Journal of Signal Processing Systems (2022) 94:147–160
https://doi.org/10.1007/s11265-021-01711-w

Online Acoustic System Identification Exploiting Kalman Filtering and an Adaptive Impulse Response Subspace Model

Thomas Haubner[1] · **Andreas Brendel**[1] · **Walter Kellermann**[1]

Received: 27 February 2021 / Revised: 30 September 2021 / Accepted: 2 October 2021 / Published online: 24 February 2022

Abstract
We introduce a novel algorithm for online estimation of Acoustic Impulse Responses (AIRs) which allows for fast convergence by exploiting prior knowledge about the fundamental structure of AIRs. The proposed method assumes that the variability of AIRs of an acoustic scene is confined to a low-dimensional manifold which is embedded in a high-dimensional space of possible AIR estimates. We discuss various approaches which exploit a training data set of AIRs, e.g., high-accuracy AIR estimates from the acoustic scene, to learn a local affine subspace approximation of the AIR manifold. The model is motivated by the idea of describing the generally nonlinear AIR manifold locally by tangential hyperplanes and its validity is verified for simulated data. Subsequently, we describe how the manifold assumption can be used to enhance online AIR estimates by projecting them onto an adaptively estimated subspace. Motivated by the assumption of manifolds being locally Euclidean, the parameters determining the adaptive subspace are learned from the nearest neighbor AIR training samples to the current AIR estimate. To assess the proximity of training data AIRs to the current AIR estimate, we introduce a probabilistic extension of the Euclidean distance which improves the performance for applications with non-white excitation signals. Furthermore, we describe how model imperfections can be tackled by a soft projection of the AIR estimates. The proposed algorithm exhibits significantly faster convergence properties in comparison to a high-performance state-of-the-art algorithm. Furthermore, we show an improved steady-state performance for speech-excited system identification scenarios suffering from high-level interfering noise and nonunique solutions.

Keywords Online Acoustic System Identification · Kalman Filter · Subspace Model · Adaptation Control

1 Introduction

The continuously increasing amount of acoustic communication devices has fueled the research on reliable speech enhancement algorithms. In this context system identification has proven to be a vital part of many state-of-the-art approaches [1, 2]. In particular online algorithms are required to cope with the large variety of acoustic environments devices are exposed to. However, even after decades of research [3–6], Online Supervised Acoustic System Identification (OSASI)-based speech enhancement algorithms are significantly challenged by interfering noise signals and their limited convergence rate. In this paper we propose a method which tackles remaining limitations of modern OSASI algorithms.

Both convergence speed and noise-robustness are usually addressed by adaptive step size-controlled Adaptive Filter (AF) algorithms [2]. Their performance decisively depends on the stochastic properties of the excitation and the noise signals [7]. In particular, for stationary white excitation signals a fast convergence speed and robust steady-state performance is achieved. This observation has led to a variety of excitation signal-dependent adaptive step size selection schemes with the most famous one being the power-normalization of the time-domain Normalized Least-Mean-Squares (NLMS) algorithm [7]. Its scalar time-domain step size has been extended to a frequency-dependent step size to cope

✉ Thomas Haubner
thomas.haubner@fau.de

Andreas Brendel
andreas.brendel@fau.de

Walter Kellermann
walter.kellermann@fau.de

[1] Multimedia Communications and Signal Processing, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Cauerstr. 7, Erlangen 91058, Germany

with the temporal correlation of many excitation signals, i.e., speech or music [8]. In particular the frequency-dependent power normalization in block processing approaches led to computationally efficient and faster converging algorithms [4, 9]. The robustness against nonstationary interfering noise sources was initially addressed by binary adaptation control, i.e., stalling the filter adaptation whenever the noise power exceeds a predefined threshold [10]. The scalar and binary decision was later extended to a frequency-dependent continuous step size control [11]. In particular the probabilistic inference of the step size by a Kalman Filter (KF) has shown great potential [12–14]. Yet, the KF performance decisively depends on an accurate estimate of the noise Power Spectral Density (PSD) [15]. Here, significant performance improvements relative to classical PSD estimators have been achieved by modern machine-learning based approaches [16, 17]. Despite the improved noise robustness, these approaches still achieve only slow convergence speed for scenarios suffering from permanently low Signal-to-Noise-Ratio (SNR), e.g., as in driving cars with open windows.

