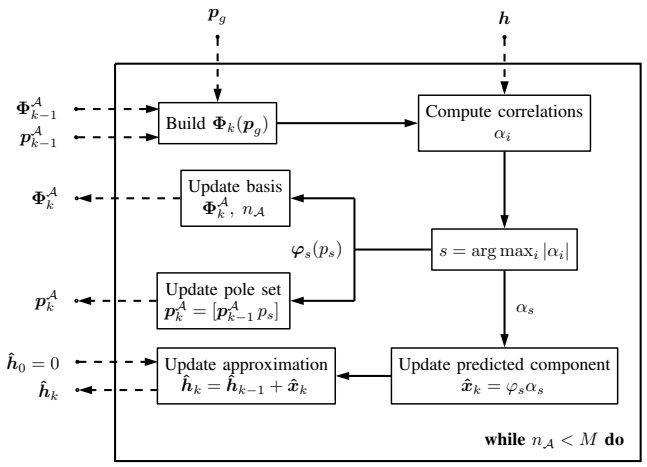


Fig. 5. The OBF-MP algorithm block diagram. Inbound dashed lines represent initial conditions and inputs, while outbound dashed lines represent outputs.

**Algorithm 1** OBF-MP algorithm

$\boldsymbol{p}_g = \{p_1, \ldots, p_G\}$ ▷ Define poles in the pole grid
$\boldsymbol{\Phi}^{\mathcal{A}}_0 = \emptyset, n_{\mathcal{A}} = 0$ ▷ Initialize set of active predictors (basis)
$\boldsymbol{p}^{\mathcal{A}}_0 = \emptyset$ ▷ Initialize the set of active poles
$\hat{\boldsymbol{h}}_0 = \mathbf{0}$ ▷ Initialize target approximation vector
$k = 1$ ▷ $k$: iteration counter
**while** $n_{\mathcal{A}} < M$ **do** ▷ $M$: desired number of OBFs
  Build $\boldsymbol{\Phi}_k(\boldsymbol{p}_g)$ ▷ Build matrix of candidate predictors $\boldsymbol{\varphi}_i$
  $s = \arg\max_i |\alpha_i|$ ▷ Find $\boldsymbol{\varphi}_i$ max. correlated with $\boldsymbol{h}$
  $\boldsymbol{p}^{\mathcal{A}}_k = [\boldsymbol{p}^{\mathcal{A}}_{k-1}\ p_s]$ ▷ Add selected pole to active pole set
  **if** $p_s$ real **then** ▷ Update basis and predicted component
    $\boldsymbol{\Phi}^{\mathcal{A}}_k = [\boldsymbol{\Phi}^{\mathcal{A}}_{k-1}\ \ \boldsymbol{\varphi}_s],\ n_{\mathcal{A}} = n_{\mathcal{A}} + 1$
    $\hat{\boldsymbol{x}}_k = \boldsymbol{\varphi}_s \alpha_s$
  **else if** $p_s$ complex **then**
    $\boldsymbol{\Phi}^{\mathcal{A}}_k = [\boldsymbol{\Phi}^{\mathcal{A}}_{k-1}\ \ \boldsymbol{\varphi}'_s\ \ \boldsymbol{\varphi}''_s],\ n_{\mathcal{A}} = n_{\mathcal{A}} + 2$
    $\hat{\boldsymbol{x}}_k = [\boldsymbol{\varphi}'_s\, \boldsymbol{\varphi}''_s][\alpha'_s\, \alpha''_s]^T$
  **end if**
  $\hat{\boldsymbol{h}}_k = \hat{\boldsymbol{h}}_{k-1} + \hat{\boldsymbol{x}}_k$ ▷ Update target approximation vector
  $k = k + 1$
**end while**

Another consequence of orthogonality is the scalability of the algorithm, from which it follows that the number of parameters of the final model structure does not have to be defined in advance. A pole and the related linear coefficient are estimated at each iteration, independently of poles selected at previous iterations. It follows that additional poles can be estimated just by running extra iterations of the algorithm, without any problem of instability or numerical ill-conditioning. This scalability property of the algorithm is also a consequence of the fact that, similarly to what was discussed at the end of Section II for PF models, also OBF models can be regarded as a way of approximating a RTF. It has been already mentioned that, for the same set of non-repeated poles, the basis functions of a PF model and the ones of an OBF model span the same approximation space, so that it is possible to convert the values of the linear parameters from one model form to the other by simply a linear transformation [5].

Following the above interpretation, the idea of the OBF-MP algorithm is to iteratively compute a sparse approximation $\hat{\boldsymbol{h}}$ of a target RIR $\boldsymbol{h}$ of length $N$ samples as a linear combination of length-$N$ OBFs, analogously to the definition of a RIR as a summation of exponentially decaying functions, independent one from each other. The OBFs are selected from a dictionary $\boldsymbol{\Phi}_k$ of candidate predictors $\boldsymbol{\varphi}_i$ $(i = 1, \ldots, D)$ and included in the basis $\boldsymbol{\Phi}^{\mathcal{A}}_k$. At each iteration $k$, $D$ candidate predictors $\boldsymbol{\varphi}_i$, orthogonal to the predictors in the current basis $\boldsymbol{\Phi}^{\mathcal{A}}_{k-1}$ constructed with the current set of active poles $\boldsymbol{p}^{\mathcal{A}}_k$, are built from $G$ poles placed arbitrarily in a grid $\boldsymbol{p}_g$ inside the upper half of the unit disc. The matrix $\boldsymbol{\Phi}_k$ has dimensions $N \times D$, with $D = \dot{m} + 2\ddot{m}$ where $\dot{m}$ and $\ddot{m}$ denote respectively the number of real poles and complex poles in the grid $\boldsymbol{p}_g$, so that $G = \dot{m} + \ddot{m}$.

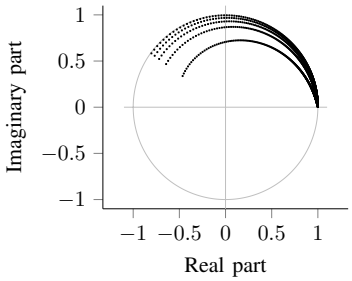


Fig. 6. The Bark-exp pole grid for the OBF-MP algorithm (here with 5 radii and 400 angles and upper angle limited to $0.8\pi$).

The OBF-MP algorithm is described in detail below, and a graphical representation is depicted in Fig. 5. First, a grid of $G$ candidate poles $\boldsymbol{p}_g$ is defined, similarly to [28]–[30], with poles distributed according to a desired frequency resolution or prior knowledge about the system. In [28]–[30], the angle and the radius of the poles were distributed either uniformly or logarithmically on the unit disc, with the latter option intended to increase the resolution at low frequencies. Here, a different pole grid is used, depicted in Fig. 6, henceforth referred to as Bark-exp grid; the radius $\rho_i$ of the poles decreases exponentially at the increase of the angle $\sigma_i$, as suggested in [23], according to $\rho_i = \varrho^{\frac{\sigma_i}{\pi}}$, with $\varrho$ the value of the radius defined at the Nyquist frequency. Regarding the values for $\varrho$, it is suggested here to set the number of radii for each angle and distribute them logarithmically in order to increase density toward the unit circle. Differently from [23], in which the angles follow a logarithmic scale, the Bark frequency scale [38] was chosen. The Bark scale further increases the resolution at low frequency, providing an effect similar to the prewarping of the RIR used in the wBU method. In this way, a higher density of poles close to the unit circle is achieved at low frequencies, allowing a more accurate approximation of energetic and narrow-bandwidth resonances, while at higher frequencies poles sparser in frequency and more distant from the unit circle provide a coarser approximation.

