



Physics-informed adaptive filtering for acoustic echo cancellation

Michele Scarpiniti ^{ID}*, Danilo Communiello ^{ID}, Aurelio Uncini ^{ID}

Department of Information Engineering, Electronics and Telecommunications (DIET), Sapienza University of Rome, Via Eudossiana 18, Rome, 00184, Italy

ARTICLE INFO

Keywords:

Adaptive filter
Physics-informed adaptive filter (PIAF)
Acoustic echo cancellation (AEC)
Room impulse response (RIR)
NLMS

ABSTRACT

This paper introduces a physics-informed adaptive filtering framework for acoustic echo cancellation (AEC). Unlike conventional adaptive algorithms that rely solely on data-driven error minimization, the proposed method incorporates physically motivated priors derived from acoustic wave propagation and room impulse response structure. The echo path estimation problem is formulated as a composite stochastic optimization task, where the instantaneous squared error is regularized by constraints encoding causality, exponential energy decay, time-weighted sparsity of early reflections, spectral smoothness, and slow temporal variation of the acoustic path. The resulting Physics-Informed Normalized Least-Mean-Squares (PI-NLMS) algorithm performs stochastic gradient descent on the regularized cost while enforcing hard causality through projection. The proposed formulation restricts adaptation to a physically plausible echo-path manifold, improving conditioning and reducing variance without substantially increasing computational complexity. Theoretical analysis establishes mean convergence conditions and characterizes the bias-variance trade-off introduced by structured regularization. Simulation results under stationary and time-varying echo paths demonstrate faster convergence, improved steady-state misalignment, and enhanced echo return loss enhancement (ERLE) compared to conventional NLMS and sparsity-aware baselines.

1. Introduction

Acoustic echo cancellation (AEC) is a fundamental component in hands-free telephony, teleconferencing systems, smart speakers, and in-vehicle communication platforms [1,2]. The objective of AEC is to identify and suppress the acoustic coupling between a loudspeaker and a microphone, thus preventing far-end echoes from degrading communication quality [3]. This task is typically formulated as an adaptive identification problem in which the acoustic echo path is modeled as an unknown finite impulse response (FIR) filter.

Classical solutions rely on adaptive filtering algorithms such as the least-mean-squares (LMS), normalized LMS (NLMS), and recursive least squares (RLS) methods [4,5]. These algorithms estimate the echo path by minimizing the instantaneous or exponentially weighted squared error between the microphone signal and the estimated echo [6]. Their low computational complexity and robustness have made them the dominant approach in practical AEC systems for several decades.

Despite their maturity, adaptive filtering remains an active research area and several recent contributions have sought to improve the robustness, convergence speed, and theoretical understanding of echo path identification. For example, advanced optimization-based formulations have recently been investigated for acoustic system identification, in-

cluding Newton-Raphson solutions to exponentially weighted least-M estimation problems [7]. Multi-channel and multi-input acoustic scenarios have also attracted increasing attention, leading to novel filter decomposition strategies for MISO acoustic echo cancellation [8]. In parallel, theoretical analyses have revisited the fundamental performance limits of adaptive echo cancellers through Cramér-Rao lower bound formulations [9], providing useful benchmarks for evaluating practical adaptive algorithms. Nevertheless, despite their widespread use, conventional adaptive filters treat echo path identification as an unconstrained linear system identification problem. However, in practice, room impulse responses (RIRs) exhibit a well-known physical structure determined by acoustic propagation and room geometry [10]. Typical RIRs are strictly causal due to finite sound propagation speed, contain a dominant direct-path component followed by sparse early reflections, and exhibit exponentially decaying late reverberation. This exponential decay behavior has been widely studied in room acoustics and can be modeled using energy decay curves and reverberation time metrics derived from the impulse response [11–13]. Additionally, the impulse response tends to vary smoothly across time and frequency due to transducer characteristics and room acoustics, and evolves slowly over time except during abrupt environmental changes. Since classical adaptive filters do not explicitly incorporate these structural properties, their

* Corresponding author.

E-mail addresses: michele.scarpiniti@uniroma1.it (M. Scarpiniti), danilo.comminiello@uniroma1.it (D. Communiello), aurelio.uncini@uniroma1.it (A. Uncini).

convergence speed, robustness to double-talk, and steady-state misalignment can degrade in high-dimensional or low signal-to-noise ratio (SNR) scenarios [14].

To address some of these limitations, several extensions of adaptive filtering have been proposed. Recognizing that early reflections often produce sparse impulse responses, proportionate adaptive algorithms were introduced to accelerate convergence for sparse systems. Methods such as proportionate NLMS (PNLMS) [15,16], improved PNLMS (IPNLMS) [17], and related variants assign coefficient-dependent step sizes that favor active taps [18,19]. Other approaches incorporate ℓ_1 or ℓ_2 regularization to promote sparse or well-conditioned solutions, linking adaptive filtering with sparse system identification and compressed sensing principles [20–22]. Subsequent works explored structured modeling of echo paths and acoustic impulse responses to better capture the temporal distribution of reflections and reverberation [57] or exploit Kronecker and tensor decomposition to handle long RIRs [23,24]. Although these methods encourage sparsity, the priors employed remain generic and are not explicitly derived from the physics of acoustic propagation.

Another active research direction focuses on improving robustness against impulsive disturbances, non-Gaussian noise, and adverse acoustic conditions. Recent examples include generalized mixed-norm maximum correntropy approaches [25], and adaptive filtering methods based on generalized correntropy criteria combined with p -norm loss functions [26]. Fractional-order adaptive algorithms have also been proposed to improve convergence and robustness in challenging vehicular acoustic environments [27]. While these approaches modify the adaptation mechanism or cost function to improve robustness, they generally remain data-driven and do not explicitly exploit prior knowledge regarding the physical structure of room impulse responses. Kalman filter-based AEC algorithms introduce temporal evolution models for the echo path, improving tracking performance in time-varying environments [28,29]. These approaches allow prior assumptions on parameter evolution, but typically rely on simple Gaussian random-walk models rather than physically motivated structural constraints.

Frequency-domain adaptive filtering has also been widely adopted to efficiently handle long echo paths [30–32], also exploiting robust subband RLS algorithms [33]. Partitioned-block frequency-domain NLMS algorithms significantly reduce computational complexity while maintaining adaptation performance [34,35]. Nevertheless, frequency-domain implementations primarily address computational scalability and do not inherently enforce structural priors on the impulse response [36].

More recently, deep learning approaches have been explored for echo suppression and nonlinear echo cancellation [37–39,58]. Neural networks can model nonlinear loudspeaker distortions and complex acoustic interactions, often operating as residual echo suppressors or post-filters following linear adaptive cancellation [40,59]. For example, deep neural networks have been proposed to control the adaptation behavior of frequency-domain adaptive filters by predicting step-size parameters directly from signal features [37]. Hybrid systems that combine adaptive filters with neural residual models have shown promising performance improvements [41,42]. Recent developments have further expanded this trend through neural Kalman filtering frameworks that integrate state-space estimation with deep neural networks [43], convolutional recurrent architectures employing multiple adaptive sub-filters for monophonic acoustic echo cancellation [44], and attention-based architectures inspired by Kolmogorov-Arnold networks for double-talk and noisy scenarios [45]. Lightweight neural post-filtering strategies have also been investigated to improve speech enhancement performance while maintaining low computational complexity [46]. However, purely data-driven models typically require large training datasets and may lack interpretability or explicit consistency with acoustic physical principles [47]. Nevertheless, these methods typically impose generic mathematical priors rather than physically grounded constraints derived from acoustic propagation principles. A systematic framework that

directly embeds physical knowledge of room impulse responses into adaptive echo path identification remains underexplored.

In parallel with these developments, physics-informed learning has emerged as an important paradigm in scientific machine learning [48, 49]. Physics-informed methods incorporate governing equations or domain knowledge directly into optimization objectives, improving generalization, robustness, and data efficiency. While this idea has been successfully applied in many scientific domains, its integration into classical adaptive signal processing, and particularly into acoustic echo cancellation, remains largely unexplored.

Existing regularized adaptive filters can be interpreted as incorporating generic mathematical priors, such as sparsity or coefficient smoothness [21,50,60], or as introducing regularization terms to improve robustness against model uncertainties and interference [51,52]. However, these regularization strategies are typically not derived from the underlying physics of acoustic propagation, and a unified framework that systematically integrates multiple physically motivated properties of RIRs, such as causality, exponential energy decay, sparse early reflections, spectral smoothness, and slow temporal variation, has not yet been fully developed for adaptive echo path identification.

1.1. Main contributions and paper organization

Overall, while recent research has focused on increasingly sophisticated adaptive filtering algorithms, robust optimization techniques, state-space estimation methods, and neural-network-based echo cancellers, the explicit incorporation of physically motivated priors describing room acoustics remains relatively unexplored. This motivates the development of physics-informed adaptive filtering frameworks that exploit prior knowledge on the structure of room impulse responses while preserving the low complexity and online adaptability of classical signal-processing methods.

Motivated by these considerations, we reinterpret acoustic echo cancellation as a physics-constrained stochastic optimization problem. Instead of searching the entire coefficient space, the adaptive filter is guided toward a physically plausible echo-path manifold defined by propagation delay, energy decay structure, sparsity patterns, and smoothness properties derived from room acoustics.

In this paper, we introduce a Physics-Informed Adaptive Filtering (PIAF) framework for AEC. The proposed approach augments the instantaneous squared-error criterion with physically motivated regularization terms and projection operators. The resulting Physics-Informed NLMS (PI-NLMS) algorithm performs stochastic gradient descent on a composite objective that integrates both error minimization and acoustic structural priors. Hard constraints, such as causality, are enforced through projection, while soft constraints, including exponential decay, spectral smoothness, and temporal continuity, are incorporated via differentiable regularization terms.

The proposed framework offers several advantages. By restricting adaptation to a physically consistent parameter space, it improves the conditioning of the optimization problem and accelerates convergence. Structured regularization also reduces steady-state variance. Moreover, the formulation is modular and extensible, allowing additional priors or hybrid integration with nonlinear residual models while maintaining computational complexity comparable to standard NLMS algorithms.

The main contributions of this work are summarized as follows:

1. We formulate acoustic echo path identification as a physics-informed stochastic optimization problem that integrates multiple physically motivated structural constraints within a unified adaptive filtering framework.
2. We derive the Physics-Informed NLMS (PI-NLMS) algorithm, combining gradient-based regularization with projection-based constraints.
3. We analyze convergence properties and characterize the bias-variance trade-off induced by structured physical priors.

4. We demonstrate through simulations under stationary and time-varying echo paths that the proposed method achieves faster convergence, improved steady-state misalignment (MSD), and enhanced echo return loss enhancement (ERLE) compared to conventional and sparsity-aware baselines.

The remainder of the paper is organized as follows. Section 2 presents the AEC model. Section 3 introduces the physics-informed optimization framework and derives the PI-NLMS algorithm, while Section 4 analyzes convergence behavior. Section 5 presents simulation results under various acoustic scenarios. Finally, Section 6 concludes the paper and outlines future research directions.

1.2. Notation

In this paper, matrices are represented by boldface capital letters, i.e. $\mathbf{A} \in \mathbb{R}^{M \times N}$, while $\text{Tr}(\cdot)$ denotes the trace of a matrix. All vectors are column vectors, denoted by boldface lowercase letters, like $\mathbf{h} \in \mathbb{R}^{L \times 1} = [h_0, h_1, \dots, h_{L-1}]^T$, where h_i denotes the i -th individual entry of $\mathbf{h}$. In recursive algorithm definition, a discrete-time subscript index n is added. For example, the weight vector, calculated according to some law, is written as $\mathbf{w}_{n+1} = \mathbf{w}_n + \Delta \mathbf{w}_n$. In the case of signal regression, vectors are indicated as $\mathbf{x}_n \in \mathbb{R}^{L \times 1} = [x(n), x(n-1), \dots, x(n-L+1)]^T$. The Euclidean norm and the transpose of a vector are denoted by $\|\cdot\|$ and $(\cdot)^T$, respectively, while the expectation of a random variable is denoted by $\mathbb{E}[\cdot]$. We also have, for a vector $\mathbf{z}$, that: $\mathbf{z}^T \mathbf{z} = \|\mathbf{z}\|^2 = \text{Tr}(\mathbf{z}\mathbf{z}^T)$. The symbol ∇ denotes the gradient operator while $\lfloor \cdot \rfloor$ denotes the floor operator, i.e., it gives as output the greatest integer less than or equal to the input.

2. Acoustic echo cancellation (AEC)

Formally, let $x(n)$ denote the far-end signal driving the loudspeaker. The acoustic echo received at the microphone can be modeled as the convolution between $x(n)$ and the room impulse response (RIR) $h(n)$, which represents the acoustic path between loudspeaker and microphone. As shown in Fig. 1, the microphone signal is therefore given by:

$$d(n) = \sum_{k=0}^{L-1} h_k x(n-k) + s(n) + v(n) \equiv \mathbf{h}_0^T \mathbf{x}_n + s(n) + v(n),$$

where $\mathbf{h}_0 = [h_0, h_1, \dots, h_{L-1}]^T$ is the vector collecting the RIR coefficients, $\mathbf{x}_n = [x(n), x(n-1), \dots, x(n-L+1)]^T$ is the input vector (or frame) collecting the last L samples, $s(n)$ is the near-end speech signal, and $v(n)$ represents background noise characterized by a variance σ_v^2 .

The objective of the acoustic echo canceler is to estimate the echo component and subtract it from the microphone signal, resulting in an estimate of near-end speech. The echo path $h(n)$ is typically long (hundreds to thousands of taps), time-varying due to speaker or device motion, and excited by highly colored input signals such as speech or music. These characteristics make AEC a challenging adaptive system identification problem.

The conventional AEC architecture is based on an adaptive finite impulse response (FIR) filter $\mathbf{w}_n \in \mathbb{R}^L$ that models the unknown acoustic path. The estimated echo is:

$$y(n) = \mathbf{w}_n^T \mathbf{x}_n,$$

and the error signal used for adaptation is:

$$e(n) = d(n) - y(n).$$

This error is the residual signal delivered to the output and drives the adaptation algorithm to update $\mathbf{w}_n$.

The most widely used adaptation algorithms in AEC include: Least Mean Squares (LMS), Normalized LMS (NLMS), and Recursive Least Squares (RLS). Among these, NLMS is particularly popular due to its low computational complexity and robustness to signal scaling. The NLMS

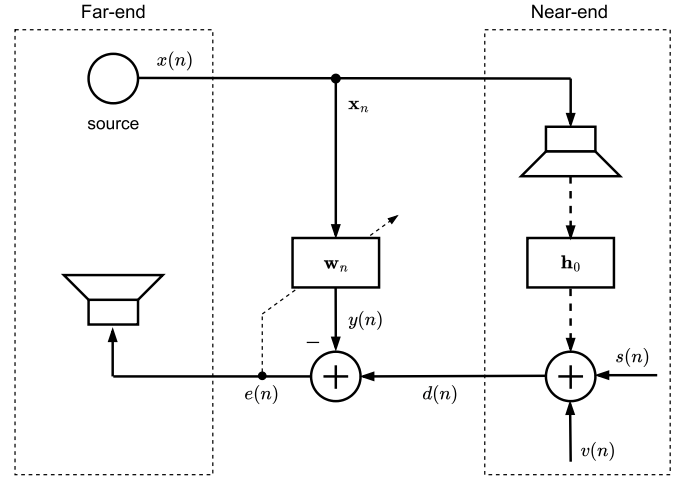


Fig. 1. Basic framework of the acoustic echo cancellation (AEC) problem.

update, obtained by minimizing the instantaneous squared error $e^2(n)$, is given by [4,5]:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \frac{\mu}{\|\mathbf{x}_n\|^2 + \epsilon} e(n) \mathbf{x}_n, \quad (1)$$

where μ is the step size and ϵ is a (small) regularization constant.