Recently, besides adaptation control, the exploitation of prior knowledge about the structure of AIRs has been successfully used to deal with slow convergence and non-robust steady-state performance [18–21]. These algorithms rely on the assumption that not all possible AIR estimates, i.e., Finite Impulse Response (FIR) filters of fixed length, are equally likely, i.e., certain regions exhibit a higher probability of representing a valid AIR. In [18] this assumption has been used by regularizing a least-squares system identification cost function with the Mahalanobis distance based on an estimated AIR covariance matrix. The extreme case that the AIRs of a considered acoustic scene can all be represented by a structured subset of the high-dimensional estimation space, i.e., FIR filters of fixed length, motivates the assumption of a low-dimensional AIR manifold [22]. Its existence is often tightly coupled to the parameter changes of an underlying physical process, e.g., location of sources and sensors or temperature changes, which govern the variability of the AIRs [23]. Noisy AIR estimates can be enhanced by projection onto the manifold, i.e., by removing the part which is not confined to the manifold. Yet due to the complex interaction of the physical parameters and the high-dimensional AIRs, an analytic manifold description is difficult to obtain. However, in many applications a device is exposed over long time periods to the same time-varying acoustic scene, e.g., a microphone array in a car cabin or a TV in a living room. This allows to collect high-accuracy AIR estimates of the OSASI algorithm during operation of the device. These estimates can serve as training data to optimize a data-driven AIR manifold model which can then be exploited to improve the performance of the OSASI algorithm in acoustically adverse conditions. Various approaches have been proposed to model an AIR manifold with the most prominent one being a global affine subspace whose parameters are estimated by Principal Component Analysis (PCA) [24]. Yet, due to the inherent trade-off between generalization to unknown AIRs and denoising of AIR estimates when choosing the subspace dimension, the global PCA approach is limited to simplistic scenarios (cf. Subsec. 3.4). This trade-off can be circumvented by locally approximating the manifold by affine subspaces that can be interpreted as tangential hyperplanes. It allows to maintain the denoising property of a low-dimensional approximation while still being capable to represent a large class of realistic acoustic scenes (cf. Subsec. 3.4). Local affine subspace models have been exploited for sound field reconstruction by sparse dictionary learning [25] and also local PCA-supported system identification algorithms [19, 21]. Besides the affine subspace-based approaches, a globally nonlinear manifold model has been proposed in [20] and was used in an offline least-squares system identification task.

In this paper, we introduce a novel algorithm which exploits an adaptive AIR subspace model for enhancing KF-based system identification algorithms. For this we discuss various data-driven local AIR manifold approximations by affine subspaces and compare their respective approximation errors for simulated data. Subsequently, we suggest to exploit the manifold assumption for improving KF-based OSASI by projecting the current AIR estimate onto an adaptively learned subspace. In contrast to state-of-the-art approaches the subspace parameters are inferred online from the K-Nearest Neighbor (KNN) AIR training samples to the current AIR estimate which allows for an improved modeling of realistic acoustic scenes. To assess this proximity we propose a novel probabilistically motivated distance measure which takes into account the convergence state of the adaptive filter. Furthermore, we modify the idea of a 'hard' projection of the noisy AIR estimate onto the AIR manifold to a soft projection which allows the algorithm to cope with model imperfections. It is shown that the proposed method improves the convergence speed of KF-based system identification algorithms and achieves higher steady-state performance in scenarios suffering from high-level interfering noise. In addition, we show improved AF performance for system identification scenarios that are challenged by nonunique Minimum Mean Square Error (MMSE) solutions [26, 27]. This problem is often faced in rendering and teleconferencing applications for which the excitation signals are composed of less sources than loudspeakers.

