KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	G. Vairetti, E. De Sena, S. H. Jensen, M. Moonen, and T. van Waterschoot (2018), **Orthonormal Basis Functions Adaptive Filters for Room Acoustic Signal Enhancement** Technical Report (to be submitted)
Archived version	Author manuscript
Published version	
Journal homepage	
Author contact	giacomo.vairetti@esat.kuleuven.be *+ 32 (0)16 321817*
IR	url in Lirias https://lirias.kuleuven.be/handle/123456789/xxxxxx

(article begins on next page)

Orthonormal Basis Functions Adaptive Filters for Room Acoustic Signal Enhancement

Giacomo Vairetti, Enzo De Sena, Søren Holdt Jensen,
Marc Moonen, and Toon van Waterschoot

***Abstract*—Room acoustic signal enhancement applications strongly rely on algorithms for identifying the acoustic response of the room (or its inverse) at one or multiple locations of the sound sources and microphones. As shown in recent studies, stable infinite impulse response (IIR) filters based on orthonormal basis functions (OBFs) (hereafter called *OBF filters*) are well-suited for approximating room responses, especially at low frequencies where the approximation is well motivated also from a physical point of view. Moreover, when used in an adaptive algorithm, the orthogonality property of OBF filters enables an analysis of their tracking performance, with theoretical results showing faster convergence compared to other fixed-poles IIR filters for a wide range of input spectra. In this paper, the theory of OBF adaptive filters is reviewed in relation to the identification of room acoustic systems. An iterative algorithm is proposed for the identification of room acoustic systems at low frequencies, able to accurately estimate the characteristic poles of the room transfer functions from white noise and speech signals. It is shown that the actual advantage compared to the use of finite impulse response filters mostly depends on the characteristics of the room itself, e.g. its dimensions and reverberation time. Finally, the applicability of OBF adaptive filters to acoustic signal enhancement algorithms is discussed by means of examples in the context of acoustic echo cancellation and room equalization.**

***Index Terms*— system identification, room acoustics, signal enhancement, IIR adaptive filters, orthonormal basis functions.**

I. Introduction

APPLICATIONS of room acoustic signal enhancement (RASE), such as acoustic echo cancellation (AEC) [1], acoustic feedback cancellation (AFC) [2], or room response equalization (RRE) [3], normally rely on the modeling and identification of the room impulse response (RIR) or its inverse at one or multiple locations of the sound sources and microphones. The most commonly used RIR model is the all-zero model, in which the model parameter values are identified as the coefficients of an finite impulse response (FIR) adaptive filter [4]. The reasons for the popularity of FIR adaptive filters are their simplicity, a well-consolidated theory, and a large variety of algorithms available for each specific application. Far less popular is the use of infinite impulse response (IIR) filters, which implies modeling a room transfer function (RTF) as a pole-zero model [5]. In theory, IIR filters would allow to reduce the number of coefficients required to adequately model the RTFs, thus providing more efficient algorithms. However, their adoption in RASE applications has been discouraged by their higher complexity, the added difficulty in the adaptive identification of the pole parameters, potential instability issues and convergence to local minima [6], and possibly by the wide-spread belief that no relevant advantage can be obtained compared to FIR filters [7], [8].

Recent years have witnessed an increasing interest into pole-zero models based on orthonormal basis functions (OBFs) [9]–[11] for modeling room acoustics [12]–[15]. The idea consists in modeling a RTF as a weighted combination of second-order resonators, whose frequencies and bandwidths are determined by the position of the pole parameters, while their amplitude is controlled by a set of linear coefficients (or weights). As for other fixed-poles IIR filters [16], advantages of using IIR filters based on OBFs (hereafter called *OBF filters*) are related to the increase in accuracy obtained by having the poles of the filter transfer function (TF) closer to the true poles of the system [9], [17], [18], while easily ensuring filter stability. In addition, the orthogonality property of OBF filters provides numerical well-conditioning and fast convergence of the filter adaptation [17], [19], [20], and is the key aspect in enabling an analysis of the error performance and of the dynamic behavior of adaptive algorithms, analogously to the FIR case.

Even though the theory is well established within the field of system identification, OBF filters have been adopted for room acoustic modeling only recently. Effective modeling algorithms have been suggested for the estimation of the poles, showing that a set of stable and accurate poles can be obtained, with the possibility of allocating frequency resolution unevenly in different regions of the spectrum [12], [14]. These algorithms have been modified for the estimation of a common set of poles from multiple RTFs [13], [15] and for modeling in subbands [15] and in time-domain partitions [21]. Although improvements in the approximation accuracy can be obtained over all-zero models on the entire audible spectrum [12], OBF models are particularly suited at low frequencies, where the RTFs are characterized by moderately overlapping and slowly decaying modes [12], [13].

Nonetheless, the use of OBF filters in RASE applications is still very limited. Some examples are found in the context of AEC [20], AFC [22] and RRE [23], [24], all relying on the availability of measured or pre-identified RIRs. The few cases in which the filter pole parameter values are directly estimated from input-output data have been found applied to the identification of acoustic echo systems. Methods have been proposed for the on-line estimation of both the poles and the linear coefficients [25]–[27], but limited to the case of an OBF filter with a single repeated pole (known as *Laguerre* and *2-parameter-Kautz* filters). The method in [27] is said to be applicable to OBF filters with non-repeated pairs of complex-conjugate poles (also known as *Kautz filters*)

using approximated expressions for the gradients [28], but this possibility has not yet been verified on realistic AEC scenarios.

The purpose of this paper is to discuss the applicability of OBF adaptive filters to RASE algorithms. In Section 2, a review of OBF models and OBF adaptive filters is provided. The focus is on the frequency domain analysis of the error performance and dynamic behavior of adaptive algorithms with respect to the characteristics of the input and noise signals. It is shown that adaptive algorithms developed for FIR filters are easily extended to the OBF filter case when poles are fixed, whereas the adaptation of the poles requires nonlinear recursive identification algorithms. An alternative version of the normalized least mean squares (NLMS) algorithm is introduced to deal with the adaptation of the linear coefficients when the filter order is small. The proposed modification, called here OBF-NLMS, normalizes the response of each OBF second-order section individually, in an analogy with transform-domain (TD) adaptive filters [29]–[33].

In Section 3, an identification algorithm is introduced, inspired by the scalable group matching pursuit (GMP) modeling algorithm, named OBF-GMP and described in [12], [13], which ameliorates the previously proposed block-based (BB)-OBF-GMP identification algorithm in [34]. The algorithm, named here stage-based (SB)-OBF-GMP, performs an iterative grid search that avoids the nonlinear problem and gives the possibility of estimating a common set of poles from multiple microphone signals to be then fixed and used in RASE applications. The newly proposed identification algorithm uses the NLMS and OBF-NLMS adaptation algorithms in order to overcome the problems of the BB algorithm in dealing with non-stationary and non-white input signals.

Section 4 provides identification results at low frequencies performed on measured and simulated RIRs. The aim of this section is to analyze the relation between the characteristics of the room, such as its dimensions and reverberation time, and the advantages over FIR filters that can be expected by using OBF adaptive filters with a common set of poles.

In Section 5 examples are given in order to illustrate possible uses of OBF adaptive filters and the proposed identification algorithm in the context of AEC and RRE. Finally, Section 6 concludes the paper.

II. OBF ADAPTIVE FILTERS

A RTF can be expressed as an infinite summation of room resonances, each with a certain frequency, bandwidth and amplitude, whose density increases with frequency [35]. It can be approximated by means of an OBF filter as a finite summation of second-order all-pole filters (or resonators), having TF as in (1), each weighted by a pair of linear amplitude coefficients.

$$P_i(z) = \frac{1}{D_i(z)} = \frac{1}{(1-p_i z^{-1})(1-p_i^* z^{-1})} \ , \qquad (1)$$

$$A_i(z) = \frac{\bar{D}_i(z)}{D_i(z)} = \frac{(z^{-1}-p_i)(z^{-1}-p_i^*)}{(1-p_i z^{-1})(1-p_i^* z^{-1})} \ , \qquad (2)$$

$$N_i^{\pm}(z) = |1 \pm p_i| \sqrt{\frac{1-|p_i|^2}{2}} (z^{-1} \mp 1) \ . \qquad (3)$$

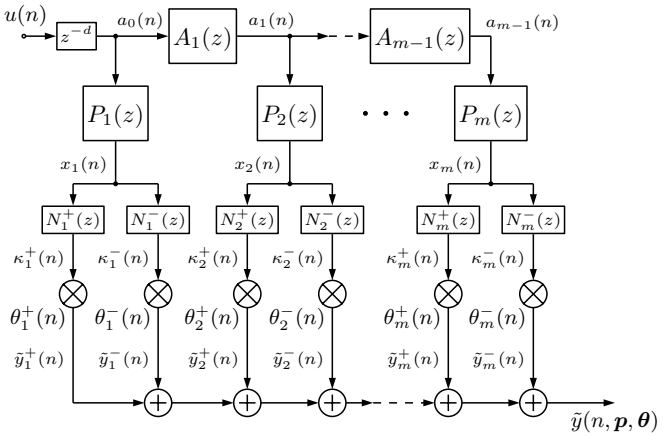

Fig. 1: The OBF adaptive filter for m pole pairs.

Each resonator is defined by a pair of complex-conjugate poles $\boldsymbol{p}_i = [p_i, p_i^*] = \rho_i e^{\pm j\vartheta_i}$, with radius $\rho_i = e^{-\zeta_i/f_s}$ ($\rho_i < 1$ for stability) related to the bandwidth ζ_i, and angle $\vartheta_i = \omega_i/f_s$ related to the resonance frequency ω_i (f_s being the sampling frequency and * indicating complex conjugation). The responses of the resonators are orthogonalized with respect to each other by means of a series of all-pass filters built from the same pole pairs $\boldsymbol{p}_i$ with TF as in (2) (the poles p_i and p_i^* are canceled by the nonminimum-phase zeros in $1/p_i$ and $1/p_i^*$ of the all-pass TF [11]), whereas a pair of first-order all-zero filters with TF as in (3) produces mutually orthonormal responses. Given that a RTF presents multiple resonances with different frequencies and a band-pass characteristic due to the band-pass response of the loudspeaker (and the anti-aliasing filter), only OBF filters with multiple complex poles are considered here. For more details about the construction of OBF models and their relation to room acoustics and other parametric models, the reader is referred to [12].

The OBF adaptive filter structure is depicted in Figure 1. Each ith resonator section has a pair of orthonormal TFs, or basis functions, given by $\Psi_i^{\pm}(z, \tilde{\boldsymbol{p}}_i)$, which are weighted by a pair of time-varying amplitude coefficients $\boldsymbol{\theta}_i(n) = [\theta_i^+(n), \theta_i^-(n)]$ (with $i = 1, \ldots, m$ and m the number of complex-conjugate pole pairs), giving the overall filter TF at discrete time-instant $n = t/f_s$

$$G(z, \boldsymbol{p}, \boldsymbol{\theta}(n)) = z^{-d} \sum_{i=1}^{m} \left[\theta_i^+(n)\Psi_i^+(z, \tilde{\boldsymbol{p}}_i) + \theta_i^-(n)\Psi_i^-(z, \tilde{\boldsymbol{p}}_i)\right]$$

$$\text{with } \ \Psi_i^{\pm}(z, \tilde{\boldsymbol{p}}_i) = N_i^{\pm}(z) P_i(z) \prod_{\iota=1}^{i-1} A_\iota(z), \qquad (4)$$

$\boldsymbol{p} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_m]^T$ a set of m pairs of complex-conjugate poles, $\boldsymbol{\theta}(n) = [\boldsymbol{\theta}_1(n), \ldots, \boldsymbol{\theta}_m(n)]^T$, both with dimensions $M \times 1$ ($M = 2m$), and $\tilde{\boldsymbol{p}}_i = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_i]^T$. Figure 2 shows the power spectrum of the basis functions $\Psi_i^{\pm}(z, \tilde{\boldsymbol{p}}_i)$ generated from a set of $m = 5$ pole pairs with different radii and angles.

The intermediate signals $\boldsymbol{\kappa}_i(n, \tilde{\boldsymbol{p}}_i) = [\kappa_i^+(n, \tilde{\boldsymbol{p}}_i), \kappa_i^-(n, \tilde{\boldsymbol{p}}_i)]$ are filtered versions of the input signal $u(n)$ (i.e. $\kappa_i^{\pm}(n, \tilde{\boldsymbol{p}}_i) = \Psi_i^{\pm}(q, \tilde{\boldsymbol{p}}_i)u(n)$, with q^{-1} the backward time-shift operator for which $q^{-1}u(n) = u(n-1)$), and the output signal of the filter

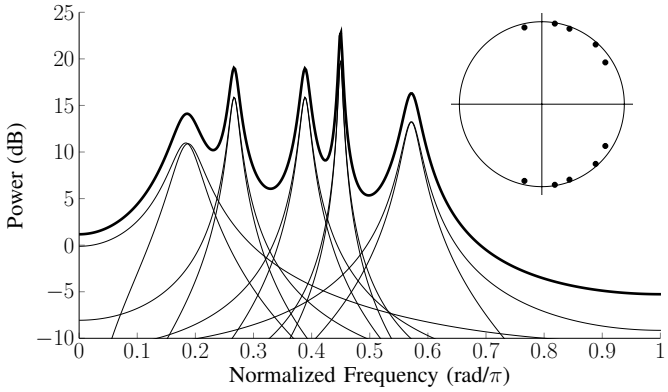

Fig. 2: The power responses of 5 pairs of basis functions generated from the pole pairs depicted in the top-right corner. The resulting $\gamma_M(\omega, \boldsymbol{p})$ in (11) is also shown (thick line).

is a weighted summation of the intermediate signals given as

$$\begin{aligned}\tilde{y}(n, \boldsymbol{p}, \boldsymbol{\theta}) &= \sum_{i=1}^{m} \left[\tilde{y}_i(n, \tilde{\boldsymbol{p}}_i, \boldsymbol{\theta}_i)\right] \\ &= \sum_{i=1}^{m} \left[\tilde{y}_i^+(n, \tilde{\boldsymbol{p}}_i, \theta_i^+) + \tilde{y}_i^-(n, \tilde{\boldsymbol{p}}_i, \theta_i^-)\right] \\ &= \sum_{i=1}^{m} \left[\kappa_i^+(n, \tilde{\boldsymbol{p}}_i)\theta_i^+(n) + \kappa_i^-(n, \tilde{\boldsymbol{p}}_i)\theta_i^-(n)\right] \ , \quad (5)\end{aligned}$$

or in vector form as $\tilde{y}(n, \boldsymbol{p}, \boldsymbol{\theta}) = \boldsymbol{\kappa}^T(n, \boldsymbol{p})\boldsymbol{\theta}(n)$, with $\boldsymbol{\kappa}(n, \boldsymbol{p}) = [\boldsymbol{\kappa}_1(n, \tilde{\boldsymbol{p}}_1), \ldots, \boldsymbol{\kappa}_m(n, \tilde{\boldsymbol{p}}_m)]^T$. Finally, a d-samples delay can be included to take the acoustic delay of the RIR into account (cfr. leftmost block in Figure 1).

The main reason for using an orthonormal model structure in place of other non-orthogonal fixed-pole ones is related to numerical considerations: even though orthogonal and non-orthogonal models with the same fixed poles span the same approximation space, it is found that the estimation of a RTF with respect to the non-orthogonal structure can be very ill-conditioned. Orthogonal model structures, instead, provide a well-conditioned estimation problem for a wide range of input power spectral density (PSD) (not only white), so that their use is said to be the only practical way of fixing the poles in a filter structure [17]. For the same reason, OBF adaptive filters also show faster convergence than other fixed-poles adaptive filters.

To keep the discussion as simple as possible, a RTF $H(z)$ is assumed to be linear and time-invariant, so that a microphone signal can be defined as

$$y(n) = H(q)u(n) + v(n), \quad (6)$$

with $v(n)$ a zero-mean additive white noise (WN) signal with variance $S_v(\omega) = \mathbb{E}\{v^2(n)\} = \sigma_v^2$, $\forall\omega$, and $u(n)$ the loudspeaker input signal having spectral density

$$S_u(\omega) = \sum_{\tau=-\infty}^{\infty} R_u(\tau)e^{-j\omega\tau}, \quad (7)$$

with covariance function $R_u(\tau) = \mathbb{E}\{u(n)u(n-\tau)\}$ ($\mathbb{E}\{\cdot\}$ denotes the expected value). The input PSD plays an important role in the behavior of OBF adaptive filters, as it will be explained later on. Indeed, the power σ_ι^2 of each intermediate signal $\kappa_\iota(n)$ (corresponding to $\kappa_i^+(n)$ for ι odd and to $\kappa_i^-(n)$ for ι even, with $i = {}^{(\iota+\iota \bmod 2)}/_2$ and $\iota = 1, \ldots, M$) is determined by the product between the input PSD and the power of the corresponding OBF frequency response $\Psi_\iota(e^{j\omega}, \tilde{\boldsymbol{p}}_i)$,

$$\mathbb{E}\{|\kappa_\iota(n)|^2\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_u(\omega)\left|\Psi_\iota(e^{j\omega}, \tilde{\boldsymbol{p}}_i)\right|^2 d\omega. \quad (8)$$

This means that an intermediate signal will have small power at a given frequency ω whenever either $S_u(\omega)$ or the power of the OBF frequency response is small. Also notice that, because of normalization,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\Psi_\iota(e^{j\omega}, \tilde{\boldsymbol{p}}_i)\right|^2 d\omega = 1\,. \quad (9)$$

Finally, it is noticed here that the total power of the intermediate signals is given by the sum of the power of each intermediate signal as

$$\begin{aligned}\mathbb{E}\{\|\boldsymbol{\kappa}(n, \boldsymbol{p})\|^2\} &= \sum_{\iota=1}^{M}\frac{1}{2\pi}\int_{-\pi}^{\pi} S_u(\omega)\left|\Psi_\iota(e^{j\omega}, \tilde{\boldsymbol{p}}_i)\right|^2 d\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} S_u(\omega)\sum_{\iota=1}^{M}\left|\Psi_\iota(e^{j\omega}, \tilde{\boldsymbol{p}}_i)\right|^2 d\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} S_u(\omega)\gamma_M(\omega, \boldsymbol{p})d\omega, \quad (10)\end{aligned}$$

$$\text{with } \gamma_M(\omega, \boldsymbol{p}) = \sum_{\iota=1}^{M}\left|\Psi_\iota(e^{j\omega}, \tilde{\boldsymbol{p}}_i)\right|^2 \quad (11)$$

a frequency-dependent term that will be essential in the analysis of the main properties of OBF adaptive filters, summarized in the remainder of the present section. For more extensive explanations, including the case for time-varying systems, the reader is referred to [9], [17]–[19].

A. Estimation accuracy: bias and variance errors

An estimate of $H(z)$ obtained using an OBF filter as in (4) with poles fixed in $\boldsymbol{p}$ produces a prediction error given by

$$\varepsilon(n, \boldsymbol{p}, \boldsymbol{\theta}) = y(n) - \tilde{y}(n, \boldsymbol{p}, \boldsymbol{\theta}) = y(n) - \boldsymbol{\kappa}^T(n, \boldsymbol{p})\boldsymbol{\theta}(n). \quad (12)$$

If N samples of the input and output signals are available and the following quadratic cost function is adopted

$$V_N(\boldsymbol{\theta}, \boldsymbol{p}) = \frac{1}{2N}\sum_{n=1}^{N}\varepsilon^2(n, \boldsymbol{\theta}, \boldsymbol{p}), \quad (13)$$

a least squares (LS) estimate $\hat{\boldsymbol{\theta}}_N$ of the linear amplitude coefficients $\boldsymbol{\theta}$ can then be found by minimizing (13) with respect to $\boldsymbol{\theta}$. The estimation error produced is made of two terms, a *bias error* $E_\beta(\omega)$ and a *variance error* $E_\nu(\omega)$, and it can be written (in the frequency domain, where $\omega \equiv e^{j\omega}$) as

$$\begin{aligned}E(\omega, \boldsymbol{p}, \hat{\boldsymbol{\theta}}_N) &= H(\omega) - G(\omega, \boldsymbol{p}, \hat{\boldsymbol{\theta}}_N) \\ &= [H(\omega) - G(\omega, \boldsymbol{p}, \boldsymbol{\theta}_o)] + [G(\omega, \boldsymbol{p}, \boldsymbol{\theta}_o) - G(\omega, \boldsymbol{p}, \hat{\boldsymbol{\theta}}_N)] \\ &= E_\beta(\omega, \boldsymbol{p}, \boldsymbol{\theta}_o) + E_\nu(\omega, \boldsymbol{p}, \boldsymbol{\theta}_o, \hat{\boldsymbol{\theta}}_N), \quad (14)\end{aligned}$$

with $\boldsymbol{\theta}_o = \lim_{N\to\infty}\hat{\boldsymbol{\theta}}_N$ the Wiener solution ($\boldsymbol{\theta}_o = \boldsymbol{R}_{\boldsymbol{\kappa}}^{-1}\boldsymbol{r}$, where $\boldsymbol{R}_{\boldsymbol{\kappa}} = \mathbb{E}\{\boldsymbol{\kappa}(n,\boldsymbol{p})\boldsymbol{\kappa}^T(n,\boldsymbol{p})\}$ is the autocorrelation matrix of the intermediate signals and $\boldsymbol{r} = \mathbb{E}\{\boldsymbol{\kappa}(n,\boldsymbol{p})y(n)\}$ is the cross-correlation vector between the intermediate signals and the microphone signal, assuming $u(n)$ and $y(n)$ being jointly wide-sense stationary stochastic processes, both with zero mean). This means that the bias error depends on the model structure chosen and on its order (i.e. the dimension of $\boldsymbol{p}$ and $\boldsymbol{\theta}$), whereas the variance error depends on the actual estimation of the parameters, and is thus influenced by the characteristics of the input and noise signals. It can be seen from (14) that both terms also depend on the pole set $\boldsymbol{p}$, which means that not only the number of poles, but also their position, affects the estimation error.

