Online adaptive identification of multichannel systems for audio applications

Guilhem Pagès, Roberto Longo, Laurent Simon, Manuel Melon

JASA ARTICLE

Online adaptive identification of multichannel systems for audio applications

Guilhem Pagès,[a] Roberto Longo, Laurent Simon, and Manuel Melon
Laboratoire d'Acoustique de l'Université du Mans (LAUM), Unité Mixte de Recherche 6613, Institut d'Acoustique–Graduate School (IA-GS), Centre National de la Recherche Scientifique, Le Mans Université, France

ABSTRACT:
Impulse responses (IRs) estimation of multi-input acoustic systems is a prerequisite for many audio applications. In this paper, an adaptive identification problem based on the Autostep algorithm is extended to the simultaneous estimation of room IRs for multiple input single output linear time invariant systems without any *a priori* information. To do so, the proposed algorithm is initially evaluated in a simulated room with several sound sources active at the same time. Finally, an experimental validation is proposed for the cases of a semi-anechoic chamber and an arbitrary room. Special attention is dedicated to the algorithm convergence behavior, considering different meta parameters settings. Results are eventually compared with the other normalized version of the least mean square algorithm.
 https://doi.org/10.1121/10.0024149
(Received 7 July 2023; revised 11 November 2023; accepted 8 December 2023; published online 8 January 2024)
[Editor: Efren Fernandez-Grande] Pages: 229–240

I. INTRODUCTION

Estimation of multichannel impulse responses (IRs) is at the core of many audio applications. Multichannel audio systems are increasingly used for sound reinforcement applications, three-dimensional (3D) sound using techniques such as wave field synthesis (WFS; Berkhout *et al.*, 1993) or higher order ambisonics (HOA; Ward and Abhayapala, 2001), or for the implementation of personalized sound zones (Choi and Kim, 2002; Hu *et al.*, 2023; Moller and Ostergaard, 2020; Vindrola *et al.*, 2021). Room response compensation or equalization (Kodrasi and Doclo, 2016; Yoshioka and Nakatani, 2012), active noise control (Hu *et al.*, 2019), or echo cancellation (Makino, 2001) may also be considered as common applications requiring multichannel IRs estimation. For these applications, joint *in situ* estimation of the IRs is a key factor in optimizing reproduction performance and/or for controlling the sound field.

In the fields of room acoustics and audio system characterization, the diversity of IR shapes (time support from a few tens of milliseconds to several seconds and energy distribution from sparse to dense) requires the adjustment of many control parameters. The development of methods for estimating multichannel IRs in a tuning-free/data-driven format is, therefore, a real challenge and the topic of this paper.

Although it is common to focus on linear time-invariant (LTI) systems in the first instance, it may be necessary for audio applications to consider estimation methods that can be extended to slowly varying systems (nonstationary environment and varying source characteristics). Furthermore, online estimation may be a constraint directly linked to the application as for sound field control. For all these reasons, the methods considered in this paper are based on adaptive filtering (Haykin, 2014).

Most of adaptive filtering techniques rely on least mean square (LMS) algorithms, for which filter coefficients are calculated by minimizing the mean square of the error signal (Haykin, 2014). The optimization process is related to the stochastic gradient descent method as the filter is adapted by considering only the error at the current time (or adaptation cycle), n. To modulate the gradient descent toward the global minimum, LMS makes use of a scalar step size parameter, μ, which depends on the physical system under test. To avoid this dependence, the normalized version of least mean square algorithm (NLMS) proposes to divide this parameter (noted as $\tilde{\mu}$) by an instantaneous estimate of the input signal energy, leading to a time-varying step size parameter, $\mu[n]$, at the adaptation cycle, n. The values of these parameters (μ for LMS and $\tilde{\mu}$ for NLMS) must be initialized according to the desired performance: a low estimation error or a fast convergence rate (the two performances being exclusive; Haykin, 2014).

To overcome the speed/error trade-off, variable step size least mean square algorithms (VS-LMS) have been developed (Bismor *et al.*, 2016). Many solutions exist to adapt the parameter, which are often specialized for a particular application. Some of these algorithms are penalized by the number of parameters to be adjusted (Bismor *et al.*, 2016). The VS-LMS algorithms of Harris *et al.* (1986) and Mikhael *et al.* (1986) propose to replace the scalar parameter, $\mu[n]$, by a vector (or a diagonal matrix), $\boldsymbol{\mu}[n]$, to tune more finely the direction of the gradient descent. Note that the first algorithm requires five parameters to be set while the second requires only one. More recently, new

[a]Email: guilhem.pages@univ-lemans.fr

algorithms, called meta gradients, propose to adapt the vector parameter, $\boldsymbol{\mu}[n]$, by gradient descent to adapt automatically to the problem under analysis. Incremental delta bar delta (IDBD; Sutton, 1992) is one of these algorithms, which, however, requires setting one parameter. To get rid of the setting of this parameter, Mahmood (2010) proposes the Autostep algorithm, which then has no parameter to set.

In this paper, the Autostep algorithm is extended to multiple input single output (MISO) systems, i.e., multichannel IR estimation. The performance of this proposed unsupervised algorithm is evaluated using simulated and measured room IRs and compared to the results of MISO-NLMS algorithm.

The paper is structured as follows: in Sec. II, the general problem is developed from a theoretical point of view. The definition of the evaluation metrics and influence of the tuning parameters are detailed in Sec. III. In Sec. IV, the Autostep algorithm is tested in a simulated environment. An experimental comparison between the NLMS and Autostep algorithms is presented in Sec. V for two different rooms. Finally, conclusions are given in Sec. VI.

Notations: Scalars, vectors, and matrices are denoted with ordinary lowercase symbols, bold lowercase symbols, and bold uppercase symbols, respectively. The superscripts $(\cdot)^\top$ and $(\cdot)^{(l)}$ denote the transposition and input l, respectively. The subscript $(\cdot)_n$ denotes the time n, $(\circ)$ and $(\oslash)$ denote the Hadamard product and element-wise division, respectively. For a vector $\boldsymbol{a}$, we denote $f[\boldsymbol{a}]$ as the vector $[f(a_i)]$ such that the function f has been applied to each element ai of the vector $\boldsymbol{a}$.

II. THEORETICAL BACKGROUND

In this section, NLMS and Autostep algorithms are presented in the context of LTI MISO systems.

A. Identification of MISO LTI system with adaptive filtering

We consider an unknown LTI MISO system, consisting in L inputs, $x^{(l)}$, and a single output, y, where all are real-valued. Throughout this paper, we consider $l \in [1, L]$, where $L > 1$ is the number of inputs. Finite impulse response (FIR) adaptive filters of order N are considered here. Consequently, we note the $N \times 1$ vector of the optimal FIR filter, i.e., the linear estimate of the IR related to input l,

$$\boldsymbol{h}^{(l)} = \left[h_0^{(l)}, h_1^{(l)}, \ldots, h_{N-1}^{(l)}\right]^\top. \tag{1}$$

Eventually, the identification block diagram is shown in Fig. 1. At current time n, the output sequence, $\tilde{y}_n$, generated by the optimal filter is defined as

$$\tilde{y}_n = \sum_{l=1}^{L}\sum_{m=0}^{N-1} h_m^{(l)} x_{n-m}^{(l)} = \sum_{l=1}^{L} \boldsymbol{h}_n^{(l)^\top} \boldsymbol{x}_n^{(l)}, \tag{2}$$

where the $N \times 1$ input vector,

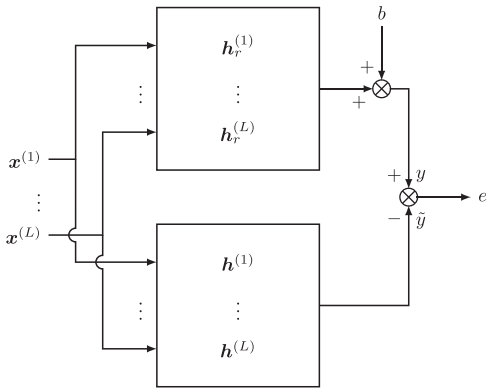

FIG. 1. Block diagram of the adaptive filtering problem, showing (upper block) real (unknown) IRs, $\boldsymbol{h}_r^{(l)}$ (where r is the subscript for real) and (lower block) recursively estimated IRs, $\boldsymbol{h}^{(l)}$. Outputs are such that y is the measured output signal corrupted by additive noise b, and $\tilde{y}$ is the output sequence of the adaptive filter.

$$\boldsymbol{x}_n^{(l)} = \left[x_n^{(l)}, x_{n-1}^{(l)}, \ldots, x_{n-N+1}^{(l)}\right]^\top, \tag{3}$$

is composed of the last N samples of input l.

Last, we write the estimation error, e_n, at discrete time n such that

$$e_n = y_n - \tilde{y}_n = y_n - \sum_{l=1}^{L} \boldsymbol{h}_n^{(l)^\top} \boldsymbol{x}_n^{(l)}. \tag{4}$$

The objective is then to estimate the IRs, $h^{(l)}$, knowing the input signals, $x^{(l)}$, and the output signal, y.

B. NLMS algorithm for MISO system identification

We recall here the main results of the NLMS algorithm in the MISO configuration (Goodwin and Sin, 2009). Introducing the step size parameter, $\tilde{\mu}^{(l)}$, of input l, the weight vector, $\boldsymbol{h}^{(l)}$, can be recursively updated as

$$\boldsymbol{h}_{n+1}^{(l)} = \boldsymbol{h}_n^{(l)} + \frac{\tilde{\mu}^{(l)}/L}{\boldsymbol{x}^{(l)^\top}\boldsymbol{x}^{(l)} + \epsilon} e_n \boldsymbol{x}_n^{(l)}, \tag{5}$$

where ϵ is a small scalar value to avoid division by zero.

To ensure the convergence, $\tilde{\mu}^{(l)}$ must satisfy

$$0 < \tilde{\mu}^{(l)} \leq 2, \tag{6}$$

as detailed in Algorithm 1.

ALGORITHM 1. MISO-NLMS.