In this paper, vectors are typeset as bold lowercase letters and matrices as bold uppercase letters with underlined symbols representing time-domain quantities. The all-zero matrix of dimension $D_1 \times D_2$ is denoted by $\mathbf{0}_{D_1 \times D_2}$, the $D \times D$-dimensional identity matrix by $\mathbf{I}_D$ and the Discrete Fourier Transform (DFT) matrix by $\mathbf{F}_D$, respectively.

Transposition and Hermitian transposition are represented by $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$, respectively, and $\otimes$ denotes the Kronecker product. Furthermore, a matrix element in the mth row and the nth column is indicated by $[\cdot]_{mn}$ and the diag$(\cdot)$ operator creates a diagonal matrix from its vector-valued argument. Finally, the proper complex Gaussian Probability Density Function (PDF) with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Psi}$ is denoted by $\mathcal{N}_c(\cdot|\boldsymbol{\mu}, \boldsymbol{\Psi})$ and the equivalency of two terms up to a constant is denoted by $\overset{\mathrm{c}}{=}$.

The remainder of this paper is structured as follows: In Sec. 2, a probabilistic signal observation model is introduced which relates the noisy observations to the unknown AIRs. Subsequently, in Sec. 3, various affine subspace approaches to locally model an AIR manifold are described and evaluated for simulated data. The fusion of the affine subspace models with a KF-based OSASI algorithm is introduced in Sec. 4. Experimental results for the proposed algorithm are shown in Sec. 5. Finally, the paper is concluded in Sec. 6.

2 Probabilistic Signal Model

We first introduce a probabilistic signal model for describing the microphone observations of a Multiple-Input Single-Output (MISO) system identification scenario with B loudspeakers and one microphone as depicted in Fig. 1. The multipath propagation from the loudspeakers to the microphone is modeled by FIR filters $\underline{\mathbf{w}}_{b,\tau} \in \mathbb{R}^L$ of length L with loudspeaker index $b = 1, \ldots, B$ which are shortly denoted as AIRs in the sequel. Note that due to computational limitations the modeled FIR filter length L can often not be chosen sufficiently large to properly describe the room acoustics. To account for this undermodeling, we assume that the modeled AIRs $\underline{\mathbf{w}}_{b,\tau}$ represent the first L taps of respective ground-truth AIRs $\underline{\mathbf{h}}_{b,\tau} \in \mathbb{R}^W$ of length $W \geq L$, i.e., $\underline{\mathbf{w}}_{b,\tau} = \mathbf{Q}_3^{\mathrm{T}} \underline{\mathbf{h}}_{b,\tau} \in \mathbb{R}^L$ with the selection matrix $\mathbf{Q}_3^{\mathrm{T}} = \left[\mathbf{I}_L \ \mathbf{0}_{L\times W-L}\right]$.

A block of microphone observations at block time index τ

$$\underline{\mathbf{y}}_\tau = \underline{\mathbf{d}}_\tau + \underline{\mathbf{n}}_\tau \in \mathbb{R}^R \tag{1}$$

is described as a linear superposition of the noise-free observation vector $\underline{\mathbf{d}}_\tau$ and the noise signal vector $\underline{\mathbf{n}}_\tau$. Each signal block, i.e., $\underline{\mathbf{y}}_\tau$, $\underline{\mathbf{d}}_\tau$, and $\underline{\mathbf{n}}_\tau$, contains R consecutive samples

$$\underline{\mathbf{y}}_\tau = \left[\underline{y}_{(\tau-1)R+1} \ \underline{y}_{(\tau-1)R+2} \ \cdots \ \underline{y}_{\tau R}\right]^{\mathrm{T}} \in \mathbb{R}^R, \tag{2}$$

with R being the frame shift. The noise-free observation block $\underline{\mathbf{d}}_\tau$ is described as the sum of B linear convolution products between the modeled AIRs $\underline{\mathbf{w}}_{b,\tau} \in \mathbb{R}^L$ and the respective loudspeaker signal blocks $\underline{\mathbf{x}}_{b,\tau}$ (cf. Fig. 1). This model can be expressed efficiently by overlap-save processing in the DFT domain [7]. For this, the most recent