3. Proposed approach

With reference to Fig. 1, consider the AEC problem:

$$d(n) = \mathbf{h}_0^T \mathbf{x}_n + v(n), \quad (2)$$

where $d(n) \in \mathbb{R}$ is the microphone signal, $\mathbf{x}_n = [x(n), x(n-1), \dots, x(n-L+1)]^T$ is the far-end signal vector, $\mathbf{h}_0 \in \mathbb{R}^L$ is the true acoustic echo path, and $v(n)$ is background noise. The objective is to estimate $\mathbf{h}_0$ using an adaptive filter $\mathbf{w}_n$.

Despite decades of research, several fundamental challenges remain when highly colored excitation signals (such as speech) are used, the environment is represented by long and sparse impulse responses, time variability is present due to environmental changes, input signal dropouts, and model mismatch. These challenges often cause slow convergence, misadjustment, or instability in classical adaptive algorithms. However, the acoustic echo path is not arbitrary, because it possesses a typical physical structure. The energy typically decays with delay, its coefficients exhibit spectral smoothness, and the system varies slowly over time.

Unlike conventional adaptive filters that minimize only the instantaneous squared error:

$$J_{\text{NLMS}}(n) = \frac{1}{2} e^2(n),$$

where the error signal is defined as:

$$e(n) = d(n) - y(n) = d(n) - \mathbf{w}_n^T \mathbf{x}_n, \quad (3)$$

and do not explicitly exploit these RIR properties, we propose minimizing a physics-informed composite cost:

$$\begin{aligned} J(n) &= \frac{1}{2} e^2(n) + \sum_{i=1}^M \lambda_i \mathcal{P}_i(\mathbf{w}_n) = \frac{1}{2} e^2(n) + \sum_{i=1}^M \Phi_i(\mathbf{w}_n) \\ &= \frac{1}{2} e^2(n) + \Phi(\mathbf{w}_n), \end{aligned} \quad (4)$$

where $\mathcal{P}_i(\cdot)$ encode physical priors and $\lambda_i \geq 0$ are regularization parameters, $\Phi_i(\mathbf{w}_n) = \lambda_i \mathcal{P}_i(\mathbf{w}_n)$, and $\Phi(\mathbf{w}_n) = \sum_{i=1}^M \Phi_i(\mathbf{w}_n)$, where M is the number of priors. This defines a constrained search over a physically plausible echo-path manifold. Incorporating structured priors or physics-informed constraints into the adaptive filtering framework can

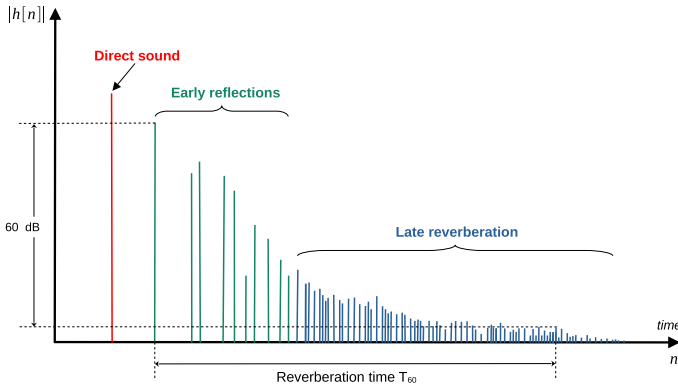


Fig. 2. Magnitude of a typical room impulse response (RIR).

improve convergence speed, robustness, and steady-state performance, especially under colored input and low signal-to-noise ratio conditions. We incorporate the physically motivated penalties described in the following subsection. We also anticipate that the considered physical priors are either quadratic by construction (e.g., exponential decay, smoothness, and slow variation constraints) or are treated through a quadratic approximation as in the case of the ℓ_1 sparsity penalty. This representation is adopted to obtain a compact formulation of the proposed physics-informed framework and to derive a tractable gradient-based adaptation rule.

3.1. Physical structure of the room impulse response and prior modeling

As described above, AEC aims at identifying the acoustic path between a loudspeaker and a microphone. This path is governed by the RIR, which is not arbitrary but obeys well-established physical laws of sound propagation in enclosed environments. As shown in Fig. 2, the RIR is composed of a first impulse characterizing the direct sound, a number of sparse early reflections, and a decaying late reverberation tail. Exploiting these physical properties motivates the structured priors introduced in the proposed Physics-Informed NLMS (PI-NLMS) algorithm. This section summarizes the main physical characteristics of RIRs and links each to the regularization terms used in the proposed method.

3.1.1. Causality

The RIR should be strictly causal due to the finite propagation speed of sound. In fact, the first impulse in the RIR, characterizing the direct sound is localized at the sample $\tau_{\min}$, corresponding to the minimum propagation delay, which depends on the distance d between the source and the microphone, and the sampling frequency f_s :

$$\tau_{\min} = \left\lceil \frac{d \times f_s}{c} \right\rceil,$$

where c is the sound speed in the room.

The causality constraint can be interpreted as a hard projection operator:

$$\mathbf{w}_n \leftarrow \Pi_C(\mathbf{w}_n),$$

where C is the convex set:

$$C = \{\mathbf{w} : w_k = 0 \text{ for } k < \tau_{\min}\}. \quad (5)$$

The projection effectively reduces the dimensionality of the parameters from L to $L - \tau_{\min}$, improving the conditioning.

3.1.2. Energy decay and exponential attenuation

In enclosed spaces, acoustic energy decays due to geometric spreading, air absorption, surface absorption, and scattering. According to

Sabine's reverberation theory [10], the energy envelope of the RIR follows approximately an exponential decay:

$$E[|h(k)|^2] \propto e^{-2kT_s/\tau} = e^{-2k\rho T_s},$$

where T_s is the sampling period and $\tau = 1/\rho$ is related to the reverberation time T_{60} . Thus, early taps contain most of the energy, while late taps are progressively weaker. This implies that large coefficients at long delays are physically unlikely. For this reason, we can adopt an exponential decay prior:

$$\mathcal{P}_{\text{decay}}(\mathbf{w}) = \frac{1}{2} \sum_{k=0}^{L-1} \delta_k w_k^2,$$

with $\delta_k = e^{\alpha k}$, where α controls the decay weight. From room acoustics and Sabine's reverberation theory, the parameter α is related to the reverberation time T_{60} (see [10] for details):

$$\alpha = \rho T_s = \frac{3 \log(10)}{T_{60}} T_s \approx \frac{6.91}{T_{60}} T_s. \quad (6)$$

This enforces a stronger penalization for late taps and an improved robustness under highly reverberant conditions. In matrix form at time instant n we have:

$$\Phi_{\text{decay}}(\mathbf{w}_n) = \lambda_{\text{decay}} \mathcal{P}_{\text{decay}}(\mathbf{w}_n) = \frac{1}{2} \mathbf{w}_n^T \Lambda_{\text{decay}} \mathbf{w}_n, \quad (7)$$

where $\Lambda_{\text{decay}} = \lambda_{\text{decay}} \Delta = \lambda_{\text{decay}} \text{diag}(\delta_k)$.

3.1.3. Sparsity of early reflections

In many practical environments, the early part of the RIR consists of a small number of dominant reflections, while late reverberation becomes diffuse and noise-like. Thus, especially in short acoustic paths, the RIR is approximately sparse. This implies that many coefficients are near zero.

For this reason, denoting with $\odot$ the element-wise multiplication, the corresponding prior is:

$$\mathcal{P}_{\ell_1}(\mathbf{w}) = \|\beta \odot \mathbf{w}\|_1 = \sum_{k=0}^{L-1} \beta_k |w_k|,$$

where $\beta_k = e^{-\eta k}$ encourages strong early sparsity but relaxes later for diffused late reverberation. The parameter η is indirectly related through the reflection density, which depends on the volume and absorption of the room, and assume the expression [10]: $\eta = \eta_0 T_{60}/(LT_s)$, with η_0 related to the early region. This prior promotes sparse early reflections and suppression of spurious taps under noisy excitation.

This is a non-quadratic penalty term. However, we can rewrite the ℓ_1 prior as a weighted ℓ_2 one:

$$|w_k| = \frac{w_k^2}{|w_k|}.$$

Therefore, we can rewrite it at time n in matrix form as:

$$\Phi_{\ell_1}(\mathbf{w}_n) = \lambda_{\ell_1} \mathcal{P}_{\ell_1}(\mathbf{w}_n) = \frac{1}{2} \mathbf{w}_n^T \Lambda_{\ell_1} \mathbf{w}_n, \quad (8)$$

where, using a small parameter ε that prevents division by zero, we set:

$$\Lambda_{\ell_1} = \lambda_{\ell_1} \text{diag}\left(\frac{\beta_k}{|w_k| + \varepsilon}\right).$$

As a remark, compared with the exact ℓ_1 penalty, the approximation produces a smoother shrinkage effect and generally results in slightly weaker sparsity promotion, particularly for coefficients close to zero. Consequently, the steady-state solution may exhibit a small reduction in sparsity but also a reduced estimation bias. In the context of room impulse response identification, where only the early reflections are strongly sparse while the late reverberation tail is typically dense, this behavior provides a favorable compromise between sparsity promotion and preservation of reverberant components.

3.1.4. Temporal smoothness of the impulse response

Although the RIR contains reflections, its fine structure is not arbitrarily oscillatory. The acoustic transfer function is band-limited by the loudspeaker and microphone responses, air propagation, and surface scattering. Therefore, the RIR coefficients vary smoothly at the scale of the sampling period. Abrupt oscillations between adjacent taps would imply unrealistically sharp frequency-domain behavior. This implies that the RIR has limited curvature.

For this reason, we can adopt the following smoothness penalty:

$$P_{ts}(\mathbf{w}) = \frac{1}{2} \sum_{k=1}^{L-1} (w_k - w_{k+1})^2,$$

or, in matrix form at time n :

$$\Phi_{ts}(\mathbf{w}_n) = \lambda_{ts} P_{ts}(\mathbf{w}_n) = \frac{1}{2} \mathbf{w}_n^\top \mathbf{\Lambda}_{ts} \mathbf{w}_n, \quad (9)$$

where:

$$\mathbf{\Lambda}_{ts} = \lambda_{ts} \mathbf{D}^\top \mathbf{D} \equiv \lambda_{ts} \mathbf{L}_s,$$

where $\mathbf{D}$ is the first-order discrete difference operator, which leads to the tridiagonal Laplacian matrix:

$$\mathbf{L}_s = \mathbf{D}^\top \mathbf{D} = \begin{bmatrix} 1 & -1 & 0 & 0 & \dots \\ -1 & 2 & -1 & 0 & \dots \\ 0 & -1 & 2 & -1 & \dots \\ \vdots & & \ddots & & 1 \end{bmatrix}$$

This corresponds to a discrete Laplacian operator, penalizing curvature ($\nabla^2 w_k \approx 2w_k - w_{k-1} - w_{k+1}$). The smoothness prior suppresses non-physical high-frequency fluctuations in the estimated impulse response, acting as a spatial low-pass filter on the tap sequence.

3.1.5. Spectral smoothness of the impulse response

The room impulse response $h[k]$ and its frequency response $H(e^{j\omega}) = \sum_{k=0}^{L-1} h[k]e^{-j\omega k}$ are related via the discrete-time Fourier transform. While the impulse response may appear irregular in time due to reflections, the acoustic transfer function is inherently spectrally smooth. This spectral smoothness arises from fundamental physical constraints of acoustic wave propagation.

Several physical mechanisms enforce smooth frequency behavior. First, loudspeakers and microphones have smooth frequency responses. Their transfer functions do not exhibit abrupt spectral variations. Thus, the overall acoustic channel is filtered by smooth transducer responses. Second, air absorption increases smoothly with frequency. The attenuation factor can be approximated as:

$$A(f) \propto e^{-\kappa(f)d},$$

where $\kappa(f)$ is a slowly varying function of frequency and d is the propagation distance. Therefore, the magnitude of the frequency response decays smoothly. Third, reflection coefficients of walls and objects vary smoothly with frequency, except at structural resonances. In typical rooms, these resonances are relatively sparse and low-Q. Thus, the global transfer function remains spectrally smooth. Fourth, late reverberation behaves approximately as a diffuse sound field, whose power spectral density is smooth. Therefore, high-frequency oscillations in $H(e^{j\omega})$ are physically unlikely.

Spectral smoothness means that $d^2 H(e^{j\omega})/d\omega^2$ is small. We can adopt the following smoothness penalty:

$$P_{ss}(\mathbf{w}) = \frac{1}{2} \sum_{k=1}^{L-1} (W_k - W_{k+1})^2,$$

where W_k represents the k -th frequency bin of $W(f)$. The spectral smoothness penalty can be expressed, in matrix form at time n , as:

$$\Phi_{ss}(\mathbf{w}_n) = \lambda_{ss} P_{ss}(\mathbf{w}_n) = \frac{\lambda_{ss}}{2} \mathbf{w}_n^\top \mathbf{F}^\top \mathbf{D}^\top \mathbf{D} \mathbf{F} \mathbf{w}_n = \frac{1}{2} \mathbf{w}_n^\top \mathbf{\Lambda}_{ss} \mathbf{w}_n, \quad (10)$$

where $\mathbf{W}_n = \mathbf{F} \mathbf{w}_n$ is the Fourier transform of $\mathbf{w}_n$, $\mathbf{F}$ is the Fourier matrix, and:

$$\mathbf{\Lambda}_{ss} = \lambda_{ss} \mathbf{F}^\top \mathbf{D}^\top \mathbf{D} \mathbf{F} = \lambda_{ss} \mathbf{F}^\top \mathbf{L}_s \mathbf{F}.$$

Physically, it suppresses artificial spectral ripples, reduces variance amplification under colored excitation, and improves robustness to input spectral nulls.

Without smoothness regularization NLMS may create oscillatory coefficient patterns, especially under colored input and when input spectrum has deep nulls. These oscillations correspond to non-physical transfer functions. The smoothness prior constrains the solution to physically plausible acoustic responses.