Require: $\tilde{\mu}^{(l)} \in]0, 2]$
Ensure: $\boldsymbol{h}_n$
$\boldsymbol{h} \leftarrow \boldsymbol{0}$
while $\{x_n, y_n\}$ **do**
  $e_n \leftarrow y_n - \sum_{l=1}^{L} \boldsymbol{h}_n^{(l)^\top} \boldsymbol{x}_n^{(l)}$
  $\boldsymbol{h}_{n+1}^{(l)} \leftarrow \boldsymbol{h}_n^{(l)} + \tilde{\mu}^{(l)} e_n \boldsymbol{x}_n^{(l)} / L(\boldsymbol{x}_n^{(l)^\top} \boldsymbol{x}_n^{(l)})$
end while

https://doi.org/10.1121/10.0024149

C. Autostep algorithm for MISO system identification

In a MISO problem, inputs can present different statistical properties. In the case of the NLMS algorithm, the values of the step size parameters, $\tilde{\mu}^{(l)}$, have to be adapted according to the *a priori* expected performances. Another solution is to consider that these parameters must be adapted according to a given update rule. This is, for instance, what is proposed by the IDBD algorithm (Sutton, 1992) for the identification of single input single output (SISO) systems, as previously mentioned in the Introduction.

Following Sutton (1992) and extending the purpose to the identification of MISO systems, we first consider the step size parameter as an $N \times 1$ vector,

$$\boldsymbol{\mu}^{(l)} = \left[\mu_0^{(l)}, \mu_1^{(l)}, \ldots, \mu_{N-1}^{(l)}\right]^\top. \tag{7}$$

The updating rule of $\boldsymbol{h}^{(l)}$ is then expressed as

$$\boldsymbol{h}_{n+1}^{(l)} = \boldsymbol{h}_n^{(l)} + \boldsymbol{\mu}_{n+1}^{(l)} \circ e_n \boldsymbol{x}_n^{(l)}, \tag{8}$$

where $\boldsymbol{\mu}_{n+1}^{(l)}$ is adapted incrementally using a gradient descent. Note that to ensure positiveness of $\boldsymbol{\mu}_{n+1}^{(l)}$, we consider (Sutton, 1992)

$$\boldsymbol{\mu}_{n+1}^{(l)} = \exp\left[\boldsymbol{\alpha}_{n+1}^{(l)}\right]. \tag{9}$$

The updating rule for the $\boldsymbol{\alpha}^{(l)}$ vector is then expressed as

$$\boldsymbol{\alpha}_{n+1}^{(l)} = \boldsymbol{\alpha}_n^{(l)} - \kappa \frac{\partial e_n^2}{\partial \boldsymbol{\alpha}^{(l)}}, \tag{10}$$

where $\kappa > 0$ is the (scalar) meta step size parameter of IDBD algorithm. The IDBD algorithm consequently estimates the best step size vector for the LMS algorithm, where its main drawback is the use of the step size parameter, κ, to be adjusted according to the problem under analysis (Sutton, 1992).

Following Mahmood *et al.* (2012), the gradient term of Eq. (10) can be expressed by a matrix, $\boldsymbol{G}_n^{(l)}$, which is approximated by a diagonal matrix, leading to

$$\boldsymbol{\alpha}_{n+1}^{(l)} = \boldsymbol{\alpha}_n^{(l)} + \kappa e_n \boldsymbol{x}_n^{(l)} \circ \boldsymbol{g}_n^{(l)}, \tag{11}$$

where $\boldsymbol{g}_n^{(l)}$ is an $N \times 1$ vector containing the diagonal elements of $\boldsymbol{G}_n^{(l)}$, the entries of which are given by (Mahmood *et al.*, 2012)

$$\boldsymbol{g}_{n+1}^{(l)} = \boldsymbol{g}_n^{(l)} - \boldsymbol{g}_n^{(l)} \circ \boldsymbol{\mu}_{n+1}^{(l)} \circ \boldsymbol{x}_n^{(l)} \circ \boldsymbol{x}_n^{(l)} + \boldsymbol{\mu}_{n+1}^{(l)} \circ e_n \boldsymbol{x}_n^{(l)}. \tag{12}$$

Inserting Eq. (11) into Eq. (9) yields

$$\boldsymbol{\mu}_{n+1}^{(l)} = \boldsymbol{\mu}_n^{(l)} \exp\left[\kappa e_n \boldsymbol{x}_n^{(l)} \circ \boldsymbol{g}_n^{(l)}\right]. \tag{13}$$

In practice, for adaptation cycle $n+1$, the IDBD algorithm successively computes Eqs. (11), (13), and (8).

At this point, the meta step size parameter, κ, remains problem dependent. To overcome this limitation, a new version of IDBD algorithm applied to SISO systems identification has been developed, which is called Autostep (Mahmood *et al.*, 2012). Introducing an auxiliary $N \times 1$ vector, $\boldsymbol{v}$, with updating rule (Mahmood *et al.*, 2012)

$$\boldsymbol{v}_{n+1}^{(l)} = \boldsymbol{v}_n^{(l)} + \gamma \boldsymbol{\mu}_n^{(l)} \circ \boldsymbol{x}_n^2 \circ \left(\left|\boldsymbol{x}_n e_n \circ \boldsymbol{\alpha}_n^{(l)}\right| - \boldsymbol{v}_n^{(l)}\right), \tag{14}$$

where γ is a forgetting factor, Eq. (13) becomes

$$\boldsymbol{\mu}_{n+1}^{(l)} = \boldsymbol{\mu}_n^{(l)} \exp\left[\kappa e_n \boldsymbol{x}_n^{(l)} \circ \boldsymbol{g}_n^{(l)} \oslash \boldsymbol{v}_{n+1}^{(l)}\right]. \tag{15}$$

Last, to avoid division by 0, Eq. (13) is modified as (see details in Appendix A)

$$\boldsymbol{\mu}_{n+1}^{(l)} = \begin{cases} \log\left[\boldsymbol{\mu}_{n+1}^{(l)}\right] & \text{if } \boldsymbol{v}_{n+1}^{(l)} = 0, \\ \log\left[\boldsymbol{\mu}_{n+1}^{(l)}\right] + \kappa e_n \boldsymbol{x}_n^{(l)} \circ \boldsymbol{g}_n^{(l)} \oslash \boldsymbol{v}_{n+1}^{(l)} & \text{else,} \end{cases} \tag{16}$$

and if $(\boldsymbol{\mu}_{n+1}^{(l)} \circ \boldsymbol{x}_n^{(l)})^\top \boldsymbol{x}_n^{(l)} > 1$, we have

$$\boldsymbol{\mu}_{n+1}^{(l)} = \frac{\boldsymbol{\mu}_{n+1}^{(l)}}{(\boldsymbol{\mu}_{n+1}^{(l)} \circ \boldsymbol{x}_n^{(l)})^\top \boldsymbol{x}_n^{(l)}}. \tag{17}$$

The pseudo-code of the multiple input single output-Autostep (MISO-AS) algorithm is reported in Algorithm 2.

ALGORITHM 2. MISO-AS.

Require: $\kappa = 1 \times 10^{-2}$;
Require: $\gamma = 1 \times 10^{-4}$;
Ensure: $\boldsymbol{h}_n$
  $\boldsymbol{h}_0 \leftarrow \boldsymbol{0}$;
  $\boldsymbol{v}_0 \leftarrow \boldsymbol{0}$;
  $\boldsymbol{\alpha}_0 \leftarrow \boldsymbol{0}$;
  $\boldsymbol{\mu}_0^{(l)} \leftarrow \dfrac{1}{NL\sigma_{x^{(l)}}^2 + 1}$;
  while $\{x_n, y_n\}$ **do**
    $e_n \leftarrow y_n - \sum_{l=1}^{L} \boldsymbol{h}_n^{(l)\top} \boldsymbol{x}_n^{(l)}$
    $\boldsymbol{v}_{n+1}^{(l)} \leftarrow \boldsymbol{v}_n^{(l)} + \gamma \boldsymbol{\mu}_n^{(l)} \circ \boldsymbol{x}_n^{(l)^2} \circ (|\boldsymbol{x}_n^{(l)} \circ e_n \boldsymbol{\alpha}_n^{(l)}| - \boldsymbol{v}_n^{(l)})$
    if $\boldsymbol{v}_{n+1}^{(l)} = 0$ **then**
      $\boldsymbol{\alpha}_n^{(l)} \leftarrow \log\{\boldsymbol{\mu}_n^{(l)}\}$
    else
      $\boldsymbol{\alpha}_n^{(l)} \leftarrow \log\{\boldsymbol{\mu}_n^{(l)}\} + \kappa e_n \boldsymbol{x}_n^{(l)} \circ \boldsymbol{g}_n^{(l)} \oslash \boldsymbol{v}_{n+1}^{(l)}$
    end if
    $\boldsymbol{\mu}_n^{(l)} \leftarrow \exp \boldsymbol{\alpha}_n^{(l)}$
    $\boldsymbol{\mu}_n^{(l)} \leftarrow \boldsymbol{\mu}_n^{(l)} \oslash \max\{\boldsymbol{\mu}_n^{(l)} \circ \boldsymbol{x}_n^{(l)^2}, 1\}$
    $\boldsymbol{h}_{n+1}^{(l)} \leftarrow \boldsymbol{h}_n^{(l)} + \boldsymbol{\mu}_n^{(l)} \circ e_n \boldsymbol{x}_n^{(l)}$
    $\boldsymbol{g}_{n+1}^{(l)} \leftarrow (1 - \boldsymbol{\mu}_n^{(l)} \circ \boldsymbol{x}_n^{(l)^2}) \boldsymbol{g}_n^{(l)} + \boldsymbol{\mu}_n^{(l)} \circ e_n \boldsymbol{x}_n^{(l)}$
  end while

III. EVALUATION METRICS AND OPTIMIZATION SETTINGS

A. Metrics

To evaluate the performance of the MISO-NLMS and MISO-AS algorithms, two main metrics are defined in this

subsection: misalignment (calculated in time domain and noted as mis) and spectral mean square error (noted as spectral MSE). To study the convergence of the MISO-AS algorithm in relation to other algorithms, the MSE calculated in the time domain is also proposed. We recall that n is the adaptation cycle.