At the first iteration, the current basis $\boldsymbol{\Phi}^{\mathcal{A}}_0$ and the set of active poles $\boldsymbol{p}^{\mathcal{A}}_0$ are empty (with the number of predictors in the basis $n_{\mathcal{A}} = 0$). Also the target approximation vector $\hat{\boldsymbol{h}}_0$, is initially set to zero. At each iteration $k$, the matrix of candidate predictors $\boldsymbol{\Phi}_k(\boldsymbol{p}_g)$ is updated according to the mixed-Kautz

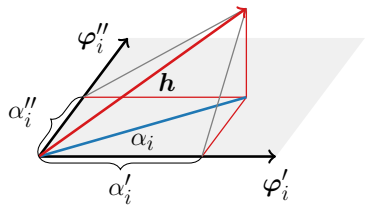


Fig. 7. Graphical interpretation of the correlation between the target RIR vector $\boldsymbol{h}$ and the predictors of a pair of complex-conjugate poles $\{p_i, p_i^*\}$.

structure in Fig. 4. The matrix $\boldsymbol{\Phi}_k$ has always dimension $N \times D$ (since an OBF model admits repeated poles, a pole that is selected by the algorithm is not removed from the pole grid $\boldsymbol{p}_g$) and its columns are the OBFs built from the poles in $\boldsymbol{p}_g$ with transfer functions as in (17) and in (18), thus orthonormal to the predictors in the current basis $\boldsymbol{\Phi}_{k-1}^{\mathcal{A}}$ built from the poles in the current active pole set $\boldsymbol{p}_{k-1}^{\mathcal{A}}$. The predictor(s) in $\boldsymbol{\Phi}_k$ that has the largest absolute correlation $\alpha_i$ with the target RIR vector $\boldsymbol{h}$ is selected and added to the basis $\boldsymbol{\Phi}_k^{\mathcal{A}}$, while the corresponding pole is included in the set of active poles $\boldsymbol{p}_k^{\mathcal{A}}$. The correlation for real and complex poles is computed in two different ways. For real poles, the correlation is the projection of $\boldsymbol{h}$ onto the predictor $\boldsymbol{\varphi}_i$ ($\alpha_i = \boldsymbol{\varphi}_i^T \boldsymbol{h}$). For a pair of complex-conjugate poles $\{p_i, p_i^*\}$ the correlation is the projection of $\boldsymbol{h}$ on the plane defined by predictors $\boldsymbol{\varphi}_i'$ and $\boldsymbol{\varphi}_i''$(see Fig. 7), which are mutually orthogonal, and is given by

$$\alpha_i = \sqrt{{\alpha_i'}^2 + {\alpha_i''}^2} = \sqrt{({\boldsymbol{\varphi}_i'}^T \boldsymbol{h})^2 + ({\boldsymbol{\varphi}_i''}^T \boldsymbol{h})^2}. \qquad (21)$$

The $k^{th}$ predicted component $\hat{\boldsymbol{x}}_k$ is obtained from the last added predictor(s) using the maximum correlation $\alpha_s$ as regression coefficient (with $s = \arg\max_i |\alpha_i|$), by $\hat{\boldsymbol{x}}_k = \boldsymbol{\varphi}_s \alpha_s$, if the selected pole is real, or by $\hat{\boldsymbol{x}}_k = [\boldsymbol{\varphi}_s' \, \boldsymbol{\varphi}_s''][\alpha_s' \, \alpha_s'']^T$, otherwise. The current target RIR approximation vector $\hat{\boldsymbol{h}}_k$ is obtained by adding the predicted component $\hat{\boldsymbol{x}}_k$ to the previous target RIR approximation vector $\hat{\boldsymbol{h}}_{k-1}$. As a consequence of its scalability property, the algorithm can terminate when the desired number $M$ of predictors in the basis is reached, or alternatively when the approximation error falls below a desired value.

### A. Algorithmic complexity analysis

Here the asymptotic computational complexity of the OBF-MP algorithm is analyzed, assuming for simplicity that only complex poles are included in the pole grid. With reference to Algorithm 1, there are two operations that determine the asymptotic behavior of the algorithm. Building the matrix $\boldsymbol{\Phi}_k$ of candidate predictors (step 7) at each iteration is the most demanding operation, which involves the generation of $D$ predictors of length $N$, which sums up to a complexity of $\mathcal{O}(3ND)$ multiplications (cfr. the expressions in (18) and Figure 4). The second operation to consider is the computation of the correlation coefficients (step 8), which is a multiplication of the matrix $\boldsymbol{\Phi}_k$ with the vector $\boldsymbol{h}$ of length $N$, which results in $\mathcal{O}(ND)$ multiplications. The computational complexity associated to vector updates and other operations is negligible. The overall complexity of the OBF-MP algorithm after $k = M/2$ iterations, is $\mathcal{O}(2MND)$ multiplications, i.e. linearly proportional to the three variables considered. In other words, the computational complexity increases linearly with the length of the impulse response, the number of candidate poles, and the number of iterations. This is comparable with the complexity of the BU method, whose most demanding operation is represented by the solution to a set of overdetermined linear equations, which implies a QR factorization of a large $N \times M$ rectangular matrix (complexity $\mathcal{O}(NM^2)$), followed by a back-substitution of a $M \times M$ triangular matrix (complexity $\mathcal{O}(M^2)$) [44]. This is performed for $I$ iterations, with the overall computational complexity of the BU method summing up to $\mathcal{O}(INM^2)$, which is quadratic with respect to the number of estimated poles $M$.

## V. Model and Filter Complexity

In this section, the complexity of the parametric models presented in the previous sections will be analyzed from two different perspectives. First, the *model complexity* (or *representation complexity*) $C_\mathrm{m}$ is considered, which is the number of parameters that is necessary to represent the system under study. Second, the *filter complexity* (or *simulation complexity*) $C_\mathrm{f}$ is considered as the number of operations that are required to obtain the filter output signal for a given input signal when the parameter values are available. While a measure often used in the literature is the *model order*, it is believed that the two concepts just proposed are less prone to misinterpretation and thus preferable for the comparison of different parametric models in terms of complexity. For simplicity, OBF models and PF models having complex-conjugate pole pairs only are considered.

*1) Model complexity:* The calculation of the model complexity $C_\mathrm{m}$ is straightforward for AZ and PZ models. By referring to (5) and (6), the number of parameters for AZ models corresponds to the number of numerator coefficients ($C_\mathrm{m} = Q + 1$), while for PZ models it is the sum of denominator and numerator coefficients ($C_\mathrm{m} = P+Q+1$). For PF models, if $P/2$ is the number of complex-conjugate poles pairs, the number of parameters required is $C_\mathrm{m} = 2P$, since each second-order section can be represented with one pole $p_i$ (which is a complex number defined by two parameters, while $p_i^*$ is given by complex conjugation) and two linear parameters (denoted by $d_{i,0}$ and $d_{i,1}$ in (12) and in Fig. 2). The same is obtained for OBF models, in which the all-pass filters and the normalization factors can be computed from the knowledge of the poles (see e.g. Fig. 4). The model complexity $C_\mathrm{m}$ is summarized in the left column of Table I.