It is quite intuitive to expect a decrease in the bias error when the poles are moved away from the origin of the z-plane closer to the true poles of the system. A result valid for a constant $S_u(\omega) = \sigma_u^2$, but extendable to other cases [17], formalizes this idea showing that the bias error tends to zero for $M \to \infty$ (more specifically, it decreases geometrically in the model order M) and that it is proportional to the distance between the K true poles $\xi_\varkappa$ of the system ($\varkappa = 1,\ldots,K$) and the M poles p_i of the model structure ($i = 1,\ldots,M$), according to

$$E_\beta(\omega,\boldsymbol{p},\boldsymbol{\theta}_o) \leq \sum_{\varkappa=1}^{K}\left|\frac{\aleph_\varkappa}{e^{j\omega}-\xi_\varkappa}\right|\prod_{i=1}^{M}\left|\frac{\xi_\varkappa - p_i}{1-p_i^*\xi_\varkappa}\right|, \tag{15}$$

with $\aleph_\varkappa$ the residues of the partial fraction expansion of $H(z)$ (see [17]).

The dependency of the asymptotic variance error on the poles is instead less intuitive. It turns out [18] that the influence of the pole location is quantified by the frequency-dependent term $\gamma_M(\omega,\boldsymbol{p})$ in (11) in such a way that the variance can be approximated as

$$\mathbb{E}\{|E_\nu(\omega,\boldsymbol{p},\boldsymbol{\theta}_o,\hat{\boldsymbol{\theta}}_N)|^2\} \approx \frac{1}{N}\frac{\sigma_v^2}{S_u(\omega)}\gamma_M(\omega,\boldsymbol{p}), \tag{16}$$

which is a generalization of the FIR case, for which $\gamma_M(\omega,\boldsymbol{p}) = M$, $\forall\omega$ [36]. The implications are that increasing the number of poles in order to reduce the bias error in a certain frequency region would also increase the sensitivity to noise in that same region. As a consequence, a good estimate of $H(z)$ would have as few poles, as close as possible to the actual poles of the RTF.

B. Adaptation of the linear coefficients

In most RASE applications, an estimate of a RTF has to be obtained adaptively, as new samples of the source and microphone signals are available. The adaptation rule for the recursive estimation of the linear filter parameter vector $\boldsymbol{\theta}(n)$ is given by

$$\hat{\boldsymbol{\theta}}(n+1) = \hat{\boldsymbol{\theta}}(n) + \boldsymbol{g}(n,\boldsymbol{p})\varepsilon(n,\boldsymbol{p},\hat{\boldsymbol{\theta}}), \tag{17}$$

with $\boldsymbol{g}(n,\boldsymbol{p})$ a gain vector. When considering the linear $\boldsymbol{\theta}$ parameters, OBF models are linear regression models, as can be seen in (5) (the regression vectors are independent from the previous estimates of the parameters). Thus, it is possible to apply standard adaptive algorithms developed for FIR filters [37], with the only difference that the regression vector for an OBF adaptive filter is represented by the vector of intermediate signals $\boldsymbol{\kappa}(n,\boldsymbol{p})$, instead of $\boldsymbol{u}(n) = [u(n), u(n-1),\ldots,u(n-M+1)]$ (the last M samples of the input signal $u(n)$). The increase in complexity is only given by the filtering of the input signal to compute $\boldsymbol{\kappa}(n,\boldsymbol{p})$, and not in the adaptation scheme itself.

Based on how the gain vector $\boldsymbol{g}(n,\boldsymbol{p})$ is computed, different algorithms are obtained: the least mean squares (LMS) algorithm is obtained for

$$\boldsymbol{g}(n,\boldsymbol{p}) = \mu\boldsymbol{\kappa}(n,\boldsymbol{p}), \tag{18}$$

with μ the step size. Normalization is usually necessary to avoid that large values in $\boldsymbol{\kappa}(n,\boldsymbol{p})$ would lead to large variations in $\hat{\boldsymbol{\theta}}(n+1)$. The vector of intermediate signals $\boldsymbol{\kappa}(n,\boldsymbol{p})$ can hence be normalized, leading to the NLMS gain vector

$$\boldsymbol{g}(n,\boldsymbol{p}) = \frac{\tilde{\mu}}{\delta + \|\boldsymbol{\kappa}(n,\boldsymbol{p})\|^2}\boldsymbol{\kappa}(n,\boldsymbol{p}), \tag{19}$$

where δ is a small regularization term to avoid instability or convergence problems resulting from poor excitation [38], [39]. Another common, yet more complex, adaptation algorithm is the recursive least squares (RLS) algorithm [16], [17], for which the gain vector is given by

$$\begin{aligned}&\boldsymbol{g}(n,\boldsymbol{p}) = \boldsymbol{\Upsilon}_{\boldsymbol{\kappa}}(n)\boldsymbol{\kappa}(n,\boldsymbol{p}),\ \text{with}\\ &\boldsymbol{\Upsilon}_{\boldsymbol{\kappa}}(n+1) = \frac{1}{\lambda}\left\{\boldsymbol{\Upsilon}_{\boldsymbol{\kappa}}(n) - \frac{\boldsymbol{\Upsilon}_{\boldsymbol{\kappa}}(n)\boldsymbol{\kappa}(n,\boldsymbol{p})\boldsymbol{\kappa}^T(n,\boldsymbol{p})\boldsymbol{\Upsilon}_{\boldsymbol{\kappa}}(n)}{\lambda + \boldsymbol{\kappa}^T(n,\boldsymbol{p})\boldsymbol{\Upsilon}_{\boldsymbol{\kappa}}(n)\boldsymbol{\kappa}(n,\boldsymbol{p})}\right\},\end{aligned} \tag{20}$$

and $\lambda = 1-\mu$ a forgetting factor and $\boldsymbol{\Upsilon}_{\boldsymbol{\kappa}}(n)$ an estimate of the inverse of the autocorrelation matrix $\boldsymbol{R}_{\boldsymbol{\kappa}} = \mathbb{E}\{\boldsymbol{\kappa}(n,\boldsymbol{p})\boldsymbol{\kappa}^T(n,\boldsymbol{p})\}$ of the intermediate signals. A trade-off between convergence and complexity is obtained with the affine projection algorithm (APA) [40], for which the update rule in (17) becomes

$$\hat{\boldsymbol{\theta}}(n+1) = \hat{\boldsymbol{\theta}}(n) + \mu\boldsymbol{K}_Q(n)\big(\boldsymbol{I}_Q + \boldsymbol{K}_Q^T(n)\boldsymbol{K}_Q(n)\big)^{-1}\boldsymbol{\varepsilon}_Q(n) \tag{21}$$

where $\boldsymbol{I}_Q$ is the $Q\times Q$ identity matrix used for regularization, $\boldsymbol{K}_Q(n) = [\boldsymbol{\kappa}(n,\boldsymbol{p}),\boldsymbol{\kappa}(n-1,\boldsymbol{p}),\ldots,\boldsymbol{\kappa}(n-Q+1,\boldsymbol{p})]$ of size $M\times Q$, and $\boldsymbol{\varepsilon}_Q(n) = [\varepsilon(n,\boldsymbol{p},\boldsymbol{\theta}),\varepsilon(n-1,\boldsymbol{p},\boldsymbol{\theta}),\ldots,\varepsilon(n-Q+1,\boldsymbol{p},\boldsymbol{\theta})]$, with Q the so-called projection order. Notice that for $Q = 1$, the APA corresponds to the NLMS algorithm. Most algorithms developed for FIR adaptive filters can be derived easily for OBF filters as well. For instance, variable step size (VSS) algorithms [41]–[43] or regularized algorithms [38] can be used for highly non-stationary signals, or the Kalman filter [17] for time-varying systems.

1) Dynamic behavior: transient and steady-state errors: As mentioned earlier, orthogonality offers the possibility of studying the error behavior and transient performance of OBF filters in adaptive algorithms and the relation to step size, noise power and input PSD. The following results have been derived in [17] for $M \to \infty$, but proved (empirically) to be valid also for small model orders.

The steady-state error (after convergence, i.e. for $n \to \infty$) is similar to the expression of the variance in (16), with the introduction of the step size μ,

$$\mathbb{E}\{|E(\omega, \boldsymbol{p}, \hat{\boldsymbol{\theta}}(\infty))|^2\} \approx \frac{\mu \sigma_v^2}{[S_u(\omega)]^r} \gamma_M(\omega, \boldsymbol{p}), \qquad (22)$$

where $r = 0$ for LMS and $r = 1$ for RLS (with $\mu = 1 - \lambda$). This expression shows that the LMS algorithm depends on the choice of the step size and on the noise variance, but it is invariant to the input PSD $S_u(\omega)$, whereas for RLS the steady-state estimation error is inversely proportional to $S_u(\omega)$; for NLMS, expression (22) with $r = 0$ is still valid, if the normalization in (19) is considered to be included in the step size μ, which depends on $S_u(\omega)$ according to (10). As for the variance, also the steady-state error in (22) depends on $\gamma_M(\omega, \boldsymbol{p})$, i.e. it increases for a larger number of poles and for larger poles densities, and it is equal to the FIR case for $\gamma_M(\omega, \boldsymbol{p}) = M$ [36].

The transient error for the LMS algorithm, i.e. the estimation error at iteration $n+1$ with respect to the error at iteration n, can be approximated as

$$\begin{aligned}\mathbb{E}\{|E(\omega, \boldsymbol{p}, \hat{\boldsymbol{\theta}}(n+1))|^2\} \approx & [1 - \mu S_u(\omega)]^2 \mathbb{E}\{|E(\omega, \boldsymbol{p}, \hat{\boldsymbol{\theta}}(n))|^2\} \\ & + \mu^2 \sigma_v^2 S_u(\omega) \gamma_M(\omega, \boldsymbol{p}), \qquad (23)\end{aligned}$$

which shows the error dependency on two terms: when $\mu S_u(\omega)$ is large, the error is reduced in the first term, but it increases based on the second term. The presence of the frequency-dependent term $\gamma_M(\omega, \boldsymbol{p})$, i.e. the number and location of the poles, affects the second term of (23), which also depends on the noise variance. Once again, the expression in (23) is a generalization of the FIR case [36].

2) Convergence rate: step size and numerical conditioning: The orthogonality property of OBF filters ensures better-behaved and faster convergence of the adaptation algorithm, compared to non-orthogonal fixed-pole adaptive filters [17]. For the LMS algorithm, the convergence speed is determined by the choice of the step size μ and by the condition number C of the intermediate signals correlation matrix $\boldsymbol{R}_{\boldsymbol{\kappa}}$, defined as the spread of its eigenvalues as $C(\boldsymbol{R}_{\boldsymbol{\kappa}}) = {}^{\lambda_{\max}}/_{\lambda_{\min}}$, with $\lambda_{\max}$ and $\lambda_{\min}$ the maximum and minimum eigenvalues, respectively [44]. Convergence is controlled by the exponential factor $(1 - \mu\lambda_{\min})^n$, with a larger value for $\lambda_{\min}$ yielding faster convergence [37], which decays to zero for $\mu < {}^1/_{\lambda_{\max}}$. It follows that the convergence rate in the mean for the LMS algorithm can be no faster than [17]

$$\left(1 - \frac{\lambda_{\min}}{\lambda_{\max}}\right)^n = \left(1 - \frac{1}{C(\boldsymbol{R}_{\boldsymbol{\kappa}})}\right)^n. \qquad (24)$$

For OBF filters with a WN input signal, i.e. with constant PSD $S_u(\omega) = \sigma_u^2$, the convergence rate is optimal as $C(\boldsymbol{R}_{\boldsymbol{\kappa}}) \approx 1$ ($\boldsymbol{R}_{\boldsymbol{\kappa}} \approx \sigma_u^2 \boldsymbol{I}$, with $\boldsymbol{I}$ the identity matrix). For colored input signals, the optimal conditioning of the correlation matrix is lost. However, by virtue of orthogonality, a bound on $C(\boldsymbol{R}_{\boldsymbol{\kappa}})$ in relation to the input PSD is derived as

$$\min_{\omega \in [-\pi, \pi]} S_u(\omega) \leq \lambda(\boldsymbol{R}_{\boldsymbol{\kappa}}) \leq \max_{\omega \in [-\pi, \pi]} S_u(\omega), \qquad (25)$$

with $\lambda(\boldsymbol{R}_{\boldsymbol{\kappa}})$ the set of eigenvalues of $\boldsymbol{R}_{\boldsymbol{\kappa}}$, so that an upper bound on the average convergence rate can be found as [17]

$$\left(1 - \frac{\min\limits_{\omega \in [-\pi, \pi]} S_u(\omega)}{\max\limits_{\omega \in [-\pi, \pi]} S_u(\omega)}\right)^n. \qquad (26)$$

This means that OBF adaptive filters are particularly robust in terms of numerical well-conditioning for a wide range of input PSD [19], so that the condition number is expected to be smaller compared to the case in which a non-orthogonal structure is used, and thus the convergence rate faster.

C. The OBF-NLMS and its analogy to TD-NLMS

As mentioned above, the NLMS algorithm is meant to avoid that large values in the regression vector would result in large variations in the parameter update. Whereas in FIR adaptive filters, this is accomplished by normalizing with respect to the power of the previous M samples of the input signal (or, by considering $\tilde{\mu} = \mu M$, with the respect to the mean value of $\|\boldsymbol{u}(n)\|^2$) [39], the normalization in the OBF filter case has a different interpretation. Indeed, the NLMS update in (19) normalizes the regression vector $\boldsymbol{\kappa}(n, \boldsymbol{p})$ with respect to the instantaneous power of all the intermediate signals at time n, or equivalently, by considering $\tilde{\mu} = \mu M$, by their mean power. This means that the update rule has no memory of previous samples of $\boldsymbol{\kappa}(n, \boldsymbol{p})$, so that large values of the input signal may result in large variations of the linear coefficients. It is possible, indeed, especially when the filter order M is very small, that the power σ_ι^2 at time n of different intermediate signals $\kappa_\iota(n)$ is similar, and so is their mean power.

For this reason an alternative version of the NLMS algorithm (named OBF-NLMS) is introduced here, which normalizes each intermediate signal $\kappa_\iota(n)$ individually based on an estimate of its power σ_ι^2. The gain vector $\boldsymbol{g}(n)$ for the OBF-NLMS is then

$$\boldsymbol{g}(n) = \mu[\delta \boldsymbol{I}_M + \hat{\boldsymbol{\Sigma}}_M(n, \boldsymbol{p})]^{-1} \boldsymbol{\kappa}(n, \boldsymbol{p}), \qquad (27)$$

with $\hat{\boldsymbol{\Sigma}}_M(n, \boldsymbol{p})$ an $M \times M$ diagonal matrix, whose ι^{th} diagonal element $\hat{\sigma}_\iota^2$ is an estimate of the the intermediate signal power σ_ι^2. The power estimates are computed using an exponential window update, implemented as a one-pole filter with pole $0 \ll \beta < 1$ (i.e. the forgetting factor) as

$$\hat{\sigma}_\iota^2(n) = \beta \hat{\sigma}_\iota^2(n-1) + (1 - \beta)|\kappa_\iota(n)|^2 \qquad (28)$$

Each linear coefficient is then updated individually as

$$\hat{\theta}_\iota(n+1) = \hat{\theta}_\iota(n) + \frac{\mu}{\delta + \hat{\sigma}_\iota^2(n)} \kappa_\iota(\tilde{\boldsymbol{p}}_i, n) \varepsilon(n, \boldsymbol{p}, \boldsymbol{\theta}). \qquad (29)$$

The only disadvantages compared to the standard NLMS are a small increase in complexity due to (28) and the requirement of a reasonable initial estimate $\hat{\sigma}_\iota^2(0)$ at the beginning of the adaptation, in order to avoid slow convergence of the power estimates. This second issue is easily overcome by computing a short-term mean on the first few samples of $|\kappa_\iota(n)|^2$ before starting to adapt the filter coefficients.

The necessity of introducing the OBF-NLMS algorithm for low model orders M is illustrated in Figure 3, showing the

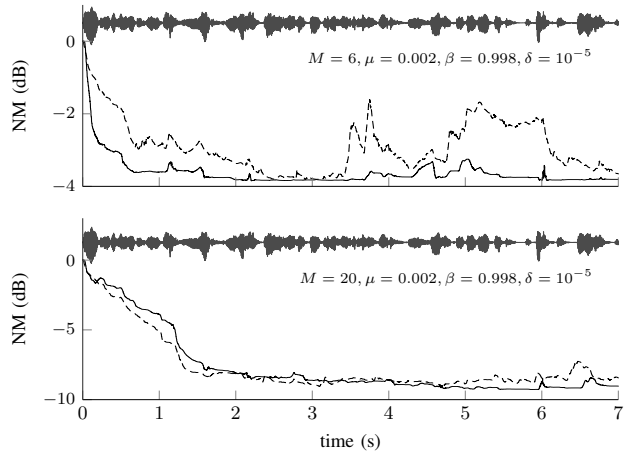

Fig. 3: The comparison between the NM of NLMS (dashed) and OBF-NLMS (solid) for $M = 6$ (top) and $M = 20$ (bottom).

identification of a low-frequency RIR simulated using the randomized image-source method (RIM) [45] (reverberation time (RT) $T_{60} = 0.25$ s, room M in Table I) with male speech input signal downsampled to $f_s = 800$ Hz. The curves represent the normalized misalignment (NM) obtained using the standard NLMS and the OBF-NLMS algorithm, which is defined as

$$\mathrm{NM}(n) = 10\log_{10}\left(\frac{\|\boldsymbol{h} - \hat{\boldsymbol{h}}(n, \boldsymbol{p}, \hat{\boldsymbol{\theta}})\|_2^2}{\|\boldsymbol{h}\|_2^2}\right) \text{ dB}, \tag{30}$$

where $\boldsymbol{h}$ is the true RIR vector of length N samples, and $\hat{\boldsymbol{h}}(n, \boldsymbol{p}, \hat{\boldsymbol{\theta}}) = \boldsymbol{\Psi}_N(\boldsymbol{p})\hat{\boldsymbol{\theta}}(n)$ is the estimated RIR at time n, with the columns of the $N \times M$ matrix $\boldsymbol{\Psi}_N(\boldsymbol{p})$ being the N-samples OBF responses to an impulsive input signal (see [12]). It can be seen in the top plot that large values of the input signal results in large variations of the steady-state error for the NLMS algorithm when the model order is small (here $M = 6$). The OBF-NLMS algorithm, on the other hand, provides faster convergence and low parameter variability in the steady-state. For higher model orders, instead, the NLMS adaptation rule in (19) becomes effective in dealing with large values of the input signal, with comparable performance at a lower complexity, as can be seen in the bottom plot for $M = 20$.

An analogy between the OBF-NLMS algorithm and the time-domain implementation of TD adaptive algorithms [29]–[33] is noticed. TD algorithms apply an orthonormal transform, such as the discrete Fourier transform (DFT) or the discrete cosine transform (DCT) to name the most common, in order to partially decorrelate the samples of the input signal vector $\boldsymbol{u}(n)$ and thus accelerate the rate of convergence of the FIR filter parameters when using the NLMS algorithm. It has been shown [30] that the convergence actually improves when, instead of normalizing all the intermediate signals (i.e. the transformed input signals) with respect to the instantaneous power of $\boldsymbol{u}(n)$ as in the standard NLMS algorithm, the normalization is performed at each intermediate signal with respect to the inverse of a short-time average of its power, as in (29). It is actually this normalization and not the transformation itself that reduces the eigenvalue spread and speeds up convergence.

Another, more conceptual, analogy between OBF adaptive filters and TD algorithms is in their interpretation as filter-banks [31], [33]: the orthonormal transforms, such as the DFT or the DCT, can indeed be seen as a parallel of band-pass filters with a uniform distribution of their central frequency. The decorrelation performance of different transforms depends mainly on the characteristics of the input PSD [31]. For instance, the DCT is particularly suitable for input signals with a low-pass characteristic. Also OBF adaptive filters can be seen as a parallel of band-pass filters which partially decorrelate the intermediate signals, with the difference that the filter central frequencies and bandwidth (i.e. the poles) are chosen with respect to the system to be identified and not to the expected input PSD, even though both the pole location and the input PSD influence the tracking behavior of the algorithm, as explained in the previous section.