1. Misalignment

The misalignment codes the Euclidean distance between two vectors. Noting $\boldsymbol{h}_r^{(l)}$ as the target IR of input l and $\boldsymbol{h}^{(l)}(n)$ as the current estimation at n, the misalignment for input l is defined as

$$\text{mis}^{(l)}(n) = ||\boldsymbol{h}_r^{(l)} - \boldsymbol{h}^{(l)}(n)||_2, \tag{18}$$

where $||\cdot||_2$ is the L^2 norm. Given M as the total number of adaptation cycles for a specific simulation, we note $\text{mis}^{(l)} \equiv \text{mis}^{(l)}(M)$ and $\overline{\text{mis}}$ is the average value over the L inputs.

2. Spectral MSE

For most applications, the MISO system needs to be estimated in a given frequency band of interest. Noting $H^{(l)}(k)$ as the discrete Fourier transform (DFT) of the target IR (i.e., the frequency response function, FRF) and $\hat{H}^{(l)}(k)$ as the estimated FRF, we define the spectral MSE to be

$$\text{mse}_{\text{DFT}}^{(l)} = \frac{\sum_{k=k_1}^{k_2} \left(|\hat{H}^{(l)}(k)| - |H^{(l)}(k)| \right)^2}{\sum_{k=k_1}^{k_2} |H^{(l)}(k)|^2}, \tag{19}$$

where k_1 and k_2 are the lower and upper bounds of the spectral lines to be considered in the frequency band, respectively, and we note $\overline{\text{mse}}_{\text{DFT}}$ as the average value over the L inputs.

3. MSE

From the error signal, e_n, of Eq. (4), the MSE is defined as the expectation of e_n^2 such that

$$\text{mse}(n) = \mathbb{E}\left[e_n^2\right]. \tag{20}$$

For a single realization, MSE is approximated by

$$\text{mse}(n) \simeq \frac{1}{K} \sum_{i=n-K+1}^{n} e_i^2, \tag{21}$$

where $K > 0$ is an integer.

B. Optimization settings

We define the cost function by

$$J = (\overline{\text{mis}})^2 + \frac{1}{L-1} \sum_{l=1}^{L} (\text{mis}^{(l)} - \overline{\text{mis}})^2, \tag{22}$$

where the choice of J is not only to estimate each IR, $h^{(l)}[n]$, with the smallest error but also to minimize the variance between all inputs l. For the Autostep algorithm, as the cost function depends on two parameters (κ and γ), we denote it by $J(\kappa, \gamma)$, whereas the cost function of the NLMS algorithm depends on a single parameter, $\tilde{\mu}$, and is denoted as $J(\tilde{\mu})$.

To better illustrate the role of κ and γ parameters in the optimization process, a first simulation in free field conditions is presented here, considering a scenario of multiple loudspeakers located at the same distance from a microphone. The IRs can be analytically modelled as (Bruneau, 2013)

$$h^{(l)}[n] = \frac{1}{4\pi d^{(l)}} \delta\left(n - \left\lfloor \frac{d^{(l)}}{cT_s} \right\rfloor\right), \tag{23}$$

where $d^{(l)}$ is the distance between loudspeaker l and microphone, T_s is the sampling period, $\delta(n)$ is the unit sample sequence, and $\lfloor \alpha \rfloor$ is the rounding down of α to the nearest integer. Then, we seek the minimum of the cost function using a Bayesian optimization approach (Snoek *et al.*, 2012). Additional simulation parameters are $L = 4$, $d^{(l)} = 1\,\text{m}\,\forall l$, celerity of sound is $c_0 = 343\ \text{ms}^{-1}$, sampling frequency is $F_s = 16$ kHz, and number of taps of the IR is $N = 100$ taps. The duration of the input signals is $T = 5$ s, which gives $M = 80\,000$ points. The cost function is estimated on 200 points. The four IRs are then convoluted with four independent unit-variance $[\sigma^{(l)^2}]$ white noises and summed. Finally, a white Gaussian distributed noise is added to the output signal, yielding a signal-to-noise ratio (SNR) of 30 dB.

Figure 2 shows $J(\kappa, \gamma)$ as a function of (κ, γ). According to Bagheri *et al.* (2016) and Casebeer *et al.* (2021), the MISO-AS algorithm also shows a weak sensitivity to the meta parameters for a similar range for (κ, γ) despite very different problems being analyzed, including SISO vs MISO. Once convergence is checked for this large range of meta parameters, the influence on the speed of convergence is investigated hereafter.

To verify this assumption and analyze the possible effects of the values of (κ, γ) on the convergence of the algorithm, the learning curves are estimated by Monte Carlo simulation over 100 trials. The vectors $\boldsymbol{h}$, $\boldsymbol{v}$, and $\boldsymbol{\alpha}$ are initialized by a zero-valued vector and all the elements of vectors

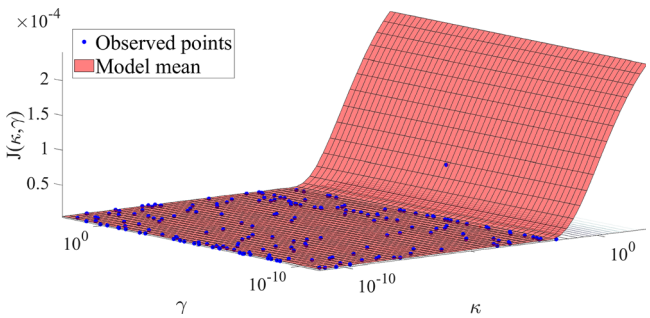

FIG. 2. (Color online) Cost function, $J(\kappa, \gamma)$, as a function of (κ, γ). Blue dots are points where J is calculated. Orange surface is Gaussian process regression model. Minimum value of J for $(\kappa, \gamma) = (0.22, 4.7 \times 10^{-10})$.

https://doi.org/10.1121/10.0024149

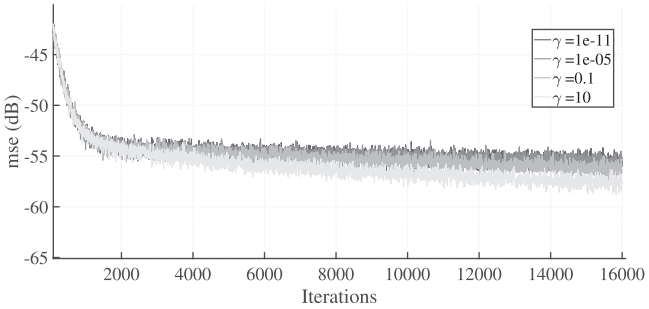

FIG. 3. MISO-AS learning curves for different values of γ with $\kappa = 0.01$.

$\boldsymbol{\mu}$ are initialized to the same value equal to $\boldsymbol{\mu}_0$ as explained in Algorithm 1. We first consider $\kappa = 0.01$ (Mahmood *et al.*, 2012). Figure 3 proposes the learning curves corresponding to four values of γ spread over a large range and shows that the value of γ has a weak influence on the convergence of the algorithm. For any γ, the learning curves decrease rapidly during the first 1000 iterations and then slowly thereafter. The higher γ is, the lower the MSE is. We next set $\gamma = 0.01$ and analyze the effects of different values for κ. As displayed in Fig. 4, the algorithm converges faster for the highest value of κ that we consider: κ driving the update equation of $\boldsymbol{\mu}$ [Eq. (10)], the higher κ is, the faster the convergence. We then choose $\kappa = 0.1$ in the following. Note that all learning curves have been obtained by averaging 100 independent Monte Carlo runs and smoothing the result with a 20 point moving average. To compare MISO-AS and MISO-NLMS algorithms in terms of sensitivity to their meta parameters, (κ, γ) for Autostep and $(\tilde{\mu})$ for NLMS, we have set $\gamma = 0.01$ as proposed in Bagheri *et al.* (2016) and Casebeer *et al.* (2021). Figure 5 shows the value of J for both algorithms. On the one hand, the MISO-AS algorithm has a low residual error for a very wide range of κ values. On the other hand, an inflection point can be noted for values of κ close to 0.1, beyond which the value of J increases significantly. On the contrary, the MISO-NLMS algorithm is more sensitive to the value of the meta parameter $\tilde{\mu}$. Although there is a very narrow range of $\tilde{\mu}$ values in $]0, 2]$, for which the algorithm offers equivalent performance to MISO-AS, for all other values, the MISO-NLMS algorithm presents a residual error 3 orders of magnitude higher than MISO-AS.

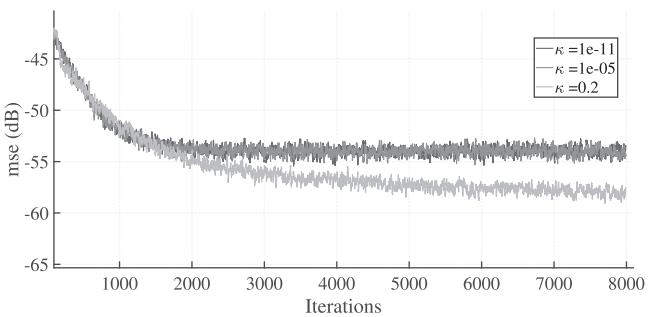

FIG. 4. MISO-AS learning curves for different values of κ with $\gamma = 0.01$.

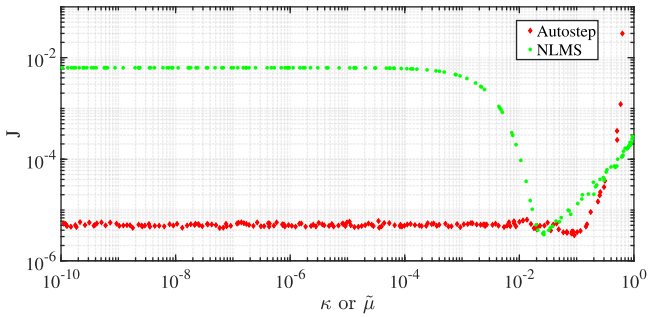

FIG. 5. (Color online) Cost function, J, as a function of κ (for $\gamma = 0.01$) or $\tilde{\mu}$. Green dots are points where J is calculated for NLMS, and red diamonds are points where J is calculated for Autostep. Minimum values of J are 3.15×10^{-6} for $\tilde{\mu} = 2.6 \times 10^{-2}$ and 3.19×10^{-6} for $\kappa = 9.11 \times 10^{-2}$.