*2) Filter complexity:* The filter complexity $C_\mathrm{f}$ is calculated here as the number of multiplications required to compute the filter output for a given input signal. For AZ and PZ models, one multiplication is required for each coefficient, so that $C_\mathrm{f} = C_\mathrm{m}$. This is true also for PF models, in which four multiplications are required for each second-order section, two for the second-order IIR filter and two for the linear parameters (see Fig. 2). In case of repeated poles, the structure has to be modified, but the number of multiplications remains the

TABLE I
MODEL AND FILTER COMPLEXITY

| model | $C_\mathrm{m}$ | $C_\mathrm{f}$ |
|---|---|---|
| AZ | $Q+1$ | $Q+1$ |
| PZ | $P+Q+1$ | $P+Q+1$ |
| PF | $2P$ | $2P$ |
| OBF | $2P$ | $3P$ |

TABLE II
DATABASE SPECIFICATIONS

| database | $V$ ($\mathrm{m}^3$) | $S$ ($\mathrm{m}^2$) | $t_\mathrm{m}$ (ms) | $T$ (s) | $f_\mathrm{Sch}$ (Hz) | RIRs |
|---|---|---|---|---|---|---|
| SMARD | 170.4 | 207.3 | 28.4 | 0.15 | 59 | 8 |
| MARDY | 208.8 | 255.6 | 28.0 | 0.45 | 93 | 9 |
| MIRD | 86.4 | 129.6 | 25.3 | 0.16 | 86 | 8 |
| | | | | 0.36 | 129 | 8 |
| | | | | 0.61 | 168 | 8 |
| SUBRIR | 62.3 | 102.1 | 24.2 | 0.5-1.5 | >180 | 24 |

same [35]. For OBF models, the normalization coefficients $c_i'$ and $c_i''$ can be combined together with the related linear parameters $\theta_i'$ and $\theta_i''$, so that the only difference between OBF models and PF models in terms of filter complexity is determined by the orthogonalization. By including the second-order all-pass filters in the structure, two more multiplications per section have to be included (assuming that the input of the all-pass filter is the output of the previous second-order IIR filter), summing up to six per section, so that the filter complexity for $P/2$ pairs $\{p_i, p_i^*\}$ is $C_\mathrm{f} = 3P$. The filter complexity $C_\mathrm{f}$ is summarized in the right column of Table I. Notice that an OBF model is more complex than a PF model. However, these two models span the same approximation space for the same set of poles, thus leading exactly to the same filter response when the optimal linear coefficients are computed using the $\ell_2$ norm in LS design. It would be then possible to convert an OBF model into a PF model with lower filter complexity, as was also suggested in [4].

## VI. SIMULATION RESULTS

The modeling performance of the OBF-MP algorithm described in Section IV was tested on $R = 41$ RIRs measured for several source-receiver positions in three different rooms with different reverberation times. The RIRs were taken from three publicly available databases, namely MARDY [45], SMARD [46], and MIRD [47]. A fourth database of 24 low-frequency RIRs, called SUBRIR [48], was used separately to evaluate the modeling performance of the algorithm in the modal frequency region, as will be discussed later in this section. Their specifications, such as the room volume $V$, the surface area $S$, the reverberation time $T$, the Schroeder frequency $f_\mathrm{Sch}$ computed as in (4), and the mixing time $t_\mathrm{m}$, are listed in Table II. According to [49], the most accurate estimate of the mixing time $t_\mathrm{m}$, i.e. the time instant at which the diffuse reverberation tail begins, is given by a formula related to the concept of mean free path length, given by $t_\mathrm{m} = 20V/S + 12$ (in ms). Notice that the MIRD database includes RIRs measured in a room where the reverberation time is controlled by means of movable acoustic panels, resulting in 3 different values of $T$ in Table II. All target RIRs are sampled at $f_\mathrm{s} = 48\,\mathrm{kHz}$ and truncated to $N = 6000$ samples. This corresponds to the shortest 'useful duration', defined as the time instant where the SNR of the measured RIR is $10\,\mathrm{dB}$ [50]. In order to compute the SNR value, the decay curve and the noise floor level were estimated with the method by Lundeby et al. [51]. Since the modeling of the delay of the RIR is not part of the scope of this paper, the direct path component was considered as the starting point of the RIR. However, a simple delay could be easily included in the model structure of the OBF model by setting the parameter $d$ in Fig. 4.

In the simulations presented in this section, an approximated response $\hat{\boldsymbol{h}}^{(r)}$, with $r = 1, \ldots, R$, was computed for each target RIR $\boldsymbol{h}^{(r)}$ using the OBF-MP algorithm. OBF models obtained with OBF-MP were compared to AZ and PZ models and to OBF models obtained with the warped BU (wBU) method suggested in [22], henceforth called OBF-wBU models. The measure used to compare the performance of different models with the same model complexity $C_\mathrm{m}$ is the normalized mean-square error (NMSE), averaged over all $R$ RIRs, which in the time domain is given by

$$h_\mathrm{NMSE}(\mathrm{dB}) = 10\log_{10}\frac{1}{R}\sum_{r=1}^{R}\frac{\|\boldsymbol{h_r} - \hat{\boldsymbol{h}}_{\boldsymbol{r}}\|_2^2}{\|\boldsymbol{h_r}\|_2^2}, \qquad (22)$$

while the average frequency response NMSE is defined as

$$H_\mathrm{NMSE}(\mathrm{dB}) = 10\log_{10}\frac{1}{R}\sum_{r=1}^{R}\frac{\|\boldsymbol{H_r} - \hat{\boldsymbol{H}}_{\boldsymbol{r}}\|_2^2}{\|\boldsymbol{H_r}\|_2^2}, \qquad (23)$$

with $\boldsymbol{H}_r$ and $\hat{\boldsymbol{H}}_r$ the Discrete Fourier Transform (DFT) of $\boldsymbol{h}_r$ and $\hat{\boldsymbol{h}}_r$, respectively. The NMSE was computed on the complete time response ($h_\mathrm{NMSE}^\mathrm{full}$), and on the early ($h_\mathrm{NMSE}^\mathrm{early}$) and late ($h_\mathrm{NMSE}^\mathrm{late}$) responses separately. The time instant separating the two parts was set to 25 ms, corresponding to the shortest mixing time $t_m$ for the three rooms considered (see Table II). Also the NMSE in the frequency response was analyzed for the frequency range between 0 Hz and 20 kHz ($H_\mathrm{NMSE}^\mathrm{full}$), as well as at low/mid frequencies between 0 and 4 kHz ($H_\mathrm{NMSE}^\mathrm{low/mid}$), and at high frequencies between 4 kHz and 20 kHz ($H_\mathrm{NMSE}^\mathrm{high}$), in order to show the differences in performance of the models in different frequency ranges. Although the Schroeder frequency in (4) would have been a more natural choice for separating the frequency range, its value in the databases considered was found to be below or just above the lower cut-off frequency of the loudspeaker used for the measurements. The upper limit of 20 kHz was chosen to avoid considering the frequency range dominated by the influence of the anti-aliasing filter.

The Bark-exp grid used in these simulations counts $G = 6000$ poles with 5 different radii distributed logarithmically with values $\varrho$ at Nyquist from 0.5 to 0.99, and 1200 different angles placed from 48 Hz to 19.2 kHz according to the Bark frequency scale [38] with Bark-warping factor $w = 0.766$. The limits on the angle were chosen to avoid approximating the response below the cut-off frequency of the loudspeakers and above the cut-off frequency of the anti-aliasing filter. As a result, the grid contains only complex poles. The reason

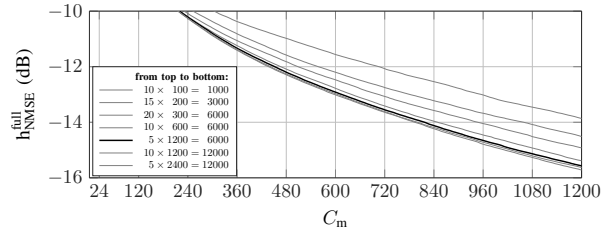


Fig. 8. The average time-domain NMSE in (22) for different pole allocations and densities of a Bark-exp grid. The darker line indicates the grid chosen for the simulations.

of such an uneven allocation of the number of radii and angles is due to the frequency resolution that is required to approximate low-frequency resonances and to the observation that increasing the resolution in the angle is more important than in the radius. Using 1200 angles provides a constant resolution of $2.5\,\text{Hz}$ below $500\,\text{Hz}$; this seems to be already a sufficient resolution, as confirmed by the results depicted in Fig. 8, showing the average full-response time-domain NMSE over a selection of 10 RIRs, computed for different allocations of radii and angles of poles in the Bark-exp grid. It should be noted in the figure that doubling the number of poles in the grid from 6000 to 12000 does not provide a significant increase in the accuracy.