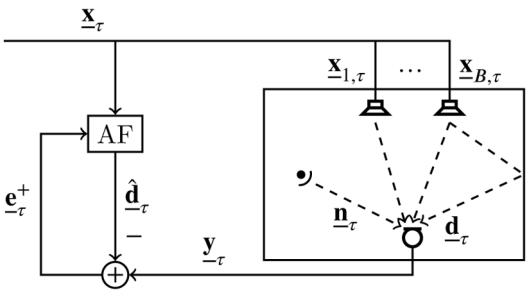

Figure 1 Identification of an acoustic MISO system.

$M = R + L$ samples of each loudspeaker are transformed to the DFT domain

$$\mathbf{X}_{b,\tau} = \mathrm{diag}\left(\mathbf{F}_M \left[\underline{x}_{b,\tau R-M+1} \ \cdots \ \underline{x}_{b,\tau R}\right]^{\mathrm{T}}\right), \tag{3}$$

and subsequently multiplied with the respective Acoustic Transfer Function (ATF) vector

$$\mathbf{w}_{b,\tau} = \mathbf{F}_M \mathbf{Q}_2 \underline{\mathbf{w}}_{b,\tau} \in \mathbb{C}^M, \tag{4}$$

where $\mathbf{Q}_2^{\mathrm{T}} = \left[\mathbf{I}_L \ \mathbf{0}_{L\times R}\right]$ is a zero-padding matrix. Afterwards, the B DFT-domain products are added up and the inverse DFT of the sum is multiplied with the constraint matrix $\mathbf{Q}_1^{\mathrm{T}} = \left[\mathbf{0}_{R\times L} \ \mathbf{I}_R\right]$, as follows:

$$\underline{\mathbf{d}}_\tau = \mathbf{Q}_1^{\mathrm{T}} \mathbf{F}_M^{-1} \sum_{b=1}^{B} \mathbf{X}_{b,\tau} \mathbf{w}_{b,\tau}. \tag{5}$$

Note that the linear convolution constraint matrix $\mathbf{Q}_1^{\mathrm{T}}$ discards the first L samples of the inverse DFT to avoid circular convolution effects [p. 351] [7]. By inserting the linear convolution model (5) into the additive signal model (1) and premultiplying with $\mathbf{F}_M \mathbf{Q}_1$, we obtain the linear frequency-domain observation model

$$\mathbf{y}_\tau = \mathbf{F}_M \mathbf{Q}_1 \underline{\mathbf{y}}_\tau = \mathbf{C}_\tau \mathbf{w}_\tau + \mathbf{n}_\tau \in \mathbb{C}^M. \tag{6}$$

Here, we used the overlap-save constrained loudspeaker signal matrix

$$\mathbf{C}_\tau = \left[\mathbf{C}_{1,\tau} \ \cdots \ \mathbf{C}_{B,\tau}\right] \in \mathbb{C}^{M\times BM}, \tag{7}$$

with $\mathbf{C}_{b,\tau} = \mathbf{F}_M \mathbf{Q}_1 \mathbf{Q}_1^{\mathrm{T}} \mathbf{F}_M^{-1} \mathbf{X}_{b,\tau} \in \mathbb{C}^{M\times M}$ and the MISO ATF vector

$$\mathbf{w}_\tau = \left[\mathbf{w}_{1,\tau}^{\mathrm{T}} \ \cdots \ \mathbf{w}_{B,\tau}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{MB}. \tag{8}$$