Eventually, Fig. 6 shows learning curves for MISO-NLMS for different values of $\tilde{\mu}$, where one of them is the optimal value regarding the cost function J of Fig. 5. A comparison with the MISO-AS learning curve ($\kappa = 0.1$ and $\gamma = 0.01$) is also proposed in Fig. 5. For the optimal value of $\tilde{\mu}$, the asymptotic value is reached much faster for MISO-AS than for MISO-NLMS. For a higher value, $\tilde{\mu} = 0.1$, the convergence for MISO-AS remains faster than that for MISO-NLMS. In the following, we consider $\tilde{\mu} = 0.1$ as a compromise between level of cost function, J, and convergence speed.

IV. ESTIMATION OF ROOM IRS BY MEANS OF NUMERICAL SIMULATIONS

Numerical simulations are performed to test MISO-NLMS and MISO-AS algorithms in the case of a room in which several sound sources are active at the same time. The room simulated here has the following dimensions: 4 m × 3 m × 2.5 m, with a reverberation time, RT60, equal to 0.5 s. Six monopole sound sources are placed along a line while a microphone is located almost in line with the second source (Fig. 7).

IRs are computed in the MATLAB environment with an image source method using the toolbox proposed by Lehmann *et al.* (2007) with a sampling rate of $F_s = 16$ kHz. The longest simulated IR has 5608 taps, which corresponds to the length of the filters to be estimated. Each source generates a zero-mean-valued unit-variance white noise,

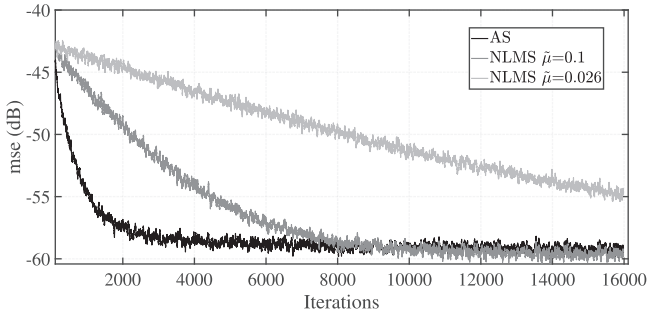

FIG. 6. Learning curves for MISO-NLMS and MISO-AS algorithms. Gray lines:are MISO-NLMS, and black line is MISO-AS. For MISO-NLMS, two different initial values of $\tilde{\mu}$ are considered. For MISO-AS, $\kappa = 0.1$ and $\gamma = 0.01$.

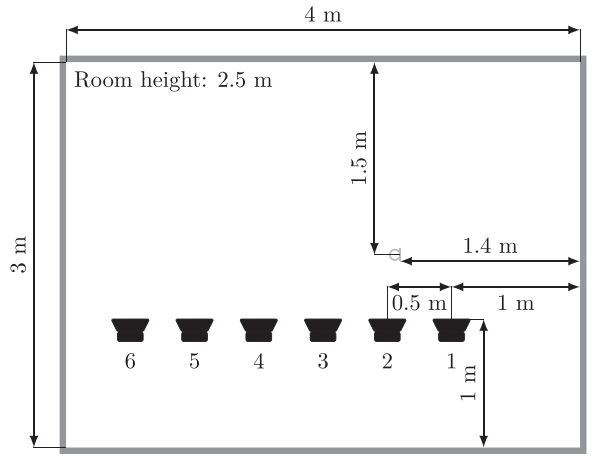

FIG. 7. Simulated setup geometry (top view), showing six loudspeakers (black) and one microphone (gray).

independent of the other source signals. Consequently, the signal measured by the microphone is the sum of the contributions from the six sources plus a white noise, independent of the input signals, resulting in a SNR of 30 dB.

In the following, MISO-NLMS and MISO-AS are compared regarding their learning curves (Fig. 8), their estimation of IRs of two loudspeakers [Figs. 9(b)–12(b)] and their global behavior (Tables I and II) after 5 s and 60 s of measurement time, respectively. For the MISO-NLMS case, we set a convergence step of $\tilde{\mu}^{(l)} = 0.1\ \forall l$, also corresponding to the initial values of $\boldsymbol{\mu}_0^{(l)}$ for the Autostep algorithm.

The learning curves are shown in Fig. 8. As for Fig. 6, MISO-AS converges rapidly to a given value (first 10 s), whereas MISO-NLMS needs many more iterations (about 60 s) to reach the performance of MISO-AS.

Figures 9(b) and 10(b) show the estimation with the MISO-NLMS algorithm of IRs of loudspeakers 1 and 6, respectively, while Figs. 11(b) and 12(b) show the estimation with the MISO-AS algorithm of the IRs of the same loudspeakers, where each of these figures displays the IRs at measurement times 5 s (top) and 30 s (bottom). It is worthy to notice that after 5 s, both algorithms still do not reach the

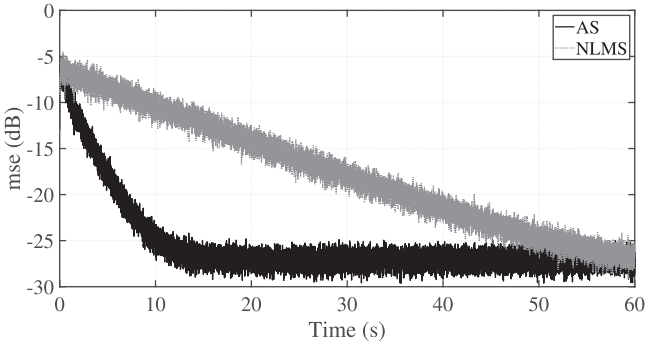

FIG. 8. Learning curves for $K = 100$ in Eq. (21), where gray denotes MISO-NLMS with $\tilde{\mu}^{(l)} = 0.1, \forall l$, and black denotes MISO-AS with $\kappa = 0.1$, $\gamma = 0.01$.

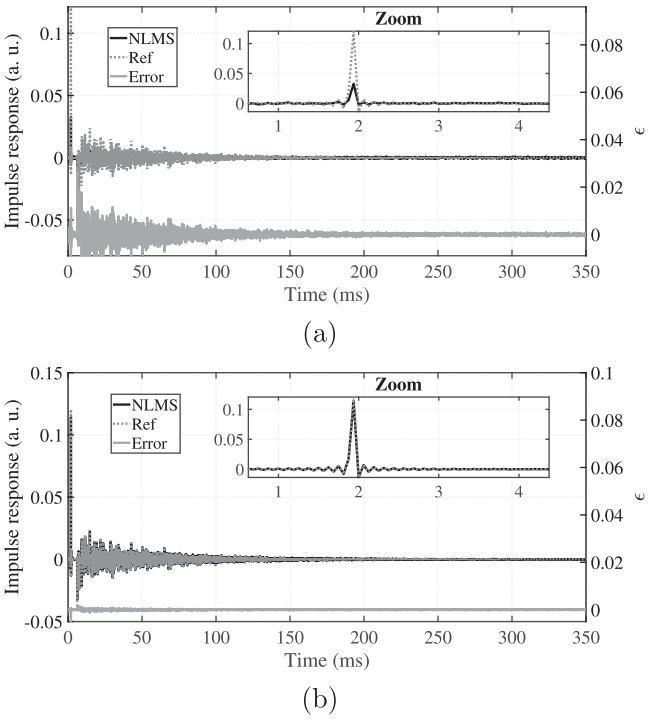

FIG. 9. IR of loudspeaker 1 in the simulated room (a) after 5 s and (b) after 60 s. Gray denotes true IR, and black denotes estimated IR with MISO-NLMS ($\tilde{\mu} = 0.1$). A zoom on the beginning of the IRs is proposed. Errors ϵ at the end of simulation are also given at the bottom of the graphs (corresponding scale on the right).

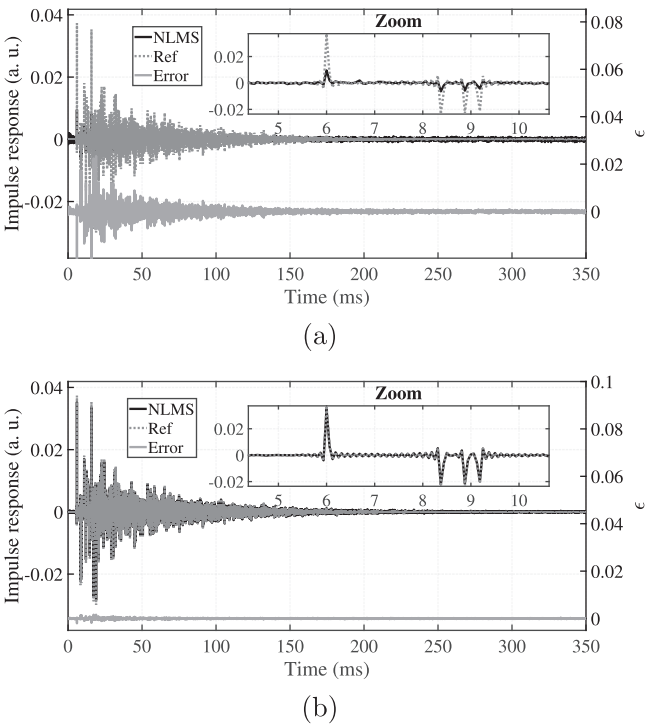

FIG. 10. IR of loudspeaker 6 in the simulated room (a) after 5 s and (b) after 60 s. Gray denotes true IR, and black denotes estimated IR with MISO-NLMS ($\tilde{\mu} = 0.1$). A zoom on the beginning of the IRs is proposed. Errors ϵ at the end of simulation are also given at the bottom of the graphs (corresponding scale on the right).