D. Adaptation of the poles

Gradient-based algorithms can also be used as well in the adaptation of the denominator coefficients of the TF of IIR filters [6]. However, difficulties arise from the fact that a pole-zero model is a nonlinear regression model, in the sense that the regressors depend on previous values of the filter coefficients (at least in the output-error approach) [6], [36], [46]. This issue is normally circumvented in direct-form pole-zero models by disregarding this dependency and treating the problem as in linear regression (in which case they are called *pseudolinear* regression models). This is not a possibility for OBF adaptive filters, the reason being that each pole pair $\boldsymbol{p}_i$ appears in the all-pass sequence of the successive $m-i$ TFs (4) and in the normalization filters as well, so that they cannot be regarded as pseudolinear regression models.

When minimizing the sum of squared errors in (13), recursive prediction error algorithms [46], or Gauss-Newton-type recursive algorithms, should be then used to try to adjust the filter coefficients. For instance, an algorithm was proposed in [25], in which linear coefficients and poles are updated in an alternating fashion using RLS and a recursive Gauss-Newton algorithm with a backtracking strategy for determining the optimal step-size, respectively. The idea of recursive nonlinear identification algorithms is to update the pole parameters $\boldsymbol{p}(n)$ along the search direction $\boldsymbol{q}(n)$ according to

$$\boldsymbol{p}(n+1) = \boldsymbol{p}(n) + \mu\boldsymbol{q}(n), \tag{31}$$

where $\boldsymbol{q}(n)$ for the quadratic cost function in (13) has the form

$$\begin{aligned}\boldsymbol{q}(n) &= 2\boldsymbol{B}_{\boldsymbol{p}}^{-1}(n)\nabla\varepsilon_{\boldsymbol{p}}(n)\varepsilon(n, \boldsymbol{p}, \boldsymbol{\theta}), \text{ with} \\ \nabla\varepsilon_{\boldsymbol{p}}(n) &= {}^{\partial\varepsilon(n,\boldsymbol{p},\boldsymbol{\theta})}/_{\partial\boldsymbol{p}(n)} = -{}^{\partial\tilde{y}(n,\boldsymbol{p},\boldsymbol{\theta})}/_{\partial\boldsymbol{p}(n)}.\end{aligned} \tag{32}$$

Different algorithms differ on how $\boldsymbol{B}_{\boldsymbol{p}}(n)$ is chosen. For instance, the steepest descent algorithm chooses $\boldsymbol{B}_{\boldsymbol{p}}(n) = \boldsymbol{I}$, while in the Gauss-Newton algorithm $\boldsymbol{B}_{\boldsymbol{p}}(n) = \nabla\varepsilon_{\boldsymbol{p}}(n)\nabla\varepsilon_{\boldsymbol{p}}^H(n)$ is an approximation of the Hessian matrix [47] ($\{\cdot\}^H$ indicating the Hermitian transpose). Common to all these algorithms is the computation of the gradient vector $\nabla\varepsilon_{\boldsymbol{p}}(n)$, i.e. the derivatives of the error with respect to the pole parameters. The expressions for the gradient with respect to

the kth pole pair $\boldsymbol{p}_k$ are obtained by computing (and leaving out the dependency of $\tilde{y}(n)$ and $\tilde{y}_i(n)$ on $\boldsymbol{p}$ and $\boldsymbol{\theta}$ for brevity)

$$\frac{\partial \tilde{y}(n)}{\partial \boldsymbol{p}_k(n)} = \sum_{i=1}^{m} \frac{\partial \tilde{y}_i(n)}{\partial \boldsymbol{p}_k(n)}, \tag{33}$$

which can be divided in three parts, as

$$\frac{\partial \tilde{y}(n)}{\partial \boldsymbol{p}_k(n)} = \sum_{i=1}^{k-1} \frac{\partial \tilde{y}_i(n)}{\partial \boldsymbol{p}_k(n)} + \frac{\partial \tilde{y}_k(n)}{\partial \boldsymbol{p}_k(n)} + \sum_{i=k+1}^{m} \frac{\partial \tilde{y}_i(n)}{\partial \boldsymbol{p}_k(n)}. \tag{34}$$

The first term is zero, since $\tilde{y}_i(n)$ for $i = 1, \ldots, k-1$ is independent from $\boldsymbol{p}_k$. Even though analytic expressions for the other two terms can be derived, the nonlinear dependency of the filter output on the pole parameters makes the expressions involved and expensive to compute. Also, the pole parameters being complex-valued, a parametrization of the poles is necessary in order to obtain real-valued expressions for the gradients. Although it would be natural to parametrize the poles in terms of their radius ρ_i and angle ϑ_i, the computation of the gradients becomes slightly simpler by considering the parameters $\zeta_i = -(p_i + p_i^*) = -2\rho_i \cos\vartheta_i$ and $\eta_i = p_i p_i^* = \rho_i^2$, for which $D_i(z)$ in (1) becomes $D_i(z) = 1 + \zeta_i z^{-1} + \eta_i z^{-2}$ (in [25] the real and imaginary part of the pole parameters were used instead). The expressions for the gradients with respect to ζ_i and η_i were derived in [28] for a slightly different realization of OBF filters and for orthogonal filters.

The full and approximated expressions for the gradients in the realization used in this paper are given in Appendix A. It can be seen that the expressions are quite complicated, especially for the third term in (34). A computationally reasonable, but suboptimal way of adapting the pole parameters is by approximating the gradients (with χ_k either ζ_k or η_k) as

$$\frac{\partial \tilde{y}(n)}{\partial \chi_k(n)} \approx \frac{\partial \tilde{y}_k(n)}{\partial \chi_k(n)}\,, \tag{35}$$

which assumes slow convergence of the parameters and poles close to the unit circle (see [28]). A further simplification is obtained by renouncing to the normality property of OBF filters by redefining $N_i^{\pm}(z) = (z^{-1} \mp 1)$, in which case the gradient becomes

$$\frac{\partial \tilde{y}(n)}{\partial \chi_k(n)} \approx \frac{\partial \tilde{y}_k(n)}{\partial \chi_k(n)} = -\frac{1}{D_k}\frac{\partial D_k}{\partial \chi_k}\tilde{y}_k(n)\,, \tag{36}$$

with $\partial D_k / \partial \zeta_k = z^{-1}$ and $\partial D_k / \partial \eta_k = z^{-2}$. This simplification of the OBF model to an orthogonal model would also allow to regard it as a pseudolinear model, with the regressors defined by the input signal $u(n)$ and the various outputs of resonators and all-pass filters (signals $a_i(n)$, $x_i(n)$ and their previous samples), where the dependency of the regressors on previous values of the parameters to be estimated is ignored. In the following, these ideas for the adaptation of the poles are not developed further. Instead, an iterative identification algorithm is proposed, which estimates the poles of one or multiple RTFs avoiding the nonlinear problem by employing a grid-based matching pursuit approach.

III. The SB-OBF-GMP identification algorithm

It was shown in the previous section that the identification of the linear coefficients of an OBF adaptive filter is governed by the same conditions and with the same implementation complexity as for the identification with a FIR filter (with differences in frequency resolution and noise sensitivity based on $\gamma_M(\omega, \boldsymbol{p})$ defined in (11)). The identification of the poles, on the other hand, is a nonlinear problem and, even though it is possible to devise recursive algorithms as discussed above, problems such as slow convergence or convergence to local minima may occur. Slow convergence is also related to the number of poles and the choice of the initial values in recursive algorithms, which should be based on some prior (usually unavailable) knowledge about the system. Another issue is the computational complexity of the recursive algorithm, especially if a backtracking strategy for the selection of the step-size in (31) is used to speed up convergence [47].

Here, a different approach to the identification of the poles is taken. Instead of adapting the pole parameters, the inherent nonlinear problem is avoided by using a grid search and by selecting poles one by one in an order-recursive fashion. The iterative algorithm proposed here is similar to the BB-OBF-GMP algorithm in [34], in which a dictionary is built by collecting N_b samples of candidate intermediate signals, with poles defined on a grid spanning a portion of the unit disc. In each block b, one pair of complex-conjugate poles $\boldsymbol{p}_b$ is selected from the grid as the one that produces the pair of intermediate signals that is mostly correlated, on average, with the last N_b samples of the prediction error signals $\varepsilon_r(n)$ produced in each acoustic channel r considered ($r = 1, \ldots, R$) using LS estimation. The pole-pair $\boldsymbol{p}_b$ is then added to the previously selected common poles in the multi-channel OBF adaptive filter, whose number m of resonator sections increases by one ($m \leftarrow m + 1$), whereas the linear coefficients for each acoustic channel are adapted with respect to each $\varepsilon_r(n)$ using the LMS update rule in (17) with gain vector as in (18).

The BB-OBF-GMP algorithm proved to be capable of accurately identifying a common set of poles from WN input signals in a single-input/multiple-outputs (SIMO) room acoustic system. The LMS algorithm, however, works well as long as the input PSD is constant and the step size is correctly chosen according to it. As a consequence, when the input signal is non-stationary and non-white, the step size may be too small or too large, and the algorithm would either converge very slowly or become erratic. Also, when the linear coefficients adapt too slowly and the size of the block is not sufficient, the prediction error signals $\varepsilon_r(n)$ compared to which the correlation with the candidate intermediate signals in the dictionary is computed, may not have reached the steady-state, resulting in a poor identification performance.

The SB-OBF-GMP algorithm proposed here introduces a series of modifications meant to deal with these issues. First, instead of collecting N_b samples of both the candidate intermediate signals and the residual signals, the correlation between them is tracked in time for a given number of samples N_s (with $N_s < N_b$ required), before one new pole pair is selected. The

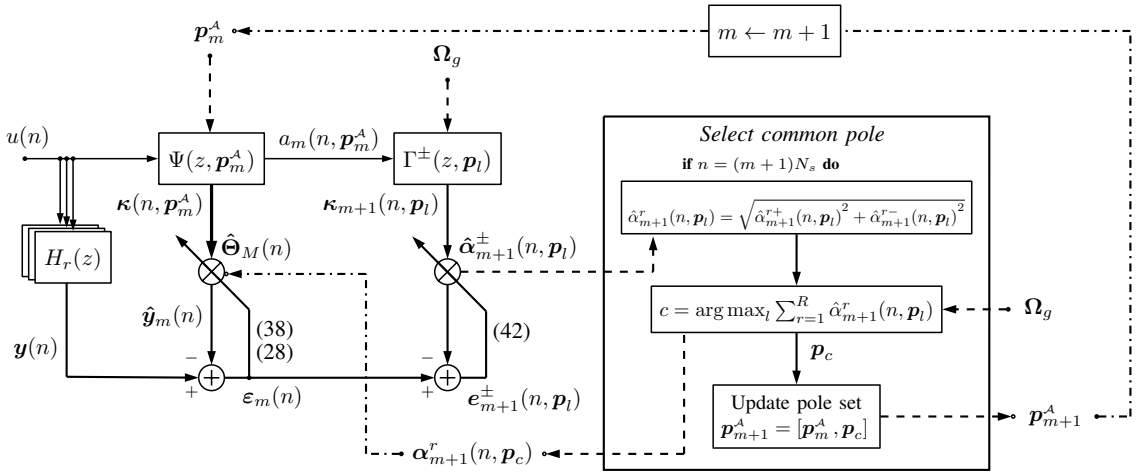

Fig. 4: The simplified schematics of the SB-OBF-GMP algorithm.

NLMS algorithm is used, as it is equivalent to an adaptation of the correlation between the error signals and the candidate intermediate signals using an exponential window update, as it will be explained later on. Another modification pertains to the introduction of the OBF-NLMS adaptation rule (27) in the multi-channel OBF adaptive filter. The reason is to deal with large values of the intermediate signals when the model order $M = 2m$ is small, i.e. when the multi-channel OBF adaptive filter has a small number m of resonator sections, as explained in Section II-C. For higher model orders, the performance of OBF-NLMS and NLMS are similar, with the latter being less expensive computationally, so that when a sufficiently large number of poles m has been included in the multi-channel OBF adaptive filter, it is possible to employ the NLMS update with gain vector as in (19).

The idea of the proposed algorithm is indeed that, since the poles are considered to be a characteristic of the room itself and thus approximately time-invariant [48], the identification of a common set of poles can be performed once for a given setup or environment at the beginning of the session of a RASE task, and then kept fixed afterwards, with RTF variations tracked by the adaptive linear coefficients only. The algorithm is designed for SIMO room acoustic systems, but it can be extended to the multiple-inputs/multiple-outputs (MIMO) case, as suggested in Section IV. It could be also used to find good initial values and to determine the order M, as required by nonlinear recursive algorithms.

A. Algorithm description

The proposed algorithm aims to build a SIMO OBF adaptive filter including one common pole pair at each stage. A *stage* is defined as the period at the end of which a pole pair $\boldsymbol{p}_c$ is selected and included in the multi-channel OBF adaptive filter. A new pole pair $\boldsymbol{p}_c$ is selected based on the correlation coefficients, which values are tracked using the NLMS algorithm for the duration of a stage, between the prediction error signals $\varepsilon_r(n)$ and the candidate intermediate signals. The pole pair in the grid associated with the pair of candidate intermediate signals with the highest correlation, averaged over the R acoustic channels and evaluated at the end of the stage, is selected and added to the active pole set $\boldsymbol{p}_{m+1}^{\mathcal{A}} = [\boldsymbol{p}_m^{\mathcal{A}}, \boldsymbol{p}_c]$ of the multi-channel OBF adaptive filter, thus adding a new resonator section. At the beginning of each new stage ($m \leftarrow m+1$), the linear coefficients $\hat{\boldsymbol{\theta}}_m^r(n) = [\theta_m^{r+}(n), \theta_m^{r-}(n)]$ associated with the intermediate signals built from the pole pair $\boldsymbol{p}_c$ selected at the previous stage, start to be adapted for each channel r, together with $\hat{\boldsymbol{\theta}}_i^r(n) = [\theta_i^{r+}(n), \theta_i^{r-}(n)]$ with $i = 1, \ldots, m-1$. In other words, the size of each rth set of linear coefficients increases by two at the end of each stage, and their value adapted using either the OBF-NLMS or the NLMS update. The algorithm stops whenever the number of selected poles m reaches a maximum value $m_{\max}$ or some stopping criterion based on the average power of the prediction error signals is satisfied. Readers uninterested in the details of the algorithm may skip the following description and resume the reading from Section III-B or even from Section IV, without compromising their understanding of the rest of the paper.

The aim of the SB-OBF-GMP algorithm is to minimize the sum of the instantaneous squared errors, over the R channels,

$$\underset{\boldsymbol{p}_m^{\mathcal{A}}, \hat{\boldsymbol{\Theta}}_M(n)}{\text{minimize}} \ \|\boldsymbol{\varepsilon}_m(n)\|^2 = \left\|\boldsymbol{y}(n) - \hat{\boldsymbol{y}}_m(n, \boldsymbol{p}_m^{\mathcal{A}}, \hat{\boldsymbol{\Theta}}_M)\right\|^2 = \left\|\boldsymbol{y}(n) - \boldsymbol{\kappa}(n, \boldsymbol{p}_m^{\mathcal{A}})^T \hat{\boldsymbol{\Theta}}_M(n)\right\|_{,}^2 \quad (37)$$

where $\boldsymbol{\varepsilon}_m(n) = [\varepsilon_m^1(n), \ldots, \varepsilon_m^R(n)]$ represents the vector of prediction error signals $\varepsilon_m^r(n)$ (with $r = 1, \ldots, R$) for the R acoustic channels at time n, $\boldsymbol{y}(n) = [y^1(n), \ldots, y^R(n)]$ is the vector of output signals and $\hat{\boldsymbol{y}}_m(n, \boldsymbol{p}_m^{\mathcal{A}}, \hat{\boldsymbol{\Theta}}_M) = [\hat{y}_m^1(n, \boldsymbol{p}_m^{\mathcal{A}}, \hat{\boldsymbol{\theta}}^1(n)), \ldots, \hat{y}_m^R(n, \boldsymbol{p}_m^{\mathcal{A}}, \hat{\boldsymbol{\theta}}^R(n))]$ the vector of estimated outputs of the multi-channel OBF adaptive filter, with the linear filter coefficient vectors defined as $\hat{\boldsymbol{\theta}}^r(n) = [\hat{\boldsymbol{\theta}}_1^r(n), \ldots, \hat{\boldsymbol{\theta}}_m^r(n)]^T$ and $\hat{\boldsymbol{\theta}}_i^r(n) = [\hat{\theta}_i^{r+}(n), \hat{\theta}_i^{r-}(n)]$. The symbol $\{\hat{\cdot}\}$ is used instead of $\{\tilde{\cdot}\}$ to indicate the fact that $\hat{\boldsymbol{y}}_m(n, \boldsymbol{p}_m^{\mathcal{A}}, \hat{\boldsymbol{\Theta}}_M)$ is an estimate. The vector of estimated output signals $\hat{\boldsymbol{y}}_m(n, \boldsymbol{p}_m^{\mathcal{A}}, \hat{\boldsymbol{\Theta}}_M)$ is obtained by the linear combination of the $M = 2m$ intermediate signals $\boldsymbol{\kappa}(n, \boldsymbol{p}_m^{\mathcal{A}}) = [\boldsymbol{\kappa}_1(n), \ldots, \boldsymbol{\kappa}_m(n)]^T$ weighted by the linear filter coefficient

Algorithm 1 SB-OBF-GMP algorithm

$\boldsymbol{\Omega}_g = \{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_L\}$ ▷ Define pole grid
$\boldsymbol{p}_0^{\mathcal{A}} = \emptyset\,, m = 0$ ▷ Initialize the set of active poles
$\boldsymbol{\varepsilon}_0(n) = \boldsymbol{y}(n),\ a_0(n) = u(n-d)$ ▷ Initialize signals
$\hat{\boldsymbol{\alpha}}_l^{\pm}(0) = \mathbf{0},\ \ l = 1, \ldots, L$ ▷ Set correlation coefficients
while $m < m_{\text{max}}$ **do** ▷ m_{max}: max number of pole pairs
Update multi-channel OBF adaptive filter
  if $m > 0$ **then**
    $\boldsymbol{\varepsilon}_m(n) = \boldsymbol{y}(n) - \boldsymbol{\kappa}(n, \boldsymbol{p}_m^{\mathcal{A}})^T \hat{\boldsymbol{\Theta}}_M(n)$ ▷ (37)
    $\hat{\sigma}_\iota^2(n) = \beta\hat{\sigma}_\iota^2(n-1) + (1-\beta)|\kappa_\iota(n)|^2$ ▷ (28)
    $\hat{\boldsymbol{\Theta}}_M(n+1) = \hat{\boldsymbol{\Theta}}_M(n)$
      $+ \mu[\delta \boldsymbol{I}_M + \hat{\boldsymbol{\Sigma}}_M(n, \boldsymbol{p}_m^{\mathcal{A}})]^{-1} \boldsymbol{\kappa}(n, \boldsymbol{p}_m^{\mathcal{A}}) \boldsymbol{\varepsilon}_m(n)$ ▷ (38)
  end if
Update candidate intermediate signals and correlation coeffs.
  $\kappa_{m+1}^{\pm}(n, \boldsymbol{p}_l) = \Gamma^{\pm}(q, \boldsymbol{p}_l) a_m(n), \quad \forall \boldsymbol{p}_l \in \boldsymbol{\Omega}_g$
  $\hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n+1, \boldsymbol{p}_l) = \lambda \hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n, \boldsymbol{p}_l)$
    $+ (1-\lambda) \frac{\kappa_{m+1}^{\pm}(n, \boldsymbol{p}_l) \boldsymbol{\varepsilon}_m(n)}{\|\boldsymbol{\kappa}_{m+1}(n, \boldsymbol{\Omega}_g)\|^2 / 2L}$ ▷ (42)
Select common pole and set variables
  if $n = (m+1)N_s$ **then**
    $\hat{\alpha}_{m+1}^r(n, \boldsymbol{p}_l) = \sqrt{\hat{\alpha}_{m+1}^{r+}(n, \boldsymbol{p}_l)^2 + \hat{\alpha}_{m+1}^{r-}(n, \boldsymbol{p}_l)^2}$ ▷ (43)
    $c = \arg\max_l \sum_{r=1}^{R} \hat{\alpha}_{m+1}^r(n, \boldsymbol{p}_l)$ ▷ (44)
    $\boldsymbol{p}_{m+1}^{\mathcal{A}} = [\boldsymbol{p}_m^{\mathcal{A}}\,,\ \boldsymbol{p}_c]$ ▷ Add $\boldsymbol{p}_c$ to active pole set
    $\sigma_\iota^2 = \|\boldsymbol{\kappa}_{m+1}(n, \boldsymbol{\Omega}_g)\|^2 / 2L$ ▷ Set power estimate
    $\hat{\boldsymbol{\theta}}_{m+1}^r(n) \leftarrow \hat{\boldsymbol{\alpha}}_{m+1}^r(n, \boldsymbol{p}_c)\ \ (r = 1, \ldots, R)$ ▷ Set new $\boldsymbol{\theta}$
    $m \leftarrow m+1$ ▷ Move to next stage
    $\hat{\boldsymbol{\alpha}}_l^{\pm}(n) = \emptyset,\ \forall l$ ▷ Reset correlation coefficients
  end if
end while

vectors $\hat{\boldsymbol{\theta}}^r(n)$, which are stacked in the $M \times R$ matrix $\hat{\boldsymbol{\Theta}}_M(n) = [\hat{\boldsymbol{\theta}}^1(n), \ldots, \hat{\boldsymbol{\theta}}^R(n)]$. The intermediates signals in $\boldsymbol{\kappa}(n, \boldsymbol{p}_m^{\mathcal{A}})$ are the output signals of the orthonormalized resonator sections built from the active pole set $\boldsymbol{p}_m^{\mathcal{A}}$ and having TFs $\Psi(z, \boldsymbol{p}_m^{\mathcal{A}}) = \{\Psi_1^{\pm}(z, \tilde{\boldsymbol{p}}_1^{\mathcal{A}}), \ldots, \Psi_m^{\pm}(z, \tilde{\boldsymbol{p}}_m^{\mathcal{A}})\}$, where $\Psi_i^{\pm}(z, \tilde{\boldsymbol{p}}_i^{\mathcal{A}})$, defined as in (4), is built from the first i poles $\tilde{\boldsymbol{p}}_i^{\mathcal{A}} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_i]^T \in \boldsymbol{p}_m^{\mathcal{A}}$.