Note that the corresponding time-domain MISO AIR vector is obtained by

$$\underline{\mathbf{w}}_\tau = \left[\underline{\mathbf{w}}_{1,\tau}^{\mathrm{T}} \ \cdots \ \underline{\mathbf{w}}_{B,\tau}^{\mathrm{T}}\right]^{\mathrm{T}} \tag{9}$$

$$= \left(\mathbf{I}_B \otimes \left(\mathbf{Q}_2^{\mathrm{T}} \mathbf{F}_M^{-1}\right)\right) \mathbf{w}_\tau \in \mathbb{R}^Q, \tag{10}$$

with $Q = LB$. Finally, the frequency-domain interfering noise block $\mathbf{n}_\tau = \mathbf{F}_M \mathbf{Q}_1 \underline{\mathbf{n}}_\tau$ is modeled as a proper complex zero-mean Gaussian random vector [28]

$$\mathbf{n}_\tau \sim \mathcal{N}_c(\mathbf{n}_\tau | \mathbf{0}_{M\times 1}, \mathbf{\Psi}_\tau^{\mathrm{N}}), \tag{11}$$

with the diagonal covariance matrix $\mathbf{\Psi}_\tau^{\mathrm{N}} \in \mathbb{C}^{M\times M}$.

3 Analysis of Acoustic Impulse Responses

In this section, we discuss various approaches to model the neighbourhood of the unknown AIR vector $\underline{\mathbf{w}}_\tau \in \mathbb{R}^Q$. We start by introducing the first-order Markov model assumption which is commonly used in KF-based system identification algorithms and discuss its limitations. Subsequently, we describe how these limitations can be mitigated by modeling an AIR manifold. Finally, we discuss various affine subspace-based approaches to locally approximate the manifold. Note that a straightforward extension of the subsequently described MISO AIR models to Multiple-Input Multiple-Output (MIMO) systems is obtained by stacking the respective MISO AIR vectors $\underline{\mathbf{w}}_\tau$ to an extended vector [21].

3.1 Acoustic Impulse Response Manifold

In [12, 28] it is suggested to describe the temporal variation of the MISO AIR vector $\underline{\mathbf{w}}_\tau$ (cf. Eqs. (9) and (10)) by a DFT-domain random walk Markov model

$$\begin{aligned} \mathbf{w}_\tau &= A\, \mathbf{w}_{\tau-1} + \Delta\mathbf{w}_\tau, \\ \Delta\mathbf{w}_\tau &\sim \mathcal{N}_c(\Delta\mathbf{w}_\tau | \mathbf{0}_{M\times 1}, \mathbf{\Psi}_\tau^{\Delta\mathrm{W}}) \end{aligned} \tag{12}$$

with the state transition coefficient A and the block-diagonal process noise covariance matrix

$$\mathbf{\Psi}_\tau^{\Delta\mathrm{W}} = \begin{bmatrix} \mathbf{\Psi}_{11,\tau}^{\Delta\mathrm{W}} & \cdots & \mathbf{0}_{M\times M} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{M\times M} & \cdots & \mathbf{\Psi}_{BB,\tau}^{\Delta W} \end{bmatrix} \in \mathbb{C}^{BM\times BM}. \tag{13}$$

As depicted in Fig. 2, the random walk model enforces a continuous temporal variation of subsequent AIR vectors $\underline{\mathbf{w}}_{\tau-1}$ and $\underline{\mathbf{w}}_\tau$, depicted as blue dots, of the OSASI application. However, as no additional knowledge about the filter coefficients is assumed, the process noise power is distributed into all directions of the high-dimensional FIR filter vector space $\mathbb{R}^Q$, as shown by the shaded gray area in Fig. 2. In contrast, the AIR vectors in the vicinity of $\underline{\mathbf{w}}_{\tau-1} \in \mathbb{R}^Q$ often populate only a subset of this space [22]. This is visualized in Fig. 2 by showing exemplary samples of the surrounding AIR vectors as black dots. On a global view this motivates the assumption that all AIR vectors are confined to a structured subset of the vector space $\mathbb{R}^Q$ which is termed AIR manifold. As manifolds are locally Euclidean [29], each neighbourhood of an AIR vector can be described by an affine subspace $\mathcal{M}_i$ of the vector space $\mathbb{R}^Q$. This is visualized in Fig. 3 where each shaded grid cell illustrates a different affine subspace $\mathcal{M}_i$. We now introduce a mathematical description of a single affine subspace which will serve as a basis for the following approaches to describe the manifold globally by patches of affine subspaces.