https://doi.org/10.1121/10.0024149

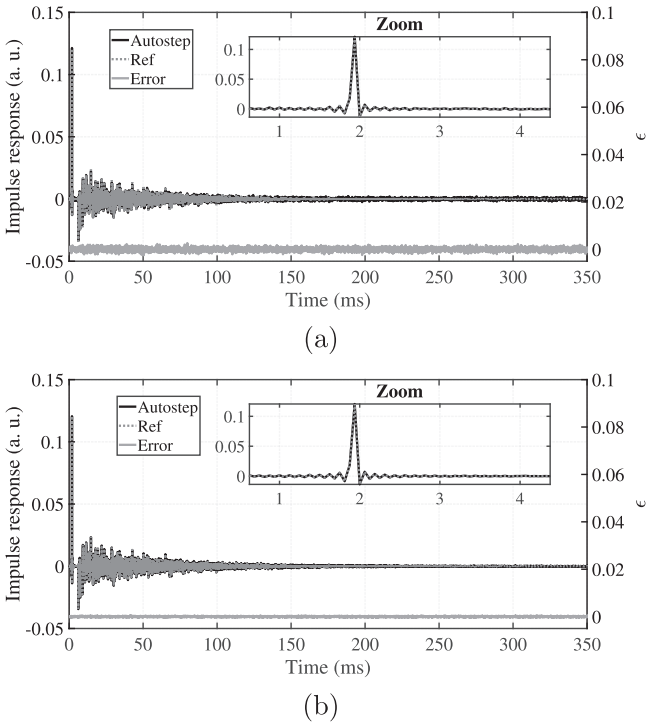

FIG. 11. IR of loudspeaker 1 in the simulated room (a) after 5 s and (b) after 60 s. True IR are in gray, whereas estimated IR with MISO-AS are in black ($\kappa = 0.1$, $\gamma = 0.01$). A zoom on the beginning of the IRs is proposed. Errors ϵ at the end of simulation are also given at the bottom of the graphs (corresponding scale on the right).

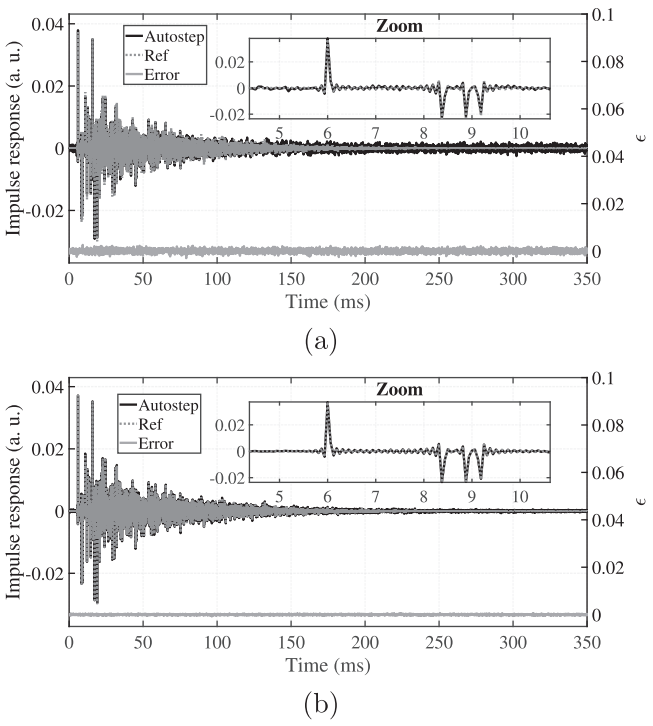

FIG. 12. IR of loudspeaker 6 in the simulated room (a) after 5 s and (b) after 60 s. Gray denotes true IR, and black denotes estimated IR with MISO-AS ($\kappa = 0.1$, $\gamma = 0.01$). A zoom on the beginning of the IRs is proposed. Errors ϵ at the end of simulation are also given at the bottom of the graphs (corresponding scale on the right).

TABLE I. Misalignment and spectral MSE in the simulated room after 5 s, where the average is over all loudspeakers, with min and max values.

	Autostep	NLMS
$\overline{\text{mis}}$ (dB)	−12.8	−8.1
min/max mis (dB)	−12.8/−12.8	−9.1/−7.0
$\overline{\text{mse}}_{\text{DFT}}$	0.08	0.87
min/max mse_{DFT}	0.05/0.09	0.85/0.89

convergence value (Fig. 8). Nevertheless, this may be sufficient for some applications requiring a rough estimation in a limited time. This is best shown by comparing Figs. 9(a) and 11(a) for loudspeaker 1 for MISO-NLMS and MISO-AS, respectively, or for loudspeaker 6 by comparing Figs. 10(a) and 12(a) for MISO-NLMS and MISO-AS, respectively. For greater clarity, a subfigure shows a zoom on the first part of the IR. One can note that the IR of the furthest loudspeaker is noisier than the IR of the closest loudspeaker due to reverberation effects. For both loudspeakers, MISO-AS gives a much lower error than MISO-NLMS, where the latter fails, in particular, to correctly approximate the first IR maximum related to the direct sound, contrarily to the MISO-AS.

The performances of algorithms MISO-NLMS and MISO-AS are much closer after 60 s as illustrated in Figs. 9(b) and 11(b) for loudspeaker 1, and in Figs. 10(b) and 12(b) for loudspeaker 6. This result is consistent with the behavior of the learning curves of Fig. 8 at 60 s of measurement time.

Last, metrics after 5 s and 60 s are given in Tables I and II, respectively (average over all loudspeakers and min and max values), to compare performances of both algorithms. For misalignment, MISO-AS performs 4.7 dB better than MISO-NLMS after 5 s, and both algorithms yields the same mean misalignment after 60 s (with a lower dispersion of values for MISO-AS). For the spectral MSE, MISO-AS error is ten and three times smaller than that obtained by MISO-NLMS after 5 s and 60s, respectively.

V. EXPERIMENTAL VALIDATION

This section presents the results derived from experimental measurements performed in two different types of rooms: a semi-anechoic chamber and a laboratory room. These two rooms are chosen to provide two very different test environments: the first one produces a fairly light mixture of the different loudspeaker signals (direct fields and reflections off the floor), and the second one produces a

TABLE II. Misalignment and spectral MSE in the simulated room after 60 s, where the average is over all loudspeakers, with min and max values.

	Autostep	NLMS
$\overline{\text{mis}}$ (dB)	−37.8	−37.5
min/max mis (dB)	−37.9/−37.7	−38.9/−35.8
$\overline{\text{mse}}_{\text{DFT}}$ $(\times 10^{-3})$	4	13.3
min/max mse_{DFT} $(\times 10^{-3})$	3.0/5.9	12.4/14.1

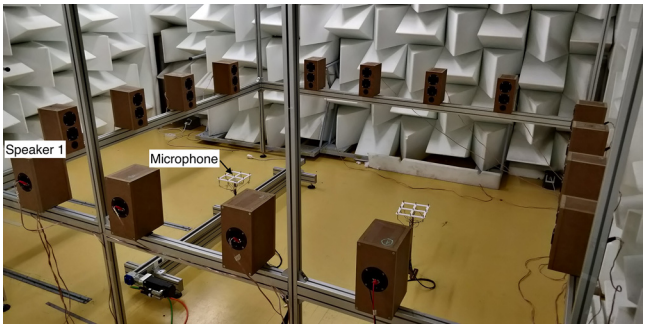

FIG. 13. (Color online) Experimental setup in the semi-anechoic chamber.

much more complex mixture (due to the many reflections off the different walls and furniture and diffraction phenomena).

A. Experimental setup

The experimental setup is controlled from an RME Madiface XT audio interface (Audio AG, Haimhausen, Germany) that allows synchronization of the generation and acquisition channels. For sound signal generation, up to 16 2-way custom-built loudspeakers are driven by a DA32 RME converter via four four-channel HPA D604 amplifiers (HPA Co., Shenzhen, China). A Brüel and Kjaer 4958 microphone (BK, Brüel & Kjær, Nærum, Denmark) is used to record the sound pressure signal via an AD32 RME converter. For all the configurations tested in this section, each IR is measured using an exponential sweep sine (ESS) signal (Farina, 2000) and estimated from the Autostep method for which the excitation signals are white noise of unit variance, independent of each other. The acquisition and reproduction are performed at a sampling frequency of 48 kHz.

B. Measurements in the semi-anechoic chamber

The first set of measurements is performed in a semi-anechoic chamber. Sixteen loudspeakers are distributed regularly around a square of side 3 m at 1 m from the ground (Fig. 13). The microphone is placed in the square center. The acquisition time is 60 s, and the IRs are composed by

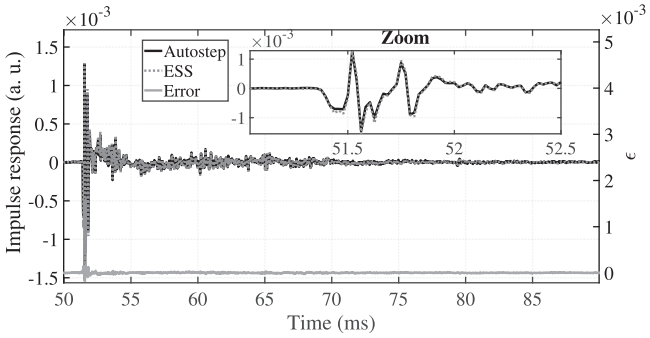

FIG. 14. IR of loudspeaker 1 in the semi-anechoic chamber. Gray denotes measured with an ESS signal, and black denotes estimated with Autostep after 60 s ($\kappa = 0.01$, $\gamma = 10$). A zoom on the beginning of the IRs is proposed. Error ϵ is also given at the bottom of the graph (corresponding scale on the right).

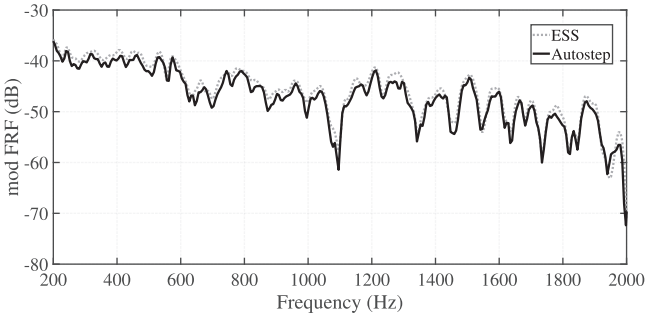

FIG. 15. FRF (magnitude) of loudspeaker 1 in the semi-anechoic chamber. Gray denotes measured with ESS technique, and black denotes estimated with Autostep after 60 s.