A schematic of the SB-OBF-GMP algorithm is depicted in Figure 4 and listed in Algorithm 1 with a slightly simplified notation, both containing elements explained below.

1) Multi-channel OBF linear filter coefficient adaptation: As already mentioned, the proposed OBF-NLMS update rule in (27) is used, at least while M is small, for the adaptation of the linear coefficients of the multi-channel OBF filter, which in matrix form for the multi-channel case becomes

$$\hat{\boldsymbol{\Theta}}_M(n+1) = \hat{\boldsymbol{\Theta}}_M(n) + \mu[\delta \boldsymbol{I}_M + \hat{\boldsymbol{\Sigma}}_M(n, \boldsymbol{p}_m^{\mathcal{A}})]^{-1} \boldsymbol{\kappa}(n, \boldsymbol{p}_m^{\mathcal{A}}) \boldsymbol{\varepsilon}_m(n)\,. \tag{38}$$

Initially ($m = 0$), the active pole set $\boldsymbol{p}_m^{\mathcal{A}}$ is empty, so that the vector of estimated output equal to all zeros ($\hat{\boldsymbol{y}}_0(n) = \mathbf{0}$) and $\boldsymbol{\varepsilon}_0(n) = \boldsymbol{y}(n)$. At the beginning of each new stage, a new pole pair $\boldsymbol{p}_c$ is added to the multi-channel OBF filter based on the selection strategy described below, and a pair of linear coefficients per each channel is included in coefficient vectors $\hat{\boldsymbol{\theta}}^r(n)$, thus augmenting the size of the square matrices $\boldsymbol{I}_M$ and $\hat{\boldsymbol{\Sigma}}_M(n, \boldsymbol{p}_m^{\mathcal{A}})$ and the number of rows of $\hat{\boldsymbol{\Theta}}_M(n)$ in (38) by two ($M \leftarrow M+2$).

2) Common-poles selection strategy: The main purpose of the SB-OBF-GMP algorithm is to identify a set of poles common to all R acoustic channels to be fixed into a multi-channel OBF adaptive filter. The poles are estimated using a matching pursuit approach. First, a grid $\boldsymbol{\Omega}_g$ of L candidate pairs of complex-conjugate poles is defined on the unit disc based on some prior knowledge of the room acoustic system or some particular desired frequency resolution [12]. For each pole pair $\boldsymbol{p}_l \in \boldsymbol{\Omega}_g$ (with $l = 1, \ldots, L$), the pair of candidate intermediate signals $\boldsymbol{\kappa}_{m+1}(n, \boldsymbol{p}_l) = [\kappa_{m+1}^{+}(n, \boldsymbol{p}_l), \kappa_{m+1}^{-}(n, \boldsymbol{p}_l)]$ is obtained as the $(m+1)$-th intermediate signals of an OBF filter built from the pole set $[\boldsymbol{p}_m^{\mathcal{A}}, \boldsymbol{p}_l]$, i.e. by filtering the input $u(n)$ with the TFs $\Psi_{m+1}^{\pm}(z, \boldsymbol{p}_l) = N_{m+1}^{\pm}(z, \boldsymbol{p}_l) P_{m+1}(z, \boldsymbol{p}_l) \prod_{i=1}^{m} A_i(z, \boldsymbol{p}_i)$, where the product corresponds to the series of m second-order all-pass filters defined by the pole pairs $\boldsymbol{p}_i \in \boldsymbol{p}_m^{\mathcal{A}}$ (cfr. Figure 1). Equivalently, $\boldsymbol{\kappa}_{m+1}(n, \boldsymbol{p}_l)$ can be computed by filtering the output of the all-pass series $a_m(n, \boldsymbol{p}_m^{\mathcal{A}}) = \prod_{i=1}^{m} A_i(q, \boldsymbol{p}_i) u(n)$ with pairs of filters having TFs $\Gamma^{\pm}(z, \boldsymbol{p}_l) = N_{m+1}^{\pm}(z, \boldsymbol{p}_l) P_{m+1}(z, \boldsymbol{p}_l)$.

The result is a parallel of L pairs of candidate intermediate signals $\boldsymbol{\kappa}_{m+1}(n, \boldsymbol{p}_l)$ with a common input $a_m(n)$, which can be stacked in a vector $\boldsymbol{\kappa}_{m+1}(n, \boldsymbol{\Omega}_g) = [\boldsymbol{\kappa}_{m+1}(n, \boldsymbol{p}_1), \ldots, \boldsymbol{\kappa}_{m+1}(n, \boldsymbol{p}_L)]^T$ of size $2L \times 1$. The idea is that, while the linear coefficients $\hat{\boldsymbol{\Theta}}_M(n)$ of the OBF filter are being adapted using the OBF-NLMS rule to minimize the power of the error signals in $\boldsymbol{\varepsilon}_m(n)$, the algorithm updates also the correlation coefficients $\hat{\alpha}_{m+1}^{r\pm}(n, \boldsymbol{p}_l)$ between each error signal $\varepsilon_m^r(n)$ and each candidate intermediate signal $\kappa_{m+1}^{\pm}(n, \boldsymbol{p}_l)$. In order to minimize the instantaneous squared error signals in $\boldsymbol{e}_{m+1}^{\pm}(n, \boldsymbol{p}_l) = [\boldsymbol{e}_{m+1}^{1\pm}(n, \boldsymbol{p}_l), \ldots, \boldsymbol{e}_{m+1}^{R\pm}(n, \boldsymbol{p}_l)]$, averaged over the R channels, produced by each candidate intermediate signal,

$$\underset{\boldsymbol{p}_l, \hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n)}{\text{minimize}} \ \left\|\boldsymbol{e}_{m+1}^{\pm}(n, \boldsymbol{p}_l)\right\|^2 = \left\|\boldsymbol{\varepsilon}_m(n) - \kappa_{m+1}^{\pm}(n, \boldsymbol{p}_l) \hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n, \boldsymbol{p}_l)\right\|^2\,, \tag{39}$$

each vector of correlation coefficients $\hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n, \boldsymbol{p}_l) = [\hat{\alpha}_{m+1}^{1\pm}(n, \boldsymbol{p}_l), \ldots, \hat{\alpha}_{m+1}^{R\pm}(n, \boldsymbol{p}_l)]$ is adapted in time. It follows from simple calculations that the LMS adaptation rule minimizing (39) can be written as

$$\hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n+1, \boldsymbol{p}_l) = \big(1 - \mu|\kappa_{m+1}^{\pm}(n, \boldsymbol{p}_l)|^2\big) \hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n, \boldsymbol{p}_l) + \mu \kappa_{m+1}^{\pm}(n, \boldsymbol{p}_l) \boldsymbol{\varepsilon}_m(n). \tag{40}$$

Normalizing the step size μ with respect to the instantaneous power of the regressor, gives (for $\lambda = 1 - \mu$)

$$\hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n+1, \boldsymbol{p}_l) = \lambda \hat{\boldsymbol{\alpha}}_{m+1}^{\pm}(n, \boldsymbol{p}_l) + (1-\lambda) \frac{\kappa_{m+1}^{\pm}(n, \boldsymbol{p}_l) \boldsymbol{\varepsilon}_m(n)}{|\kappa_{m+1}^{\pm}(n, \boldsymbol{p}_l)|^2}, \tag{41}$$

which is recognized as an exponential window update of the normalized instantaneous correlation between $\kappa_{m+1}^{\pm}(n, \boldsymbol{p}_l)$ and $\boldsymbol{\varepsilon}_m(n)$ with forgetting factor λ.

The update rule in (41) is not immune to large values of the intermediate signals $\kappa^{\pm}_{m+1}(n, \boldsymbol{p}_l)$, which would result in large variations of the estimates $\hat{\boldsymbol{\alpha}}^{\pm}_{m+1}(n+1, \boldsymbol{p}_l)$. Since tracking the power of each $\kappa^{\pm}_{m+1}(n, \boldsymbol{p}_l)$ as proposed for the OBF-NLMS would be too expensive, an effective solution is proposed, which consists of normalizing the instantaneous correlations by the instantaneous average of the overall power of all candidate intermediate signals, giving

$$\hat{\boldsymbol{\alpha}}^{\pm}_{m+1}(n+1, \boldsymbol{p}_l) = \lambda \hat{\boldsymbol{\alpha}}^{\pm}_{m+1}(n, \boldsymbol{p}_l) + (1-\lambda) \frac{\kappa^{\pm}_{m+1}(n, \boldsymbol{p}_l) \boldsymbol{\varepsilon}_m(n)}{\|\boldsymbol{\kappa}_{m+1}(n, \boldsymbol{\Omega}_g)\|^2 / 2L}. \tag{42}$$

The average of the instantaneous power is in this case a good estimator of the actual power of the candidate intermediate signals, given that the poles in the grid are numerous and distributed within a wide frequency range. This is the same argument to explain the comparable performance of the NLMS algorithm with respect to OBF-NLMS for higher model orders (cfr. Figure 3).

At the end of each stage, i.e. at sample $n = (m+1)N_s$, the selection of the pole pair is performed. Given that (42) is a correlation update, the choice is based on which pair of candidate intermediate signals is mostly correlated on average with the error signals $\boldsymbol{\varepsilon}_m(n)$. The correlation of each pair of intermediate signals for the r-th channel is computed as

$$\hat{\alpha}^{r}_{m+1}(n, \boldsymbol{p}_l) = \sqrt{\hat{\alpha}^{r+}_{m+1}(n, \boldsymbol{p}_l)^2 + \hat{\alpha}^{r-}_{m+1}(n, \boldsymbol{p}_l)^2}. \tag{43}$$

A pole pair $\boldsymbol{p}_{m+1} \leftarrow \boldsymbol{p}_c$ is then chosen from the grid as the one giving intermediate signals with the highest average correlation on the R acoustic channels as

$$c = \arg\max_{l} \sum_{r=1}^{R} \hat{\alpha}^{r}_{m+1}(n, \boldsymbol{p}_l) \tag{44}$$

and added to the active pole set $\boldsymbol{p}^{\mathcal{A}}_{m+1} = [\boldsymbol{p}^{\mathcal{A}}_{m}, \boldsymbol{p}_{m+1}]$ of the OBF adaptive filter. The linear filter coefficients $\hat{\boldsymbol{\theta}}^{r}_{m+1}(n) = [\hat{\theta}^{r+}_{m+1}(n), \hat{\theta}^{r-}_{m+1}(n)]$ are set equal to the correlation coefficients $\hat{\boldsymbol{\alpha}}^{r}_{m+1}(n, \boldsymbol{p}_c) = [\hat{\alpha}^{r+}_{m+1}(n, \boldsymbol{p}_c), \hat{\alpha}^{r-}_{m+1}(n, \boldsymbol{p}_c)]$, so that they are already close to their steady-state value. Moreover, if the OBF-NLMS is used, the power estimates for the newly added poles are set equal to $\sigma^2_{\iota} = \|\boldsymbol{\kappa}_{m+1}(n, \boldsymbol{\Omega}_g)\|^2 / 2L$ (with ι equal to $M+1$ and $M+2$). Finally, the algorithm moves to the next stage ($m \leftarrow m+1$), all the correlation coefficients are reset to zero, and another pole pair is estimated as described above, until a desired number of poles $m = m_{\max}$ has been selected or some other stopping criterion based on the error in (37) is satisfied.

B. Algorithm evaluation

The final goal of the SB-OBF-GMP algorithm is to determine a set of poles, common to multiple acoustic paths, which are close to the true poles of the system considered. Since this knowledge, as well as the knowledge of the optimal pole parameters of an OBF filter, is usually not available, the algorithm is evaluated with respect to the OBF-GMP algorithm [12], [13]. Also, it should be noticed that the poles identified do not have a one-to-one correspondence with the true poles. In many cases, due to the mode overlapping, the finite resolution of the grid, and the iterative nature of the algorithm, one pole pair with slightly smaller radius is selected in the vicinity of two or more true pole pairs, so that on average the distance between the estimated poles and the true poles is reduced, as well as the bias error in (15).

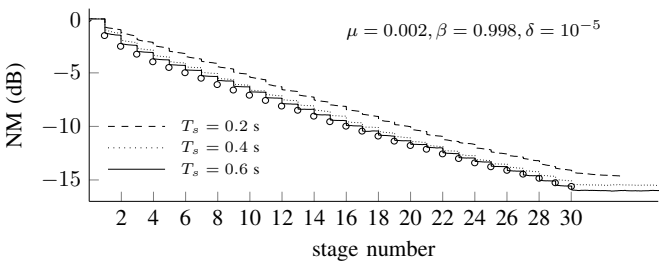

Fig. 5: The averaged NM for the OBF-GMP algorithm (○), which has access to the measured RIRs, and for the SB-OBF-GMP with different stage lengths T_s. Room M, $T_{60} = 0.5$ s, WN input signals.

Most estimation methods, including the OBF-GMP algorithm, for estimating the pole parameters of an OBF filter rely on the availability of measured RIRs. Thus, the purpose of this evaluation is to verify whether the proposed identification algorithm is able to achieve the same approximation error that is obtained with the OBF-GMP algorithm using the same pole grid. Obtaining comparable results is a confirmation of the performance of the SB-OBF-GMP algorithm, given that estimating the parameters from input-output signals is more involved than from measured RIRs.

1) Identification from white noise signals: The evaluation is performed with respect to the identification of a set of simulated RIRs generated at sampling frequency $f_s = 800$ Hz, using the RIM [45] (see Section IV for details) at one loudspeaker and 4 microphones positions ($R = 4$, room M, T_{60}= 0.5 s). First, the RIRs are modeled using the OBF-GMP algorithm [12], [13]. The pole grid used has 400 angles uniformly placed from 1 Hz to 399 Hz and 5 radii logarithmically distributed from 0.7 to 0.9925, summing up to $L = 2000$ candidate pole pairs. The NM in (30), averaged over the 4 RIRs, is computed for each estimation of a new pole pair and depicted in Figure 5 using ○. Then, the SB-OBF-GMP algorithm is tested for 10 realizations of an input WN sequence with unit variance, using the same pole grid described above. The NM is averaged over the 4 RIRs and over the different realizations of the input signal. Three different values for the stage length $T_s = N_s/f_s$ have been used, with a new pole pair added to the active pole set every 0.2, 0.4, and 0.6 seconds, corresponding to the three curves of the averaged NM in Figure 5. It is seen that results very similar to the OBF-GMP algorithm can be obtained already with stages of $T_s = 0.4$ s (a significant improvement with respect to the BB-OBF-GMP algorithm [34] requiring blocks of more than 2 s).

2) Identification from speech signals: The identification from non-stationary and non-white signals, such as speech, is more challenging. One difficulty is related to the slower convergence rate due to the non-constant input PSD, as described in Section II-B, for which longer stages may be required to reduce the misalignment. Moreover, since OBF-NLMS and

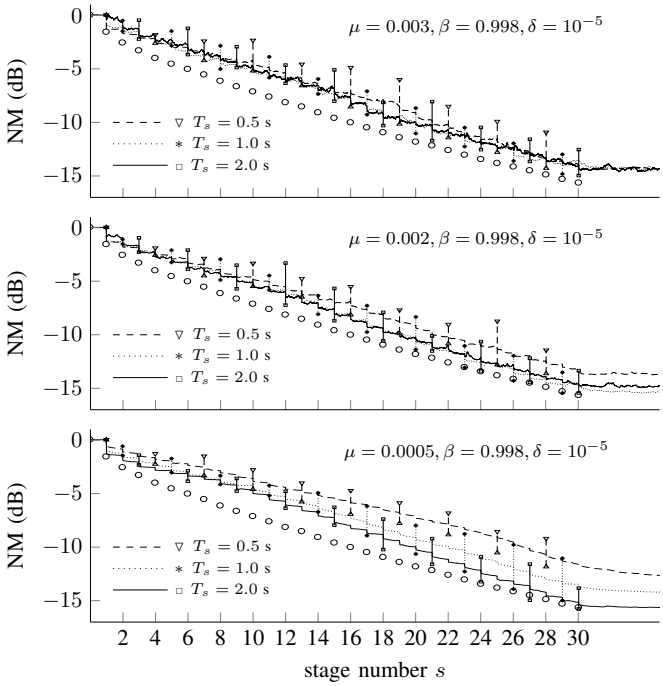

Fig. 6: The averaged NM for the OBF-GMP algorithm (○) and for the SB-OBF-GMP with different step sizes μ and different stage lengths T_s. Room M, $T_{60} = 0.5$ s, male speech input signal (librivox).

NLMS are used to counteract problems related to the non-stationarity of speech, the non-constant input PSD increases the steady-state error (see (22)), both for the linear coefficients in $\hat{\boldsymbol{\Theta}}_M$ and the correlation coefficients in $\hat{\boldsymbol{\alpha}}^{\pm}_{m+1}(n, \boldsymbol{p}_l)$. Convergence rate can be increased using a larger step size μ (see (23)), at the expense of a larger steady-state error, which also leads to higher misalignment and less accurate pole identification. On the other hand, greater accuracy is achieved by reducing the step size, which however requires longer stages for the coefficients to converge.

In any case, regardless of the stage length T_s chosen, the short-term frequency spectrum within one stage is usually far from flat, resulting in an uneven excitation of the frequency range of interest. In a given stage, some of the candidate OBF TFs are not sufficiently excited for the corresponding correlation coefficient to converge sufficiently fast. It follows that the pole selection is influenced by the frequency content of the input signal in the current stage, so that deviations from the behavior of the modeling algorithm are normally expected.

Another issue is the long-term frequency range excited by a speech signal at low frequencies. The voiced speech of an adult male typically has fundamental frequency between 85 and 180 Hz, whereas that of an adult female between 165 and 255 Hz [49], so that a speech signal rarely has sufficient power to excite the lower modes of the system. Relatively small rooms already have modal frequencies well below the cut-off frequency of a speech signal. Thus, the capabilities of the algorithm of identifying the system from speech signals at low frequencies depend on both the characteristics of the signal itself and of the modal characteristics of the room response, as discussed further in Section IV.

To illustrate these concepts, the SB-OBF-GMP algorithm is tested on 10 long sequences of male speech taken from an audiobook "A Tramp Abroad" by Mark Twain[1], downsampled to $f_s = 800$ Hz, where the silent portions of the signals were removed using a voice activity detection algorithm [50], [51]. The same pole grid as in the WN case is used. Three different stage lengths $T_s = \{0.5, 1, 2\}$ s have been chosen, corresponding to the three curves of the NM, averaged over the 4 RIRs and over the 10 different input signal sequences (the vertical lines correspond to the range between the maximum and minimum values). The experiment is repeated for three different values of the step size μ, corresponding to the three plots in Figure 6. In the top plot ($\mu = 0.003$), roughly the same result, with similar deviations, is obtained for different stage lengths, even though the performance of the OBF-GMP algorithm is not attained. This bias is mostly due to the large steady-state error, and not due to the low convergence rate, given that a longer stage does not provide almost any improvement. As already mentioned, a lower misalignment is achieved by employing a smaller step size, thus reducing the parameter variability. The convergence rate, however, decreases, so that improvements are obtained only for stages long enough to allow the coefficients to converge. An overall improvement can be noticed in the middle plot, especially for $T_s = 1$ s and $T_s = 2$ s. By further reducing the step size, the parameter variability decreases, but also the convergence rate, so that a further reduction in the misalignment may be difficult even for long stages, as suggested from the bottom plot of Figure 6 for $\mu = 0.0005$ and $T_s = 2$ s. Extensive simulations on different materials show that, in general, a good trade-off between convergence rate and steady-state error is provided by a step size between $\mu = 0.002$ and $\mu = 0.001$, which gives reasonably low misalignment for a stage length T_s below 1 s.