3.1.6. Slow temporal variation of the acoustic path

The acoustic path changes due to possibly: device movement, speaker movement, temperature variations, and minor environmental changes. However, these changes occur slowly relative to the sampling rate. Therefore:

$$\mathbf{w}_n \approx \mathbf{w}_{n-1}.$$

This implies that large frame-to-frame variations are physically unlikely. The corresponding prior is:

$$P_{slow}(\mathbf{w}_n) = \frac{1}{2} \|\mathbf{w}_n - \mathbf{w}_{n-1}\|^2.$$

Slow variation prior pulls weights toward previous estimate. Thus, PI-NLMS incorporates a physically justified state-space evolution model. This prior can be rewritten as:

$$\begin{aligned} \Phi_t(\mathbf{w}_n) &= \lambda_t P_{slow}(\mathbf{w}_n) = \frac{\lambda_t}{2} \mathbf{w}_n^\top \mathbf{w}_n - \lambda_t \mathbf{w}_{n-1}^\top \mathbf{w}_n + \frac{\lambda_t}{2} \mathbf{w}_{n-1}^\top \mathbf{w}_{n-1} \\ &= \frac{1}{2} \mathbf{w}_n^\top \mathbf{\Lambda}_t \mathbf{w}_n - \mathbf{b}^\top \mathbf{w}_n + a, \end{aligned} \quad (11)$$

where we set:

$$\mathbf{\Lambda}_t = \lambda_t \mathbf{I}, \quad \mathbf{b} = \lambda_t \mathbf{w}_{n-1} \quad \text{and} \quad a = \frac{\lambda_t}{2} \mathbf{w}_{n-1}^\top \mathbf{w}_{n-1}.$$

3.1.7. Total prior

The total prior used in PI-NLMS is therefore:

$$\begin{aligned} \Phi(\mathbf{w}_n) &\equiv \sum_i \lambda_i P_i(\mathbf{w}_n) = \sum_i \Phi_i(\mathbf{w}_n) \\ &= \frac{1}{2} \mathbf{w}_n^\top (\mathbf{\Lambda}_{decay} + \mathbf{\Lambda}_{\ell_1} + \mathbf{\Lambda}_{ts} + \mathbf{\Lambda}_{ss} + \mathbf{\Lambda}_t) \mathbf{w}_n - \mathbf{b}^\top \mathbf{w}_n + a \\ &= \frac{1}{2} \mathbf{w}_n^\top \mathbf{\Lambda} \mathbf{w}_n - \mathbf{b}^\top \mathbf{w}_n + a, \end{aligned} \quad (12)$$

where $i \in \{\text{decay}, \text{ts}, \text{ss}, \ell_1, t\}$ corresponds to each of the five considered priors and $\mathbf{\Lambda} = \mathbf{\Lambda}_{decay} + \mathbf{\Lambda}_{\ell_1} + \mathbf{\Lambda}_{ts} + \mathbf{\Lambda}_{ss} + \mathbf{\Lambda}_t$.

3.2. Algorithm derivation

To derive the proposed PI-NLMS algorithm, we compute the gradient of the composite cost function in (4):

$$\nabla J(n) = \frac{1}{2} \nabla e^2(n) + \nabla \Phi(\mathbf{w}_n). \quad (13)$$

The gradient of the first term is the standard LMS one:

$$\frac{1}{2} \nabla e^2(n) = \frac{1}{2} 2 e(n) \frac{\partial e(n)}{\partial \mathbf{w}_n} = e(n) \frac{\partial(d(n) - \mathbf{w}_n^\top \mathbf{x}_n)}{\partial \mathbf{w}_n} = -e(n) \mathbf{x}_n. \quad (14)$$

For the gradients of the total prior in (12), we obtain the following results:

$$\nabla \Phi(\mathbf{w}_n) \equiv \mathbf{g}_{phys}(n) = \mathbf{\Lambda} \mathbf{w}_n - \mathbf{b}. \quad (15)$$

Therefore, by using the normalized step-size:

$$\mu_n = \frac{\mu}{\|\mathbf{x}_n\|^2 + \epsilon}, \quad (16)$$

and taking into account the gradient terms in (14) and (15), the proposed PI-NLMS update becomes:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu_n e(n) \mathbf{x}_n - \mu_n \mathbf{g}_{phys}(n). \quad (17)$$

Then we apply the causality projection in (5), which simply enforces:

$$w_k(n+1) = 0, \quad \forall k < \tau_{\min}, \quad (18)$$

given a known minimum propagation delay $\tau_{\min}$. The proposed Physics-Informed NLMS (PI-NLMS) algorithm is summarized in Algorithm 1.

Algorithm 1 The proposed PI-NLMS algorithm.

Input: $x[n]$, $d[n]$, μ , ϵ , λ_i , $\tau_{\min}$

Initialize: $\mathbf{w}_0 = [1, 0, 0, \dots, 0]$

for $n = 0, 1, \dots, N-1$ **do**

 Form the input vector $\mathbf{x}_n$

 Compute output: $y(n) = \mathbf{w}_n^T \mathbf{x}_n$

 Compute error: $e(n) = d(n) - y(n)$

 Compute normalized step: $\mu_n = \mu / (\|\mathbf{x}_n\|^2 + \epsilon)$

 Compute physics gradients: $\mathbf{g}_{phys}(n) = \sum_i \lambda_i \nabla \mathcal{P}_i$

 Update the filter: $\mathbf{w}_{n+1} = \mathbf{w}_n + \mu_n e(n) \mathbf{x}_n - \mu_n \mathbf{g}_{phys}(n)$

 Enforce causality: $w_k(n+1) = 0, \forall k < \tau_{\min}$

end for

Return: $\mathbf{w}_{n+1}, e(n)$

Importantly, the proposed PI-NLMS algorithm preserves the simplicity and scalability of NLMS while incorporating physically motivated priors that improve convergence behavior and robustness in challenging acoustic environments.

3.3. Computational complexity

This section evaluates the computational cost of the proposed PI-NLMS algorithm. Let L denote the adaptive filter length. The analysis focuses on the number of multiplications and additions required per iteration, which are the dominant operations in real-time acoustic echo cancellation systems.

The standard NLMS algorithm performs the following operations at each iteration: evaluation of the filter output, which requires L multiplications and $L-1$ additions; evaluation of the signal power $\mathbf{x}_n^T \mathbf{x}_n + \epsilon$, which requires L multiplications and L additions; computation of the error $e(n)$, which requires 1 addition; and the weight update, which requires L multiplications and L additions. Thus, the total computational cost of NLMS is $3L$ multiplications and $3L$ additions.

The proposed PI-NLMS algorithm introduces five physically motivated priors (decay prior, ℓ_1 sparsity prior, temporal smoothness prior, spectral smoothness prior, and slow-variation prior). Each prior contributes with an additional gradient term in the weight update. The decay prior corresponds to a diagonal weighting matrix applied to the coefficient vector, which requires L multiplications. The sparsity regularization involves evaluating the sign function and scaling by a constant, resulting in approximately L multiplications. The temporal smoothness prior penalizes first-order differences between adjacent filter taps, requiring approximately L multiplications and $2L$ additions. Similarly, the spectral smoothness prior corresponds to a first-order difference but applied to adjacent filter bins in the frequency domain; hence, its evaluation requires, if the FFT algorithm is applied, approximately $L/2 \log_2 L$ complex multiplications and $L \log_2 L$ complex additions. Finally, the slow-variation prior penalizes the difference between consecutive weight vectors and requires L multiplications and L additions.

Combining the NLMS baseline with the additional regularization terms, the total computational cost of PI-NLMS becomes approximately $7L + L/2 \log_2 L$ multiplications and $6L + L \log_2 L$ additions per iteration, as summarized in Table 1. Let us note that the spectral smoothness prior represents the most computationally demanding component of the algorithm, accounting for a significant fraction of the additional

Table 1

Computational complexity of the proposed PI-NLMS algorithm.

Term	Multiplications	Additions
NLMS	$3L$	$3L$
Decay	L	0
Sparsity	L	0
Temporal smooth	L	$2L$
Spectral smooth	$\frac{L}{2} \log_2 L$	$L \log_2 L$
Slow variation	L	L
Total w/ spectral smooth	$7L + \frac{L}{2} \log_2 L$	$6L + L \log_2 L$
Total w/o spectral smooth	$7L$	$6L$

operations, and leading to an asymptotical complexity of $\mathcal{O}(L \log_2 L)$. In cases where computational resources are limited, the spectral smoothness prior can be omitted. In this case, despite the additional regularization terms, the proposed algorithm maintains linear asymptotic computational complexity $\mathcal{O}(L)$ with respect to the filter length. Therefore, the algorithm remains suitable for real-time acoustic echo cancellation applications where the filter length can range from several hundred to several thousand taps. Nevertheless, the overall computational burden remains modest compared with more complex adaptive filtering techniques such as affine projection algorithms or recursive least squares methods.

If necessary, further complexity reductions can be achieved by evaluating the regularization term every K iterations rather than at every sample. Alternatively, the regularization may be activated through an event-triggered strategy, for example when the residual error exceeds a predefined threshold. Such approaches reduce the average computational burden while preserving most of the benefits of the proposed physics-informed regularization and are particularly suitable for slowly varying echo paths. Moreover, the proposed framework can be naturally integrated into partitioned-block frequency-domain adaptive filtering implementations commonly used in practical AEC systems, enabling efficient processing of long room impulse responses while retaining the benefits of the proposed physical priors [4].

4. Convergence analysis

In this section, we analyze the mean and mean-square behavior of the proposed PI-NLMS algorithm. The analysis extends classical NLMS convergence results to the regularized stochastic optimization framework.

Let $d(n)$ be the desired signal in (2), where $\mathbf{h}_0 \in \mathbb{R}^L$ is the true echo path, $\mathbf{x}_n \in \mathbb{R}^L$ is a zero-mean input vector, and $v(n)$ is zero-mean white noise with variance σ_v^2 and independent of $x(n)$. Let us also define $\mathbf{R}_x = \mathbb{E}[\mathbf{x}_n \mathbf{x}_n^T]$ the auto-correlation matrix of the input signal $\mathbf{x}_n$. For analytical tractability, we assume that:

1. $\mathbf{x}_n$ is independent of $\mathbf{w}_n$ and $v(n)$ (independence assumption);
2. μ_n is sufficiently small (small step-size approximation).

4.1. Mean convergence analysis

Define the weight error vector:

$$\tilde{\mathbf{w}}_n = \mathbf{h}_0 - \mathbf{w}_n. \quad (19)$$

By using the fact that $d(n) = \mathbf{x}_n^T \mathbf{h}_0 + v(n)$ and $y(n) = \mathbf{w}_n^T \mathbf{x}_n$, we have:

$$\begin{aligned} \mathbf{x}_n e(n) &= \mathbf{x}_n (d(n) - y(n)) = \mathbf{x}_n (\mathbf{x}_n^T \mathbf{h}_0 + v(n) - \mathbf{x}_n^T \mathbf{w}_n) \\ &= \mathbf{x}_n \mathbf{x}_n^T (\mathbf{h}_0 - \mathbf{w}_n) + \mathbf{x}_n v(n) = \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n + \mathbf{x}_n v(n). \end{aligned} \quad (20)$$

Hence, taking the expectation of both sides of (20) and invoking the standard independence assumption between the current input vector and the weight-error vector leads to:

$$\mathbb{E}[e(n) \mathbf{x}_n] = \mathbf{R}_x \mathbb{E}[\tilde{\mathbf{w}}_n]. \quad (21)$$

On the other side, from (12) and (15) we have:

$$\mathbf{g}_{phys}(n) \equiv \nabla \Phi(\mathbf{w}_n) = \Lambda \mathbf{w}_n - \mathbf{b}, \quad \text{and} \quad \nabla \mathbf{g}_{phys}(n) \equiv \nabla^2 \Phi(\mathbf{w}_n) = \Lambda \quad (22)$$

Using (17), (19), and the first equation in (22), we obtain:

$$\begin{aligned} \tilde{\mathbf{w}}_{n+1} &= \mathbf{h}_0 - \mathbf{w}_{n+1} = \mathbf{h}_0 - (\mathbf{w}_n + \mu_n e(n) \mathbf{x}_n - \mu_n \Lambda \mathbf{w}_n + \mu_n \mathbf{b}) \\ &= \tilde{\mathbf{w}}_n - \mu_n e(n) \mathbf{x}_n + \mu_n \Lambda \mathbf{w}_n - \mu_n \mathbf{b} \\ &= \tilde{\mathbf{w}}_n - \mu_n e(n) \mathbf{x}_n + \mu_n \Lambda (\mathbf{h}_0 - \tilde{\mathbf{w}}_n) - \mu_n \mathbf{b}, \end{aligned}$$

where the last line has been derived by using the fact that from (19) we have $\mathbf{w}_n = \mathbf{h}_0 - \tilde{\mathbf{w}}_n$. Taking expectations of both sides and using the result in (21), after some manipulation yields:

$$\mathbb{E}[\tilde{\mathbf{w}}_{n+1}] = (\mathbf{I} - \mu_n \mathbf{R}_x - \mu_n \Lambda) \mathbb{E}[\tilde{\mathbf{w}}_n] + \mu_n \mathbf{c}, \quad (23)$$

where $\mathbf{c} = \Lambda \mathbf{h}_0 - \mathbf{b}$. In order (23) be convergent, it is necessary that:

$$\rho(\mathbf{I} - \mu_n (\mathbf{R}_x + \Lambda)) < 1,$$

where $\rho(\cdot)$ denotes the spectral radius. This is accomplished if:

$$\left| \lambda_{\max} [\mathbf{I} - \mu_n (\mathbf{R}_x + \Lambda)] \right| < 1 \quad \Rightarrow \quad 0 < \mu_n \lambda_{\max} [\mathbf{R}_x + \Lambda] < 2,$$

where $\lambda_{\max}[\mathbf{A}]$ corresponds to the maximum eigenvalue of the matrix $\mathbf{A}$. The previous result is obtained due to the identity matrix $\mathbf{I}$, since in this case the eigenvalues of matrix $\mathbf{I} - \mathbf{G}$ are $1 - \lambda_i[\mathbf{G}]$ under appropriate ordering. Therefore, the PI-NLMS algorithm is stable in mean if:

$$0 < \mu_n < \frac{2}{\lambda_{\max} [\mathbf{R}_x + \Lambda]}. \quad (24)$$

Compared to standard NLMS, the effective correlation matrix becomes: $\mathbf{R}_{\text{eff}} \equiv \mathbf{B} = \mathbf{R}_x + \Lambda$. Hence, the regularization term increases eigenvalues, by improving conditioning and potentially allowing larger effective stability margins. Thus, structured physical priors can improve convergence behavior.

The recursion in (23) can be further solved using the approach in [53] as follows:

$$\mathbb{E}[\tilde{\mathbf{w}}_n] = (\mathbf{I} - \mu_n \mathbf{R}_x - \mu_n \Lambda)^n \mathbb{E}[\tilde{\mathbf{w}}_0] + \mu_n \sum_{p=0}^{n-1} (\mathbf{I} - \mu_n \mathbf{R}_x - \mu_n \Lambda)^{n-1-p} \mathbf{c}, \quad (25)$$

where $\mathbf{w}_0$ denotes the initial value of $\mathbf{w}_n$. Since the eigenvalues of the first term in (25) are smaller than one if (24) is satisfied and the second term is a geometric series, the steady-state solution of (23) is:

$$\tilde{\mathbf{w}}_{\infty} = \lim_{n \rightarrow \infty} \mathbb{E}[\tilde{\mathbf{w}}_n] = \mu_n (\mu_n \mathbf{R}_x + \mu_n \Lambda)^{-1} \mathbf{c} = (\mathbf{R}_x + \Lambda)^{-1} \mathbf{c}. \quad (26)$$

Therefore, the steady-state solution is:

$$\mathbf{w}_{\infty} = \mathbf{h}_0 - (\mathbf{R}_x + \Lambda)^{-1} \mathbf{c}. \quad (27)$$

When we have no priors (i.e. $\Lambda = \mathbf{0}$ and $\mathbf{b} = 0$, and therefore $\mathbf{c} = \Lambda \mathbf{h}_0 - \mathbf{b} = 0$) (27) reduces to $\mathbf{w}_{\infty} = \mathbf{h}_0$, otherwise a biased solution will be found in general. Hence, the structured prior introduces a controlled bias (toward physically plausible subspace), and a reduced variance in steady-state misadjustment, as shown in next sub-section. If the true echo path satisfies the imposed physical structure (e.g., exponential decay), the bias becomes negligible.

This analysis has been performed by considering the normalized step-size μ_n in (16). We can also consider directly the non-normalized step-size μ by replacing $\mathbf{R}_x$ with:

$$\tilde{\mathbf{R}}_x = \mathbb{E} \left[\frac{\mathbf{x}_n \mathbf{x}_n^T}{\|\mathbf{x}_n\|^2 + \epsilon} \right].$$

For white input, $L \gg 1$, and ϵ negligible, we have the following approximation: $\tilde{\mathbf{R}}_x \approx \mathbf{R}_x / L$.