5000 coefficients (about 100 ms). This duration is set to be greater than the IR duration of the loudspeakers for all tested positions.

The estimated IR of loudspeaker 1 (see Fig. 13) after 60 s is plotted in Fig. 14 and the corresponding FRF (magnitude) is shown in Fig. 15. For comparison purposes, Fig. 16 shows the FRF (magnitude) estimation with the MISO-NLMS algorithm vs the ESS algorithm. The MISO-AS based estimations in time and frequency domains show a very good agreement with the corresponding ESS technique, whereas the MISO-NLMS based estimation provides slightly weaker results.

Metrics after 60 s are given in Table III (average over all loudspeakers and min and max values) to compare performances of MISO-NLMS and MISO-AS algorithms. Regarding the misalignment, MISO-AS registers a 2.6 dB better performance than MISO-NLMS. One can also notice that MISO-AS better performs in terms of spectral MSE (error reduced by a factor of 1.4) and dispersion (measured through the difference between max and min values), reducing it by a factor of 2.7.

C. Measurements in the laboratory room

The second series of measurements is performed in a room of the laboratory whose reverberation time, RT30 (around 0.8 s), is relatively constant in the range 100 Hz–5 kHz. Eight loudspeakers are placed in a row on a workbench, and a BK 1/2″ 4192 microphone is located in

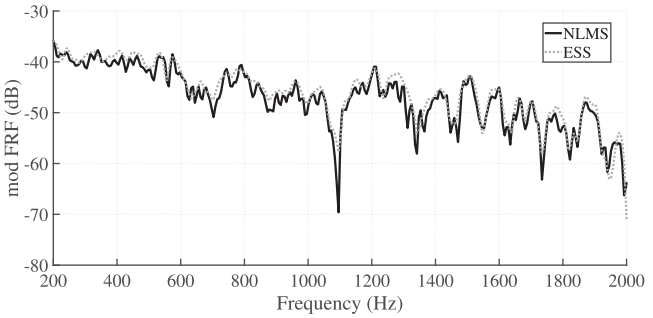

FIG. 16. FRF (magnitude) of loudspeaker 1 in the semi-anechoic chamber. Gray denotes measured with ESS technique, and black denotes estimated with NLMS ($\tilde{\mu} = 0.1$) after 60 s.

https://doi.org/10.1121/10.0024149

TABLE III. Misalignment and spectral MSE in the semi-anechoic chamber after 60 s, where the average is over all loudspeakers, with min and max values.

	Autostep	NLMS
$\overline{\text{mis}}$ (dB)	−63.3	−60.7
min/max mis (dB)	−66.3/−61.7	−62.0/−59.8
$\overline{\text{mse}}_{\text{DFT}}$	0.053	0.075
min/max mse_{DFT}	0.045/0.066	0.054/0.11

the middle of the room in line with the first loudspeaker (Fig. 17).

The IRs after 6 min of loudspeaker 1 (closest to the microphone) and loudspeaker 8 (farthest from the microphone) are shown in Figs. 18 and 19, respectively. As loudspeaker 1 mainly captures the direct field, results are similar to those obtained in the semi-anechoic chamber. However, it presents a longer convergence time due to the increased complexity of the acoustic environment. For loudspeaker 8, the IR is characterized by a longer decay due to the decrease in the ratio between the direct and reverberated fields. It can be observed that the residual error is higher when compared to the residual error of loudspeaker 1.

The corresponding FRF (magnitude) of loudspeakers 1 and 8 are, respectively, given in Figs. 20 and 21 after 6 min. Once again, the MISO-AS based estimation shows a very good agreement with the reference FRF based on the ESS technique. For comparison purposes, the very same FRFs, estimated with the MISO-NLMS based algorithm, are proposed in Figs. 22 and 23. While the overall shape of the FRF modulus is correctly estimated for each loudspeaker and both algorithms, the errors are greater for the MISO-NLMS based algorithm than for the MISO-AS based algorithm as confirmed through the $\overline{\text{mse}}_{\text{DFT}}$ values given in Table IV.

Metrics after 6 min are given in Table IV (average over all loudspeakers and min and max values) for comparing the performances of both algorithms. Regarding the misalignment, MISO-AS has a 2.2 dB better performance than

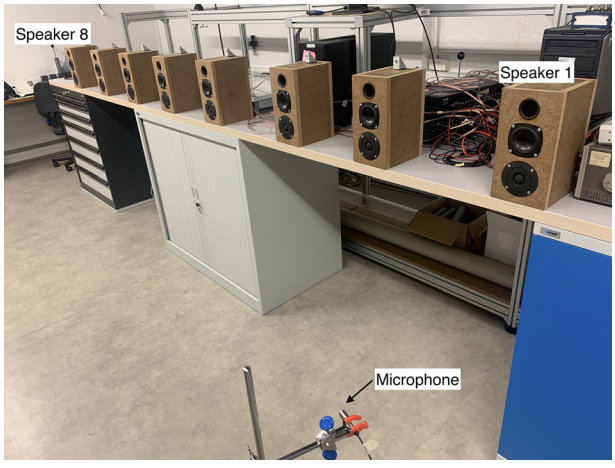

FIG. 17. (Color online) Experimental setup in the laboratory room.

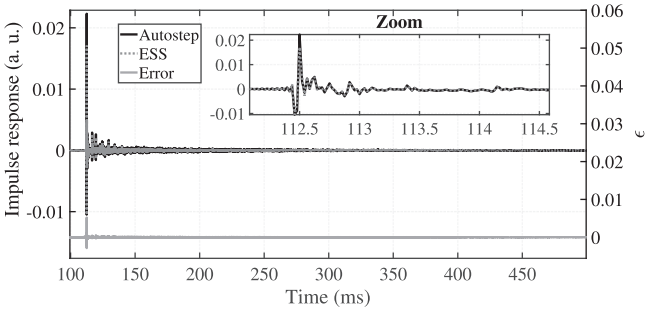

FIG. 18. IR of loudspeaker 1 in the laboratory room. Gray denotes measured with an ESS signal, and black denotes estimated with Autostep after 6 min ($\kappa = 0.01$, $\gamma = 10$). A zoom on the beginning of the IRs is proposed. Error ϵ is also given at the bottom of the graph (corresponding scale on the right).

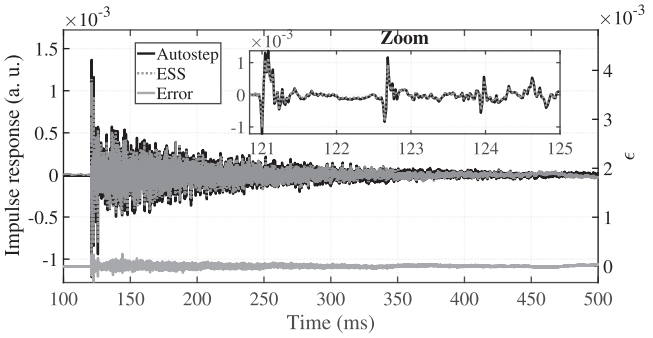

FIG. 19. IR of loudspeaker 8 in the laboratory room. Gray denotes measured with an ESS signal, and black denotes estimated with Autostep after 6 min ($\kappa = 0.01$, $\gamma = 10$). A zoom on the beginning of the IRs is proposed. Error ϵ is also given at the bottom of the graph (corresponding scale on the right).

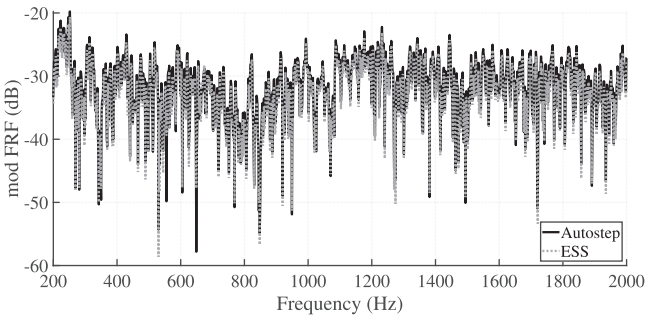

FIG. 20. FRF (magnitude) of loudspeaker 1 in the laboratory room. Gray denotes measured with ESS technique, and black denotes estimated with Autostep after 6 min.

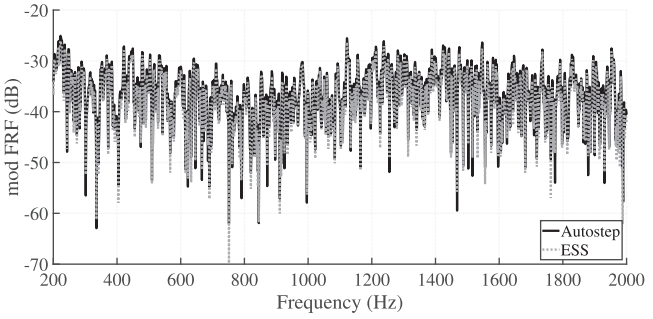

FIG. 21. FRF (magnitude) of loudspeaker 8 in the laboratory room. Gray denotes measured with ESS technique, and black denotes estimated with Autostep after 6 min.

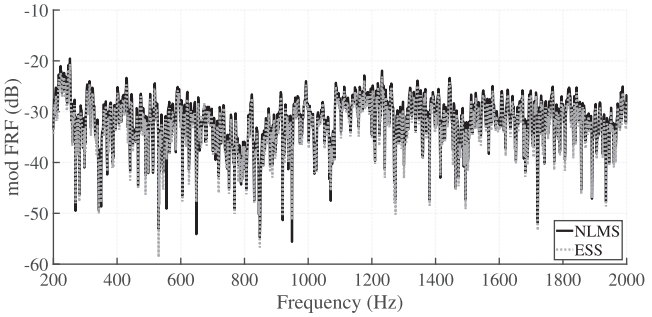

FIG. 22. FRF (magnitude) of loudspeaker 1 in the laboratory room. Gray denotes measured with ESS technique, and black denotes estimated with NLMS after 6 min.