3.2 Affine Subspace Model

An affine subspace $\mathcal{M}_i$ of dimension D_i is defined by:

$$\mathcal{M}_i := \{\underline{\mathbf{w}}_\tau^{\mathrm{p}} \in \mathbb{R}^Q | \underline{\mathbf{w}}_\tau^{\mathrm{p}} = \bar{\underline{\mathbf{w}}}_i + \underline{\mathbf{V}}_i \underline{\boldsymbol{\beta}}_\tau,\ \underline{\boldsymbol{\beta}}_\tau \in \mathbb{R}^{D_i}\}. \tag{14}$$

It is parametrized by its offset $\bar{\underline{\mathbf{w}}}_i \in \mathbb{R}^Q$ and its basis matrix $\underline{\mathbf{V}}_i \in \mathbb{R}^{Q\times D_i}$. The vector $\underline{\boldsymbol{\beta}}_\tau$ represents the coordinates of the affine subspace element $\underline{\mathbf{w}}_\tau^{\mathrm{p}}$ in the basis spanned by the columns of $\underline{\mathbf{V}}_i$.

An orthogonal projection of an arbitrary AIR vector $\underline{\mathbf{w}}_\tau \in \mathbb{R}^Q$ onto the affine subspace $\mathcal{M}_i$ is given by the mapping [30]

$$\underline{\mathbf{w}}_\tau^{\mathrm{p}} = \left[\left(\underline{\mathbf{w}}_{1,\tau}^{\mathrm{p}}\right)^{\mathrm{T}} \ \cdots \ \left(\underline{\mathbf{w}}_{B,\tau}^{\mathrm{p}}\right)^{\mathrm{T}} \right]^{\mathrm{T}} \tag{15}$$

$$= f_{\mathcal{M}_i}(\underline{\mathbf{w}}_\tau) = \bar{\underline{\mathbf{w}}}_i + \underline{\mathbf{L}}_i \left(\underline{\mathbf{w}}_\tau - \bar{\underline{\mathbf{w}}}_i\right) \in \mathbb{R}^Q \tag{16}$$

with the rank-D_i projection matrix

$$\underline{\mathbf{L}}_i = \underline{\mathbf{V}}_i \left(\underline{\mathbf{V}}_i^{\mathrm{T}} \underline{\mathbf{V}}_i\right)^{-1} \underline{\mathbf{V}}_i^{\mathrm{T}} \in \mathbb{R}^{Q\times Q}. \tag{17}$$

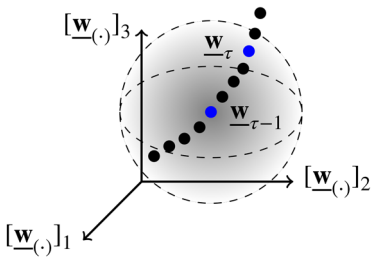

Figure 2 Exemplary AIR vectors of length $Q = 3$ of an acoustic scene with blue dots representing subsequent AIR vectors of the OSASI application. The shaded gray ball, delimited with dashed contour lines, depicts an exemplary process noise covariance matrix around $\underline{\mathbf{w}}_{\tau-1}$.

Here, it is assumed that the columns in $\underline{\mathbf{V}}_i$ are linearly independent. Figure 3 shows the projection of the AIR vector $\underline{\mathbf{w}}_\tau$ onto an exemplary AIR manifold which is locally represented by the affine subspace $\mathcal{M}_i$.