Another example is presented showing the same experiment performed on a male speech sequence taken from the EBU-SQAM database [52] (see top plot of Figure 7). In this case, the SB-OBF-GMP algorithm with stage length 1 s, is able to obtain similar results in terms of NM as the OBF-GMP algorithm[2]. What differentiates this case from the previous ones is the approximately flat long-term PSD above 80 Hz, which also corresponds to the cut-off frequency of the frequency response of the room. To conclude, the algorithm is tested for a female speech sequence from the same database. As mentioned above, the fundamental frequency of female speech is normally above 165 Hz. As a consequence, not enough signal power is available below that frequency, and the system cannot be fully identified. This is shown in the bottom plot of Figure 7, where, after a certain number of stages, adding a new pole pair does not significantly reduce the NM.

IV. Identification vs. room characteristics

In all system identification tasks, the first step is to decide which model is the most adequate for the problem at hand.

[1]publicly available at http://librivox.org/a-tramp-abroad-by-mark-twain/, MP3-files at 128kbps were converted to WAV.

[2]the SB-OBF-GMP algorithm gives a small improvement over the OBF-GMP algorithm in some cases because the iterative pole selection strategy of both algorithms is optimal only in relation to the poles already selected.

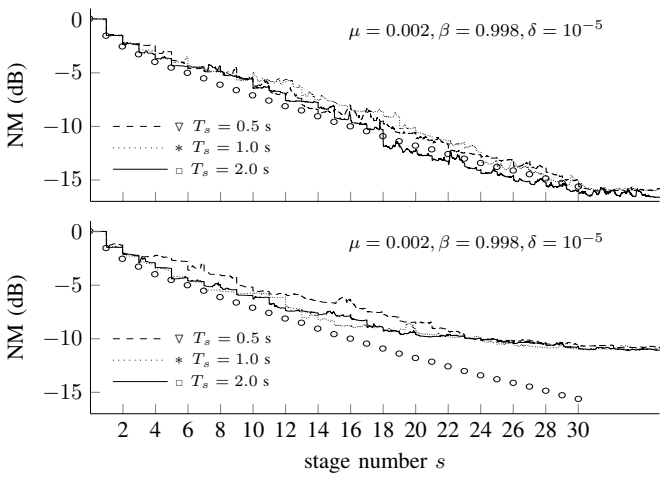

Fig. 7: The averaged NM for the OBF-GMP algorithm (○) and for the SB-OBF-GMP with different stage lengths T_s. Room M, $T_{60} = 0.5$ s, male (top) and female (bottom) speech input signal (EBU-SQAM).

Trying out different models is normally cumbersome and in some cases not even possible. It is then important to have an indication about which model to use, based on some prior knowledge of the system. For applications in room acoustics, the prior information that may be available regards the room dimensions, the characteristics of the surfaces in the room and/or the reverberation time.

The question regarding the adequateness of IIR models in RASE applications (especially AEC) is a recurrent topic in the literature [7], [8], [53]. In these works, one common conclusion is that the advantage given by the superior modeling capabilities of IIR filters over FIR filters is not significant enough to justify their use, given the higher filter complexity and the added difficulty in the approximation/identification process. This result was explained by observing that a RTF is composed of a large number of resonances, and since an IIR filter models each resonance by a second-order TF, the number of filter parameters that is required to accurately identify the system may be not far from the required number of filter taps in a FIR filter.

In [12], it was shown by simulation results on a number of RIRs measured in different rooms that the use of OBF filters gives an advantage over FIR filters, even when the increased filter complexity is accounted for; this advantage is more significant at low frequencies, where a RTF normally presents sharper resonances and a lower degree of modal overlap [35]. The actual advantage of using OBF filters is then dependent on these two elements, which in turn are related to the room dimensions, the reverberation time, and the number of poles used.

In this section, simulation results are shown in order to analyze the use of OBF and FIR adaptive filters in relation to the characteristics of the room at low frequency. Moreover, the estimation of poles common to multiple acoustic channels, which brings computational savings [12], [13] (since each loudspeaker signal is filtered by the same OBF filter, having microphone-dependent linear coefficients), is evaluated at positions in the room (validation microphones) others than the ones at which the poles are estimated (training microphones). The analysis is first performed on simulated scenarios, and then verified on measured RIRs.

TABLE I: RIM simulated room specifications

room	V (m^3)	$N_{f_s/2}$	T_{60} (s)	f_{Sch} (Hz)	$N_{f_{\text{Sch}}}$
S	17	113	0.15, 0.25, 0.50	188, 243, 343	12, 25, 71
M	50	332	0.25, 0.50, 0.75	141, 200, 245	15, 42, 76
L	300	1993	0.50, 0.75, 1.00	82, 100, 115	17, 31, 47

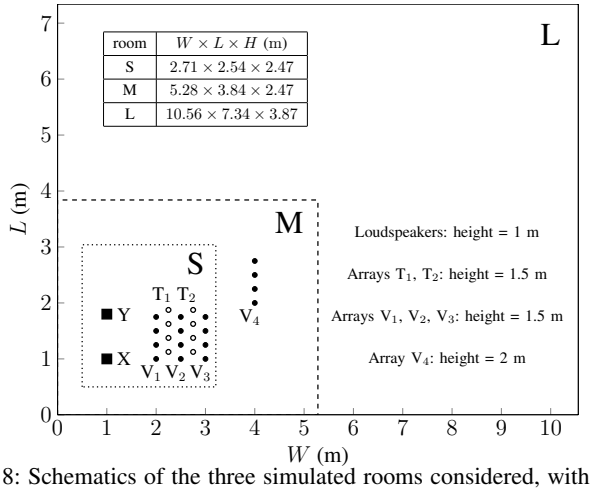

Fig. 8: Schematics of the three simulated rooms considered, with the relative position of loudspeakers (■), training microphones (○), and validation microphones (●).

A. Simulated rooms

Three different rooms were simulated at $f_s = 800$ Hz using the RIM [45] with a random displacement of 1 cm, each room with three different RTs, resulting in 9 different cases. The characteristics of the rooms are given in Table I, listing the room volume V, the theoretical number of modes [35] below half the sampling frequency,

$$N_{f_s/2} \approx \frac{4\pi}{3} V \left(\frac{f_s/2}{c} \right)^3 \tag{45}$$

(with $c = 343\,\text{m/s}$ the sound velocity), the 3 different RTs T_{60} per each room, their corresponding Schroeder frequency $f_{\text{Sch}} \approx 2000\sqrt{T_{60}/V}$, and the theoretical number of modes $N_{f_{\text{Sch}}}$ below the Schroeder frequency. The particular choice of the sampling frequency was made to include the highest Schroeder frequency among the 9 cases considered.

Multiple loudspeakers and microphones are distributed in a fixed configuration in the rooms as illustrated in Figure 8, also showing the room width W, length L and height H. Full dots represent the validation microphones, consisting of four arrays V_a ($a = 1, \ldots, 4$) of 4 microphones each, with inter-microphone spacing of 25 cm (note that V_4 cannot be used for Room S). Empty dots represent the training microphones, consisting of two arrays T_b ($b = 1, \ldots, 2$) of 4 microphones each. The first 3 validation arrays and the training arrays form a 1 m×1 m area, that will be referred at as *sweet-spot*, whereas the fourth validation array is the *isolated* array. In both training and validation, two different WN input signals are fed to

two loudspeakers X and Y, indicated with a full square. The microphone noise was set very low (SNR $= 60$ dB) so not to influence the performance of the identification. It is assumed that the loudspeaker-microphone distances are known, so that the acoustic delay parameter d_r can be set in the filter for each rth acoustic channel[3].

In the training, a set $\boldsymbol{p}_m^{\mathcal{A}}$ of common poles of the multichannel OBF adaptive filter is identified from the 8 microphone signals of arrays T_b using the SB-OBF-GMP algorithm with $T_s = 0.5$ s and variables defined as in Section III-B1. When the identification is performed from two (or more) loudspeakers, the two loudspeaker signals have to be filtered individually by two SIMO OBF filters in parallel. If the input signals are uncorrelated, as for different realizations of WN, the MIMO system can be correctly identified from the microphone signals, and a set of poles common to all the MIMO acoustic channels considered can be obtained with the pole selection strategy of the SB-OBF-GMP algorithm with very minor modifications. The identification from correlated signals is discussed in Section V-A in the framework of stereophonic acoustic echo cancellation (SAEC). Even though the pole grid could be chosen according to some prior knowledge of the room and its RT, here the same grid has been used in all conditions. The pole grid has $L = 4000$ pole pairs, with 400 angles uniformly placed from 1 Hz to 399 Hz (giving approximately 1 Hz of resolution), and 10 radii logarithmically distributed from 0.85 to 0.9925. These limits for the radius ρ are chosen to include the minimum and the maximum values of the RT T_{60} considered, according to the formula [35]

$$\rho = 10^{\frac{-3}{T_{60} f_s}} \tag{46}$$

The lower bound ($\rho_{\min} = 0.85$), corresponding to relatively wide resonances with a decay time of 50 ms, is meant to allow for the approximation of the superposition of multiple resonances.

In the validation, the estimated common poles $\boldsymbol{p}_m^{\mathcal{A}}$ are fixed in the multi-channel OBF filter and only the linear coefficients are adapted from the microphone signals of both training and validation arrays. The adaptation on the training and validation arrays using the same signals is performed also with FIR adaptive filters, whose performance is compared to that of the OBF adaptive filters with the same number $M = 2m$ of adaptive linear coefficients. For both filters, the NLMS algorithm in (19) is used, as the case of very low orders requiring the OBF-NLMS algorithm is not considered here. The validation on each array for a specific value of M is repeated 3 times, each time using 2 different realizations of WN (30 s long, one per each loudspeaker) as the input signals.

The NM, as defined in (30), is computed for each acoustic channel for the loudspeaker-microphone pairs in the specific configuration (2 loudspeakers, 1 array), and then averaged over the number of acoustic channels and the 3 repetitions,

$$\overline{\text{NM}}_{\text{Z}}(\infty) = 10\log_{10}\left(\frac{1}{3R_{\text{Z}}}\sum_{r=1}^{R_{\text{Z}}}\sum_{k=1}^{3}\frac{\|\boldsymbol{h}_r^k - \hat{\boldsymbol{h}}_r^k(\infty, \boldsymbol{p}_m^{\mathcal{A}}, \hat{\boldsymbol{\theta}}_r)\|_2^2}{\|\boldsymbol{h}_r^k\|_2^2}\right) \text{dB}, \tag{47}$$

[3]in a practical implementation, the d_r-samples delays are applied to the intermediate signals before they are fed to the multichannel linear coefficients.

where $\hat{\boldsymbol{h}}_r(\infty, \boldsymbol{p}_m^{\mathcal{A}}, \hat{\boldsymbol{\theta}}_r)$ indicates the RIR estimated at steady-state at the end of the kth repetition for a set of poles $\boldsymbol{p}_m^{\mathcal{A}}$ identified during training, and Z indicates the array used in the validation (e.g. T_b or V_a) with a total of R_{Z} channels. In practice, the NM at steady-state is computed by averaging the misalignment obtained on the last 10 s of the validation signal, when the parameters have normally converged, as

$$\frac{\|\boldsymbol{h}_r^k - \hat{\boldsymbol{h}}_r^k(\infty, \boldsymbol{p}_m^{\mathcal{A}}, \hat{\boldsymbol{\theta}}_r)\|_2^2}{\|\boldsymbol{h}_r^k\|_2^2} \approx \frac{1}{10 f_s}\sum_{n=20 f_s}^{30 f_s}\frac{\|\boldsymbol{h}_r^k - \hat{\boldsymbol{h}}_r^k(n, \boldsymbol{p}_m^{\mathcal{A}}, \hat{\boldsymbol{\theta}}_r)\|_2^2}{\|\boldsymbol{h}_r^k\|_2^2}. \tag{48}$$

Also the average convergence time $\overline{\text{CT}}_{\text{Z}}$ is computed per each array Z as the time instant at which the average NM reaches +2 dB above $\overline{\text{NM}}_{\text{Z}}(\infty)$. These quantities are illustrated in Figure 9.

In the following analysis, four measures are considered:

(i) the average NM on the training sets (speakers X and Y, and arrays T_b, $R_{T_b} = 8$) calculated for the validation microphone signals, using OBF and FIR adaptive filters.

$$\overline{\text{NM}}_{\text{T}}(\infty) = \frac{1}{2}\sum_{b=1}^{2}\overline{\text{NM}}_{\text{T}_b}(\infty), \tag{49}$$

(ii) the distance in dB between $\overline{\text{NM}}_{\text{T}}(\infty)$ and the average NM using OBF adaptive filters on the validation sets,

$$\overline{\Delta\text{NM}}_{\text{TV}}(\infty) = \frac{1}{N_a}\sum_{a=1}^{N_a}\left[\overline{\text{NM}}_{\text{T}}(\infty) - \overline{\text{NM}}_{\text{V}_a}(\infty)\right] \text{dB}, \tag{50}$$

with $\overline{\text{NM}}_{\text{V}_a}(\infty)$ computed using (47) on set a (speakers X and Y, and array V_a, $R_{V_a} = 8$) and N_a the number of validation sets considered in the average,

(iii) the distance in dB between the average NM for the validation sets obtained using OBF adaptive filters and FIR adaptive filters,

$$\overline{\Delta\text{NM}}_{\text{VF}}(\infty) = \frac{1}{N_a}\sum_{a=1}^{N_a}\left[\overline{\text{NM}}_{\text{V}_a}(\infty) - \overline{\text{NM}}_{\text{F}_a}(\infty)\right] \text{dB}, \tag{51}$$

with $\overline{\text{NM}}_{\text{F}_a}(\infty)$ computed using (47) for FIR adaptive filters on set a (speakers X and Y, and array V_a),

(iv) the average convergence rate, defined as the ratio between the NM at +2 dB above the steady-state and the convergence time CT, averaged over multiple arrays,

$$\overline{\text{CR}} = \frac{1}{N_a}\sum_{a=1}^{N_a}\left[\frac{\overline{\text{NM}}_{\text{Z}_a}(\infty) + 2}{\overline{\text{CT}}_{\text{Z}_a}}\right] \text{dB/s}, \tag{52}$$

calculated on the validation sets for OBF and FIR adaptive filters ($\text{Z}_a = \{\text{V}_a, \text{F}_a\}$).

For all these averaged measures, also the minimum and maximum values are computed and plotted as vertical bars in the plots that follow. The analysis is performed on the 9 room configurations (3 rooms, 3 RTs each), first on the arrays in the sweet-spot, and then separately on the isolated array.

1) Evaluation at the sweet-spot: The results for the measures, computed as in (49)-(52) for the training and the first 3 validation sets, are given in Figure 10, where each row corresponds to a measure and each column to one room. The first row, instead, shows the magnitude responses (loudspeaker

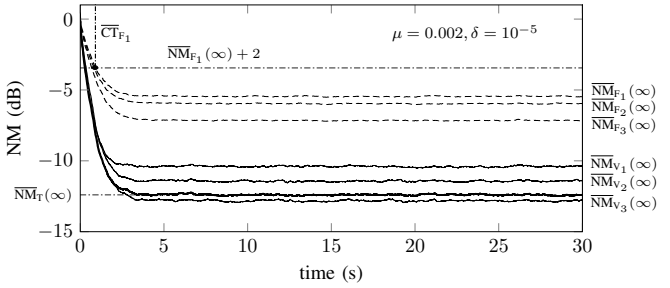

Fig. 9: Illustration of the quantities used in the analysis measures: the NM computed on the training microphones (——), and on the validation arrays using OBF (——) and FIR (- - -) filters. Room M, $T_{60} = 0.75$ s, $M = 60$.

X, first microphone of array T_2) for the 3 rooms with 3 different values of the RT, with the vertical lines indicating the corresponding Schroeder frequency.

(i) *Misalignment at training positions:* Looking at the curves in the second row of Figure 10, the first thing that can be noticed is that the average NM reduces consistently for increasing numbers M of the linear coefficients. This is of course expected, since the common poles are identified at the training positions based on the minimization of the average power of the residuals, which is also the reason for the small differences between the NM at the two training arrays (see the 'short' bars in the plots).

Also, the absolute reduction in the NM is dependent on the RT: a shorter RT implies wider resonances and a more prominent modal overlap, so that a smaller number of poles is required to approximate the RTF (e.g. notice in the top-left plot the reduced number of spectral peaks for $T_{60} = 0.15\,\mathrm{s}$ compared to $T_{60} = 0.5\,\mathrm{s}$). However, a shorter RT also implies a faster time decay of the RIR, which means that an FIR filter as well requires less parameters to attain a given NM. The NM thus reduces with the RT both for OBF and FIR filters for increasing M, with the former showing a stronger reduction, which will be analyzed further on the validation sets.

Another factor influencing the absolute reduction in the NM is the volume of the room, and more precisely the number of modes $N_{f_s/2}$ in the RTF. Indeed, large rooms have a large number of modes in a fixed frequency range and thus a higher modal density. However, they normally also have a long RT, so that the overlap between modal resonances is only partial (at least at low frequencies), as can be noticed in the top-right plot. This means that, for a given RT and a given number of poles, the absolute NM that is achievable depends on the room dimensions, with larger rooms having higher NM (e.g. compare the three rooms for $T_{60} = 0.5\,\mathrm{s}$ and $M = 120$).

An alternative way of quantifying the difference in performance of OBF and FIR filters is to compare the number of coefficients that is required for them to achieve a predefined value of the NM. As an example, consider the medium room with $T_{60} = 0.75\,\mathrm{s}$ at $\overline{\mathrm{NM}}_{\mathrm{T}}(\infty) = -18\,\mathrm{dB}$: in this case, OBF filters require $M = 90$ linear coefficients, while FIR filters $M = 180$.

(ii) *Misalignment at validation positions:* The assessment of the NM at the training positions is useful to analyze the performance of OBF filters for the identification algorithm used. In practice, however, the receiver may change position in space, so that it is of interest analyzing the degradation in performance at other positions when the poles are kept fixed. The variations in the NM in positions others than those at which the poles were computed is quantified here by means of the measure defined in (50). The third row in Figure 10 shows that the degradation introduced by moving the receivers within the sweet-spot area is limited to 1 dB (with ± 1 dB variability), irregardless of the room characteristics. This is a good indication that the estimated poles can be considered common, at least in the neighborhood of the training positions.

The validation of the identification results at a position away from the training positions is considered later on. Moreover, the tracking performance, i.e. the ability of the adaptive algorithm to track variations in the RTF due, for instance, to the receiver changing position during adaptation, is analyzed in the AEC example in Section V-A.

(iii) *Comparison between OBF and FIR filters:* The third measure considered is intended to verify the actual advantage that is given by OBF filters over FIR filters at validation positions within the sweet-spot, which is the actual situation in case poles are estimated at some training positions and kept fixed when used at other positions. Starting from small rooms, it can be seen that the gap between OBF and FIR filters, defined as in (51), increases for increasing M. However, it can be noticed that, for short RT, this gap is reduced above a certain value of M, the reason being that all the main resonances of the RTF have been already modeled, so that adding one more pole pair reduces the NM less than adding two more taps to an FIR filter.

The same trend is noticed for medium rooms, which have a larger number of resonant modes in the frequency range considered. As already mentioned, a longer RT implies sharper resonances and the need of a larger number of poles to achieve a small NM. For this reason the absolute reduction in the NM is less than in small rooms (as is the case at the training positions). It can be also seen that for longer RT (see $T_{60} = 0.75\,\mathrm{s}$), the gap increases more slowly for increasing M. This can be ascribed to the lower degree of modal overlap and the consequent increase in the number of resonances that have to be modeled. This fact is even more noticeable for large rooms, where the number of modes becomes even larger. Even though the number of required coefficients for FIR filters increases (given the longer decay time of the RIRs), a large number of coefficients is required for OBF filters as well, so that the difference between the two types of filters becomes less significant (which is the main argument against the use of IIR filters in the literature [7], [8]).