MISO-NLMS. For the spectral MSE, the error is 1.5 times smaller for MISO-AS than for MISO-NLMS and the dispersion (measured through the difference between max and min values) is 2.3 times smaller for MISO-AS than for MISO-NLMS.

VI. CONCLUSION

In this paper, the MISO-AS algorithm is proposed for estimating the IRs of a LTI MISO system. This algorithm is tested numerically on a simulated room and experimentally on two different types of rooms: a semi-anechoic chamber and standard laboratory room. The performance obtained is better than that of the MISO-NLMS algorithm, thanks to the adaptability of the step size parameter of the MISO-AS algorithm. Indeed, depending on the configuration, a gain of 0.3–4.7 dB on the misalignment and a gain of a factor of 1.4–10 on the spectral MSE are obtained. In addition, less dispersion is observed from one loudspeaker to another. However, the convergence speed of the algorithm is directly linked to the reverberation time of the room under test: the higher the reverberation time, the longer the IRs to be estimated and the slower the convergence. To conclude, it should be stressed that the low dependence of the MISO-AS algorithm on its meta parameters could open up new perspectives in system identification, in particular the characterization of nonstationary environments, including linear time-varying MISO systems.

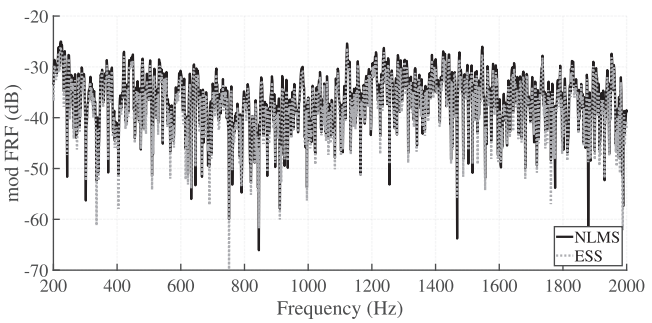

FIG. 23. FRF (magnitude) of loudspeaker 8 in the laboratory room. Gray denotes measured with ESS technique, and black denotes estimated with NLMS after 6 min.

TABLE IV. Misalignment and spectral MSE in the semi-anechoic chamber after 6 min, where the average is over all loudspeakers, with min and max values.

	Autostep	NLMS
$\overline{\text{mis}}$ (dB)	−47.6	−45.4
min/max mis (dB)	−54.4/−41.9	−48.7/−41.2
$\overline{\text{mse}}_{\text{DFT}}$	0.071	0.11
min/max mse_{DFT}	0.058/0.08	0.09/0.14

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX A: STABILITY CONDITIONS

To ensure the stability of the Autostep algorithm, it is necessary to make some modifications to the update equation (15). First, for real-world data, the values of input signal x can vary by large amounts, whereas normalization does not immediately take these variations into account. To ensure stability in this case, we temporarily limit the ratio such that

$$\boldsymbol{v}_{n+1}^{(l)} = \max\left(\boldsymbol{v}_{n+1}^{(l)}, |e_n \boldsymbol{x}_n^{(l)} \circ \boldsymbol{g}_n^{(l)}|\right). \tag{A1}$$

Second, introducing the normalization vector, $\boldsymbol{v}^{(l)}$, in the denominator may cause stability problems. In particular, it is essential to avoid division by zero. Mahmood *et al.* (2012) suggested replacing all zero elements of the vector $\boldsymbol{v}^{(l)}$ by one:

$$\boldsymbol{v}_{n+1}^{(l)} = 1 \ \text{ if } \boldsymbol{v}_{n+1}^{(l)} = 0. \tag{A2}$$

Finally, a sufficient condition on the step size parameter for the stability is a positive inner product of successive gradients (Mahmood *et al.*, 2012). In our case, this condition becomes

$$\left(\frac{\partial\left(y_n - \boldsymbol{h}_{n+1}^{(l)\top}\boldsymbol{x}_{n+1}^{(l)}\right)^2}{\partial \boldsymbol{h}_{n+1}^{(l)}}\right)^{\top}\left(\frac{\partial\left(y_n - \boldsymbol{h}_{n}^{(l)\top}\boldsymbol{x}_{n}^{(l)}\right)^2}{\partial \boldsymbol{h}_{n}^{(l)}}\right) \geq 0. \tag{A3}$$

After derivation, we obtain

$$\left(\left(y_n - \boldsymbol{h}_{n+1}^{(l)\top}\boldsymbol{x}_{n+1}^{(l)}\right)\boldsymbol{x}_{n+1}^{(l)}\right)^{\top}\left(\left(y_n - \boldsymbol{h}_{n}^{(l)\top}\boldsymbol{x}_{n}^{(l)}\right)\boldsymbol{x}_{n}^{(l)}\right) \geq 0, \tag{A4}$$

or equivalently

$$\left(\boldsymbol{\mu}_{n+1}^{(l)} \circ \boldsymbol{x}_n^{(l)}\right)^{\top}\boldsymbol{x}_n^{(l)} \leq 1. \tag{A5}$$

If Eq. (A5) is not checked, we apply the rule given in Eq. (17).

https://doi.org/10.1121/10.0024149

APPENDIX B: AUTOSTEP CONVERGENCE PROPERTIES

The convergence properties of the Autostep algorithm for MISO system identification are proposed in this appendix. We consider the error vector, $\boldsymbol{\epsilon}_{n+1}^{(l)}$, defined as the difference between the real IR, $\boldsymbol{h}_r^{(l)}$, and the estimated IR at current time $n+1$ (see Fig. 1) such that

$$\boldsymbol{\epsilon}_{n+1}^{(l)} = \boldsymbol{h}_r^{(l)} - \boldsymbol{h}_{n+1}^{(l)}. \tag{B1}$$

Then, from the output noisy signal y_n,

$$y_n = \sum_{l=1}^{L} \boldsymbol{h}_r^{(l)} \boldsymbol{x}_n^{(l)} + b_n, \tag{B2}$$

where $b_n \sim \mathcal{N}(0, \sigma_b^2)$, we obtain

$$\boldsymbol{\epsilon}_{n+1}^{(l)} = \boldsymbol{h}_r^{(l)} - \boldsymbol{h}_n^{(l)} - \boldsymbol{\mu}_n^{(l)} \circ e_n \boldsymbol{x}_n^{(l)}. \tag{B3}$$

As will be demonstrated in Appendix C, $\boldsymbol{I}$ is the identity matrix, and the error can be written as

$$\begin{aligned}\boldsymbol{\epsilon}_{n+1}^{(l)} = &\left(\boldsymbol{I} - \boldsymbol{\mu}_n^{(l)} \circ \left(\boldsymbol{x}_n^{(l)} \boldsymbol{x}_n^{(l)^\top}\right)\right)\boldsymbol{\epsilon}_n^{(l)} \\ &- \sum_{\substack{j=1 \\ j \neq l}}^{L} \left(\boldsymbol{\mu}_n^{(l)} \circ \boldsymbol{x}_n^{(j)} \boldsymbol{x}_n^{(l)^\top} \boldsymbol{\epsilon}_n^{(j)}\right) - \boldsymbol{\mu}_n^{(l)} \circ b_n \boldsymbol{x}_n^{(l)}.\end{aligned} \tag{B4}$$

The expectation value of the error is then given by

$$\begin{aligned}\mathbb{E}[\boldsymbol{\epsilon}_{n+1}^{(l)}] = &\left(\boldsymbol{I} - \mathbb{E}\left[\boldsymbol{\mu}_n^{(l)} \circ (\boldsymbol{x}_n^{(l)} \boldsymbol{x}_n^{(l)^\top})\right]\right)\mathbb{E}[\boldsymbol{\epsilon}_n^{(l)}] \\ &- \sum_{\substack{j=1 \\ j \neq l}}^{L} \mathbb{E}\left(\boldsymbol{u}_n^{(l)} \circ \boldsymbol{x}_n^{(j)} \boldsymbol{x}_n^{(l)^\top} \boldsymbol{\epsilon}_n^{(j)}\right) \\ &- \mathbb{E}\left[\boldsymbol{u}_n^{(l)} \circ b_n \boldsymbol{x}_n^{(l)}\right].\end{aligned} \tag{B5}$$

Assuming that the vectors, $\boldsymbol{\mu}$, are asymptotically independent (and consequently uncorrelated) from the input signals, $\boldsymbol{x}$ (Saeed, 2017),

$$\mathbb{E}\left[\boldsymbol{\mu}_n^{(i)} \boldsymbol{x}_n^{(j)}\right] = \mathbb{E}\left[\boldsymbol{\mu}_n^{(i)}\right]\mathbb{E}\left[\boldsymbol{x}_n^{(j)}\right] \quad \forall (i,j) \in [1, L]^2, \tag{B6}$$

then leads to

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\epsilon}_{n+1}^{(l)}\right] = &\left(\boldsymbol{I} - \mathbb{E}\left[\boldsymbol{\mu}_n^{(l)}\right] \circ \mathbb{E}\left[\boldsymbol{x}_n^{(l)} \boldsymbol{x}_n^{(l)^\top}\right]\right)\mathbb{E}\left[\boldsymbol{\epsilon}_n^{(l)}\right] \\ &- \sum_{\substack{j=1 \\ j \neq l}}^{L} \mathbb{E}\left[\boldsymbol{\mu}_n^{(l)}\right] \circ \mathbb{E}\left[\boldsymbol{x}_n^{(j)} \boldsymbol{x}_n^{(l)^\top}\right]\mathbb{E}\left[\boldsymbol{\epsilon}_n^{(j)}\right] \\ &- \mathbb{E}\left[\boldsymbol{\mu}_n^{(l)}\right] \circ \mathbb{E}\left[b_n \boldsymbol{x}_n^{(l)}\right].\end{aligned} \tag{B7}$$