3.3 Affine Subspace Parameter Estimation

We now discuss how to estimate the parameters of a single affine subspace $\mathcal{M}_i$, i.e., its offset $\bar{\underline{\mathbf{w}}}_i$ and basis matrix $\underline{\mathbf{V}}_i$, from a training data set $\mathcal{U}$ including K AIR vectors $\underline{\mathbf{w}}_\kappa^{\mathrm{tr}}$ with training sample index $\kappa = 1, \ldots, K$. The superscript $(\cdot)^{\mathrm{tr}}$ is chosen to label the respective AIR vectors $\underline{\mathbf{w}}_\kappa^{\mathrm{tr}}$ as training data samples. As the affine subspace model $\mathcal{M}_i$ should only represent a local approximation of the manifold (cf. Fig. 3), its parameters are also only learned from a subset $\mathcal{U}_i$ of the full training data set $\mathcal{U}$, i.e., $\mathcal{U}_i \subseteq \mathcal{U}$. The indicator variable

$$z_{\kappa i} := \begin{cases} 1 \text{ if } \underline{\mathbf{w}}_\kappa^{\mathrm{tr}} \in \mathcal{U}_i, \\ 0 \text{ if } \underline{\mathbf{w}}_\kappa^{\mathrm{tr}} \notin \mathcal{U}_i \end{cases} \tag{18}$$

describes the assignment of the training samples to the respective local training data set $\mathcal{U}_i$ with cardinality $K_i = \sum_{\kappa=1}^{K} z_{\kappa i}$. The choice of $z_{\kappa i}$ will be discussed in detail in Subsecs. 3.4 and 4.2. The offset of the affine subspace $\mathcal{M}_i$ is estimated as arithmetic average

$$\bar{\underline{\mathbf{w}}}_i = \frac{1}{K_i} \sum_{\kappa=1}^{K} z_{\kappa i} \underline{\mathbf{w}}_\kappa^{\mathrm{tr}} \tag{19}$$

of the local training data set $\mathcal{U}_i$. For computing the basis matrix $\underline{\mathbf{V}}_i$ we first estimate the local AIR covariance matrix

$$\underline{\mathbf{R}}_i = \frac{1}{K_i - 1} \sum_{\kappa=1}^{K} z_{\kappa i} (\underline{\mathbf{w}}_\kappa^{\mathrm{tr}} - \bar{\underline{\mathbf{w}}}_i)(\underline{\mathbf{w}}_\kappa^{\mathrm{tr}} - \bar{\underline{\mathbf{w}}}_i)^{\mathrm{T}}. \tag{20}$$

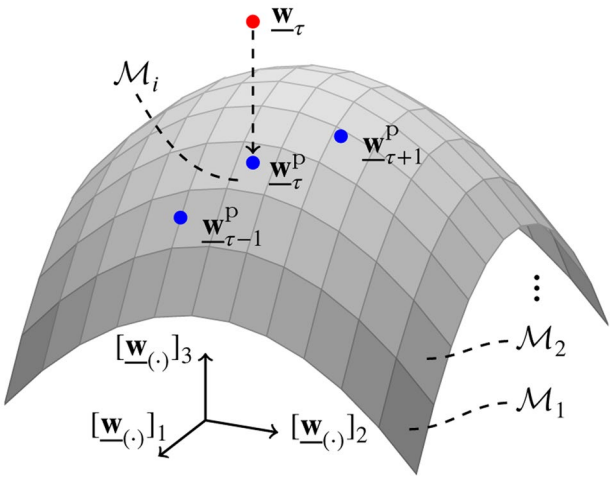

Figure 3 Projection of an AIR vector $\underline{\mathbf{w}}_\tau$ of length $Q = 3$ onto an exemplary AIR manifold. Each shaded grid cell represents an affine subspace $\mathcal{M}_i$ of dimension $D_i = 2$ which is locally tangential to the manifold.

Subsequently, the basis matrix $\underline{\mathbf{V}}_i$ is determined by the eigenvectors corresponding to the D_i largest eigenvalues, i.e., the principal components of the covariance matrix $\underline{\mathbf{R}}_i$ [24]. Note that, due to the broadband nature of the AIR vector (10), the covariance matrix $\underline{\mathbf{R}}_i$ describes the correlation between different AIRs, i.e., $\underline{\mathbf{w}}_{1,\tau}, \ldots, \underline{\mathbf{w}}_{B,\tau}$ as well as the correlation of different taps of one AIR $\underline{\mathbf{w}}_{b,\tau}$. We now discuss how the number of affine subspaces I and the selection of the indicator variables $z_{\kappa i}$ results in different approaches to approximate the AIR manifold.