Another factor influencing the gap between the performances of OBF and FIR filters is the finite resolution of the pole grid of the SB-OBF-GMP algorithm used in the training phase. Even though the grid-based approach is effective in finding a good low-order approximation of the RTFs, it happens that a mode is not efficiently modeled by a single pole pair in the OBF filter because the true pole of the system has values for the angle and/or radius in-between the values used in the grid. Thus, this lack of resolution may lead to a reduced gap

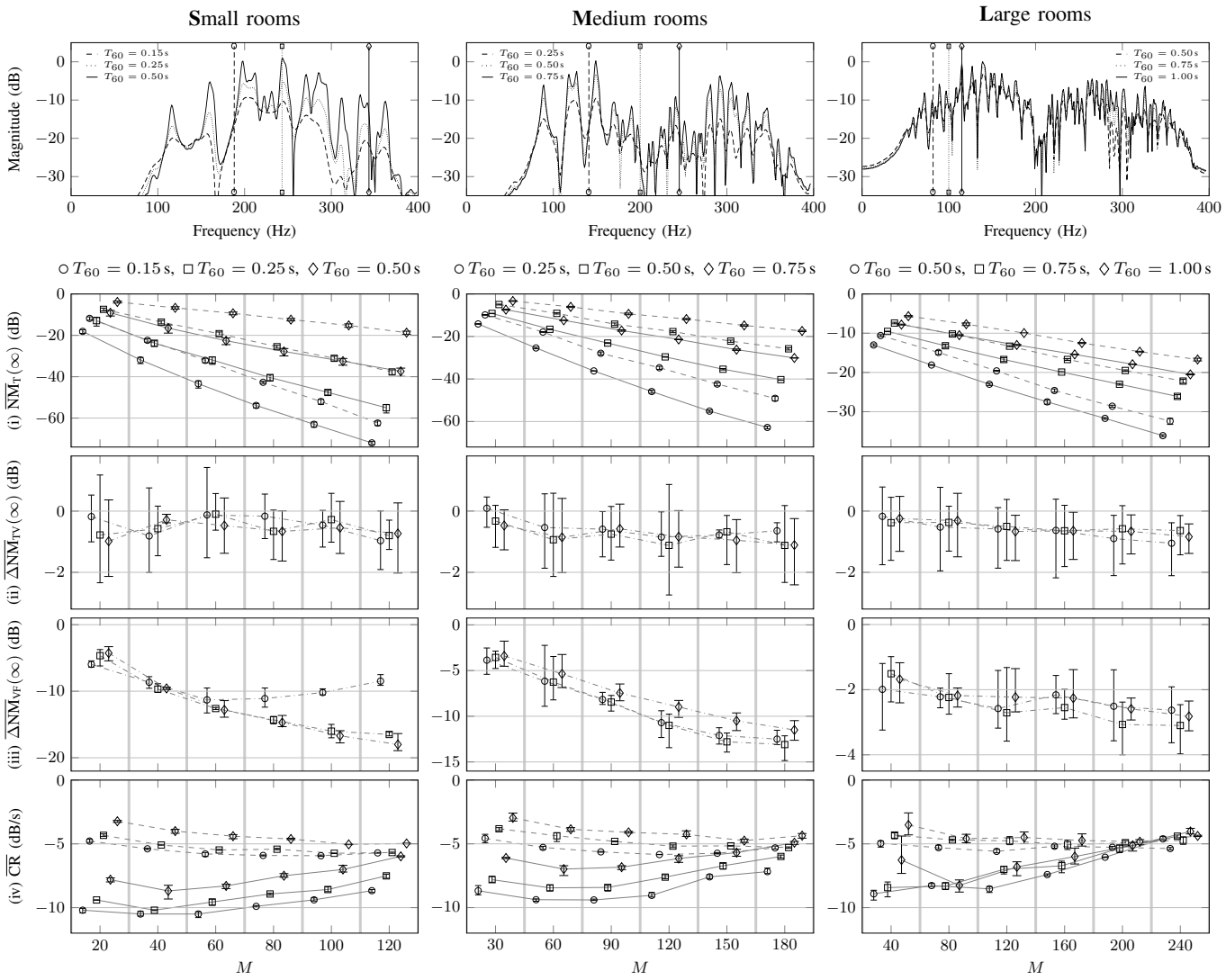

Fig. 10: *Sweet-spot*: identification results for measures in (49-52), using OBF (solid) and FIR (dashed) filters with different orders M, and for the 9 cases considered (3 rooms, 3 RTs). Top row: example magnitude responses and corresponding f_{Sch} (vertical lines).

between the performances of OBF and FIR filters, especially when the modal resonances are sharper and more numerous, as happens in large rooms.

(iv) *Convergence rate:* The convergence rate (CR) measure in (52) is meant to assess the relation between the NM at steady-state and the time it takes for the adaptation algorithm to attain it. The CR is clearly dependent on the step size μ, so that a faster CR is achieved for a larger μ, at the expense of a larger steady-state error (see Section II-B). As shown in Section II-B, the estimation error for OBF adaptive filters between two successive iterations is governed by the term $\gamma_M(\omega, \boldsymbol{p})$ in (23), with $\gamma_M(\omega, \boldsymbol{p}) = M$ for FIR filters. In these simulations, the step size in the NLMS adaptation rule in (19) was kept fixed at $\mu = 0.002$, so that $\tilde{\mu} = \mu M$ increases with increasing values of M. It is worth recalling that the loudspeaker signals are in this case unit-variance WN sequences (i.e. $S_u(\omega) = 1,\ \forall\omega$), and that the microphone noise variance is quite low ($\sigma_v^2 \approx 10^{-6}$).

From the last row of Figure 10 it can be seen that the CR is basically constant for FIR filters, with possibly a slightly lower rate for small model orders (the first samples of a downsampled RIR correspond to early reflections, which are large in absolute value and so more iterations are needed for the coefficients to converge from zero). Since $\tilde{\mu}$ increases for increasing M, the first term of the error in (23) must decrease, whereas the second term increases. It follows that, for the CR to be constant, the two terms must compensate each other.

The faster convergence rate for OBF filters at low values of M can then be partially explained by the fact that $\gamma_M(\omega, \boldsymbol{p}) < M$ for most values of the frequency, so that the second term in (23) remains small. By adding more poles, the NM reduces, but $\gamma_M(\omega, \boldsymbol{p})$ increases, at least at those frequencies where the pole density is high. The other reason can be found in how the NM decreases for increasing M. Whereas, apart from early reflections, the NM for FIR filters tends to decrease linearly (on a logarithmic scale), OBF filters

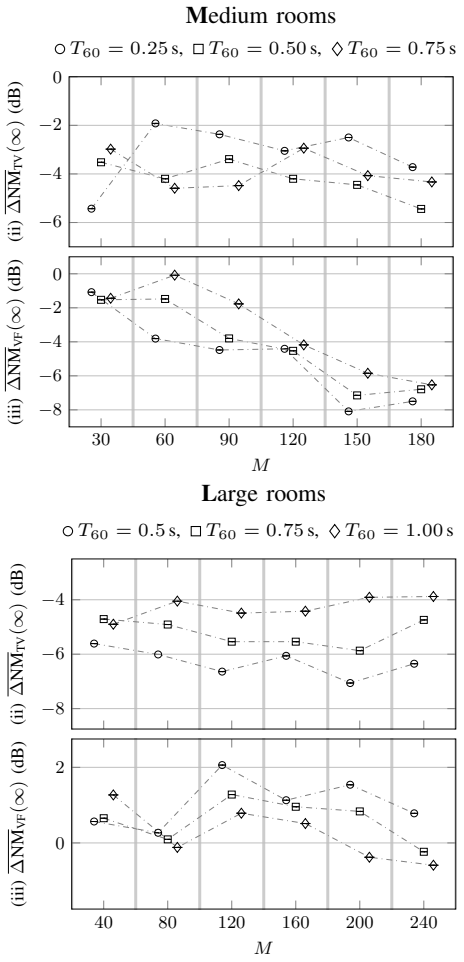

Fig. 11: *Isolated array*: identification results for measures in (50-51), using OBF and FIR filters with different orders M, and for the 6 cases considered (2 rooms, 3 RTs).

show a decreasing exponential trend of the NM (see [12], [13] for more convincing examples than in Figure 10). This is a consequence of the fact that with OBF filters, and especially when iterative procedures are used, the most dominant room modes are modeled first, so that the relative reduction in the NM decreases with increasing M, and consequently the CR.

2) Evaluation at the isolated array: The above analysis focused on receivers relatively close to the training positions. Here, validation is performed on an array of microphones V_4 placed at roughly 2 m from the training microphones. The results are shown in Figure 11. Even in this case, the difference in the NM between training and validation are basically independent on the room characteristics, but larger than in the sweet-spot. The degradation is even more prominent for larger rooms (around 5.5 dB), where the fact of having more and sharper resonances increases the chances of a higher variability in the spatial distribution of the modes in the room (i.e. the finite number of common poles selected at the sweet-spot may not correspond, at least in part, to the set of dominant modes in the isolated array). The analysis of the comparison between OBF and FIR filters at validation positions done above is still valid in this case, with the difference that the gap is smaller due to the higher NM at V_4. The trends seen in Figure 10 (fourth row) are recognizable also in this case, with the gap between OBF and FIR filters getting larger for increasing M in the medium room case, and no substantial difference for large rooms, with FIR filters slightly outperforming OBF filters. It follows that common poles estimated in a restricted area of a room do not correspond to all the poles of the system, so that more microphones should be placed around the room in order to extend the sweet-spot area. This is not a surprising result, whose implications are discussed later on in Section VI

B. Real room (SMARD database)

The same experiment and the same analysis described above in the case of simulated RIRs is repeated here for RIRs measured in a real rectangular room of dimension $7.34\,\mathrm{m} \times 8.09\,\mathrm{m} \times 2.87\,\mathrm{m}$ with a measured RT below 400 Hz of $T_{60} = 0.3$ s ($V = 209\,\mathrm{m}^3$, $N_{fs/2} = 1387$, $f_{\mathrm{Sch}} = 93$ Hz, $N_{f_{\mathrm{Sch}}} = 17$) [54]. Two loudspeakers and two orthogonal arrays were used in the simulations, the training set consisting of two loudspeakers and 8 microphones from the first array (inter-microphone distance ≈15 cm) and the 2 validation sets consisting of the same two loudspeakers and 4 microphones from the second array[4] The same room and configurations were also simulated using the RIM, so to compare the results with respect to the real room. Already from the comparison of the magnitude response at the top of Figure 12 (microphone 4Y, first array), it is clear that the simulated response does not match the actual response very closely, the main reason being the fact that in the simulated responses the same homogeneous broadband reflection coefficient was used for all the surfaces. The result is that the real response shows some strong and sharp resonances (probably due to a lack of absorption at low frequency) and a high degree of modal overlap at higher frequency, whereas the simulated response has sharp resonances also above 200 Hz.

The same kind of analysis as above is performed also in this case. The NM at training positions shows characteristics similar to previous simulation results, with a distance between OBF and FIR filters around -6 dB. For the simulated room, the NM decreases more slowly as a results of the larger number of spectral peaks to be modeled, as already discussed. The degradation at validation positions is around 3 dB, but with a larger variability if compared to the simulated room. The OBF-FIR gap (fourth row) increases only slightly for increasing M; for the real room, the reason being that, apart from few slowly decaying modes below 100 Hz, the response decays quite fast, so that it can be approximated with a relatively small number of coefficients of an FIR filter as well. For the simulated room the gap is only 1 dB, because the large number of spectral peaks makes the advantage of OBF over FIR filters less significant, as it was the case for the large rooms in the simulated environment discussed above. Finally, the CR results are in accordance with previous analysis, with OBF filters showing high efficiency especially for small M, i.e. when the most prominent modes are approximated with a small number of poles, such that $\gamma_M(\omega, \boldsymbol{p})$ has small values.

[4]more specifically, configurations 1000 and 1100 for the training set, and 1001 and 1101 for the validation sets (see http://www.smard.es.aau.dk/).

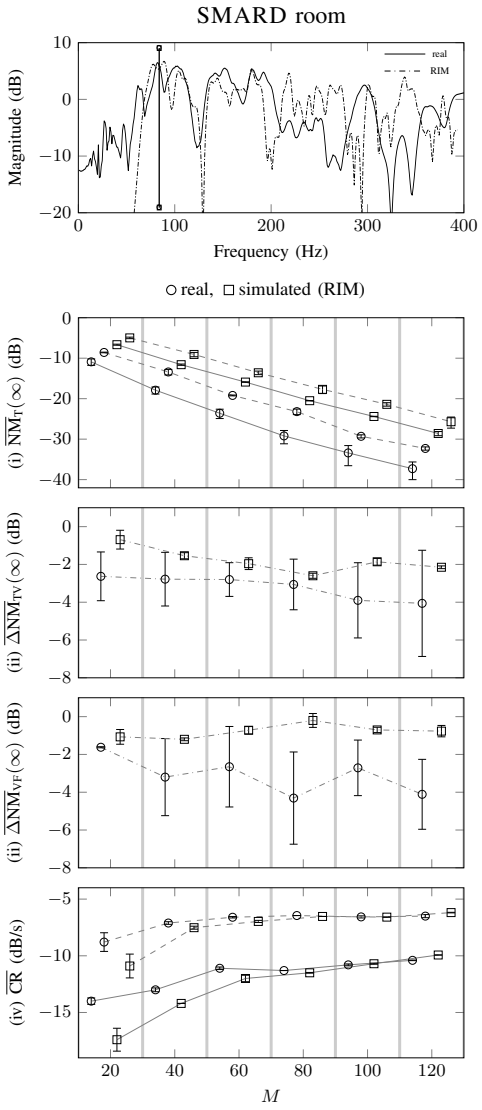

Fig. 12: *SMARD*: identification results for measures in (49-52), using OBF (solid) and FIR (dashed) filters with different orders M, and for the measured and simulated SMARD room. Top: example magnitude response and corresponding f_{Sch} (vertical line).

V. Applications in acoustic signal enhancement

In the previous section it has been shown that the use of OBF adaptive filters can bring an advantage compared to FIR filters in terms of identification accuracy and convergence behavior, depending on the characteristics of the room, and that common poles estimated at given positions inside a room can be considered valid at other positions in their vicinity. The aim of the current section is to show, by means of examples, how OBF adaptive filters can be practically used in RASE applications. It has been discussed in Section II how OBF filters with fixed poles can be seen as a generalization of FIR filters, also possessing similar properties in terms of coefficient adaptation. The only difference consists in the location of the fixed poles, which influences the estimation accuracy, but also convergence and the variability of the coefficients at steady-state. It follows that most of the tasks encountered in RASE applications can be tackled with the same strategies usually employed with FIR filters.

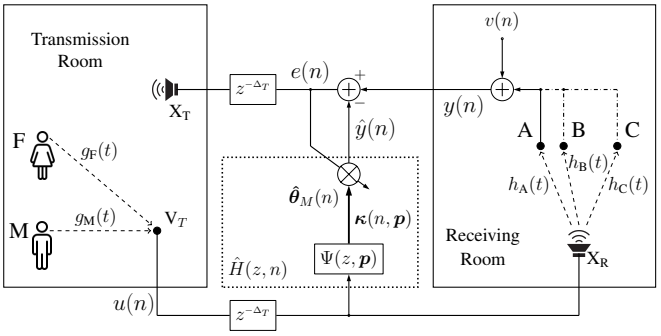

Fig. 13: Schematics of the AEC scenario using an OBF filter.

Even though the application examples presented in this section, as well as the previous analysis, focus on low frequencies, the extension to higher frequencies is straightforward, especially if a subband processing approach is used. The discussion about the use of OBF filters at higher frequencies is postponed to Section VI. In the following, examples are provided for two RASE tasks, namely AEC and RRE.

A. *Acoustic echo cancellation* (AEC)

The first scenario, depicted in Figure 13, considers a simple monophonic acoustic echo canceler. A male speaker M is talking in the transmission room (small RIM room, $T = 0.15$ s, position X, see Figure 8), and his voice signal $s_{\mathrm{M}}(t)$ is convolved with RIR $g_{\mathrm{M}}(t)$ and picked up by microphone $\mathrm{V_T}$ (first microphone, array V_3). The resulting (sampled and delayed by Δ_T samples) signal $u(n-\Delta_T)$ is reproduced in the receiving room (SMARD room, configurations 1000 and 1001) through loudspeaker $\mathrm{X_R}$. The reproduced speech signal is then convolved with RIR $h_{\mathrm{A}}(t)$, picked up by the microphone at position A (corresponding to microphone 1Y in the first orthogonal array) and, finally, the (sampled and delayed) microphone signal $y(n)$ is transmitted back through loudspeaker $\mathrm{X_T}$ to the transmission room, where it is perceived as echo. At a certain time t_1, the male speaker stops talking, whereas the female speaker F starts talking (position Y). After some time, at t_2, while F is still speaking, the microphone in the receiving room moves from position A to position B (microphone 3Y, first array) and then, at time t_3, to position C (microphone 7X, second array), so that the acoustic path in the receiving rooms first becomes $h_{\mathrm{B}}(t)$, and then $h_{\mathrm{C}}(t)$. Finally, at time t_4, the female voice stops, and the male speaker resumes talking, with the microphone in the receiving room still at position C. An external noise source is present in the receiving room, which produces a WN signal $v(t)$, assumed uncorrelated to the speech signal $u(t)$ and distributed in space (i.e. not localized), with $= 30$ dB SNR.

The aim of AEC is to identify the acoustic path in the receiving room so as to allow the removal of the echo from the microphone signal $y(n)$. For this purpose, the loudspeaker signal $u(n)$ is processed by a filter $\hat{H}(z,n)$, either a FIR or an OBF adaptive filter, whose linear coefficients are adapted using NLMS, thus producing the signal $\hat{y}(t)$. The poles of the OBF

adaptive filters are first identified using the SB-OBF-GMP algorithm from the male speech sequence [52] with voice activity detection (VAD) activated and using the configuration described in Section IV-B (only one sound source, maximum radius in the grid $\rho = 0.975$, and stage duration $T_s = 1\,\text{s}$). A set of 30 pole pairs is identified ($M = 60$) and then fixed in the OBF echo canceler, counting $m = M/2$ resonator sections having TFs $\Psi(z, \boldsymbol{p}_m^{\mathcal{A}}) = \{\Psi_1^{\pm}(z, \tilde{\boldsymbol{p}}_1^{\mathcal{A}}), \ldots, \Psi_m^{\pm}(z, \tilde{\boldsymbol{p}}_m^{\mathcal{A}})\}$. Two cases are considered for the order of the FIR echo canceler, $M_F = 60$ and $M_F = 80$, with the latter determined such that the steady-state NM at the training positions is roughly equal to the one obtained by the OBF filters in the identification (i.e. -17 dB). The step size is set to $\mu = 0.002$ for both the OBF and the FIR echo canceler, giving comparable convergence rate for both.

One drawback of using OBF filters in AEC consists in the necessity of estimating the acoustic delay parameter d (cfr. Figure 1). Different methods are available in the literature, from single-channel algorithms based on (generalized) cross-correlation or on adaptive filters, to more sophisticated multi-channel algorithms (see [55] for an overview). For simplicity, the acoustic delays of the acoustic paths used in the following examples (selected from the SMARD database) are assumed to be known and applied to both OBF and FIR filters.

In addition to the difficulties described in Section III-B about identifying the room acoustic system from speech signals, such as non-stationarity and the high-pass characteristics of speech, an extra difficulty is represented here by the presence of reverberation in the loudspeaker signal $u(n)$, which has a negative impact on the accuracy of the identification of the poles, compared to the case of an 'anechoic' loudspeaker signal, as can be seen in Figure 14. Reverberation in the transmission room introduces additional coloration in the loudspeaker signal, thus increasing the condition number C of the signal correlation matrix, which leads to slower convergence of the parameters (see Section II-B). Moreover, the spectral properties of the RTF of the transmission room have an influence on the excitation characteristics of the loudspeaker signal, so that an RTF in the receiving room may not be fully identifiable. In the case considered, the magnitude response of the Small simulated room (see top-left plot of Figure 10, $T = 0.15\,\text{s}$) presents small energy below 100 Hz and above 370 Hz, plus a number of deep anti-resonances. It follows that the system is not excited enough in those regions, thus making the identification more difficult. This is confirmed by the lack of OBFs in some parts of the spectrum, as seen in the bottom plot of Figure 14 showing the power responses of the OBFs built from the estimated poles $\boldsymbol{p}$ and the resulting shape of the $\gamma_M(\omega, \boldsymbol{p})$ factor in (11) compared to the two constant factors $\gamma_{M_F}(\omega) = M_F$ considered for the FIR echo cancelers. Nevertheless, as long as the characteristics of the speech signal are similar in terms of excitation to those of the speech signal used in the pole identification, the set of poles estimated in the training phase allows to achieve good cancellation performances, as shown in the following.