The second term of the right-hand side of Eq. (B7) is zero because input signals are mutually independent of each other. The third term is also zero because the additive noise is zero-mean valued. Finally, this yields

$$\mathbb{E}\left[\boldsymbol{\epsilon}_{n+1}^{(l)}\right] = \left(\boldsymbol{I} - \mathbb{E}\left[\boldsymbol{\mu}_n^{(l)}\right] \circ \boldsymbol{R}_x^{(l)}\right)\mathbb{E}\left[\boldsymbol{\epsilon}_n^{(l)}\right], \tag{B8}$$

where $\boldsymbol{R}_x^{(l)}$ is the correlation matrix of input l. Noting that the correlation matrix can be factorized as

$$\boldsymbol{R}_x^{(l)} = \boldsymbol{Q}^{(l)} \Lambda^{(l)} \boldsymbol{Q}^{(l)^\top}, \tag{B9}$$

where $\boldsymbol{Q}^{(l)}$ is the matrix of orthonormal eigenvectors and $\Lambda^{(l)}$ is the diagonal matrix of eigenvalues, we then get

$$\mathbb{E}\left[\bar{\boldsymbol{\epsilon}}_{n+1}^{(l)}\right] = \left(\boldsymbol{I} - \mathbb{E}\left[\boldsymbol{\mu}_n^{(l)}\right] \circ \Lambda^{(l)}\right)\mathbb{E}\left[\bar{\boldsymbol{\epsilon}}_n^{(l)}\right], \tag{B10}$$

where $\bar{\boldsymbol{\epsilon}}_n^{(l)} = \boldsymbol{Q}^{(l)^\top} \boldsymbol{\epsilon}_n^{(l)}$. This results in a convergence condition of $\boldsymbol{\mu}_n^{(l)}$ as

$$0 < \mathbb{E}\left[\boldsymbol{\mu}_n^{(l)}\right] < \frac{2}{\lambda_{\max}^{(l)}}, \tag{B11}$$

where $\lambda_{\max}^{(l)}$ is the largest eigenvalue of the correlation matrix $\boldsymbol{R}_x^{(l)}$. From this condition, it can be deduced that

$$\mathbb{E}\left[\boldsymbol{\epsilon}_n^{(l)}\right] \underset{n \to +\infty}{\to} 0 \Rightarrow \mathbb{E}\left[\boldsymbol{h}_n^{(l)}\right] \underset{n \to +\infty}{\to} \boldsymbol{h}_r^{(l)}. \tag{B12}$$

APPENDIX C: ϵ ERROR CALCULATION

Starting with Eq. (B3) and replacing e_n by its expression given in (4) gives

$$\boldsymbol{\epsilon}_{n+1}^{(l)} = \boldsymbol{h}_r^{(l)} - \boldsymbol{h}_n^{(l)} - \boldsymbol{\mu}_n^{(l)} \circ \left(y_n - \sum_{j=1}^{L} \boldsymbol{h}_n^{(j)^\top} \boldsymbol{x}_n^{(j)}\right) \boldsymbol{x}_n^{(l)}. \tag{C1}$$

Then, replacing y_n according to Eq. (B2) leads to

$$\begin{aligned}\boldsymbol{\epsilon}_{n+1}^{(l)} = &\boldsymbol{\epsilon}_n^{(l)} - \boldsymbol{\mu}_n^{(l)} \circ \left[\sum_{j=1}^{L} \left(\boldsymbol{h}_r^{(j)^\top} - \boldsymbol{h}_n^{(j)^\top}\right) \boldsymbol{x}_n^{(j)}\right] \boldsymbol{x}_n^{(l)} \\ &- \boldsymbol{\mu}_n^{(l)} b_n \circ \boldsymbol{x}_n^{(l)}.\end{aligned} \tag{C2}$$

Isolating the terms related to input l from the other terms as

$$\begin{aligned}\boldsymbol{\epsilon}_{n+1}^{(l)} = &\boldsymbol{\epsilon}_n^{(l)} - \boldsymbol{\mu}_n^{(l)} \circ \left(\boldsymbol{\epsilon}_n^{(l)^\top} \boldsymbol{x}_n^{(l)}\right) \boldsymbol{x}_n^{(l)} - \boldsymbol{\mu}_n^{(l)} \\ &\circ \left(\sum_{\substack{j=1 \\ j \neq l}}^{L} \boldsymbol{\epsilon}_n^{(j)^\top} \boldsymbol{x}_n^{(j)}\right) \boldsymbol{x}_n^{(l)} - \boldsymbol{\mu}_n^{(l)} b_n \circ \boldsymbol{x}_n^{(l)},\end{aligned} \tag{C3}$$

factorizing by $\boldsymbol{\epsilon}_n^{(l)}$ and rearranging the terms, we finally get

$$\boldsymbol{\epsilon}_{n+1}^{(l)} = \left(\boldsymbol{I} - \boldsymbol{\mu}_n^{(l)} \circ \left(\boldsymbol{x}_n^{(l)} \boldsymbol{x}_n^{(l)^\top}\right)\right)\boldsymbol{\epsilon}_n^{(l)} - \sum_{\substack{j=1 \\ j \neq l}}^{L} \left(\boldsymbol{\mu}_n^{(l)} \circ \boldsymbol{x}_n^{(j)} \boldsymbol{x}_n^{(l)^\top} \boldsymbol{\epsilon}_n^{(j)}\right) - \boldsymbol{\mu}_n^{(l)} \circ b_n \boldsymbol{x}_n^{(l)}. \quad \text{(C4)}$$

Bagheri, S., Thill, M., Koch, P., and Konen, W. (**2016**). "Online adaptable learning rates for the game Connect-4," IEEE Trans. Comput. Intell. AI Games **8**(1), 33–42.

Berkhout, A. J., De Vries, D., and Vogel, P. (**1993**). "Acoustic control by wave field synthesis," J. Acoust. Soc. Am. **93**, 2764–2778.

Bismor, D., Czyz, K., and Ogonowski, Z. (**2016**). "Review and comparison of variable step-size LMS algorithms," Int. J. Acoust. Vib. **21**(1), 24–39.

Bruneau, M. (**2013**). *Fundamentals of Acoustics* (Wiley-ISTE, London, UK).

Casebeer, J., Bryan, N. J., and Smaragdis, P. (**2021**). "Auto-DSP: Learning to optimize acoustic echo cancellers," in *2021 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY (IEEE, New York), pp. 291–295.

Choi, J.-W., and Kim, Y.-H. (**2002**). "Generation of an acoustically bright zone with an illuminated region using multiple sources," J. Acoust. Soc. Am. **111**(4), 1695–1700.

Farina, A. (**2000**). "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *AES 108th Convention*, Paris, France (AES Inc., New York), pp. 1–24.

Goodwin, G. C., and Sin, K. S. (**2009**). *Adaptive Filtering Prediction and Control* (Dover, New York).

Harris, R., Chabries, D., and Bishop, F. (**1986**). "A variable step (VS) adaptive filter algorithm," IEEE Trans. Acoust, Speech, Signal Process. **34**(2), 309–316.

Haykin, S. S. (**2014**). *Adaptive Filter Theory*, 5th ed. (Pearson, Upper Saddle River, NJ).

Hu, M., Shi, L., Zou, H., Christensen, M. G., and Lu, J. (**2023**). "Sound zone control with fixed acoustic contrast and simultaneous tracking of acoustic transfer functions," J. Acoust. Soc. Am. **153**(5), 2538–2544.

Hu, M., Xue, J., and Lu, J. (**2019**). "Online multi-channel secondary path modeling in active noise control without auxiliary noise," J. Acoust. Soc. Am. **146**(4), 2590–2595.

Kodrasi, I., and Doclo, S. (**2016**). "Joint dereverberation and noise reduction based on acoustic multi-channel equalization," IEEE/ACM Trans. Audio. Speech. Lang. Process. **24**(4), 680–693.

Lehmann, E. A., Johansson, A. M., and Nordholm, S. (**2007**). "Reverberation-time prediction method for room impulse responses simulated with the image-source model," in *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY (IEEE, New York), pp. 159–162.

Mahmood, A. (**2010**). "Automatic step-size adaptation in incremental supervised learning," M.S. thesis, University of Alberta, Alberta, Canada, available at https://doi.org/10.7939/R31G9D.

Mahmood, A. R., Sutton, R. S., Degris, T., and Pilarski, P. M. (**2012**). "Tuning-free step-size adaptation," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan (IEEE, New York), pp. 2121–2124.

Makino, S. (**2001**). "Stereophonic acoustic echo cancellation: An overview and recent solutions," Acoust. Sci. Technol. **22**(5), 325–333.

Mikhael, W., Wu, F., Kazovsky, L., Kang, G., and Fransen, L. (**1986**). "Adaptive filters with individual adaptation of parameters," IEEE Trans. Circuits Syst. **33**(7), 677–686.

Moller, M. B., and Ostergaard, J. (**2020**). "A moving horizon framework for sound zones," IEEE/ACM Trans. Audio. Speech. Lang. Process. **28**, 256–265.

Saeed, M. O. B. (**2017**). "A unified analysis approach for LMS-based variable step-size algorithms," Arab. J. Sci. Eng. **42**(7), 2809–2816.

Snoek, J., Larochelle, H., and Adams, R. P. (**2012**). "Practical Bayesian optimization of machine learning algorithms," in *Proceedings of the 25th International Conference on Neural Information Processing Systems* Lake Tahoe, NV (Curran Associates Inc., New York), Vol. 2, p. 2951–2959.

Sutton, R. (**1992**). "Adapting bias by gradient descent: An incremental version of Delta-Bar-Delta," in *Proceedings of the 10th National Conference on Artificial Intelligence*, San Jose, CA (AAAI Press, Washington, DC), pp. 171–176.

Vindrola, L., Melon, M., Chamard, J.-C., and Gazengel, B. (**2021**). "Use of the filtered-*x* least-mean-squares algorithm to adapt personal sound zones in a car cabin," J. Acoust. Soc. Am. **150**(3), 1779–1793.

Ward, D., and Abhayapala, T. (**2001**). "Reproduction of a plane-wave sound field using an array of loudspeakers," IEEE Trans. Speech Audio Process. **9**(6), 697–707.

Yoshioka, T., and Nakatani, T. (**2012**). "Generalization of multi-channel linear prediction methods for blind MIMO impulse response shortening," IEEE Trans. Audio. Speech. Lang. Process. **20**(10), 2707–2720.