4.2. Mean-square behavior

In this section, we aim at analyzing the mean-square behavior in terms of steady-state mean square deviation (MSD) and excess mean-square error (EMSE), which are defined respectively as:

$$\text{MSD} = \mathbb{E}[\|\tilde{\mathbf{w}}_n\|^2] \quad \text{and} \quad \text{EMSE} = \mathbb{E}[e_a^2(n)],$$

where $e_a(n)$ is the *a priori* error defined as:

$$e_a(n) = \tilde{\mathbf{w}}_n^T \mathbf{x}_n. \quad (28)$$

In addition, we also investigate the relationship between the EMSE and the steady-state value of the echo return loss enhancement (ERLE), defined in (52).

4.2.1. Steady-state mean square deviation (MSD)

Assume first the unbiased case: $\mathbf{b} = \Lambda \mathbf{h}_0$, so that bias term $\mathbf{c} = \Lambda \mathbf{h}_0 - \mathbf{b}$ in (23) disappears. Let us define $\mathbf{P}_n = \mathbb{E}[\tilde{\mathbf{w}}_n \tilde{\mathbf{w}}_n^T]$ the covariance matrix of weight error vector $\tilde{\mathbf{w}}_n$. Using (23), we have the recursion:

$$\mathbf{P}_{n+1} = \mathbf{A} \mathbf{P}_n \mathbf{A}^T + \mu_n^2 \mathbf{Q}, \quad (29)$$

where:

$$\mathbf{A} = \mathbf{I} - \mu_n (\mathbf{R}_x + \Lambda)$$

and

$$\mathbf{Q} = \mathbb{E}[\mathbf{x}_n v(n) v(n) \mathbf{x}_n^T] = \sigma_v^2 \mathbf{R}_x, \quad (30)$$

since $v(n)$ is independent of $\mathbf{x}_n$.

For simplicity of notation, let us pose $\mathbf{B} = \mathbf{R}_x + \Lambda$ and observe that both $\mathbf{A}$ and $\mathbf{B}$ are symmetric matrices (i.e., $\mathbf{A}^T = \mathbf{A}$ and $\mathbf{B}^T = \mathbf{B}$). Hence the covariance matrix recursion (29) can be written as:

$$\begin{aligned} \mathbf{P}_{n+1} &= (\mathbf{I} - \mu_n \mathbf{B}) \mathbf{P}_n (\mathbf{I} - \mu_n \mathbf{B})^T + \mu_n^2 \mathbf{Q} = (\mathbf{P}_n - \mu_n \mathbf{B} \mathbf{P}_n) (\mathbf{I} - \mu_n \mathbf{B})^T + \mu_n^2 \mathbf{Q} \\ &= \mathbf{P}_n - \mu_n (\mathbf{B} \mathbf{P}_n + \mathbf{P}_n \mathbf{B}) + \mu_n^2 \mathbf{B} \mathbf{P}_n \mathbf{B} + \mu_n^2 \mathbf{Q} \approx \mathbf{P}_n - \mu_n (\mathbf{B} \mathbf{P}_n + \mathbf{P}_n \mathbf{B}) \\ &\quad + \mu_n^2 \mathbf{Q}. \end{aligned} \quad (31)$$

Last approximation is justified by the fact that, if $\mu_n \ll 1$, $\|\mu_n^2 \mathbf{B} \mathbf{P}_n \mathbf{B}\| \ll \|\mu_n \mathbf{B} \mathbf{P}_n\|$, resorting to the small step-size approximation. This is a standard approximation in adaptive filtering analysis that enables a tractable closed-form characterization of the steady-state MSD.

At steady-state for $n \rightarrow \infty$, since the system has converged, we have: $\mathbf{P}_{n+1} = \mathbf{P}_n = \mathbf{P} \equiv \mathbf{P}$. Therefore, from previous relationship in (31), we obtain the following discrete Lyapunov equation:

$$\mathbf{B} \mathbf{P} + \mathbf{P} \mathbf{B} = \mu_n \mathbf{Q}. \quad (32)$$

Now, diagonalize the matrix $\mathbf{B}$ by computing the eigenvalue decomposition:

$$\mathbf{B} = \mathbf{U} \mathbf{\Gamma} \mathbf{U}^T, \quad (33)$$

where $\mathbf{U}$ is the matrix of eigenvectors and $\mathbf{\Gamma} = \text{diag}(\gamma_i)$ is the matrix that collects the eigenvalues γ_i , for $i = 1, \dots, L$, on the main diagonal. Let us also consider the transformed matrix:

$$\mathbf{P}' = \mathbf{U}^T \mathbf{P} \mathbf{U}. \quad (34)$$

By using (33), the Lyapunov Eq. (32) can be rewritten as:

$$\mathbf{U} \mathbf{U}^T \mathbf{P} + \mathbf{P} \mathbf{U} \mathbf{U}^T = \mu_n \mathbf{Q}.$$

Multiply this equation for $\mathbf{U}^T$ from the left side and for $\mathbf{U}$ from the right side, obtaining:

$$\begin{aligned} \mathbf{U}^T (\mathbf{U} \mathbf{U}^T \mathbf{P} + \mathbf{P} \mathbf{U} \mathbf{U}^T) \mathbf{U} \\ = \mu_n \mathbf{U}^T \mathbf{Q} \mathbf{U} \quad \Rightarrow \quad \mathbf{U}^T \mathbf{U} \mathbf{U}^T \mathbf{P} \mathbf{U} + \mathbf{U}^T \mathbf{P} \mathbf{U} \mathbf{U}^T \mathbf{U} = \mu_n \sigma_v^2 \mathbf{U}^T \mathbf{R}_x \mathbf{U}, \end{aligned}$$

where the last identity is obtained by the definition of $\mathbf{Q}$ in (30). Using now (34) and remembering that the eigenvector matrix is orthonormal (i.e., $\mathbf{U}^T \mathbf{U} = \mathbf{I}$), we obtain:

$$\mathbf{\Gamma} \mathbf{P}' + \mathbf{P}' \mathbf{\Gamma} = \mu_n \sigma_v^2 \mathbf{R}'_x, \quad (35)$$

where $\mathbf{R}'_x = \mathbf{U}^T \mathbf{R}_x \mathbf{U}$ is the transformed autocorrelation matrix.

We can now consider the per mode components:

$$(\mathbf{\Gamma} \mathbf{P}')_{ij} = \gamma_i P'_{ij} \quad \text{and} \quad (\mathbf{P}' \mathbf{\Gamma})_{ij} = \gamma_j P'_{ij},$$

for $i = 1, 2, \dots, L$ where G_{ij} denotes the entry of row i and column j of matrix $\mathbf{G}$. Therefore, the per mode version of Lyapunov Eq. (35) becomes:

$$(\gamma_i + \gamma_j) P'_{ij} = \mu_n \sigma_v^2 R'_{x,ij}. \quad (36)$$

For the MSD = $\text{Tr}(\mathbf{P})$, we have to consider only the diagonal entries, i.e., $i = j$. Hence, (36) becomes:

$$2\gamma_i P'_{ii} = \mu_n \sigma_v^2 R'_{x,ii},$$

from which we derive:

$$P'_{ii} = \frac{\mu_n \sigma_v^2}{2} \frac{R'_{x,ii}}{\gamma_i}, \quad (37)$$

where γ_i are the eigenvalues of the matrix $\mathbf{R}_x + \Lambda$.

For analytical tractability, and to further decompose the previous equation, we assume that the input covariance matrix $\mathbf{R}_x$ and the regularization operator Λ are jointly diagonalizable. Although this assumption does not strictly hold in general, it is commonly adopted in adaptive filtering analysis and becomes accurate under standard approximations, such as large filter lengths and Toeplitz-to-circulant equivalence. This allows decoupling the dynamics into independent modes and provides useful insight into the effect of the proposed regularization. Since Γ diagonalizes $\mathbf{B} = \mathbf{R}_x + \Lambda$, and making the assumption that these two matrices are jointly diagonalizable, we have $\gamma_i = \lambda_{x,i} + \lambda_{\Lambda,i}$, where $\lambda_{x,i}$ and $\lambda_{\Lambda,i}$ are the i -th eigenvalues of $\mathbf{R}_x$ and Λ , respectively. In addition, in this case we also have $R'_{x,ii} = \lambda_{x,i}$. By using these results, (36) can be rewritten as follows:

$$P'_{ii} = \frac{\mu_n \sigma_v^2}{2} \frac{\lambda_{x,i}}{\lambda_{x,i} + \lambda_{\Lambda,i}}. \quad (38)$$

Therefore, the final expression for the MSD is:

$$\text{MSD} = \text{Tr}(\mathbf{P}_\infty) = \sum_{i=1}^L P'_{ii} = \frac{\mu_n \sigma_v^2}{2} \sum_{i=1}^L \frac{\lambda_{x,i}}{\lambda_{x,i} + \lambda_{\Lambda,i}}. \quad (39)$$

When the input signal is white ($\mathbf{R}_x = \sigma_x^2 \mathbf{I}$) and matrix Λ can be considered diagonal ($\Lambda = \lambda \mathbf{I}$), the MSD becomes:

$$\text{MSD} = \frac{\mu_n L \sigma_v^2 \sigma_x^2}{2(\sigma_x^2 + \lambda)}. \quad (40)$$

If $\lambda = 0$ (standard NLMS), for white input signals we have:

$$\text{MSD}_{NLMS} = \frac{\mu_n L \sigma_v^2}{2}.$$

If $\lambda > 0$, the MSD is reduced by the factor $\sigma_x^2/(\sigma_x^2 + \lambda)$. So larger λ provides lower MSD, while too large λ gets higher bias (see (27)).

4.2.2. Steady-state excess mean-square error (EMSE)

From the definition of EMSE and for the independence assumption, we have:

$$\text{EMSE} = \mathbb{E}[e^2(n)] = \mathbb{E}[(\tilde{\mathbf{w}}_n^T \mathbf{x}_n)^T \cdot (\tilde{\mathbf{w}}_n^T \mathbf{x}_n)] = \mathbb{E}[\mathbf{x}_n^T \tilde{\mathbf{w}}_n \tilde{\mathbf{w}}_n^T \mathbf{x}_n] = \text{Tr}(\mathbf{R}_x \mathbf{P}).$$

Therefore, by using the results found for the steady-state MSD, we have:

$$\text{EMSE} = \text{Tr}(\mathbf{R}_x \mathbf{P}) = \sum_{i=1}^L \lambda_{x,i} P'_{ii},$$

which using (38) leads to the final result:

$$\text{EMSE} = \frac{\mu_n \sigma_v^2}{2} \sum_{i=1}^L \frac{\lambda_{x,i}^2}{\lambda_{x,i} + \lambda_{\Lambda,i}}. \quad (41)$$

Again, in the case of white input and diagonal Λ , (41) simplifies as:

$$\text{EMSE} = \frac{\mu_n L \sigma_v^2 \sigma_x^4}{2(\sigma_x^2 + \lambda)}. \quad (42)$$

If $\lambda = 0$ (standard NLMS), for white input signals we have:

$$\text{EMSE}_{NLMS} = \frac{\mu_n L \sigma_v^2 \sigma_x^2}{2}.$$

If $\lambda > 0$, the EMSE is reduced again by the factor $\sigma_x^2/(\sigma_x^2 + \lambda)$. Thus, a larger value of λ reduces the EMSE, whereas excessively large values increase the estimation bias (see (27)).

The Λ matrix increases effective eigenvalues of $\mathbf{R}_{\text{eff}} \equiv \mathbf{B} = \mathbf{R}_x + \Lambda$. So the eigenvalue spread is reduced, the noise amplification is reduced, and the EMSE decreases.

4.2.3. Steady-state echo return loss enhancement (ERLE)

In this section, we derive an interesting relationship between the steady-state ERLE and the EMSE derived in previous section. Let us compute separately the terms in the numerator and denominator of the ERLE defined in (52). For the numerator, by using the independence assumption and (2), we have:

$$\begin{aligned} \mathbb{E}[d^2(n)] &= \mathbb{E}[(\mathbf{h}_0^T \mathbf{x}_n + v(n))(\mathbf{h}_0^T \mathbf{x}_n + v(n))] = \mathbb{E}[\mathbf{h}_0^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{h}_0] + \mathbb{E}[v^2(n)] \\ &= \mathbf{h}_0^T \mathbf{R}_x \mathbf{h}_0 + \sigma_v^2 \approx \mathbf{h}_0^T \mathbf{R}_x \mathbf{h}_0. \end{aligned}$$

Similarly, for the denominator, by using the independence assumption and the definition of the error signal $e(n) = d(n) - y(n) = \mathbf{h}_0^T \mathbf{x}_n + v(n) - \tilde{\mathbf{w}}_n^T \mathbf{x}_n = \tilde{\mathbf{w}}_n^T \mathbf{x}_n + v(n)$, we have:

$$\begin{aligned} \mathbb{E}[e^2(n)] &= \mathbb{E}[(\tilde{\mathbf{w}}_n^T \mathbf{x}_n + v(n))^T (\tilde{\mathbf{w}}_n^T \mathbf{x}_n + v(n))] = \mathbb{E}[\mathbf{x}_n^T \tilde{\mathbf{w}}_n \tilde{\mathbf{w}}_n^T \mathbf{x}_n] + \mathbb{E}[v^2(n)] \\ &= \text{Tr}(\mathbf{R}_x \mathbf{P}) + \sigma_v^2 = \text{EMSE} + \sigma_v^2. \end{aligned}$$

Therefore, we obtain the following equation for the steady-state ERLE:

$$\text{ERLE} = \frac{\mathbb{E}[d^2(n)]}{\mathbb{E}[e^2(n)]} = \frac{\mathbf{h}_0^T \mathbf{R}_x \mathbf{h}_0}{\text{EMSE} + \sigma_v^2}, \quad (43)$$

where EMSE is defined in (41). For white input ($\mathbf{R}_x = \sigma_x^2 \mathbf{I}$) and normalized RIR ($\|\mathbf{h}_0\|^2 = 1$), (43) can be rewritten as:

$$\text{ERLE} = \frac{\sigma_x^2}{\text{EMSE} + \sigma_v^2}, \quad (44)$$

where EMSE is defined in (42). Hence, for normalized RIRs, the ERLE tends to approach the SNR.

4.3. The case of biased solution

In case of the bias term $\mathbf{c} = \Lambda \mathbf{h}_0 - \mathbf{b} \neq 0$ in (23) does not disappear, the previous equations for MSD and EMSE are modified. To analyze the covariance, in this case it is convenient to define the centered weight error vector:

$$\mathbf{z}_n = \tilde{\mathbf{w}}_n - \tilde{\mathbf{w}}_\infty, \quad (45)$$

since now $\tilde{\mathbf{w}}_\infty \neq 0$ due to bias in estimating $\mathbf{h}_0$ (see (27)).