The wBU method, also using Bark-warping factor $w = 0.766$, was slightly modified in order to avoid numerical instabilities. Although it has been proved in [24] that the BU method provides a stable IIR filter by conversion of an FIR filter, we observed cases where the solution with the minimum conversion error contains poles outside the unit circle. In the same paper, instabilities were noticed in those cases where the FIR filter was maximum-phase. However, in those cases, the conversion error was supposed to be high, so that it was always possible to find a stable solution with low conversion error. In [22], numerical limitations were observed in the wBU method and in the computation of the roots of the poles for a number of parameters $C_{\text{m}}$ above 300. Since higher values of $C_{\text{m}}$ were considered in our simulations, the wBU method was modified by choosing the first stable solution with minimum conversion error. The number of iterations was set to 100. For PZ models, the STMCB method [25] with $P = Q + 1$ parameters has been used, with the initial estimate obtained with Prony's method [52]. In order to reduce the number of unstable solutions, only three iterations were executed.

Fig. 9 presents simulation results comparing the performance of the different models for varying model complexity $C_{\text{m}}$. The MATLAB code for generating the results presented in this section is available online[1]. The NMSE produced by OBF models obtained with OBF-MP was computed at each iteration, while for other models the NMSE was computed only for given values of $C_{\text{m}}$. In the bottom row of the figure, the occurrences of unstable solutions given by the wBU method and the STMCB method are reported. For the wBU method, the first stable solution was used, as described above, while for PZ models, unstable solutions were removed from the calculation of the average NMSE. It is clear that both methods suffer from instability due to ill-conditioning above certain values of $C_{\text{m}}$.

In the left column of Fig. 9, results for the NMSE in the time domain defined in (22) are given for the complete response, for the early reflections and for the late reverberation. The plot on top shows that OBF models provide in general a better approximation of the target RIR over $N$ samples, with OBF-MP outperforming AZ models even in the approximation of the early response (middle plot), except when AZ models achieve perfect modeling (at $C_{\text{m}} = 1200$, $h^{\text{early}}_{\text{NMSE}}(\text{dB}) = -\infty$ for AZ models). Focusing on OBF models, OBF-MP shows an overall improvement over OBF-wBU, with the former having a better performance in the early part of the response and the latter performing better in the late part (bottom plot).

The plots in the right column of Fig. 9 show results in the frequency domain. Results in the frequency range between $0$ and $4\,\text{kHz}$ show a clear improvement in the approximation of the low/mid frequencies given by OBF models, with OBF-MP and OBF-wBU having a similar performance for small $C_{\text{m}}$, but with an increased accuracy for increasing $C_{\text{m}}$ provided by the wBU method. This does not imply a degraded performance in the higher part of the spectrum, where OBF models give an error comparable to the error of AZ and PZ models, with OBF-MP providing an improvement over OBF-wBU and the other models, as can be seen in the plot at the bottom (this improvement is less visible than above, given the larger frequency range considered).

In general, differences in the performance of OBF-MP and OBF-wBU are a result of the inherent discretization of the OBF-MP algorithm; its limited resolution prevents the OBF-MP algorithm from perfectly matching the frequency and bandwidth of some magnitude peaks. As a consequence, these peaks are approximated using poles with a slightly shorter radius, which corresponds to a larger bandwidth and a shorter decay of the time response; which is the reason why OBF-wBU shows better performances at low frequencies and in the late response. On the other hand, OBF-MP has a higher resolution at higher frequencies and, as a result, a better performance in that frequency region and in the approximation of the early response.

These results can be visualized on the approximated frequency magnitude responses of the example in Fig. 10, showing the more accurate approximation of low-frequency resonances provided by OBF models with a Bark-scale resolution compared to AZ and PZ models, with the OBF model obtained with OBF-wBU performing better than the OBF-MP, for the reason explained above. However, a large error is introduced by OBF-wBU at high frequencies, while OBF-MP is able to better approximate the envelope of the magnitude response. Looking at the selected pole sets for the different models, some differences can be observed: for PZ models, the poles are evenly distributed in the entire Nyquist interval, while for OBF models, the poles are mostly concentrated in the low frequencies (and closer to the unit circle), with a larger concentration in the very low frequencies for OBF-wBU models. It should be noted that, while the wBU method

[1] https://lirias.kuleuven.be/handle/123456789/581178

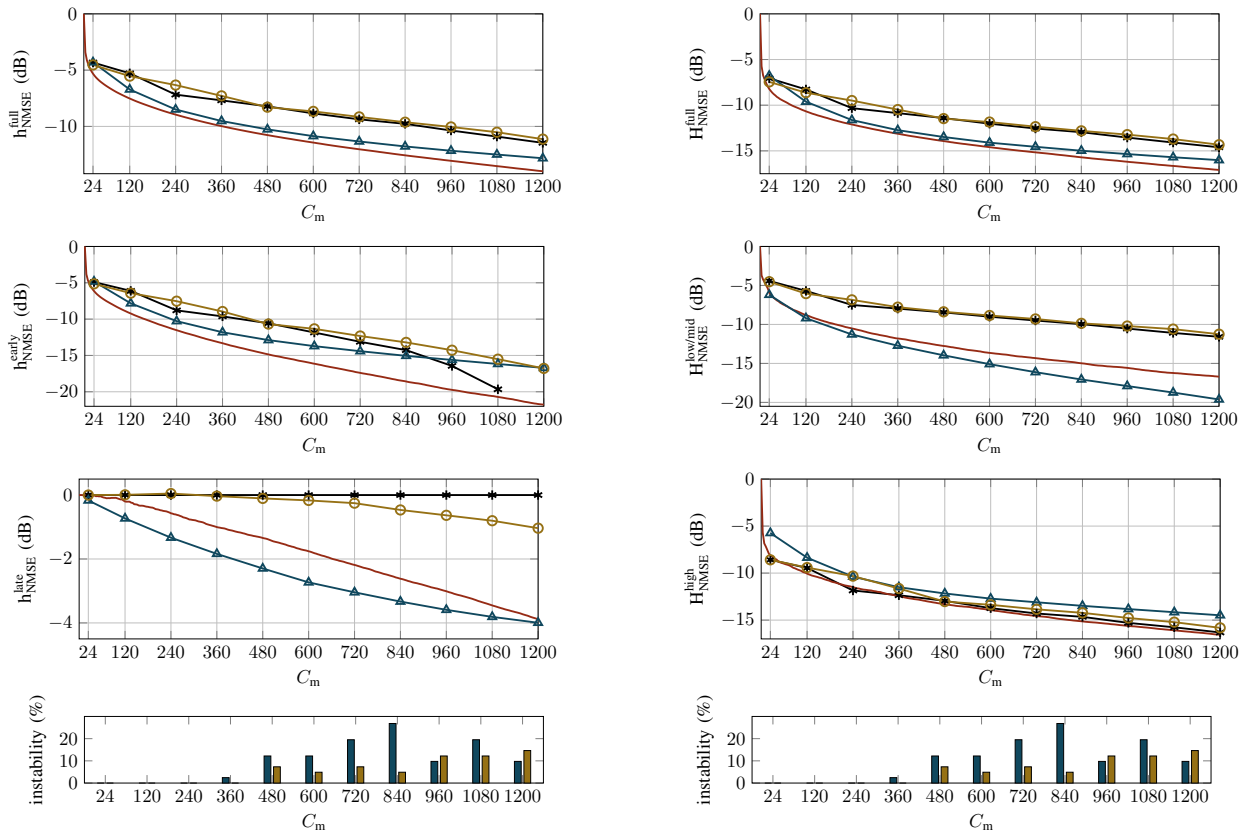


Fig. 9. The average NMSE vs. the model complexity $C_m$. (left) The average time-domain NMSE in (22) for the entire response (top) and for the early (middle) and late response (bottom). (right) The average frequency-domain NMSE in (23) for the entire frequency range (top) and for the frequency regions $[0, 4]$ kHz (middle) and $[4, 20]$ kHz (bottom). AZ models (∗), PZ models (○), OBF models obtained with OBF-wBU (△) and with OBF-MP (—). At the bottom, occurrences of unstable solutions for OBF models obtained with OBF-wBU (left bars) and PZ models (right bars) are reported (same plot on both columns).

allows to control the frequency resolution only by means of the warping parameter, the pole grid of the OBF-MP algorithm offers more flexibility in the selection of the candidate poles and the possibility of incorporating prior knowledge about the characteristics of the room.