Apart from the NM in (30), a common measure to evaluate the effectiveness of an acoustic echo canceler is the echo return loss enhancement (ERLE), defined as the ratio of the power

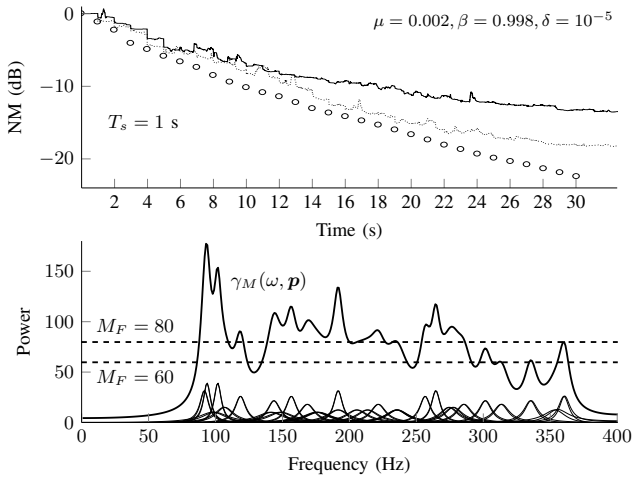

Fig. 14: Top: the NM for the OBF-GMP algorithm (○) and for the SB-OBF-GMP with 'anechoic' (dotted) and reverberated (solid) speech input signal (EBU-SQAM) in the AEC scenario. Bottom: The power responses of the 30 pairs of basis functions generated from the estimated pole set $\boldsymbol{p}$, and the resulting $\gamma_M(\omega, \boldsymbol{p})$ and $\gamma_{M_F}(\omega) = M_F$.

of the microphone signal in the receiving room and the power of the residual signal after cancellation,

$$\text{ERLE}(n) = 10 \log_{10} \left(\frac{S_y(n)}{S_e(n)} \right) \text{dB}, \tag{53}$$

where estimates of $S_y(n)$ and $S_e(n)$ of the two signals are obtained by low-pass filtering their instantaneous power. Figure 15 shows the NM and the ERLE obtained in the scenario described above using the standard NLMS for adapting the coefficients of both types of canceler. It can be seen that, when $M_F = M = 60$, the FIR canceler initially converges quickly to low values of NM, but then it starts diverging with the passing of time. This is a consequence of the order M_F of the FIR canceler being much shorter than the effective length of the RIR. Indeed, FIR filters using standard adaptive algorithms provide a biased estimate with a large variance when the system is undermodeled [56] (i.e. for RIRs truncated to a small number of samples), so that the unmodeled part of the RIR (i.e. its 'tail') contributes to an increase in the NM [57] and to a reduction of the ERLE. Increasing the order of the FIR canceler to $M_F = 80$ reduces this effect and a performance comparable to that of the OBF canceler with $M = 60$ is achieved. The OBF canceler, due to its IIR, is indeed less prone to misalignment problems related to a low filter order.

To test the robustness to undermodeling in the AEC case, the order of both cancelers is reduced to $M = M_F = 40$ (for the OBF filter only the first 20 estimated pole pairs are used). The results are shown in Figure 16, showing on one hand the aggravated misalignment problem for the FIR canceler, and, on the other hand, the well-behaved dynamic properties and the good misalignment performance of the OBF canceler even at a lower order. Also, it is worth noticing the difference in the ERLE compared to the previous case. It follows that OBF filters represent a more robust choice than FIR filters in AEC, especially when the number of adaptive coefficients is required to be low.

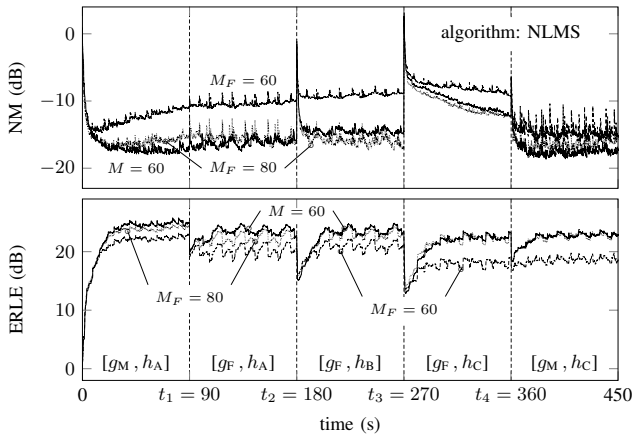

Fig. 15: The NM (top) and the ERLE (bottom) for the AEC scenario, using an OBF filter of order M and FIR filters of order M_F.

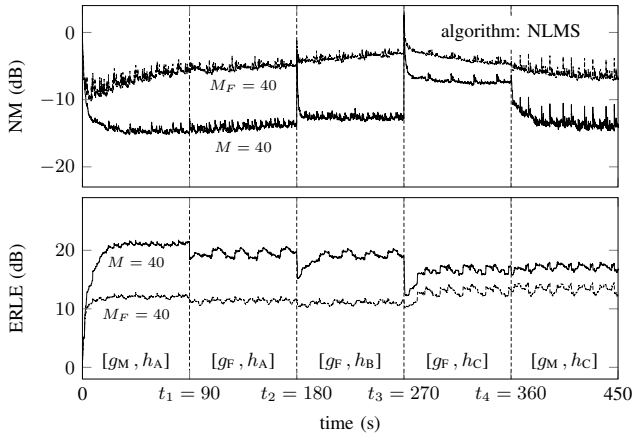

Fig. 16: The NM (top) and the ERLE (bottom) for the AEC scenario, using an OBF filter and a FIR filter, both of order $M = M_F = 40$.

Moreover, as discussed in Section II, OBF filters can be regarded as a generalization of FIR filters, so that most of the algorithms developed for the latter can be modified to work with the former as well (in most of the cases just substituting the expression of the gradient vector, as seen in Section II-B). The problem of slow convergence, normally originating from the poor excitation properties of the speech signals, can be addressed in different ways, such as using VSS algorithms (see [41]–[43] for an overview and examples), regularization techniques [38], or more complex adaptation algorithms. As an example, the scenario above is considered to evaluate the use of the APA [40] with projection order $Q = 8$ ($\mu = 0.005$, $\delta = 10^{-4}$). It is shown in Figure 17 how the increase in complexity improves the speed of convergence for both cancelers compared to when NLMS is used (see Figure 15), but all the other considerations still apply.

As a final note, the behavior of the echo cancellation algorithms when abrupt changes occur in the scenario is analyzed. First, notice at t_1, when the male speaker stops and the female speaker starts talking in the transmission room, that the degradation in both NM and ERLE is limited, with the coefficients of the canceler quickly adapting to the varied excitation characteristics of the speech signal. Modified

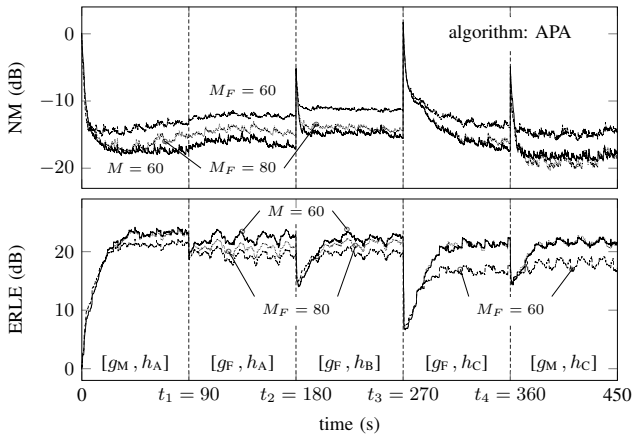

Fig. 17: The NM (top) and the ERLE (bottom) for the AEC scenario, using an OBF filter of order M and FIR filters of order M_F (APA).

conditions in the receiving room, instead, degrade the NM and the ERLE more dramatically, especially in the case (at t_3) when the receiver position moves to the second array. This is mostly due to the fact that, being at low frequencies, the RTF at position C presents more differences compared to position A and B, which belong to the same array. Instead, the common poles estimated on the first array are still valid at position C, as confirmed by the low NM in the last period, when the male speaker resumes talking.

It follows that a monophonic acoustic echo canceler needs to reconverge especially when large changes in the receiving room occur, whereas less problems are encountered when the speaker and/or its location varies in the transmission room. This is not true in SAEC [57], [58], where two microphones are located in the transmission room and two loudspeakers in the receiving room, in which case other issues have to be considered. The fact that the two loudspeaker signals $u_1(t)$ and $u_2(t)$ originate from a common source (the speech signal), convolved with two RIRs g_{M_1} and g_{M_2} in the transmission room, results in a high degree of their cross-correlation. In turn, this produces a very ill-conditioned covariance matrix, which leads to high misalignment and slow convergence [57], [59]. The misalignment can be reduced by using higher orders of the echo canceler, at the expense of slower convergence rate and higher ill-conditioning. Conversely, ill-conditioning is reduced if the order of the cancelers is much lower than the effective length of the RIRs in the transmission room, at the expense of a higher misalignment.

The IIR nature of OBF filters could help in achieving lower misalignment with a lower number of adaptive coefficients, thus meeting the second requirement more easily. However, it is often the case that using an IIR echo canceler does not solve the ill-conditioning problem to an acceptable degree. Moreover, the robustness to coloration of OBF filters and their good numerical conditioning properties discussed in Section II only refer to the monophonic case, whereas the ill-conditioning still remains in the stereophonic case. Thus, it is necessary to introduce a pre-processing stage with the aim of reducing the cross-correlation of the stereo signals. Common approaches achieve partial decorrelation by intro-

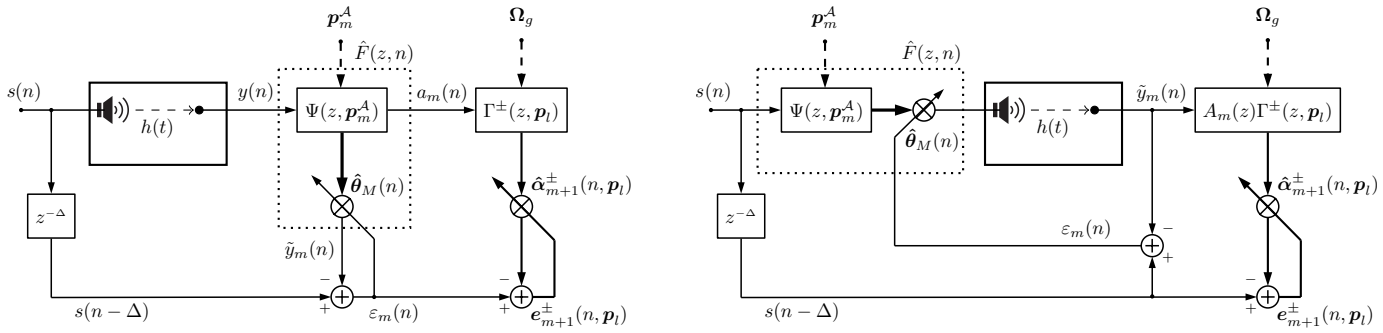

Fig. 18: The simplified schematics of the off-line (left) and the on-line (right) methods for pole estimation of an OBF equalizer.

ducing psychoacoustically-masked noise [58], by applying a non-linear function, such as the one proposed in [57], or by pre-processing the input signals with a pair of time-varying all-pass filters [60]. This last approach, with parameters as proposed in [61] for low frequency subbands (chosen in such a way not to introduce perceptible speech degradation), has proven to be an effective decorrelation method for the low-frequency case considered here. When the two loudspeaker signals are successfully decorrelated, poles of a multichannel OBF adaptive filter can be estimated similarly to the cases presented above, and the echo cancellation performance is comparable to the monophonic AEC.

B. Room response equalization (RRE)

The other RASE application considered is the equalization of RTFs (see [3] for an overview). In general terms, the aim of RRE is to correct the RTF for deviations from a desired target response, so as to improve the quality of the sound reproduced in the room or of the signal captured by a microphone. Digital filters are used to modify the frequency spectrum of the source signal before it is sent to the loudspeaker or after it is captured by a microphone, such that the spectrum of the equalized microphone signal is as close as possible to the source signal spectrum.

The use of OBF filters for minimum and nonminimum-phase equalization of room responses was proposed in [23]. It was suggested that the possibility of fixing the poles of the equalizer allows to obtain a desired frequency resolution and potentially reduce the number of filter parameters necessary for a given degree of equalization. The advantage of fixing the poles is the ability to control the resolution of the equalizer not only in terms of frequency distribution (the angle of the poles), but also in terms of frequency selectivity (the radius of the poles), so that an equalizer can be designed, for instance, to correct for sharp resonances and notches in low frequencies (by having dense poles with large radius) and to correct for the overall envelope of the magnitude response at higher frequencies. Another possibility is to limit the angle of the poles, with the pole with the lowest angle determining the high-pass cut-off frequency of the equalized response. Different pole placement strategies were proposed in [23], from the simple distribution of poles on the unit disc based on a desired frequency resolution, to the poles obtained from modeling an estimate of the inverse RTF. When the equalizer is designed from the minimum-phase inverse of a measured single-point RTF, the equalizer can only compensate for the magnitude room response. Moreover, a measured RTF has to be available or at least identified in advance. In [24], a blind equalization method is presented in which the RTF is first estimated as an FIR filter using a method relying on higher-order statistics. Then, the poles of the equalizer are placed in correspondence with the zeros of the RTF which are closest to the unit circle or obtained by an FIR-to-IIR conversion of the equalizer impulse response.

The example presented in the following focuses on adaptive single-input/single-output (SISO) equalization. The two scenarios considered, depicted in Figure 18, are those in which the signal to be processed by the equalization filter is either the microphone signal $y(n)$ (left scheme) or, alternatively, the source signal $s(n)$ before it is sent to the loudspeaker (right scheme). The coefficients of the equalizer with TF $\hat{F}(z,n)$ are adapted based on the residual error signal $\varepsilon(n)$ between the equalized microphone signal $\tilde{y}(n)$ and, assuming a flat target response, the delayed source signal. The equalization problem can be formalized as

$$\text{minimize } \mathbb{E}\{\varepsilon^2(n)\} = \mathbb{E}\{\big(s(n-\Delta)-\tilde{y}(n)\big)^2\}, \qquad (54)$$

where Δ is a modeling delay and $\tilde{y}(n) = H(q,n)\hat{F}(q,n)s(n)$ is the equalized microphone signal, i.e. the source signal filtered by the equalizer and convolved with the RTF at time n, $H(z,n)$. In case the equalizer is an OBF filter with a predefined set of poles $\boldsymbol{p}$, having TF $\hat{F}(z,n)$ defined as in (4), the minimization is performed by adapting the linear coefficients $\hat{\boldsymbol{\theta}}_i(n)$ (see Figure 18).

Here we propose two direct equalization filter design methods (i.e. not requiring a priori inversion of a RTF) which directly estimate the poles of the equalizer. Both methods use a modified version of the SB-OBF-GMP algorithm. The first method is an off-line design procedure that aims at estimating the poles by post-equalizing the microphone signal $y(n)$. The setting is typical of equalization intended to dereverberate the microphone signal, in which case equalization is not performed in the room during the estimation. The SB-OBF-GMP pole estimation algorithm can be used in this case, with the microphone signal $y(n)$ entering the OBF filter instead of the loudspeaker signal (compare the left schematics of Figure 18

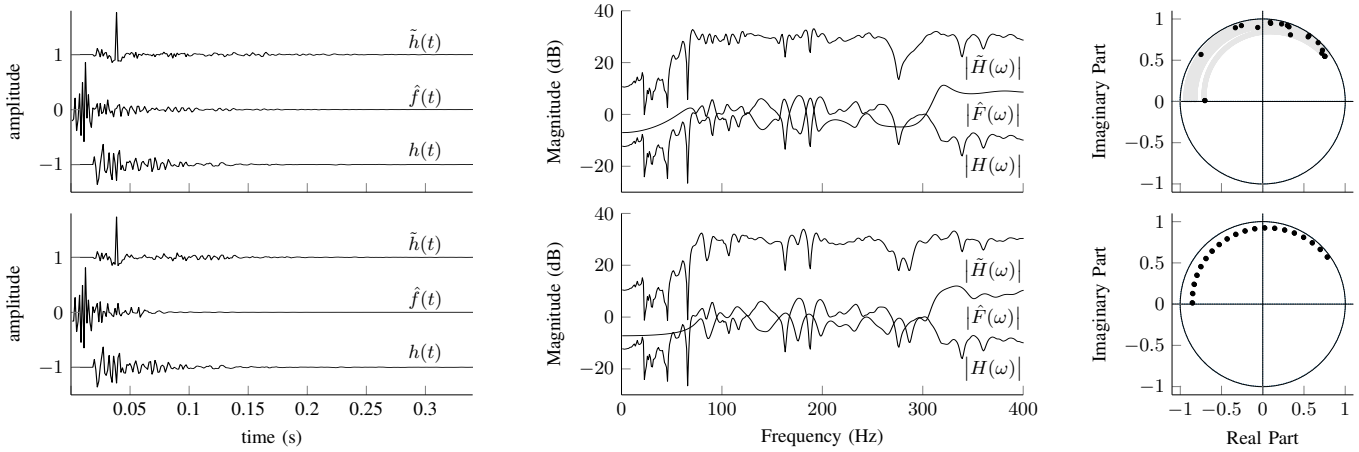

Fig. 19: Equalization example results using poles estimated with the proposed off-line method (top row) and using a fixed configuration of 20 pole pairs (bottom row). The plots show (left) the RIR $h(t)$, the equalizer impulse response $\hat{f}(t)$, the equalized RIR $\tilde{h}(t)$, and (center) their respective magnitude frequency responses $|H(\omega)|$, $|\hat{F}(\omega)|$, and $|\tilde{H}(\omega)|$ (the latter shifted by 20 dB).

with Figure 4). The pole selection strategy is unmodified, given that in this case the coefficients $\hat{\alpha}^{\pm}_{m+1}(n, \boldsymbol{p}_l)$ represent an estimate of the correlation between the residual error signal $\varepsilon_m(n)$ produced by the current OBF equalizer with m resonator sections having TFs $\Psi(z, \boldsymbol{p}^{\mathcal{A}}_m) = \{\Psi^{\pm}_1(z, \tilde{\boldsymbol{p}}^{\mathcal{A}}_1), \ldots, \Psi^{\pm}_m(z, \tilde{\boldsymbol{p}}^{\mathcal{A}}_m)\}$ and the output of the series of m all-pass filters processed by the candidate OBFs $\Gamma^{\pm}_i(z, \boldsymbol{p}_l)$ in the dictionary. If desired, once the poles are estimated, the equalizer can be moved in front of the loudspeaker, in a pre-equalization setup.

The second method is instead an on-line design procedure in which the equalizer built from the set of active (already selected) poles is placed in front of the loudspeaker, thus performing the actual equalization task in the room already during the pole estimation, while the pole selection is still performed at the microphones side (see right schematics of Figure 18). However, in this case, the regression vector required for the adaptation of the linear coefficients of the OBF filter (which would correspond to the intermediate signals convolved with the RIR) is not available, as they cannot be retrieved from the microphone signal. An option, to be verified, would be to compute the regression vector using an estimate of the RIR obtained at the previous iteration of the algorithm. Another drawback is related to the first stages of the algorithm, when the OBF equalizer has only few poles. In this case, the equalizer response being a linear combination of sparse resonant responses, the microphone signal, instead of being equalized, will present a resonant response, with a resulting degradation of the reproduced sound quality. A more useful application of the on-line method then consists in starting with a pre-equalizer built from a predefined set of poles (either chosen arbitrarily or pre-estimated with the off-line method), with the possibility of adding new poles while equalization is performed in the room.

Apart from the possibility of determining the order of the equalizer (by interrupting the algorithm when the desired equalization is achieved), the advantage of estimating the poles compared to distributing them in a fixed configuration consists in having frequency resolution determined by the optimization process and not by an initial choice. This is useful in particular when the order of the equalization filter is constrained to a small value (not too small, as said above), so that at each stage the one pole pair (among those available in the grid) for which the equalization error reduces the most is selected. When the number of poles increases, the algorithm tends to distribute them more evenly in the whole frequency range, so that estimating the poles may not bring a significant advantage over a fixed configuration.

As an example, equalization results are shown in the low frequency range of a RTF taken from the SMARD database (microphone 7Y, first array) [54]. The grid counts $L = 2000$ poles with 400 different angles distributed uniformly between 80 Hz and 399 Hz. The radius of the poles decreases exponentially at the increase of the angle [12], [23] according to $\rho_i = \varrho^{\frac{\vartheta_i}{\pi}}$, where 5 values for ϱ (the radius defined at the Nyquist frequency) were distributed logarithmically between 0.7 and 0.925, such that pole density toward the unit circle is increased (see shaded area in top-right plot of Figure 19). In this way, the equalizer will be able to correct for sharper deviations at lower frequencies and provide a smoother correction at higher frequencies. The equalization results of the OBF equalizer with $m = 20$ pole pairs estimated with the off-line method ($T_s = 0.5$ s) are compared with a fixed-pole OBF equalizer with $m = 20$ pole pairs distributed in the same way, but only for $\varrho = 0.85$ (bottom-right plot of Figure 19). The modeling delay was set to 35 samples and both equalizers are adapted with NLMS ($\mu = 0.002$, $\delta = 10^{-5}$).