Let us define $\mathbf{Z}_n = \mathbb{E}[\mathbf{z}_n \mathbf{z}_n^T]$ the covariance matrix of the centered weight error vector $\mathbf{z}_n$. We can write, similarly to (29), the following recursion:

$$\mathbf{Z}_{n+1} = \mathbf{A} \mathbf{Z}_n \mathbf{A}^T + \mu_n^2 \mathbf{Q}, \quad (46)$$

where again $\mathbf{Q} = \sigma_v^2 \mathbf{R}_x$. Since from (45) we have $\tilde{\mathbf{w}}_n = \mathbf{z}_n + \tilde{\mathbf{w}}_\infty$, we can derive:

$$\mathbf{P}_n = \mathbb{E}[\tilde{\mathbf{w}}_n \tilde{\mathbf{w}}_n^T] = \mathbb{E}[(\mathbf{z}_n + \tilde{\mathbf{w}}_\infty)(\mathbf{z}_n + \tilde{\mathbf{w}}_\infty)^T] = \mathbf{Z}_n + \tilde{\mathbf{w}}_\infty \tilde{\mathbf{w}}_\infty^T,$$

since $\mathbb{E}[\mathbf{z}_n] = 0$. Therefore, we can derive:

$$\text{MSD} = \text{Tr}(\mathbf{P}_\infty) = \underbrace{\text{Tr}(\mathbf{Z}_\infty)}_{\text{variance}} + \underbrace{\|\tilde{\mathbf{w}}_\infty\|^2}_{\text{bias}^2} = \text{Tr}(\mathbf{Z}_\infty) + \|(\mathbf{R}_x + \Lambda)^{-1} \mathbf{c}\|^2. \quad (47)$$

This is the well-known bias-variance tradeoff. By considering now the per mode components and omitting the prime sign for brevity, (38) is modified as:

$$P_{ii} = P_{ii}^{\text{var}} + P_{ii}^{\text{bias}} = \frac{\mu_n \sigma_v^2}{2} \frac{\lambda_{x,i}}{\lambda_{x,i} + \lambda_{\Lambda,i}} + \left(\frac{c_i}{\lambda_{x,i} + \lambda_{\Lambda,i}} \right)^2, \quad (48)$$

where c_i is the i -th component of the vector $\mathbf{c}$. We can observe that the first term P_{ii}^{var} in (48) depends of the noise variance σ_v^2 , while the second term P_{ii}^{bias} is constant. Therefore, the per mode components behave as:

$$P_{ii} = C_{1,i} \sigma_v^2 + C_{2,i}. \quad (49)$$

Similarly, for the EMSE and considering (26), we derive:

$$\begin{aligned} \text{EMSE} &= \text{Tr}(\mathbf{R}_x \mathbf{P}_\infty) = \text{Tr}(\mathbf{R}_x (\mathbf{Z}_\infty + \tilde{\mathbf{w}}_\infty \tilde{\mathbf{w}}_\infty^\top)) = \text{Tr}(\mathbf{R}_x \mathbf{Z}_\infty) + \tilde{\mathbf{w}}_\infty^\top \mathbf{R}_x \tilde{\mathbf{w}}_\infty \\ &= \text{Tr}(\mathbf{R}_x \mathbf{Z}_\infty) + \mathbf{c}^\top (\mathbf{R}_x + \Lambda)^{-1} \mathbf{R}_x (\mathbf{R}_x + \Lambda)^{-1} \mathbf{c}. \end{aligned} \quad (50)$$

Hence, although bias affects only the mean of the estimate, it contributes additively to the MSD and EMSE.

In summary, when the regularization is not perfectly matched to the true system, the adaptive filter converges to a biased solution. This results in an additional deterministic component in the MSD and EMSE expressions, which dominates the performance at high SNR and may produce a performance saturation.

5. Results

This section evaluates the performance of the proposed Physics-Informed NLMS (PI-NLMS) in comparison with conventional adaptive filtering methods, in particular with the standard NLMS [4] and the proportionate NLMS (PNLMS) [15]. The comparison with NLMS and PNLMS is particularly relevant because the proposed PI-NLMS is derived as a regularized extension of the NLMS framework and retains the same first-order adaptation mechanism. Therefore, these algorithms constitute the most appropriate baselines for isolating and quantifying the benefits introduced by the proposed physics-informed priors.

The objectives are to assess convergence speed and steady-state performance evaluated in terms of misalignment or Mean Square Deviation (MSD), defined as:

$$\text{MSD}(n) = 10 \log_{10} \left(\mathbb{E} \left[\|\mathbf{h}_0 - \mathbf{w}_n\|^2 \right] \right) = 10 \log_{10} \left(\mathbb{E} \left[\|\tilde{\mathbf{w}}_n\|^2 \right] \right), \quad (51)$$

and Echo Return Loss Enhancement (ERLE):

$$\text{ERLE}(n) = 10 \log_{10} \frac{\mathbb{E}[d^2(n)]}{\mathbb{E}[e^2(n)]}. \quad (52)$$

The microphone signal is generated as in (2), where the true echo path $\mathbf{h}_0$ is both a real-world and synthetically generated to emulate a realistic room impulse response. Specifically, we use the Open AIR dataset¹ [54], from which we selected the St. George's Episcopal Church in Nashville, Tennessee, measured in medium position. This RIR has been originally recorded at 96,000 Hz sampling rate. We resampled it at $f_s = 8,000$ Hz and truncated it at $L = 1024$ samples (see Fig. 3a), while the synthetic RIR has been generated with the acousticRoomResponse tool of MATLAB at $f_s = 8,000$ Hz and with $L = 1024$ samples (see Fig. 3b). The impulse response is then normalized such that $\|\mathbf{h}_0\|^2 = 1$, ensuring that ERLE comparisons are not affected by arbitrary scaling of the echo path.

Simulations have been performed by using an input far-end signal $x(n)$ and an additive near-end white Gaussian noise $v(n)$ with zero mean and SNR varying in range $[-10, 30]$ dB with step of 5 dB. Two types of far-end signals $x(n)$ are considered: a white Gaussian noise with zero mean and unit variance with $N = 20,000$ samples and a female speech signal resampled at 8 kHz with a total of $N = 22,400$ samples.

For the hyper-parameters, we selected $\mu = 0.5$ and $\epsilon = 10^{-6}$ for both the environments. For the decay factors, accordingly to (6), we set $\alpha = 0.01$ for the synthetic RIR and $\alpha = 0.005$ for the Church RIR. For the other parameters, in the case of synthetic RIR, we set: $\lambda_{\text{decay}} = 10^{-3}$, $\lambda_{\ell_1} = \lambda_t = 10^{-6}$, $\lambda_{\text{ts}} = \lambda_{\text{ss}} = 10^{-5}$, and we have estimated $\tau_{\min} = 10$ samples. In the case of Church RIR, we set: $\lambda_{\text{decay}} = 10^{-4}$, $\lambda_{\ell_1} = \lambda_t = \lambda_{\text{ts}} = \lambda_{\text{ss}} = 10^{-6}$, and we have estimated $\tau_{\min} = 140$ samples. The values of these regularizing parameters have been selected through a grid-search procedure over predefined physically meaningful intervals, as shown in

Section 5.2. The step-size for the NLMS algorithm is set to $\mu_{\text{NLMS}} = 0.5$, while that for the PNLMS is set to $\mu_{\text{PNLMS}} = 0.4$, choosing also $\rho = 10^{-2}$ and $\delta = 10^{-6}$ (refer to [15]). The regularization parameter is, also for these algorithms, set to $\epsilon = 10^{-6}$.

5.1. Performance under white noise excitation

In the first set of experiments, we evaluate the MSD for both the considered RIRs and at different levels of SNR by using the white noise excitation. As an example, Fig. 4 shows the MSD curves under white noise excitation in both environments for the compared approaches in the case of SNR = 20 dB. The complete set of results are detailed in Table 2, which reports the steady-state performance in terms of MSD and ERLE for the considered algorithms under white Gaussian excitation and varying SNR conditions, for both synthetic and measured (Church) RIRs. Several important observations can be drawn.

First, all algorithms exhibit the expected monotonic improvement in both MSD and ERLE as the SNR increases. This behavior is consistent with the theoretical analysis, since the steady-state error is driven by the noise variance σ_v^2 , and therefore scales proportionally with the SNR. In particular, the approximately linear trend in dB scale confirms the validity of the derived EMSE expressions. Across all SNR values, the proposed PI-NLMS consistently outperforms NLMS and PNLMS in terms of MSD, with gains of approximately 2–3 dB for the synthetic RIR and slightly lower but still consistent improvements for the Church RIR. This confirms that the incorporation of physics-informed priors effectively reduces the steady-state coefficient error. The improvement is slightly larger for the synthetic RIR than for the Church RIR. This can be explained by the fact that the synthetic RIR shows a better alignment with imposed priors, while the Church RIR shows a more diffuse reverberation.

The case of SNR = 20 dB is particularly meaningful, as it corresponds to a realistic operating condition in acoustic echo cancellation, where background noise is present but not dominant. At this SNR level, as it can be seen from Table 2 and Fig. 4, the proposed PI-NLMS achieves, for the synthetic RIR an MSD of -27.43 dB against -24.80 dB for NLMS (about 2.6 dB gain) and, for the Church RIR, an MSD of -26.33 dB against -24.69 dB (about 1.6 dB gain).

While the MSD improvement is clearly visible, the ERLE gain remains modest. This is fully consistent with theory, since MSD depends on the entire coefficient error vector, including weakly excited modes, while ERLE depends on the projection of the error onto the input signal. The improvement in terms of ERLE are less evident for high SNRs. In fact, considering that ERLE depends on EMSE (see (43) and (44)), and this is dominated, at high SNR, by high-energy modes that have large $\lambda_{x,i}$ terms (see (41)), hence reducing the impact of the regularizing priors. In general, PI-NLMS mainly improves low-energy modes, hence improving MSD, by keeping ERLE quite constant. This highlights the complementary nature of MSD and ERLE as performance metrics and underscores the advantage of physics-informed regularization in controlling the full coefficient error rather than only its projection onto the input signal.

Accordance between the simulated MSD and EMSE with respect to the corresponding theoretical values in (39) and (41) is shown in Figs. 5a and 5b, respectively. Both figures highlight a quite perfect accordance between theory and simulated values, validating the theory under white noise excitation.

5.2. Regularizing parameter selection and sensitivity analysis

The regularization parameters of the proposed PI-NLMS algorithm were selected through a grid-search procedure over predefined physically meaningful intervals. In particular, the search ranges were chosen to cover weak, moderate, and strong regularization regimes, while avoiding values that either make the corresponding prior negligible or dominate the data-driven NLMS update. The

¹ <https://www.openairlib.net>

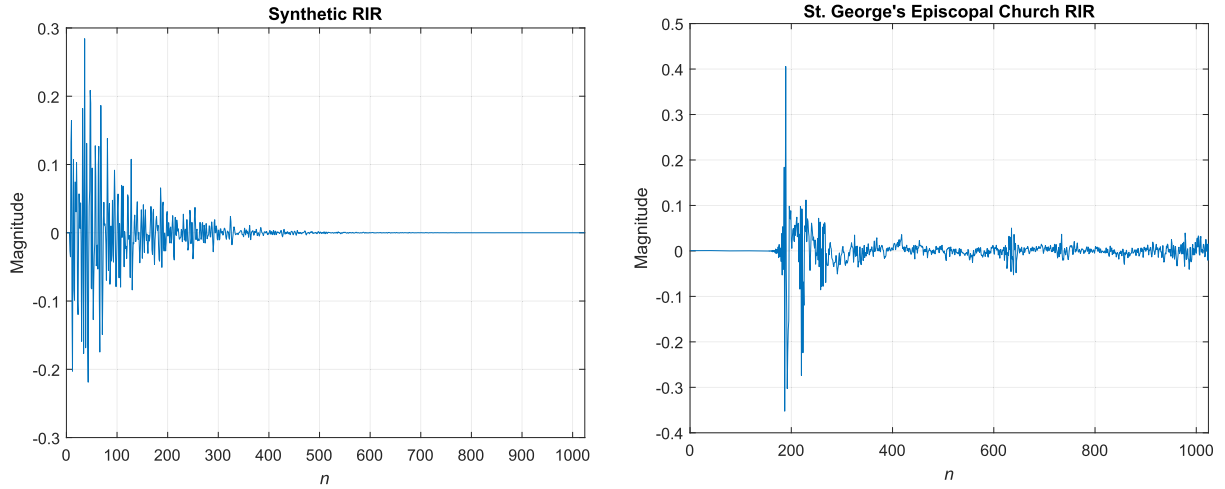


Fig. 3. The considered synthetic (left) and Church (right) RIRs.

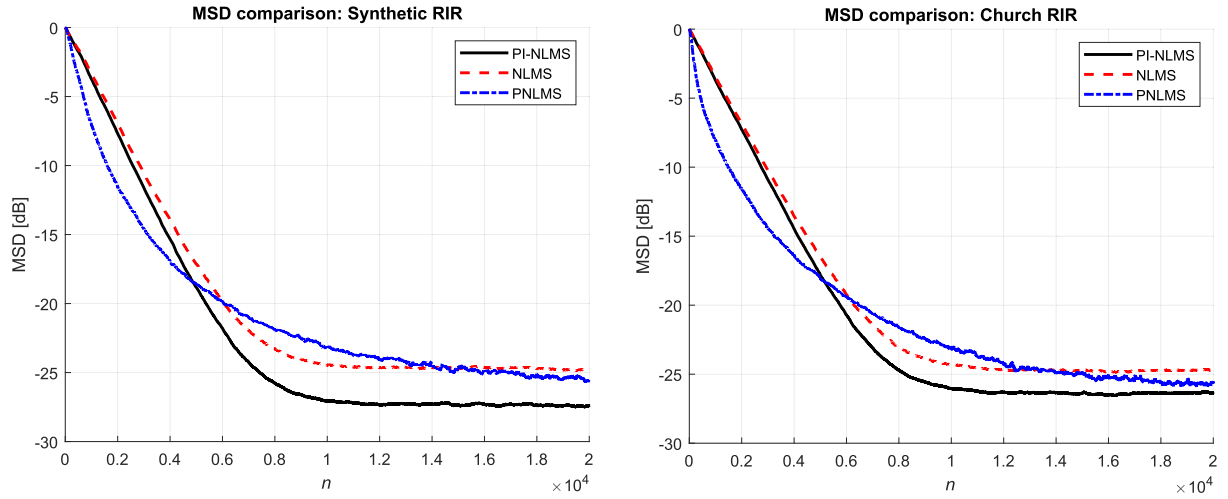


Fig. 4. MSD under white noise excitation for the synthetic RIR (left) and the Church RIR (right) with SNR = 20 dB.

Table 2

Main results in terms of MSD and ERLE at different SNRs for the synthetic and Church RIRs under white noise excitation.