As discussed in Section V, another important aspect to consider when comparing different parametric models is their filter complexity $C_f$. While the filter complexity $C_f$ for AZ and PZ models equals the model complexity $C_m$, OBF models require extra computations (cf. Table I). As an illustrative example, Fig. 11 shows the error of the different models as a function of the filter complexity $C_f$, and should be compared to the top-left plot of Fig. 9. The corresponding model complexity $C_m$ for OBF models is reported on the axis underneath. It can be seen that using OBF models gives a smaller average NMSE compared to AZ and PZ models also in terms of filter complexity. Given the equivalence in terms of filter response between OBF models and PF models, the number of multiplications for OBF models can be reduced by using a PF implementation (in which case $C_f$ equals $C_m$).

In order to perform the same kind of analysis in the modal region, similar simulations were run on the SUBRIR database [48], [53]. The SUBRIR database is a collection of RIRs measured at low frequency using a subwoofer as a source. Here, a subset of $R = 24$ RIRs measured with a Genelec 7050B subwoofer (with frequency range 25-120 Hz) and a B&K 4133 (½") microphone was used. The RIRs were downsampled to $f_s = 800$ Hz and truncated to 1.5 s (corresponding to the maximum reverberation time, as reported in Table II and in [48]).

The OBF-MP grid used is this case has 600 angles uniformly placed from 0 to $\pi$ (the Bark scale below 500 Hz has uniform resolution) and 10 radii logarithmically distributed from 0.75 to 0.999, while the BU method is applied without prewarping. The top plot of Fig. 12 shows the error of the different models as a function of the filter complexity $C_f$, similarly to Fig. 11. As in the previous examples, although OBF-MP models and OBF-BU models perform similarly, the BU method leads to numerical conditioning problems and to unstable solutions for values of the model complexity as low as $C_m = 240$. PZ models were not considered here, as the STMCB algorithm provided unstable solutions almost in every situation. In this case, the improvement obtained with OBF models with respect to AZ models is more accentuated than in the previous examples. The reason for this is that in the modal region the number of room resonances is low and the models based on OBFs provide a more meaningful approximation of a RTF than AZ models, as discussed in Section III.

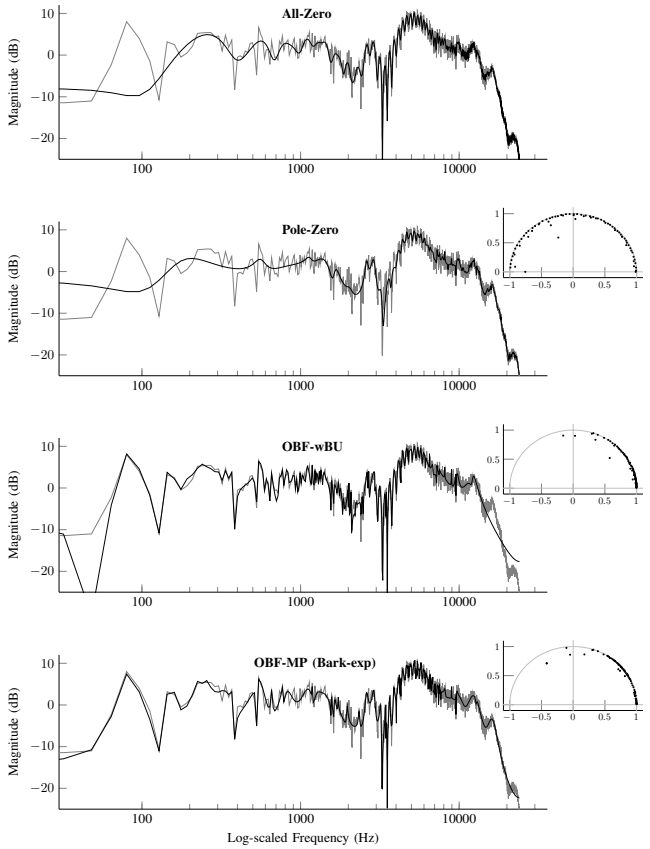


Fig. 10. Approximated magnitude responses for (from top to bottom) an AZ model, a PZ model, OBF models with the wBU method and with the proposed method using a $5 \times 1200$ Bark-exp pole grid, together with the corresponding selected pole set ($C_\mathrm{m} = 300$). The target response (from MARDY) is shown in gray.

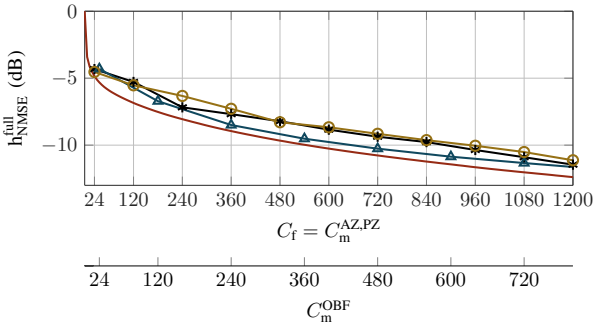


Fig. 11. The average time-domain NMSE in (22) for the entire response for different values of filter complexity $C_\mathrm{f}$. AZ (∗), PZ (○), OBF-wBU (△) and OBF-MP (—) models. The filter complexity corresponds to $C_\mathrm{m}$ for AZ and PZ models, while the corresponding values of $C_\mathrm{m}$ for OBF models are shown in the additional axis.

The comparison of the performance of OBF-MP and OBF-BU in the frequency region of the loudspeaker, instead, presents significant differences (bottom plot). Our interpretation is that the ill-conditioning problems of the BU method are worsened by the fact that the spectrum above 130 Hz contains only noise. The result is that also poles above that frequency are estimated. The poles selected by the well-conditioned

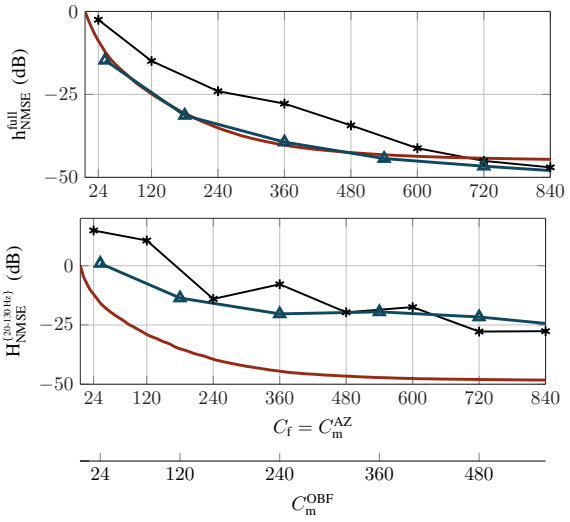


Fig. 12. SUBRIR database. (top) The average time-domain NMSE in (22) for the entire response w.r.t. the filter complexity $C_\mathrm{f}$. (bottom) The average frequency-domain NMSE in (23) between 20 Hz and 130 Hz w.r.t. the filter complexity $C_\mathrm{f}$. AZ (∗), OBF-BU (△) and OBF-MP (—) models.