From Figure 19 it can be seen that the resolution below 200 Hz is higher for the equalizer with estimated poles, with sharper peaks and dips in the equalizer magnitude response $|\hat{F}(\omega)|$. The result is a flatter equalized response $|\tilde{H}(\omega)|$, also with reduced depth of the anti-resonances. At higher frequencies, only 2 poles are estimated, against 9 in the fixed configuration, which explains the lower resolution in the former above 250 Hz. Also notice in both cases that the response is not equalized below 80 Hz, corresponding to the lowest frequency allowed for the poles. Concerning the time

domain, the presence of some estimated poles with large radius translates to a longer response of the equalization filter $\hat{f}(t)$ and a different distribution of the error in the equalized response $\tilde{h}(t)$ after the main pulse; for the case with estimated poles, this post-ringing is longer but with lower amplitude compared to the case with predefined poles. Also notice in both cases the presence of a significant pre-ringing effect, due to the low order of the equalizer and the short modeling delay. Both pre-ringing and post-ringing can be reduced by employing higher filter orders and longer modeling delays.

The extension to adaptive multi-point (SIMO) equalization [62], is straightforward for poles already fixed in the equalizer. The estimation of the poles using the two methods suggested is still possible, but presents limitations, mainly of a complexity nature. In the off-line configuration, poles could be estimated for each acoustic channel independently, but this would imply running the full algorithm in parallel for each microphone signal. For the on-line setup, a set of poles common to all channels could be computed, but still each microphone signal would need to be processed by the all-pass filter $A_m(z)$ and by the $2L$ candidate OBFs $\Gamma^{\pm}(z, \boldsymbol{p}_l)$. Moreover, the idea of having an equalizer with common poles would be motivated by the debatable assumption that not only resonances, but also anti-resonances are a characteristics of the room, independent of the source and receiver positions.

VI. Discussion

The main idea of representing a RTF by means of a pole-zero model is that a better approximation can be obtained by reducing the distance between the model poles and the true poles of the room acoustic system. From a different viewpoint, employing an IIR filter may help reducing the number of filter coefficients compared to the sampled truncated RIR representation inherent to modeling with FIR filters. The applicability of IIR filters in place of FIR filters in RASE applications has always been a matter of debate. A number of works [7], [8] concluded that the advantages are normally out-weighted by the increased difficulties encountered in the estimation and adaptation of the filter parameters, such as convergence to local minima or instability. It has been discussed in the first part of this paper that IIR adaptive filters based on OBFs make the problem of instability easy to keep under control (the information about the pole radius being readily available) and, by virtue of orthogonality, provide a well-conditioned estimation problem under a wide range of conditions. As a result, OBF adaptive filters show the fastest convergence among fixed-poles IIR filters. Moreover, OBF adaptive filters can be seen as a generalization of FIR filters, so that, when poles are fixed, most filter adaptation algorithms developed for FIR filters can be easily applied to the OBF case as well.

When approaching a RASE problem, the first thing to be considered is whether OBF adaptive filters can bring an advantage compared to the common FIR filters. Since stability and convergence are not an issue, the choice should be made with respect to different criteria, based on the specific application scenario. First, the characteristics of the room play an important role in defining the expected performance in the identification of the room acoustic system using OBF filters. To summarize the findings from Section IV, an OBF filter can provide a significant advantage when it is able to approximate the RTF with a small number of filter parameters, i.e. when the RTF is characterized by either a small number of sparse resonances (such as at low frequency in small rooms) or a significant modal overlap (such as in rooms with high absorption). As the volume of the room increases, and with that the RT, the number of relatively sharp modes increases as well, and consequently the number of spectral peaks that have to be modeled. These results were observed in the low-frequency band considered (0-400 Hz), for which the modal density can be high, but possibly not high enough to be able to approximate the superposition of multiple resonances with a small number of poles. Moreover, real rooms present non-homogeneous characteristics at different frequency regions, as noticed in the SMARD database example. Thus, simulation results presented should be used just as an indication, as the performance in real scenarios may differ due to a number of different reasons.

Second, the potentially expected reduction in the filter order and the use of a common set of poles may not only bring computational savings, but also help in overcome some of the problems encountered in RASE applications, such as undermodeling, as suggested in the AEC example. It has been shown that common poles estimated at low frequencies in a limited area of the room provide a good approximation also at locations in the surrounding area, but less so in other parts of the room. It follows that the estimation of the poles might have to be performed in a wider area, with ensuing issues pertaining to the spatial sampling of the microphone setup. This may not be too much of an issue at higher frequency, where the higher modal overlap reduces the variability of the modal distribution in space, as suggested above. Third, the possibility of fixing the poles in the filter allows to achieve a desired frequency resolution, either by fixed configurations of the poles or by estimation algorithms, as suggested in the RRE example. This is a feature not encountered in FIR filters, unless warping techniques are used.

However, the use of OBF filters is beneficial only provided that the poles are well estimated, or at least fixed in a meaningful way according to some prior knowledge or some desired properties of the filter. The estimation of the poles is then possibly the most critical issue for a wide-spread adoption of OBF filters in RASE applications. Whereas stable and effective modeling algorithms are already available, also for common poles (see [12] and Section I), the identification from input-output signals is more challenging. The identification algorithm proposed in this work proved to be effective in estimating a common set of poles for SIMO and MIMO systems, also from speech signals. Nonetheless, the identification is a time and resource-consuming task, not only with the proposed method, but also when recursive algorithms are applied.

It follows that in some situations, it may not be possible to identify the poles from input-output signals during the actual RASE task. In these cases, one option is to estimate common poles in advance, and then keep them fixed during the task. Another possibility is to first obtain a solution for the problem

at hand using already available methods for FIR filters, and then convert, with one of the modeling methods cited, the obtained filter to an OBF filter with reduced order, whose linear coefficients can be adapted to track RTF variations.

Future work should then address the identification problem. Gradient-based recursive algorithms, as discussed in Section II-D, should be investigated further. The proposed algorithm could be used to obtain good initial values for the pole parameters, so that only few iterations of the recursive algorithm should suffice. This would probably overcome the limitations imposed by the discrete resolution of the pole grid and attain improved performance, at least in some situations (e.g. when modes are sharp and only moderately overlapping). The actual improvement achievable by refining the pole parameters should be verified, also taking into account the necessity of adapting the pole parameters to track possible changes in the room acoustic system.

Modeling and identification at higher frequency bands should be also assessed. It can be expected at higher frequencies that the increased modal density and the increased absorption would result in favorable conditions for the applicability of OBF filters. Subband modeling seems to be a promising direction [15], but more in-depth analysis is required to understand if the potential advantages over FIR filters are consistent enough to justify the additional processing.

VII. Conclusion

In this paper, the use of OBF adaptive filters in room acoustic system identification and RASE applications has been considered. Since some of the problems typical of IIR filters are not encountered in the OBF case, the choice between an OBF and a FIR filter should be made based on application requirements and the characteristics of the application scenario. One aim of this paper was to provide the reader with the knowledge and some of the necessary tools to make an informed decision in this regard.

The main properties of OBF adaptive filters have been reviewed, with a focus on the error performance and dynamic behavior of filter adaptation algorithms; in this context, a modified version of the NLMS algorithm (analogous to the adaptation rule in TD algorithms and named OBF-NLMS) has been suggested to deal with issues at very low model orders. An identification algorithm has been proposed, capable of estimating a set of common poles for a MIMO room acoustic system, from both WN and speech signals. The algorithm has been used to identify the RTFs at low frequencies in different scenarios for real and simulated rooms, highlighting the relation between the characteristics of the room, such as its volume and it reverberation time, and the expected performance of OBF and FIR filters. Although the analysis has been performed at low frequencies, which is already of interest in applications such as RRE or SAEC, the methods and algorithms presented are applicable at higher frequency as well, especially if a subband approach is adopted.

Finally, examples in the context of two RASE applications were given. In the AEC example, it has been shown that OBF filters can provide good identification and cancellation performances already for small model orders, for which FIR filters may encounter undermodeling problems. In the RRE example, two methods were presented to directly estimate the poles of an equalizer in a SISO scenario, useful to allocate frequency resolution where it is more necessary, and to keep the order of the equalizer as low as possible. To conclude, OBF adaptive filters represent a useful and flexible tool for approaching most of the problems encountered in RASE applications, whose adoption should be considered based on each specific case, and whose possibilities may still have to be explored.

Appendix A
Gradient expressions for the poles

Here the full and approximated expressions for the gradients with respect to the kth pole pair defined as $\boldsymbol{p}_k = [\zeta_k, \eta_k] = [-2\rho_k \cos\vartheta_k, \rho_k^2]$ are given. After simple calculations and some reordering (and omitting the time index n and the q operator, s.t. $D_k \equiv D_k(n,q)$ and $\chi_k \equiv \chi_k(n) = \{\zeta_k, \eta_k\}$), the partial derivatives appearing in the second and third terms of equation (34) are given by

$$\frac{\partial \tilde{y}_k(n)}{\partial \chi_k} = \left(\theta_k^+ \frac{\partial N_k^+}{\partial \chi_k} + \theta_k^- \frac{\partial N_k^-}{\partial \chi_k}\right) x_k(n) - \frac{1}{D_k}\frac{\partial D_k}{\partial \chi_k}\tilde{y}_k(n)\,, \tag{55}$$

$$\frac{\partial \tilde{y}_i(n)}{\partial \chi_k} = \left(\theta_i^+ N_i^+ + \theta_i^- N_i^-\right) \frac{\prod_{j=k+1}^{i-1} A_j}{D_i} \frac{\partial \bar{D}_k}{\partial \chi_k} x_k(n) - \frac{1}{D_k}\frac{\partial D_k}{\partial \chi_k}\tilde{y}_i(n)\,, \tag{56}$$

where $x_k(n) = P_k \prod_{j=1}^{k-1} A_j u(n)$ (the output of the kth resonator) and $\tilde{y}_i(n), \forall i$, are readily available signals. By noticing in (3) that $|1 \pm p_k| = \sqrt{1 \mp \zeta_k + \eta_k}$, the partial derivatives in (55) and (56) with respect to ζ_k are given as

$$\begin{aligned} \frac{\partial N_k^+}{\partial \zeta_k} &= -\frac{\sqrt{c}}{2\sqrt{a}}(z^{-1} - 1)\,, \\ \frac{\partial N_k^-}{\partial \zeta_k} &= \frac{\sqrt{c}}{2\sqrt{b}}(z^{-1} + 1)\,, \\ \frac{\partial D_k}{\partial \zeta_k} &= z^{-1}\,, \quad \frac{\partial \bar{D}_k}{\partial \zeta_k} = z^{-1}\,, \end{aligned} \tag{57}$$

with $a = 1 - \zeta_k + \eta_k$, $b = 1 + \zeta_k + \eta_k$ and $c = {}^{(1-\eta_k)}/_2$, whereas the same expressions with respect to η_k are given as

$$\begin{aligned} \frac{\partial N_k^+}{\partial \eta_k} &= \frac{1}{2}\left[\frac{\sqrt{c}}{\sqrt{a}} - \frac{\sqrt{a}}{2\sqrt{c}}\right](z^{-1} - 1)\,, \\ \frac{\partial N_k^-}{\partial \eta_k} &= \frac{1}{2}\left[\frac{\sqrt{c}}{\sqrt{b}} - \frac{\sqrt{b}}{2\sqrt{c}}\right](z^{-1} + 1)\,, \\ \frac{\partial D_k}{\partial \eta_k} &= z^{-2}\,, \quad \frac{\partial \bar{D}_k}{\partial \eta_k} = 1\,. \end{aligned} \tag{58}$$

The first term of expression (56), however, can be computationally expensive, especially for small k. For this reason, a simplification has been proposed in [28] assuming slow convergence of the parameters and poles close to the unit circle, which allows to approximate the gradient only by the term in (55).

REFERENCES

[1] C. Breining, P. Dreiscitel, E. Hansler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, "Acoustic echo control. an application of very-high-order adaptive filters," *IEEE Signal Process. Mag.*, vol. 16, no. 4, pp. 42–69, 1999.

[2] T. van Waterschoot and M. Moonen, "Fifty years of acoustic feedback control: State of the art and future challenges," *Proc. IEEE*, vol. 99, no. 2, pp. 288–327, 2011.

[3] S. Cecchi, A. Carini, and S. Spors, "Room response equalization - a review," *Appl. Sci.*, vol. 8, no. 1, p. 16, 2017.

[4] 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.

[5] 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.

[6] J. J. Shynk, "Adaptive IIR filtering," *IEEE ASSP Mag.*, vol. 6, no. 2, pp. 4–21, 1989.

[7] 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.

[8] S. Gudvangen and S. Flockton, "Modelling of acoustic transfer functions for echo cancellers," *IEE Proc.-Vis Image Signal Process.*, vol. 142, no. 1, pp. 47–51, 1995.

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

[10] 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.

[11] 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.

[12] G. Vairetti, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. van Waterschoot, "A scalable algorithm for physically motivated and sparse approximation of room impulse responses with orthonormal basis functions," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 7, pp. 1547–1561, 2017.

[13] 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. 10th Eur. Congr. Expo. Noise Control Eng. (EURONOISE '15)*, 2015, pp. 149–154.

[14] 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.

[15] S. Hashemgeloogerdi and M. Bocko, "High precision robust modeling of long room responses using wavelet transform," in *Proc. 2017 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '17)*, 2017, pp. 356–360.

[16] G. Williamson and S. Zimmermann, "Globally convergent adaptive IIR filters based on fixed pole locations," *IEEE Trans. Signal Process.*, vol. 44, no. 6, pp. 1418–1427, 1996.

[17] B. Ninness and J. Gómez, "Frequency domain analysis of tracking and noise performance of adaptive algorithms," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1314–1332, 1998.

[18] B. Ninness, H. Hjalmarsson, and F. Gustafsson, "The fundamental role of general orthonormal bases in system identification," *IEEE Trans. Automat. Control*, vol. 44, no. 7, pp. 1384–1406, 1999.

[19] B. Ninness and H. Hjalmarsson, "Model structure and numerical properties of normal equations," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 48, no. 4, pp. 425–437, 2001.

[20] A. Kaelin, A. Lindgren, and G. Moschytz, "Simplified adaptive IIR filters based on optimized orthogonal prefiltering," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 42, no. 5, pp. 326–333, 1995.

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

[22] S. Hashemgeloogerdi and M. Bocko, "Inherently stable weighted least-squares estimation of common acoustical poles with the application in feedback path modeling utilizing a Kautz filter," *IEEE Signal Process. Lett.*, vol. 25, no. 3, pp. 368–372, 2018.

[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] M. Campi, R. Leonardi, and L. A. Rossi, "Generalized super-exponential method for blind equalization using Kautz filters," in *Proc. IEEE Signal Process. Workshop Higher-Order Stat. (SPW-HOS '99)*, 1999, pp. 107–111.

[25] L. S. Ngia, "Separable nonlinear least-squares methods for efficient off-line and on-line modeling of systems using Kautz and Laguerre filters," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 48, no. 6, pp. 562–579, 2001.

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

[27] L. Salama and J. Cousseau, "Efficient echo cancellation based on an orthogonal adaptive IIR realization," in *Proc. SBT/IEEE Int. Telecommun. Symp. (ITS'98)*, vol. 2, 1998, pp. 434–437.

[28] J. Cousseau, P. Diniz, G. Sentoni, and O. Agamennoni, "On orthogonal realizations for adaptive IIR filters," *Int. J. Circuit Theory Appls.*, vol. 28, no. 5, pp. 481–500, 2000.

[29] S. Narayan, A. Peterson, and M. Narasimha, "Transform domain LMS algorithm," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 31, no. 3, pp. 609–615, 1983.

[30] J. Lee and C. Un, "Performance of transform-domain LMS adaptive digital filters," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 34, no. 3, pp. 499–510, 1986.

[31] B. Farhang-Boroujeny and S. Gazor, "Selection of orthonormal transforms for improving the performance of the transform domain normalised LMS algorithm," in *IEE Proc. F (Radar Signal Process.)*, vol. 139, no. 5, 1992, pp. 327–335.

[32] S. Gazor and B. Farhang-Boroujeny, "Quantization effects in transform-domain normalized LMS algorithm," *IEEE Trans. Circuits Syst. II: Analog Digital Signal Process.*, vol. 39, no. 1, pp. 1–7, 1992.

[33] F. Beaufays, "Transform-domain adaptive filters: An analytical approach," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 422–431, 1995.

[34] G. Vairetti, S. H. Jensen, E. De Sena, M. Moonen, M. Catrysse, and T. van Waterschoot, "Multichannel identification of room acoustic systems with adaptive filters based on orthonormal basis functions," in *Proc. 2016 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '16)*, 2016, pp. 16–20.

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

[36] S. Gunnarsson and L. Ljung, "Frequency domain tracking characteristics of adaptive algorithms," *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. 37, no. 7, pp. 1072–1089, 1989.

[37] S. Haykin, *Adaptive filter theory*. Pearson Education India, 2008.

[38] T. van Waterschoot, G. Rombouts, and M. Moonen, "Optimally regularized adaptive filtering algorithms for room acoustic signal enhancement," *Signal Process.*, vol. 88, no. 3, pp. 594–611, 2008.

[39] D. R. Morgan and S. G. Kratzer, "On a class of computationally efficient, rapidly converging, generalized NLMS algorithms," *IEEE Signal Process. Lett.*, vol. 3, no. 8, pp. 245–247, 1996.

[40] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electron. Commun. Jpn. (Part I: Commun.)*, vol. 67, no. 5, pp. 19–27, 1984.

[41] A. Mader, H. Puder, and G. U. Schmidt, "Step-size control for acoustic echo cancellation filters–an overview," *Signal Process.*, vol. 80, no. 9, pp. 1697–1719, 2000.

[42] J.-K. Hwang and Y.-P. Li, "Variable step-size LMS algorithm with a gradient-based weighted average," *IEEE Signal Process. Lett.*, vol. 16, no. 12, pp. 1043–1046, 2009.

[43] Y. Zhang, N. Li, J. A. Chambers, and Y. Hao, "New gradient-based variable step size LMS algorithms," *EURASIP J. Advances Signal Process.*, vol. 2008, p. 105, 2008.

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

[45] E. De Sena, N. Antonello, M. Moonen, and T. van Waterschoot, "On the modeling of rectangular geometries in room acoustic simulations," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 4, pp. 774–786, 2015.

[46] L. Ljung and T. Söderström, *Theory and practice of recursive identification*. MIT press, 1983.

[47] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. Springer, 2006, ch. 3 Line Search Methods, pp. 30–65.

[48] 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.

[49] I. R. Titze, *Principles of voice production*. Englewood Cliffs, Prentice Hall, 1994.

[50] J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," *IEEE Signal Process. Lett.*, vol. 6, no. 1, pp. 1–3, 1999.

[51] M. Brookes, "VOICEBOX: A speech processing toolbox for MATLAB." Imperial College, Software Library, 2011. [Online]. Available: http://www.ee. imperial.ac.uk/hp/staff/dmb/voicebox/voicebox.html

[52] G. Waters, "Sound quality assessment material – recordings for subjective tests: Users handbook for the EBU–SQAM compact disk," European Broadcasting Union (EBU), Tech. Rep., 1988.

[53] M. Mboup and M. Bonnet, "On the adequateness of IIR adaptive filtering for acoustic echo cancellation," in *Proc. 6th Eur. Signal Process. Conf. (EUSIPCO '92)*, vol. 92, no. 1, 1992, pp. 111–114.

[54] J. K. Nielsen, J. R. Jensen, S. H. Jensen, and M. G. Christensen, "The single- and multichannel audio recordings database (SMARD)," in *Proc. 2014 Int. Workshop Acoust. Signal Enhancement (IWAENC '14)*, 2014, pp. 40–44.

[55] J. Chen, J. Benesty, and Y. A. Huang, "Time delay estimation in room acoustic environments: an overview," *EURASIP J. Advances Signal Process.*, vol. 2006, no. 1, p. 026503, 2006.

[56] G. Rombouts, T. van Waterschoot, K. Struyve, P. Verhoeve, and M. Moonen, "Identification of undermodelled room impulse responses," in *Proc. 2005 Int. Workshop Acoust. Echo Noise Control (IWAENC '05)*, Eindhoven, The Netherlands, September 2005, pp. 153–156.

[57] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 2, pp. 156–165, 1998.

[58] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic acoustic echo cancellation - an overview of the fundamental problem," *IEEE Signal Process. Lett.*, vol. 2, no. 8, pp. 148–151, 1995.

[59] T. Gänsler and J. Benesty, "Stereophonic acoustic echo cancellation and two-channel adaptive filtering: an overview," *Int. J. Adapt. Control Signal Process*, vol. 14, no. 6, pp. 565–586, 2000.

[60] M. Ali, "Stereophonic acoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation," in *Proc. 1998 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '98)*, vol. 6. IEEE, 1998, pp. 3689–3692.

[61] T. Hoya, J. A. Chambers, and P. A. Naylor, "Low complexity ϵ-NLMS algorithms and subband structures for stereophonic acoustic echo cancellation," in *Proc. 1999 Int. Workshop Acoust. Echo Noise Control (IWAENC '99)*, 1999, pp. 36–39.

[62] S. J. Elliott and P. A. Nelson, "Multiple-point equalization in a room using adaptive digital filters," *J. Audio Eng. Soc.*, vol. 37, no. 11, pp. 899–907, 1989.