		SNR	-10 dB	-5 dB	0 dB	5 dB	10 dB	15 dB	20 dB	25 dB	30 dB
Synthetic RIR											
PI-NLMS	MSD		1.57	-2.48	-7.44	-12.47	-17.45	-22.43	-27.43	-32.11	-36.30
	ERLE		-0.08	0.71	2.50	5.71	9.92	14.65	19.52	24.50	29.31
NLMS	MSD		5.26	0.24	-4.69	-9.71	-14.71	-19.78	-24.80	-29.78	-34.70
	ERLE		-0.49	0.30	2.08	5.29	9.49	14.26	19.14	24.15	29.09
PNLMS	MSD		1.81	-1.58	-5.86	-11.04	-16.08	-20.76	-25.45	-29.45	-32.37
	ERLE		-0.18	0.40	2.13	5.44	9.73	14.44	19.26	24.04	28.38
Church RIR											
PI-NLMS	MSD		1.65	-2.54	-6.62	-11.42	-16.61	-21.53	-26.33	-31.28	-35.51
	ERLE		-0.21	0.59	2.39	5.54	9.80	14.48	19.32	24.34	29.15
NLMS	MSD		5.38	0.22	-4.84	-9.65	-14.81	-19.87	-24.69	-29.78	-34.71
	ERLE		-0.52	0.28	2.09	5.23	9.49	14.20	19.04	24.09	29.02
PNLMS	MSD		1.95	-2.16	-6.38	-11.05	-16.00	-20.89	-25.69	-30.33	-34.00
	ERLE		-0.33	0.52	2.27	5.44	9.72	14.44	19.27	24.24	28.90

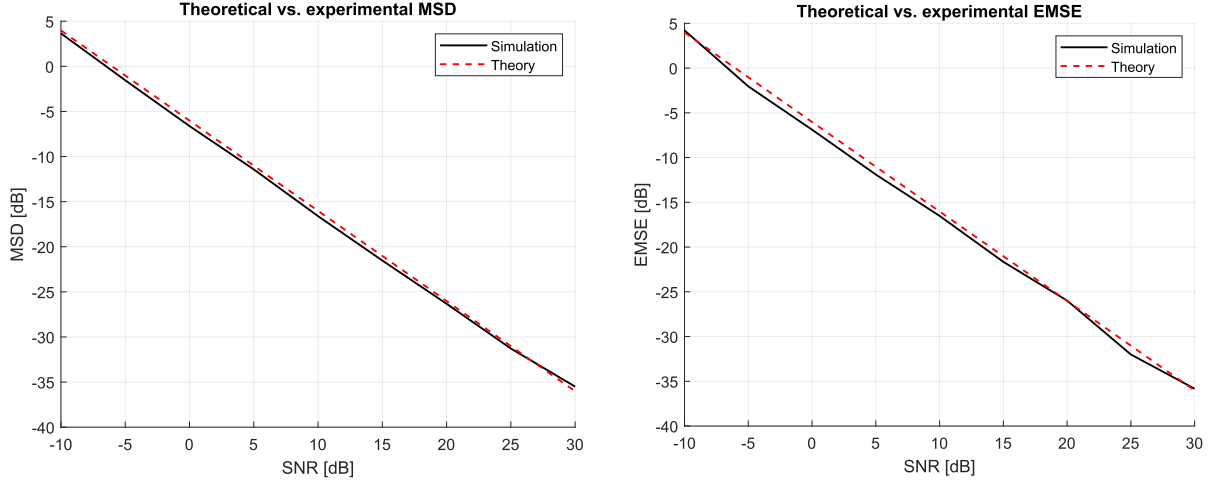


Fig. 5. Comparisons between simulated and theoretical MSD (left) and EMSE (right) at different levels of SNR for the Church RIR under white noise excitation.

same search intervals were used for all considered acoustic scenarios, and the final values were selected according to the best average MSD obtained on a validation configuration. Specifically, we used the grid $\{10^{-1}, 5 \times 10^{-2}, 10^{-2}, 5 \times 10^{-3}, 3 \times 10^{-3}, 5 \times 10^{-4}, 10^{-4}, 5 \times 10^{-5}, 10^{-5}\}$ for λ_{decay} and the grid $\{10^{-2}, 10^{-3}, 5 \times 10^{-4}, 10^{-4}, 5 \times 10^{-5}, 10^{-5}, 5 \times 10^{-6}, 10^{-6}, 5 \times 10^{-7}, 10^{-7}\}$ for all the other parameters (i.e., λ_{ℓ_1} , λ_{ts} , λ_{ss} , and λ_t).

To assess the robustness of the selected parameters, we performed a sensitivity analysis by varying pairs of regularization weights while keeping the remaining parameters fixed at their nominal values. Specifically, two heatmaps were generated and shown in Fig. 6 for the case of synthetic RIR at the SNR = 20 dB. The first reports the steady-state MSD as a function of the exponential decay weight λ_{decay} and the temporal-smoothness weight λ_{ts} . The second reports the MSD as a function of the sparsity weight λ_{ℓ_1} and λ_{ts} . These pairs were selected because they represent the most relevant trade-offs in the proposed framework: decay and temporal smoothness control the global structure and stability of the estimated RIR, while sparsity and temporal smoothness jointly affect early-reflection modeling and coefficient fluctuations.

The resulting heatmaps in Fig. 6 show that the proposed algorithm is not critically dependent on a single isolated parameter configuration. Instead, a broad region of low MSD is observed around the selected values, indicating that the proposed parameter setting is robust to moderate variations of the regularization weights. Very small values of the parameters reduce the effect of the physical priors and make the algorithm closer to standard NLMS, while excessively large values over-constrain the solution and increase the bias, leading to a degraded MSD. Therefore, the selected values provide a compromise between variance reduction and bias control. Similar results are obtained with the Church RIR, not shown here for space constraints.

5.3. Performance under speech signal excitation

In next experiment, we test the proposed approach under the speech signal excitation that introduces non-stationarity and colored input statistics. Specifically, Fig. 7 presents ERLE results using the female speech for both the synthetic and Church RIRs in the case of 20 dB of SNR. The complete set of results in terms of MSD and ERLE values for different SNRs and algorithms are detailed in Table 3. Specifically, Table 3 reports the steady-state performance under speech excitation, which represents a significantly more challenging and realistic scenario compared to white Gaussian input. The results clearly highlight the impact of input coloration on adaptive filtering performance and the benefits of the proposed PI-NLMS.

First, all algorithms exhibit a general improvement in both MSD and ERLE as the SNR increases. However, compared to the white input case, the overall performance is significantly degraded, particularly at low and moderate SNRs. This behavior is consistent with theoretical expectations, as speech signals exhibit a highly colored spectrum, leading to a large eigenvalue spread of the input covariance matrix $\mathbf{R}_x$. As discussed in the analysis, this results in poor excitation of several modes, which in turn increases the steady-state MSD. Across all SNR values, the proposed PI-NLMS consistently outperforms NLMS and PNLMS, with substantially larger gains than in the white input case. In particular, the MSD improvement is now much more pronounced, especially at low and moderate SNRs, while the ERLE gains are also significant, unlike the white input scenario. This confirms that the proposed physics-informed regularization is particularly effective when the input signal is poorly conditioned, as it compensates for the lack of excitation in weak modes.

The case of SNR = 20 dB is especially relevant, as it represents a realistic operating point for acoustic echo cancellation with speech signals. For the synthetic RIR, at 20 dB, the MSD is -13.94 dB, against -11.89 dB for NLMS and -12.82 dB for PNLMS (about 2 dB of gain with respect to NLMS). The ERLE is 15.36 dB, against 12.33 dB for NLMS and 11.46 dB for PNLMS (about 3 dB of gain with respect to PNLMS). Similar considerations can be applied to the Church RIR.

Let us observe that we got much larger ERLE gains than white input. This is justified by the fact that, under speech excitation, eigenvalues $\lambda_{x,i}$ are highly non-uniform, producing general lower results. However, the presence of the priors (refers to $\gamma_i = \lambda_{x,i} + \lambda_{\Lambda,i}$ in (39)) reduces the effective condition number $\kappa = \lambda_{\text{max}}/\lambda_{\text{min}}$ to $\kappa' = \gamma_{\text{max}}/\gamma_{\text{min}}$, improving the ERLE. Hence, PI-NLMS improves both MSD and ERLE under speech excitation, unlike the white signal case.

Accordance between the simulated MSD and EMSE with respect to the corresponding theoretical values in (39) and (41) is shown in Figs. 8a and 8b, respectively. In this case, the autocorrelation matrix $\mathbf{R}_x$ and squared norm of inputs in (16) are computed recursively on several batches of input. Both figures highlight a good accordance between theory and simulated values, for low and moderate SNR.

At high SNR values, the MSD deviates from the linear trend predicted by the noise-limited analysis and exhibits a saturation behavior, as predicted by (48) (see the dashed-dot blue line in Fig. 8). As described in Section 4.3, this effect is due to the presence of a bias term introduced by the regularization, which becomes dominant as the noise variance decreases. Specifically, MSD can be split into two terms (by summing over all modes in (49)): one proportional to the noise variance σ_v^2 and the other related to bias and independent of noise, in equation $\text{MSD} = C_1 \sigma_v^2 + C_2$. At low SNR, the first term dominates (linear behavior), while at high SNR the constant term dominates (tending to

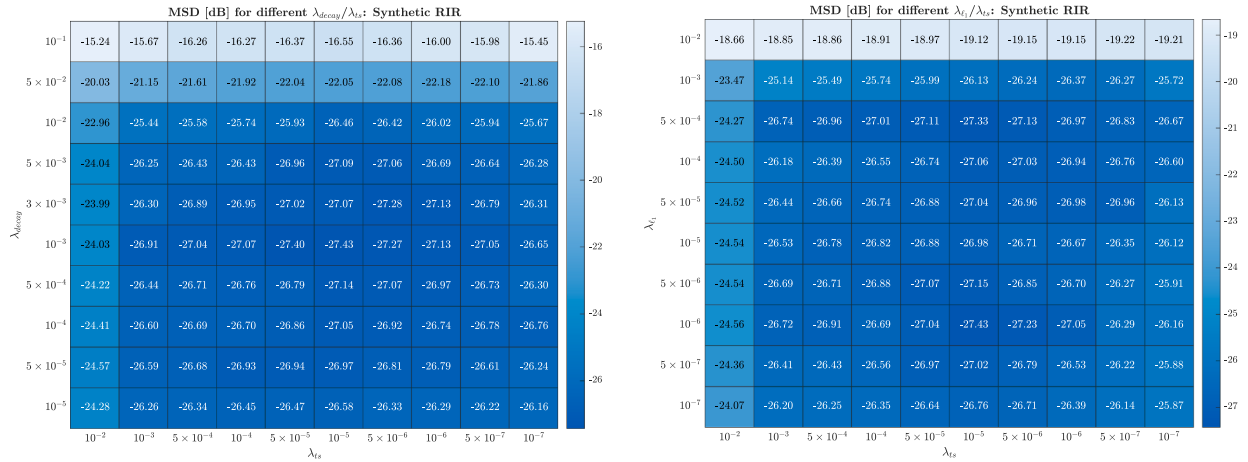


Fig. 6. MSD at different values of the λ_{decay} and λ_{ts} (left) and λ_e and λ_{ts} regularizing parameters (right) for the synthetic RIR at SNR = 20 dB.

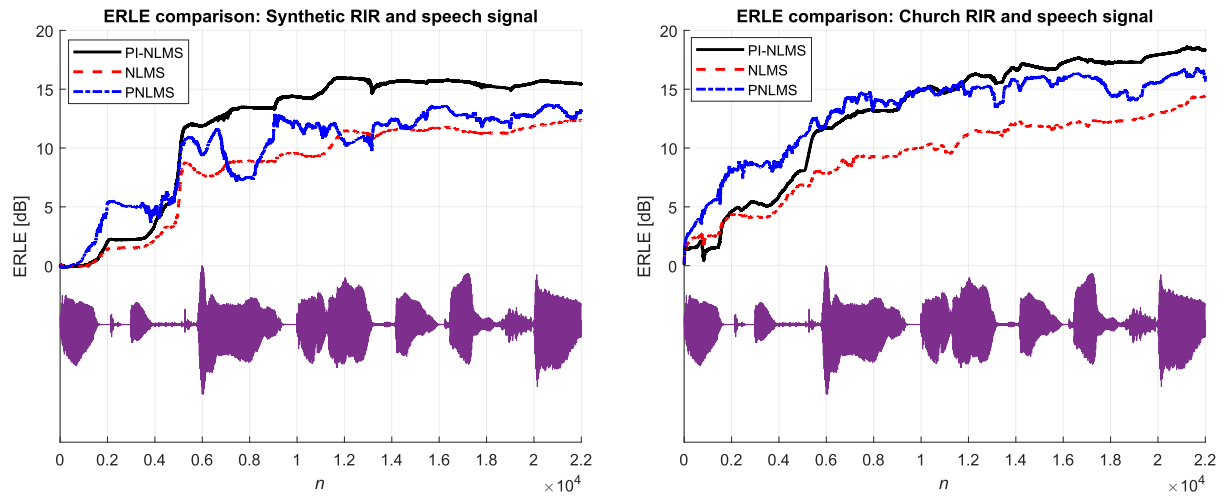


Fig. 7. ERLE under speech excitation for the synthetic (left) and Church (right) RIRs with SNR = 20 dB.

plateau). As a result, the algorithm operates in a bias-limited regime, where further increases in SNR do not translate into performance improvements. This behavior is consistent with the classical bias-variance trade-off and is particularly evident in the proposed PI-NLMS due to the presence of physics-informed priors. Compared to the white noise case, the plateau behavior of the MSD is more pronounced under speech excitation due to the highly colored nature of the input signal. In this case, the input covariance matrix exhibits a large eigenvalue spread, with many modes being weakly excited. As a result, the noise-driven component of the error vanishes rapidly with increasing SNR, while the residual bias, introduced by the regularization and the lack of excitation, remains dominant. Consequently, the algorithm enters a bias-limited regime at significantly lower SNR values compared to the white input case, leading to an earlier and more evident saturation of the MSD.

In summary, under speech excitation, the proposed PI-NLMS demonstrates significantly improved performance compared to NLMS and PNLMS, with gains that are substantially larger than those observed under white input conditions. This behavior is due to the highly colored nature of speech signals, which induces a large eigenvalue spread and leads to poor excitation of several modes. In this regime, the physics-informed regularization effectively compensates for the lack of excitation, stabilizes the adaptation, and reduces the steady-state error across all modes, despite the saturation performance at higher SNRs.

5.4. Comparison with APA and RLS

To further assess the performance of the proposed PI-NLMS algorithm, additional comparisons were conducted against the Affine Projection Algorithm (APA) and the Recursive Least Squares (RLS) algorithm [4,5]. These methods were selected because they represent two widely adopted adaptive filtering approaches characterized by faster convergence properties than NLMS-based techniques, particularly when the input signal exhibits strong temporal correlation, as is commonly the case for speech excitation.

It should be emphasized that APA and RLS are not direct counterparts of the proposed PI-NLMS. While PI-NLMS belongs to the class of low-complexity stochastic gradient algorithms, APA exploits multiple input vectors at each iteration and RLS performs a second-order adaptation based on the inverse input correlation matrix. Consequently, both methods generally achieve faster convergence at the expense of, especially the RLS one, a significantly higher computational complexity.

In experiments we set the projection order of APA to $K = 4$ and the step-size to $\mu_{\text{APA}} = 0.2$, while for RLS algorithm we set the forgetting factor to $\lambda_{\text{RLS}} = 0.999$ and the inverse of covariance matrix is initialized as $\mathbf{P} = \delta \mathbf{I}$ using $\delta = 1$. For the meaning of these parameters and details about APA and RLS, refer to [4,5].

Fig. 9a compares the proposed PI-NLMS with APA and RLS for the synthetic RIR at SNR = 20 dB. The MSD curves under white-noise excitation show that RLS provides the fastest initial convergence, as expected

Table 3

Main results in terms of MSD and ERLE at different SNRs for the synthetic and Church RIRs under speech signal excitation.