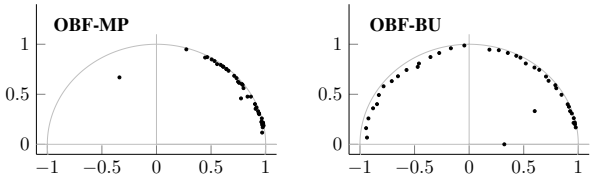


Fig. 13. SUBRIR database. The set of 40 complex-conjugate pole-pairs ($C_\mathrm{m} = 160$) obtained with OBF-MP (left) and the BU method (right) in the approximation of one RIR ($f_s = 800$ Hz).

OBF-MP algorithm, even though the poles in the grid are placed from 0 to $\pi$, are instead well concentrated within the range of the loudspeaker response, as shown in Fig. 13.

## VII. Conclusion and Future Work

The use of OBF models for obtaining a compact and accurate approximation of a target RIR has been motivated by the desirable properties derived from orthogonality, such as an improved numerical conditioning in the estimation of the numerator parameters of the transfer function for a fixed denominator. However, also OBF models are nonlinear in the parameters, so that the estimation of the poles is still a nonlinear problem. The state-of-the-art technique, the BU method, based on an FIR to IIR conversion which exploits the orthogonality property of OBF models, has some restrictions.

In this paper, the novel algorithm, termed OBF-MP, has been studied and compared to the BU method in terms of modeling performance. Simulation results for RIRs measured in different rooms showed that OBF models are able to achieve a reduction in the approximation error compared to conventional parametric models for the same model and filter complexity, provided that the estimation of the pole parameters is accurate. Although the two algorithms considered for the

estimation of the poles in an OBF model seem to have similar modeling capabilities, they present many differences. While the BU method suffers from numerical conditioning problems and instability, the OBF-MP algorithm always delivers stable and well-conditioned OBF model estimates. Indeed, the OBF-MP algorithm bypasses the nonlinear problem of estimating the poles of an OBF model by defining a set of candidate stable poles and by selecting a complex-conjugate pole pair at each iteration based on the correlation between the target RIR and the basis functions built from the candidate poles. Orthogonality of the basis functions assures that this operation is numerically well-conditioned.

Moreover, while the BU-method requires the number of poles to be determined before estimation, the OBF-MP algorithm is scalable, in the sense that a new pair of complex-conjugate poles can be estimated independently of the poles estimated at previous iterations. Scalability turned out to be related to the analogy between OBF models and the definition of the RIR as an infinite summation of exponentially decaying sinuosoids independent from each other. The OBF-MP algorithm follows this interpretation by creating an approximation of the target RIR by adding a pair of OBFs and reducing the approximation error at each iteration. Differences in the performance between OBF-MP and OBF-BU in different time and frequency regions are a consequence of the approach to the pole estimation problem. While the BU method does not make any assumption on the position of the poles and controls the frequency resolution only by prewarping of the target RIR, the grid of candidate poles of the OBF-MP algorithm is an important design choice that adds a layer of flexibility. Any desired frequency resolution could be obtained, motivated by prior knowledge about the acoustics of the room or by application requirements. In this paper, the Bark-exp grid was introduced to provide an accurate approximation at low frequencies, following the physical interpretation described above. However, the Bark-exp grid provides low resolution at high frequency, so that for larger model complexities, i.e. after the dominant modes and the spectral envelope have been approximated, pole grids with higher resolution at high frequencies can become more efficient. A possibility to overcome this issue could be to refine the estimation of the poles at each iteration using numerical optimization methods. This possibility and the inclusion of prior knowledge about the system in the estimation problem is left for future work.

The computational complexity of the OBF-MP algorithm is determined by the length of the target RIR sequence, the number of poles in the grid and the number of model parameters, and it is comparable with the algorithmic complexity of the BU method. Different approaches have been presented for reducing the complexity of the BU method and overcome its limitations, such as subband modeling [21], polyphase design and successive segmentation in the time domain [27]. It is believed that such extensions could be applied to the OBF-MP algorithm as well. Another interesting aspect is the possibility of exploiting the concept of common acoustical poles, as considered e.g. in [11] and [54]. The OBF-MP algorithm was modified in [53] in order to estimate a common set of poles from measurements taken for different source-receiver positions inside a room. It has been shown that a significant reduction in the number of parameters necessary to model the RTF for different source-receiver positions can be achieved. A block-based version of the OBF-MP algorithm has been proposed in [55] and applied in [56] to the estimation of the poles of an adaptive IIR filter based on OBFs from input-output data of a SIMO room acoustic system. Results show that poles can be accurately estimated from white input-output data as well, offering a reduced approximation error compared to FIR filters, with the same convergence rate and complexity of the adaptation algorithm for the linear coefficients, but an improved robustness to the variability of the RTF for different source-receiver positions. Further research will focus on understanding the relation between the estimated common poles and the acoustic characteristics of the room, and on estimating the poles from nonstationary input-output data.

## ACKNOWLEDGMENT

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of (i) the FP7-PEOPLE Marie Curie Initial Training Network ‘Dereverberation and Reverberation of Audio, Music, and Speech (DREAMS)’ (http://www.dreams-itn.eu/), funded by the European Commission under Grant Agreement no. 316969, (ii) KU Leuven Research Council CoE PFV/10/002 (OPTEC), (iii) Research Project FWO no. G.0763.12 ‘Wireless Acoustic Sensor Networks for Extended Auditory Communication’, and (iv) was supported by a Postdoctoral Fellowship (F+/14/045) of the KU Leuven Research Fund. The scientific responsibility is assumed by the authors. The authors would like to thank the reviewers for their insightful and constructive comments on the paper.

## REFERENCES

[1] J. Mourjopoulos and M. A. Paraskevas, “Pole and zero modeling of room transfer functions,” *J. Sound Vib.*, vol. 146, no. 2, pp. 281–302, 1991.

[2] G. Long, D. Shwed, and D. Falconer, “Study of a pole-zero adaptive echo canceller,” *IEEE Trans. Circuits Syst.*, vol. 34, no. 7, pp. 765–769, 1987.

[3] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-time Signal Processing (2Nd Ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1999.

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

[5] ——, “Audio equalization with fixed-pole parallel filters: An efficient alternative to complex smoothing,” in *Preprints Audio Eng. Soc. Conv. 128*, 2010.

[6] ——, “Computationally efficient nonlinear Chebyshev models using common-pole parallel filters with the application to loudspeaker modeling,” in *Preprints Audio Eng. Soc. Conv. 130*, 2011.

[7] B. Bank and J. O. Smith, III, “A delayed parallel filter structure with an FIR part having improved numerical properties,” in *Preprints Audio Eng. Soc. Conv. 136*, 2014.

[8] J. Rämö, V. Välimäki, and B. Bank, “High-precision parallel graphic equalizer,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 12, pp. 1894–1904, 2014.

[9] H. Kuttruff, *Room acoustics*, 5th ed. Spon Press, 2009.

[10] F. Jacobsen and P. Juhl, *Fundamentals of general linear acoustics*. John Wiley & Sons Ltd, 2013.

[11] 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 Trans. Speech Audio Process.*, vol. 7, no. 6, pp. 709–717, 1999.

[12] A. P. Liavas and P. A. Regalia, "Acoustic echo cancellation: Do IIR models offer better modeling capabilities than their FIR counterparts?" *IEEE Trans. Signal Process.*, vol. 46, no. 9, pp. 2499–2504, 1998.

[13] P. Heuberger, P. van den Hof, and B. Wahlberg, *Modelling and Identification with Rational Orthogonal Basis Functions*. Springer, 2005.

[14] P. S. C. Heuberger, P. M. J. Van den Hof, and O. Bosgra, "A generalized orthonormal basis for linear dynamical systems," *IEEE Trans. Autom. Control*, vol. 40, no. 3, pp. 451–465, 1995.

[15] B. Ninness and F. Gustafsson, "A unifying construction of orthonormal bases for system identification," *IEEE Trans. Automatic Control*, vol. 42, no. 4, pp. 515–521, 1997.

[16] G. Oliveira, A. Da Rosa, R. Campello, J. Machado, and W. Amaral, "An introduction to models based on Laguerre, Kautz and other related orthonormal functions - part I: linear and uncertain models," *Int. J. Model. Ident. Control*, vol. 14, pp. 121–132, 2011.

[17] G. W. Davidson and D. D. Falconer, "Reduced complexity echo cancellation using orthonormal functions," *IEEE Trans. Circuits Syst.*, vol. 38, no. 1, pp. 20–28, 1991.

[18] L. S. Ngia, "Recursive identification of acoustic echo systems using orthonormal basis functions," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 3, pp. 278–293, 2003.

[19] T. Paatero, M. Karjalainen, and A. Härmä, "Modeling and equalization of audio systems using Kautz filters," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP'01)*, vol. 5. IEEE, 2001, pp. 3313–3316.

[20] T. Paatero and M. Karjalainen, "New digital filter techniques for room response modeling," in *Proc. 21$^{st}$ Int. Conf. Audio Eng. Soc.: Architectural Acoustics and Sound Reinforcement*. AES, 2002.

[21] T. Paatero, "An audio motivated hybrid of warping and kautz filter techniques," in *Proc. 11$^{th}$ Eur. Signal Process. Conf. (EUSIPCO'02),* Toulouse, France, 2002, pp. 627–630.

[22] T. Paatero and M. Karjalainen, "Kautz filters and generalized frequency resolution: Theory and audio applications," *J. Audio Eng. Soc.*, vol. 51, no. 1/2, pp. 27–44, 2003.

[23] M. Karjalainen and T. Paatero, "Equalization of loudspeaker and room responses using Kautz filters: Direct least squares design," *EURASIP J. Adv. Sig. Pr.*, vol. 2007, no. 1, pp. 185–185, 2007.

[24] H. Brandenstein and R. Unbehauen, "Least-squares approximation of FIR by IIR digital filters," *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 21–30, 1998.

[25] K. Steiglitz and L. McBride, "A technique for the identification of linear systems," *IEEE Trans. Automatic Control*, vol. 10, pp. 461–464, 1965.

[26] A. Härmä, M. Karjalainen, L. Savioja, V. Välimäki, U. K. Laine, and J. Huopaniemi, "Frequency-warped signal processing for audio applications," *J. Audio Eng. Soc.*, vol. 48, no. 11, pp. 1011–1031, 2000.

[27] T. Paatero, "Modeling of long and complex responses using Kautz filters and time-domain partitions," in *Proc. 12$^{th}$ Eur. Signal Process. Conf. (EUSIPCO'04),* Vienna, Austria, 2004, pp. 313–316.

[28] G. Vairetti, T. van Waterschoot, M. Moonen, M. Catrysse, and S. H. Jensen, "Sparse linear parametric modeling of room acoustics with orthonormal basis functions," in *Proc. 22$^{nd}$ Eur. Signal Process. Conf. (EUSIPCO'14),* Lisbon, Portugal, 2014.

[29] ——, "An automatic model-building algorithm for sparse approximation of room impulse responses with orthonormal basis functions," in *Proc. Int. Workshop on Acoustic Signal Enhancement (IWAENC 2014)*, Antibes-Juan Les Pins, 2014.

[30] R. Mignot, G. Chardon, and L. Daudet, "Low frequency interpolation of room impulse responses using compressed sensing," *IEEE Trans. Audio Speech Lang. Process.*, vol. 22, no. 1, pp. 205–216, 2014.

[31] G. Chardon and L. Daudet, "Optimal subsampling of multichannel damped sinusoids," in *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2010 IEEE*. Jerusalem, Israel: IEEE, 2010, pp. 25–28.

[32] M. Schroeder and K. Kuttruff, "On frequency response curves in rooms. comparison of experimental, theoretical, and Monte Carlo results for the average frequency spacing between maxima," *J. Acoust. Soc. Am.*, vol. 34, no. 1, pp. 76–80, 1962.

[33] T. van Waterschoot, "KU Leuven ESAT Speech Lab room impulse response database," KU Leuven, Belgium, Tech. Rep. ESAT-STADIUS TR 15-74, Aug. 2015. [Online]. Available: ftp://ftp.esat.kuleuven.be/pub/sista/vanwaterschoot/abstracts/15-74.html

[34] M. Karjalainen, P. A. A. Esquef, P. Antsalo, A. Mäkivirta, and V. Välimäki, "AR/ARMA analysis and modeling of modes in resonant and reverberant systems," in *Audio Eng. Soc. Conv. 112*, 2002.

[35] J. O. Smith, *Introduction to Digital Filters with Audio Applications*. online book, 2007 edition, available: http://ccrma.stanford.edu/∼jos/filters/, accessed Oct. 2014.

[36] T. van Waterschoot and M. Moonen, "A pole-zero placement technique for designing second-order IIR parametric equalizer filters," *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 8, pp. 2561–2565, 2007.

[37] B. Wahlberg, "System identification using Laguerre models," *IEEE Trans. Autom. Control*, vol. 36, no. 5, pp. 551–562, 1991.

[38] J. O. Smith and J. S. Abel, "Bark and ERB bilinear transforms," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 6, pp. 697–708, 1999.

[39] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, 2010.

[40] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Stat. Soc. B. Met.*, pp. 267–288, 1996.

[41] S. Mallat and Z. Zhang, "Adaptive time-frequency decomposition with matching pursuits," in *Time-Frequency and Time-Scale Analysis, 1992., Proceedings of the IEEE-SP International Symposium*. IEEE, 1992, pp. 7–10.

[42] Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Conf. Rec. 27$^{th}$ Asilomar Conf. on Signals, Systems, Comput.*, 1993, pp. 40–44.

[43] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[44] G. H. Golub and C. F. Van Loan, *Matrix computations*, 3rd ed. Baltimore and London: The Johns Hopkins University Press, 1996.

[45] J. Y. Wen, N. D. Gaubitch, E. A. Habets, T. Myatt, and P. A. Naylor, "Evaluation of speech dereverberation algorithms using the MARDY database," in *Proc. Int. Workshop on Acoustic Signal Enhancement (IWAENC 2006)*, Paris, France, 2006.

[46] J. K. Nielsen, J. R. Jensen, S. H. Jensen, and M. G. Christensen, "The single-and multichannel audio recordings database (SMARD)," in *Proc. Int. Workshop on Acoustic Signal Enhancement (IWAENC 2014)*, Antibes-Juan Les Pins, France, 2014, pp. 40–44.

[47] E. Hadad, F. Heese, P. Vary, and S. Gannot, "Multichannel audio database in various acoustic environments," in *IEEE Proc. Int. Workshop on Acoustic Signal Enhancement (IWAENC 2014)*, Antibes-Juan Les Pins, 2014, pp. 313–317.

[48] G. Vairetti, N. Kaplanis, S. Bech, E. De Sena, M. Moonen, and T. van Waterschoot, "The subwoofer room impulse response (SUBRIR) database," *J. Audio Eng. Soc.*, vol. 65, no. 05, May 2017. [Online]. Available: https://lirias.kuleuven.be/handle/123456789/572970

[49] A. Lindau, L. Kosanke, and S. Weinzierl, "Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses," in *Preprints Audio Eng. Soc. Conv. 128*, 2010.

[50] M. Karjalainen, P. Antsalo, A. Mäkivirta, T. Peltonen, and V. Välimäki, "Estimation of modal decay parameters from noisy response measurements," *J. Audio Eng. Soc.*, vol. 50, no. 11, pp. 867–878, 2002.