	SNR	-10 dB	-5 dB	0 dB	5 dB	10 dB	15 dB	20 dB	25 dB	30 dB
Synthetic RIR										
PI-NLMS	MSD	5.82	2.76	-2.13	-6.56	-10.16	-12.76	-13.94	-14.29	-14.47
	ERLE	-0.53	-0.13	0.96	3.41	6.90	11.23	15.36	18.61	20.57
NLMS	MSD	12.51	8.53	5.39	0.59	-4.24	-8.47	-11.89	-13.89	-14.08
	ERLE	-4.30	-4.09	-2.65	-0.84	3.39	7.33	12.33	16.68	19.82
PNLMS	MSD	6.65	4.16	0.05	-2.10	-6.07	-8.18	-12.82	-13.94	-14.38
	ERLE	-0.10	-0.08	-0.06	1.18	2.86	6.87	11.46	15.60	18.55
Church RIR										
PI-NLMS	MSD	10.60	5.66	-0.51	-4.41	-8.84	-12.79	-15.82	-17.18	-17.95
	ERLE	-2.28	-1.17	1.39	4.91	9.33	14.04	18.24	21.21	23.18
NLMS	MSD	15.65	10.65	5.56	0.66	-4.01	-8.06	-11.33	-12.90	-13.82
	ERLE	-3.98	-3.05	-0.54	3.37	7.74	11.43	14.58	15.55	16.29
PNLMS	MSD	10.95	5.99	-0.28	-1.67	-5.54	-8.55	-13.75	-14.84	-16.56
	ERLE	-2.11	-1.13	1.14	3.84	7.74	11.21	15.64	17.81	18.48

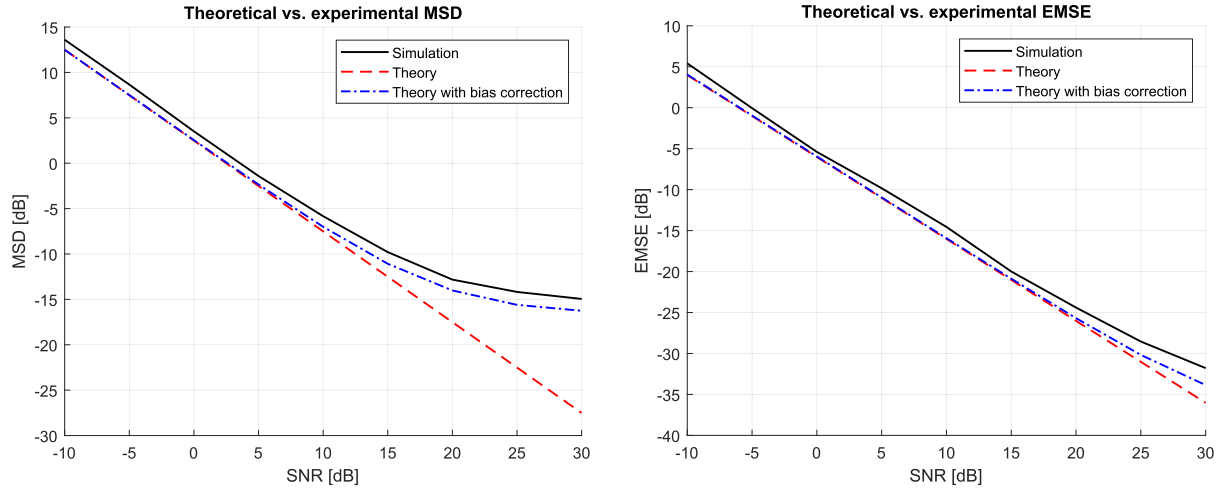


Fig. 8. Comparisons between simulated and theoretical MSD (left) and EMSE (right) at different level of SNR for the Church RIR under speech excitation.

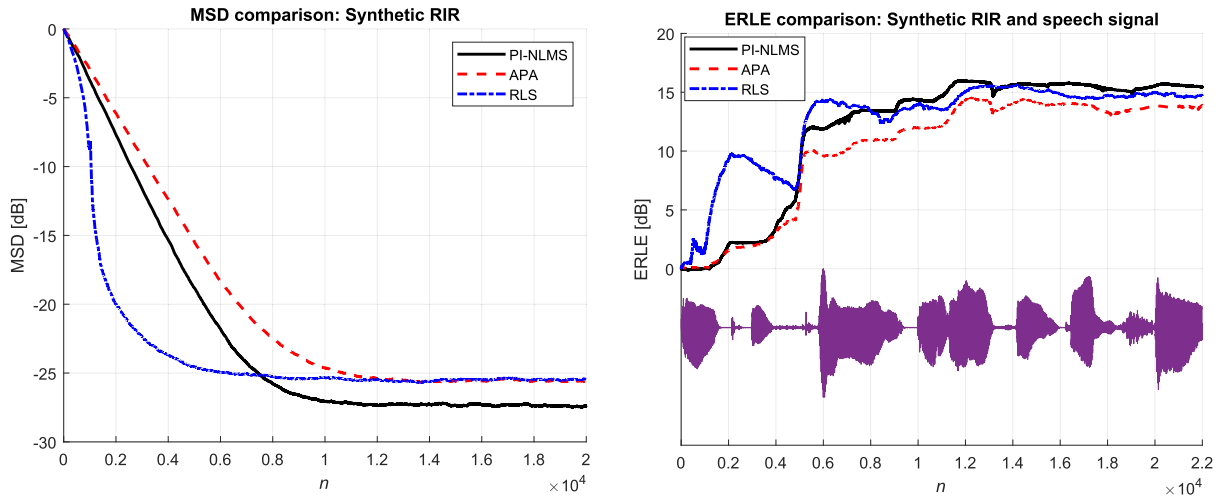


Fig. 9. Comparisons of the proposed PI-NLMS with respect to APA and RLS in the case of synthetic RIR with SNR = 20 dB: MSD under white noise excitation (left) and ERLE under speech excitation (right).

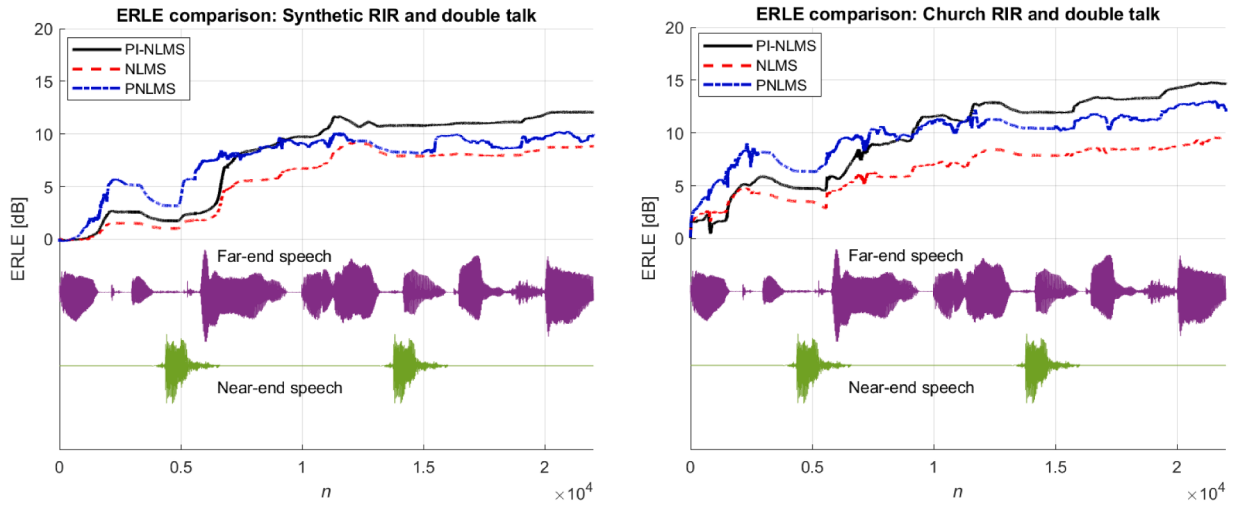


Fig. 10. ERLE under speech excitation for the synthetic (left) and Church (right) RIRs with SNR = 20 dB and double-talk conditions.

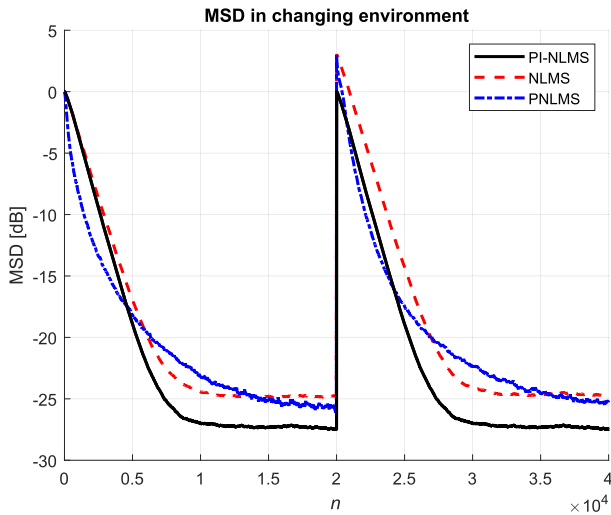


Fig. 11. MSD obtained in a changing environment, where at iteration $n = 20,000$ the RIR is abruptly changed.

from its second-order nature and use of the inverse input correlation matrix. APA also exhibits a smooth and rapid convergence behavior by exploiting multiple past input vectors. In contrast, the proposed PI-NLMS converges more gradually, reflecting its first-order adaptation mechanism. Nevertheless, after convergence, PI-NLMS achieves the lowest steady-state MSD, outperforming both APA and RLS by approximately 1.7 dB. This result suggests that the incorporation of physically motivated priors effectively reduces the variance of the estimated coefficients and guides the adaptation process toward physically plausible solutions. Consequently, the proposed regularization partially compensates for the lack of second-order adaptation while preserving a substantially lower computational complexity than RLS, whose complexity scales as $\mathcal{O}(L^2)$.

Fig. 9b reports the ERLE comparison under speech excitation for the same synthetic RIR and SNR condition. In this more realistic scenario, the colored nature of speech makes the identification problem more challenging and typically slows down the adaptation process. As expected, RLS exhibits the fastest initial ERLE growth, while PI-NLMS shows a slower transient due to its stochastic-gradient nature. However, the proposed method achieves the highest steady-state ERLE, providing approximately 1 dB improvement over both APA and RLS after convergence. These results indicate that the proposed physical priors improve

not only the accuracy of the identified echo path but also the effectiveness of the resulting echo cancellation process under correlated excitation.

Overall, the proposed PI-NLMS achieves a favorable performance-complexity trade-off, combining the low computational burden of NLMS-class algorithms with a steady-state performance that is competitive with, and in this scenario slightly superior to, more computationally demanding adaptive filtering approaches.

5.5. Performance under double-talk conditions

To evaluate the robustness of the proposed approach in realistic acoustic environments, additional experiments were conducted under double-talk conditions using both the synthetic and the measured Church RIRs. A female near-end speech signal was added during a predefined time interval while maintaining an overall SNR of 20 dB. To prevent coefficient divergence during double talk, a correlation-based double-talk detector (DTD) was employed [55,56]. When double talk was detected, the adaptation process was temporarily suspended, while the regularization terms remained active.

Double talk represents a particularly challenging condition for acoustic echo cancellation because the near-end speech component is uncorrelated with the far-end reference signal and may therefore corrupt the adaptation process. In conventional adaptive filters, this often results in coefficient drift, increased misalignment, and slower recovery after the end of the double-talk period.

The obtained results, shown in Fig. 10a and b, demonstrate that the proposed PI-NLMS maintains stable operation throughout the double-talk interval for both considered room impulse responses. As expected, the presence of near-end speech leads to a slight reduction in the achievable steady-state ERLE compared with the single-talk case. Nevertheless, the observed degradation is limited to approximately 2–3 dB, demonstrating that the combination of double-talk detection and physics-informed regularization effectively mitigates coefficient drift and preserves the quality of the echo-path estimate. Moreover, the ERLE temporarily decreases during the presence of near-end speech due to the additional signal component entering the microphone signal. However, the correlation-based DTD effectively limits the adaptation of the filter coefficients during double talk, thereby preventing significant degradation of the echo-path estimate.

In addition to the protection provided by the DTD, the proposed physics-informed regularization contributes to preserving a physically consistent impulse-response structure. Consequently, the decrease in ERLE during double talk remains limited and the recovery after the

disappearance of the near-end speech is faster than that observed for conventional adaptive filtering approaches, such as the NLMS. This behavior is particularly evident for the Church RIR, whose long reverberation tail makes the identification problem more sensitive to coefficient perturbations.

Overall, the results indicate that the combination of double-talk detection and physics-informed regularization provides improved robustness against near-end speech interference. The DTD prevents severe coefficient drift during double-talk intervals, while the proposed physical priors facilitate rapid re-convergence and maintain a physically plausible echo-path estimate, resulting in improved post-double-talk performance.

5.6. Performance under nonstationary scenario

Finally, the tracking capability of the considered algorithms is assessed in a nonstationary scenario. To this end, a sequence of 40,000 white noise samples is used as input, while an abrupt change in the echo path is introduced at sample $n = 20,000$. Specifically, the acoustic environment is modified by replacing the Church RIR with the synthetic RIR, thereby emulating a sudden variation in the acoustic conditions. The resulting MSD curves are reported in Fig. 11. As expected, all algorithms exhibit a sharp performance degradation at the change point due to the mismatch between the adaptive filter and the new echo path.

A closer inspection of the re-convergence behavior reveals clear quantitative differences. Immediately after the change, all methods experience a similar MSD increase, but the proposed PI-NLMS achieves a significantly faster decay. In particular, PI-NLMS reduces the MSD below -20 dB within approximately 5000 samples after the change, whereas NLMS and PNLMS require roughly 7000 samples to reach the same level. Moreover, the steady-state MSD attained by PI-NLMS after re-convergence is around -27.5 dB, which represents an improvement of approximately 2–3 dB over NLMS and PNLMS, respectively. These results indicate not only a faster tracking speed but also a lower misadjustment level.

This behavior can be attributed to the presence of physics-informed priors, which effectively regularize the adaptation and prevent large deviations in poorly excited modes, thereby accelerating convergence after abrupt changes. In contrast, NLMS exhibits slower adaptation due to its sensitivity to eigenvalue spread, while PNLMS benefits from proportionate updates but lacks the additional structural constraints provided by the proposed approach. Overall, the results confirm that PI-NLMS provides superior tracking performance, achieving both faster re-convergence and improved steady-state accuracy in dynamically varying environments.

6. Conclusion

This paper introduced a Physics-Informed Adaptive Filtering (PIAF) framework for acoustic echo cancellation. By reformulating echo path identification as a structured stochastic optimization problem, the proposed Physics-Informed NLMS (PI-NLMS) algorithm integrates physically motivated priors, such as causality, exponential energy decay, time-weighted sparsity of early reflections, spectral smoothness, and slow temporal variation, into the adaptive filtering process. Unlike conventional data-driven adaptive algorithms that operate over an unconstrained coefficient space, the proposed method restricts adaptation to a physically plausible echo-path manifold. Convergence analysis demonstrated that structured regularization modifies the effective correlation matrix, improving conditioning and reducing steady-state variance while introducing a controlled bias aligned with acoustic physics.

Simulation results under stationary and time-varying echo paths confirmed that the proposed approach achieves faster convergence, lower steady-state misalignment, improved ERLE, and enhanced robustness compared to conventional NLMS, proportionate algorithms, and standard sparsity-aware methods. Importantly, these gains are obtained with

modest additional computational complexity, preserving real-time feasibility for practical acoustic echo cancellation systems.

Future work may extend the framework to a proximal version to handle sparsity constraints without subgradients and extend it to frequency-domain adaptive filtering implementations for long impulse responses, incorporate nonlinear loudspeaker distortion modeling within a hybrid linear-nonlinear architecture, and investigate adaptive selection of physics regularization parameters based on environmental conditions. Moreover, theoretical extensions to nonstationary and double-talk scenarios, as well as experimental validation in real-world scenarios, constitute promising directions for further research.