[51] A. Lundeby, T. E. Vigran, H. Bietz, and M. Vorländer, "Uncertainties of measurements in room acoustics," *Acta Acust. united Ac.*, vol. 81, no. 4, pp. 344–355, 1995.

[52] C. S. Burrus and T. W. Parks, "Time domain design of recursive digital filters," *IEEE Trans. Audio Electroacoust.*, vol. 18, no. 2, pp. 137–141, 1970.

[53] G. Vairetti, E. De Sena, T. van Waterschoot, M. Moonen, M. Catrysse, N. Kaplanis, and S. H. Jensen, "A physically motivated parametric model for compact representation of room impulse responses based on orthonormal basis functions," in *Proc. of the 10$^{th}$ Eur. Congr. and Expo. on Noise Control Eng. (EURONOISE 2015)*, Maastricht, The Netherlands, 2015.

[54] Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 2, pp. 320–328, 1994.

[55] G. Vairetti, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. van Waterschoot, "Room acoustic system identification using orthonormal basis function models," in *Proc. 60$^{th}$ Int. Conf. Audio Eng. Soc.: Dereverberation and Reverberation of Audio, Music and Speech*. Leuven, Belgium: AES, 2016.

[56] ——, "Multichannel identification of room acoustic systems with adaptive IIR filters based on orthonormal basis functions," in *IEEE Int. Conf. on Acoust. Speech Signal Process. (ICASSP '16)*. Shanghai, China: IEEE, 2016, pp. 16–20.



**Giacomo Vairetti** received the B.Sc. in 2010 and the M.Sc. (cum laude) in 2012, both in Computer Engineering at Politecnico di Milano (Italy). He is currently pursuing a Ph.D. in Electrical Engineering at KU Leuven (Belgium), where he was a Marie Curie Fellow.

He was a visiting student at the Signal Processing and Acoustics Dept. of Aalto University (Finland) in 2012 and at the Electronic Systems Dept. of Aalborg University (Denmark) in 2014.

His research interests are in signal processing and system identification applied to room acoustic modeling, sound synthesis, and audio reproduction.



**Enzo De Sena** (S'11-M'14) received the B.Sc. in 2007 and M.Sc. (cum laude) in 2009, both from the Universita degli Studi di Napoli "Federico II" (Italy) in Telecommunication Engineering. In 2013, he received the Ph.D. degree in Electronic Engineering from Kings College London (UK), where he was also a Teaching Fellow from 2012 to 2013.

Between 2013 and 2016 he was a Postdoctoral Research Fellow at the KU Leuven (Belgium). Since September 2016 he is a Lecturer in Audio at the Institute of Sound Recording at the University of Surrey (UK). He held visiting positions at Stanford University (USA), Aalborg University (Denmark) and Imperial College London (UK). He is a former Marie Curie Fellow.

His current research interests include room acoustics modelling, surround sound, microphone beam forming and binaural modelling. He is a member of IEEE, EURASIP, and the Acoustical Society of America. For more information, see www.desena.org.



**Michael Catrysse** holds a M.Sc. degree in Electronic Engineering (1998) and a Ph.D. degree in Electronic Engineering (2004) from the KU Leuven (Belgium).

Michael joined Televic in 2013 as Director Technology & Innovation. Before joining Televic, Michael worked as R&D manager and plant manager with the Picanol Group (Belgium). Prior to this, he held positions as researcher (FMTC, Belgium), innovation consultant and project manager (Centexbel, Belgium) and innovation manager (Domo). Michael Catrysse has been involved in several nationally and European funded research projects, both as a participant and a project coordinator. He is (co)author of more than 30 publications.



**Søren Holdt Jensen** (S'87-M'88-SM'00) received the M.Sc. degree in electrical engineering from Aalborg University (AAU), Denmark, in 1988, and the Ph.D. degree (in signal processing) from the Technical University of Denmark (DTU) in 1995.

He is Full Professor in Signal Processing at Aalborg University. Before joining the Electronic Systems Dept. (AAU), he was with the Telecommunications Laboratory of Telecom Denmark, Ltd, Copenhagen; the Electronics Institute of Technical University of Denmark; the Scientific Computing Group of Danish Computing Center for Research and Education (UNI•C), Lyngby; the Electrical Engineering Dept. of KU Leuven, Belgium; and the Center for PersonKommunikation (CPK) of AAU. His current research interest are in statistical signal processing, numerical algorithms, optimization engineering, machine learning, and digital processing of acoustic, audio, communication, image, multimedia, speech, and video signals.

He is co-author of the textbook *Software-Defined GPS and Galileo Receiver—A Single-Frequency Approach*, Birkhäuser, Boston, USA, also translated to Chinese: National Defence Industry Press, China. Prof. Jensen has been Associate Editor for the IEEE Transactions on Signal Processing, IEEE/ACM Transactions on Audio, Speech and Language Processing, Elsevier Signal Processing, and EURASIP Journal on Advances in Signal Processing. He is a recipient of an individual European Community Marie Curie (HCM: Human Capital and Mobility) Fellowship, former Chairman of the IEEE Denmark Section and the IEEE Denmark Sections Signal Processing Chapter (founder and first chaiman). He is member of the Danish Academy of Technical Sciences (ATV) and has been member of the Danish Council for Independent Research (2011–2016) appointed by Danish Ministers of Science.



**Marc Moonen** (M'94-SM'06-F'07) is a Full Professor at the Electrical Engineering Dept. of KU Leuven, where he is heading a research team working in the area of numerical algorithms and signal processing for digital communications, wireless communications, DSL and audio signal processing.

He received the 1994 KU Leuven Research Council Award, the 1997 Alcatel Bell (Belgium) Award (with Piet Vandaele), the 2004 Alcatel Bell (Belgium) Award (with Raphael Cendrillon), and was a 1997 Laureate of the Belgium Royal Academy of Science. He received journal best paper awards from the IEEE Transactions on Signal Processing (with Geert Leus and with Daniele Giacobello) and from Elsevier Signal Processing (with Simon Doclo).

He was chairman of the IEEE Benelux Signal Processing Chapter (1998-2002), a member of the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications, and President of EURASIP (European Association for Signal Processing, 2007-2008 and 2011-2012).

He has served as Editor-in-Chief for the EURASIP Journal on Applied Signal Processing (2003-2005), Area Editor for Feature Articles in IEEE Signal Processing Magazine (2012-2014), and has been a member of the editorial board of IEEE Transactions on Circuits and Systems II, IEEE Signal Processing Magazine, Integration-the VLSI Journal, EURASIP Journal on Wireless Communications and Networking, and Signal Processing. He is currently a member of the editorial board of EURASIP Journal on Advances in Signal Processing.



**Toon van Waterschoot** (S'04-M'12) received the M.Sc. (2001) and Ph.D. (2009) degrees in Electrical Engineering, both from KU Leuven, Belgium, where he is currently a tenure-track Assistant Professor.

He has previously held teaching and research positions with the Antwerp Maritime Academy, the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT), and the Research Foundation - Flanders (FWO) in Belgium, with Delft University of Technology in The Netherlands, and with the University of Lugano in Switzerland. His research interests are in signal processing, machine learning, and numerical optimization, applied to acoustic signal enhancement, acoustic modeling, audio analysis, and audio reproduction.

He has been serving as an Associate Editor for the Journal of the Audio Engineering Society (AES) and for the EURASIP Journal on Audio, Music, and Speech Processing, and as a Guest Editor for Elsevier Signal Processing. He is a Member of the Board of Directors of the European Association for Signal Processing (EURASIP) and a Member of the IEEE Audio and Acoustic Signal Processing Technical Committee (AASP-TC). He was the General Chair of the 60th AES International Conference in Leuven, Belgium (2016), and has been serving on the Organizing Committee of the European Conference on Computational Optimization (EUCCO 2016) and the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2017). He is a member of EURASIP, IEEE, ASA, and AES.