CRedit authorship contribution statement

Michele Scarpiniti: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization; **Danilo Communiello:** Writing – review & editing, Investigation, Formal analysis, Data curation; **Aurelio Uncini:** Writing – review & editing, Validation, Supervision, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- [1] D. Yang, Echo cancellation algorithms for tele-communication system, in: International Conference on Electronics, Electrical, and Information Engineering (ICEEIE 2025), 14003, 2025, p. 140030J. <https://doi.org/10.1117/12.3090101>
- [2] A. Saremi, B. Ramkumar, G. Ghaffari, Z. Gu, An acoustic echo canceller optimized for hands-free speech telecommunication in large vehicle cabins, EURASIP J. Audio Speech Music Process. 2023 (1) (2023) 39. <https://doi.org/10.1186/s13636-023-00305-7>
- [3] J. Benesty, T. Gansler, D.R. Morgan, M.M. Sondhi, S.L. Gay, Advances in Network and Acoustic Echo Cancellation, Springer, 2001.
- [4] A. Uncini, Fundamentals of Adaptive Signal Processing, Springer, 2015.
- [5] A.H. Sayed, Adaptive Filters, Wiley, 2008.
- [6] A. Deb, A. Kar, M. Chandra, A technical review on adaptive algorithms for acoustic echo cancellation, in: 2014 International Conference on Communication and Signal Processing, 2014, pp. 041–045. <https://doi.org/10.1109/ICCSP.2014.6949795>
- [7] L. Zhang, H. He, J. Chen, Y. Yu, J. Benesty, A Newton-Raphson solution to the exponentially weighted least M-estimate formulation for acoustic system identification, Appl. Acoust. 231 (2025) 110460. <https://doi.org/10.1016/j.apacoust.2024.110460>
- [8] Z. Cao, T. Yan, J. Chen, J. Chen, J. Benesty, A MISO acoustic echo cancellation algorithm based on a two-layer filter decomposition, Signal Process. 232 (2025) 109892. <https://doi.org/10.1016/j.sigpro.2025.109892>
- [9] Z. Zheng, Z. Shao, Y. Yu, L. Lu, S. Gao, Cramér–Rao lower bound of adaptive filtering algorithms for acoustic echo cancellation, Signal Process. 238 (2026) 110111. <https://doi.org/10.1016/j.sigpro.2025.110111>
- [10] H. Kuttruff, Room Acoustics, Spon Press, 4th edition edition, 2000.
- [11] D.G. Janković, M. Ćirić, A. Pantić, Automated estimation of the truncation of room impulse response by applying a nonlinear decay model, J. Acoust. Soc. Am. 139 (3) (2016) 1047–1057. <https://doi.org/10.1121/1.4941657>
- [12] L. Wang, Y. Lu, Z. Gao, K. Li, J. Huang, Y. Kong, S. Okada, BERP: A blind estimator of room parameters for single-channel noisy speech signals, IEEE Trans. Audio Speech Lang. Process. 33 (2025) 2215–2231. <https://doi.org/10.1109/TASLP.2025.3574849>
- [13] R. Badeau, Unified stochastic reverberation modeling, in: 2018 26th European Signal Processing Conference (EUSIPCO), 2018, pp. 2175–2179. <https://doi.org/10.23919/EUSIPCO.2018.8553562>
- [14] I. Hassani, A. Benallal, R. Bendoumia, Double-talk robust fast converging and low complexity algorithm for acoustic echo cancellation in teleconferencing system, in: M. Hatti, (Eds.), Artificial Intelligence and Heuristics for Smart Energy Efficiency in Smart Cities, Springer International Publishing, 2022, pp. 409–420. https://doi.org/10.1007/978-3-030-92038-8_41
- [15] D. Duttweiler, Proportionate normalized least-mean-squares adaptation in echo cancellers, IEEE Trans. Speech Audio Process. 8 (5) (2000) 508–518. <https://doi.org/10.1109/89.861368>
- [16] C. Paleologu, J. Benesty, S. Ciochină, A robust proportionate NLMS algorithm, IEEE Signal Process. Lett. 17 (3) (2010) 237–240.

- [17] J. Benesty, S.L. Gay, An improved PNLMS algorithm, in: 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2, 2002, pp. II-1881–II-1884. <https://doi.org/10.1109/ICASSP.2002.5744994>
- [18] M. Hamidia, A. Amrouche, Improved variable step-size NLMS adaptive filtering algorithm for acoustic echo cancellation, Digit. Signal Process. 49 (2016) 44–55. <https://doi.org/10.1016/j.dsp.2015.10.015>
- [19] Y.A. Huang, J. Benesty, J. Chen, *Acoustic MIMO Signal Processing*, Springer, 2006.
- [20] Y. Chen, Y. Gu, A.O. Hero, Sparse LMS for system identification, in: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, pp. 3125–3128. <https://doi.org/10.1109/ICASSP.2009.4960286>
- [21] K. Kumar, R. Pandey, M. Karthik, S.S. Bhattacharjee, N.V. George, Robust and sparsity-aware adaptive filters: a review, Signal Process. 189 (2021) 108276. <https://doi.org/10.1016/j.sigpro.2021.108276>
- [22] Y.-Y. Chen, J.-H. Wang, P.-C. Chan, K.-W. Liang, Z.-Y. Wang, J.-C. Wang, Optimizing acoustic echo cancellation with variable step size in adaptive filtering, in: 2024 International Conference on Advanced Technologies for Communications (ATC), 2024, pp. 329–332. <https://doi.org/10.1109/ATC63255.2024.10908122>
- [23] L.-M. Dogariu, J. Benesty, C. Paleologu, S. Ciochină, Identification of room acoustic impulse responses via Kronecker product decompositions, IEEE/ACM Trans. Audio Speech Lang. Process. 30 (2022) 2828–2841. <https://doi.org/10.1109/TASLP.2022.3202128>
- [24] R.-A. Otopoleanu, L.-M. Dogariu, C. Paleologu, J. Benesty, C.-L. Stanciu, R.-L. Costea, Recursive least-squares algorithm based on a fourth-order tensor decomposition for acoustic echo cancellation, Mathematics 14 (5) (2026) 812. <https://doi.org/10.3390/math14050812>
- [25] G. Li, H. Zhang, S. Wang, G. Wang, J. Zhao, Generalized mixed-norm maximum correntropy for robust adaptive filtering, Appl. Acoust. 228 (2025) 110382. <https://doi.org/10.1016/j.apacoust.2024.110382>
- [26] G. Li, G. Wang, F. Huang, J. Zhao, H. Zhang, Novel robust filtering algorithms employing generalized correntropy within p-norm loss, Appl. Acoust. 243 (2026) 111112. <https://doi.org/10.1016/j.apacoust.2025.111112>
- [27] Y. Li, D. Sun, Y. Zhu, Y. Miao, C. Yu, Y. Fu, Fractional-order hyperbolic tangent adaptive algorithm for vehicular acoustic echo cancellation, Appl. Acoust. 242 (2026) 111069. <https://doi.org/10.1016/j.apacoust.2025.111069>
- [28] S. Wang, Z. Lai, Y. Lei, A sparse extended Kalman filtering based on laplace prior for real-time state-parameter identification, Mech. Syst. Signal Process. 241 (2025) 113500. <https://doi.org/10.1016/j.ymssp.2025.113500>
- [29] Q. Fu, L. Wang, Q. Xie, Y. Zhou, An improved adaptive iterative extended Kalman filter based on variational Bayesian, Appl. Sci. 14 (4) (2024) 1393. <https://doi.org/10.3390/app14041393>
- [30] J. Franzen, T. Fingscheidt, An efficient residual echo suppression for multi-channel acoustic echo cancellation based on the frequency-domain adaptive Kalman filter, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 226–230. <https://doi.org/10.1109/ICASSP.2018.8462488>
- [31] J.M. Gil-Cacho, T. van Waterschoot, M. Moonen, S.H. Jensen, A frequency-domain adaptive filter (FDAF) prediction error method (PEM) framework for double-talk-robust acoustic echo cancellation, IEEE/ACM Trans. Audio Speech Lang. Process. 22 (12) (2014) 2074–2086. <https://doi.org/10.1109/TASLP.2014.2351614>
- [32] Y.-S. Park, J.-H. Chang, Frequency domain acoustic echo suppression based on soft decision, IEEE Signal Process. Lett. 16 (1) (2009) 53–56. <https://doi.org/10.1109/LSP.2008.2008571>
- [33] T. Bahraini, A.N. Sadigh, Proposing a robust RLS based subband adaptive filtering for audio noise cancellation, Appl. Acoust. 216 (2024) 109755. <https://doi.org/10.1016/j.apacoust.2023.109755>
- [34] Z.M. Šarić, I.I. Papp, D.D. Kukolj, I. Velikić, G. Velikić, Partitioned block frequency domain acoustic echo canceller with fast multiple iterations, Digit. Signal Process. 27 (2014) 119–128. <https://doi.org/10.1016/j.dsp.2014.01.008>
- [35] G. Enzner, R. Martin, P. Vary, G. Enzner, R. Martin, P. Vary, Partitioned residual echo power estimation for frequency-domain acoustic echo cancellation and postfiltering, Eur. Trans. Telecommun. 13 (2) (2002) 103–114. <https://doi.org/10.1002/ett.4460130205>
- [36] H. Zhao, Y. Gao, Y. Zhu, Robust subband adaptive filter algorithms-based mixture correntropy and application to acoustic echo cancellation, IEEE/ACM Trans. Audio Speech Lang. Process. 31 (2023) 1223–1233. <https://doi.org/10.1109/TASLP.2023.3250845>
- [37] T. Haubner, A. Brendel, W. Kellermann, End-to-end deep learning-based adaptation control for linear acoustic echo cancellation, IEEE/ACM Trans. Audio Speech Lang. Process. 32 (2024) 227–238. <https://doi.org/10.1109/TASLP.2023.3325923>
- [38] M. K. S. I. V., Attention-driven echo cancellation: a novel transformer-based approach for robust acoustic echo and noise cancellation, Results Eng. 26 (2025) 105039. <https://doi.org/10.1016/j.rineng.2025.105039>
- [39] Y. Huang, W. Qin, Z. Li, Q. Zhang, Time-frequency dual-domain attention for acoustic echo cancellation, J. Supercomput. 81 (5) (2025) 739. <https://doi.org/10.1007/s11227-025-07200-2>
- [40] Z. Chen, X. Xia, C. Chen, X. Wang, Y. Leng, L. Chen, R. Togneri, Y. Xiao, P. Ding, S. Song, P. Zhang, A two-stage progressive neural network for acoustic echo cancellation, in: Proc. Interspeech 2023, 2023, pp. 795–799. <https://doi.org/10.21437/Interspeech.2023-521>
- [41] S. Xu, C. He, B. Yan, M. Wang, A multi-stage acoustic echo cancellation model based on adaptive filters and deep neural networks, Electronics 12 (15) (2023) 3258. <https://doi.org/10.3390/electronics12153258>
- [42] C. Zhang, X. Zhang, A robust and cascaded acoustic echo cancellation based on deep learning, in: Proc. Interspeech 2020, 2020, pp. 3940–3944. <https://doi.org/10.21437/Interspeech.2020-1260>
- [43] E. Seidel, G. Enzner, P. Mowlaee, T. Fingscheidt, Neural Kalman filters for acoustic echo cancellation: comparison of deep neural network-based extensions, IEEE Signal Process. Mag. 41 (6) (2024) 24–38. <https://doi.org/10.1109/MSP.2024.3449557>
- [44] V.R. Tadi, A. Kar, Fully convolutional recurrent network with multiple sub-filters for monophonic acoustic echo cancellation, Digit. Signal Process. 176 (2026) 106068. <https://doi.org/10.1016/j.dsp.2026.106068>
- [45] M. K. S. I. V., A novel adaptive Kolmogorov Arnold sparse masked attention model with multi-loss optimization for acoustic echo cancellation in double-talk noisy scenario, Comput. Speech Lang. 93 (2025) 101786. <https://doi.org/10.1016/j.csl.2025.101786>
- [46] T. Sun, X. Rong, D. Wang, Y. Hu, K. Chen, J. Lu, Exploiting lightweight neural post-filtering for directional speech enhancement, Appl. Acoust. 239 (2025) 110844. <https://doi.org/10.1016/j.apacoust.2025.110844>
- [47] A. Ivry, I. Cohen, B. Berdugo, Deep adaptation control for acoustic echo cancellation, in: 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2022), 2022, pp. 741–745. <https://doi.org/10.1109/ICASSP43922.2022.9746557>
- [48] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019) 686–707. <https://doi.org/10.1016/j.jcp.2018.10.045>
- [49] G.E. Karniadakis, I.G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning, Nat. Rev. Phys. 3 (6) (2021) 422–440. <https://doi.org/10.1038/s42254-021-00314-5>
- [50] T. van Waterschoot, G. Rombouts, M. Moonen, Towards optimal regularization by incorporating prior knowledge in an acoustic echo canceller, in: International Workshop on Acoustic Signal Enhancement (IWAENC 2005), 2005, pp. 157–160.
- [51] R.-A. Otopoleanu, J. Benesty, C. Paleologu, C.-L. Stanciu, L.-M. Dogariu, S. Ciochină, A practical regularized recursive least-squares algorithm for robust system identification, in: 2025 33rd European Signal Processing Conference (EUSIPCO), 2025, pp. 1417–1421. <https://doi.org/10.23919/EUSIPCO63237.2025.11226097>
- [52] C.-L. Stanciu, C. Anghel, L.-M. Dogariu, I.-D. Ficiu, L. Stanciu, Regularized affine projection algorithm for stereophonic acoustic echo cancellation, in: 2024 Advanced Topics on Measurement and Simulation (ATOMS), 2024, pp. 319–322. <https://doi.org/10.1109/ATOMS60779.2024.10921557>
- [53] N. Bershad, P. Celka, J.-M. Vesin, Stochastic analysis of gradient adaptive identification of nonlinear systems with memory for Gaussian data and noisy input and output measurements, IEEE Trans. Signal Process. 47 (3) (1999) 675–689. <https://doi.org/10.1109/78.747775>
- [54] S. Shelley, D.T. Murphy, OpenAIR: an interactive auralization web resource and database, in: AES 129th Convention, San Francisco, CA, USA, 2010, p. 8226.
- [55] J. Benesty, D. Morgan, J. Cho, A new class of doubletalk detectors based on cross-correlation, IEEE Trans. Speech Audio Process. 8 (2) (2000) 168–172. <https://doi.org/10.1109/89.824701>
- [56] T. Gänslar, J. Benesty, The fast normalized cross-correlation double-talk detector, Signal Process. 86 (6) (2006) 1124–1139. <https://doi.org/10.1016/j.sigpro.2005.07.035>
- [57] M. Von Berg, P.P. Schwörer, L. Prinz, J. Steffens, Analysis of physical and perceptual properties of room impulse responses: development of an online tool, in: Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum 2023, 2023, pp. 1647–1654. <https://doi.org/10.61782/fa.2023.0480>
- [58] A. Fan, W. Yang, Z. Wang, Y. Xu, Deep echo cancellation algorithm based on time-frequency domain combination, in: 2024 4th International Conference on Neural Networks, Information and Communication Engineering (NNICE), 2024, pp. 997–1001. <https://doi.org/10.1109/NNICE61279.2024.10498402>
- [59] H. Chen, G. Chen, K. Chen, J. Lu, Nonlinear residual echo suppression based on dual-stream DPRNN, EURASIP J. Audio Speech Music Process. 2021 (1) (2021) 35. <https://doi.org/10.1186/s13636-021-00221-8>
- [60] A. Favrot, C. Faller, F. Kuech, Modeling late reverberation in acoustic echo suppression, in: International Workshop on Acoustic Signal Enhancement (IWAENC 2012), 2012, pp. 1–4.