3.4 Local Training Data Estimation

In [21] it is proposed to cluster the training data into I disjoint local data sets $\mathcal{U}_i$ by the k-means algorithm [31, 32]. Subsequently, each local training data set $\mathcal{U}_i$ is used to compute a local affine subspace $\mathcal{M}_i$ by the PCA approach described in Sec. 3.3. A special case of this local PCA model is the classical global PCA which is obtained by setting the number of local training data sets I and the respective indicator variables $z_{\kappa i}$ with $\kappa = 1, \ldots, K$ to one. We now evaluate the validity of this model for a typical MISO acoustic rendering scenario. The simulation parameters are summarized in Sec. 5 and the models are learned from $K = 5000$ training data samples. As evaluation measure we use the logarithmic average system mismatch

$$\Upsilon_\tau = 10 \log_{10} \left(\frac{1}{B} \sum_{b=1}^{B} \frac{||\underline{\mathbf{h}}_{b,\tau} - \mathbf{Q}_3 \underline{\mathbf{w}}_{b,\tau}^{\mathrm{p}}||^2}{||\underline{\mathbf{h}}_{b,\tau}||^2} \right) \tag{21}$$

between the ground-truth AIRs $\underline{\mathbf{h}}_{b,\tau} \in \mathbb{R}^W$ and the respective affine subspace projections $\underline{\mathbf{w}}_{b,\tau}^{\mathrm{p}}$ (cf. Eqs. (9), (15) and (16)) of the optimum L-tap AIR approximations $\underline{\mathbf{w}}_{b,\tau} = \mathbf{Q}_3^{\mathrm{T}} \underline{\mathbf{h}}_{b,\tau}$ (cf. Sec. 2). Note that the zero-padding matrix $\mathbf{Q}_3$ in Eq. (21) accounts for the severe undermodeling, i.e., $L \ll W$, that is common to many acoustic system identification applications (cf. Sec. 2).

Fig. 4 shows the average system mismatch $\bar{\Upsilon}$ which results from the projection of 500 ground-truth MISO AIR vectors $\underline{\mathbf{h}}_\tau^{\mathrm{T}} = [\underline{\mathbf{h}}_{1,\tau}^{\mathrm{T}}, \ldots, \underline{\mathbf{h}}_{B,\tau}^{\mathrm{T}}]$ onto a global affine subspace model (*Global Proj.* ($I = 1$)) and a mixture of subspaces model (*Mixture Proj.* ($I = 40$)) with $I = 40$ clusters in dependence of the subspace dimension D_i. Note that for the mixture model the affine subspace with the lowest system mismatch Υ_τ is selected. In addition, the average system mismatch without subspace projection, i.e., $\underline{\mathbf{w}}_\tau^p = \underline{\mathbf{w}}_\tau$, is plotted as benchmark (*Oracle GT*). We conclude from Fig. 4 that for the considered scenario the global affine subspace assumption holds only coarsely. Due to the high variability of the AIRs representing the acoustic scene, a large subspace dimension is required to attain a reasonable average system mismatch $\bar{\Upsilon}$. This limits the denoising capability of

the respective projection (cf. Subsec. 4.3). In contrast, the mixture approach attains for small subspace dimensions D_i already a much lower average system mismatch $\bar{\Upsilon}$. However, as the affine subspaces $\mathcal{M}_i$ are only representative for samples close to the offset vector $\underline{\bar{\mathbf{w}}}_i$, its modeling capability depends decisively on the number of clusters I. This limitation can be mitigated by using more clusters. In the extreme case of using the same number of clusters as training data samples, i.e., $I = K$, each test sample is approximated by the best fitting Nearest Neighbor (NN) training sample. Motivated by the property of manifolds being locally Euclidean, the squared Euclidean distance

$$d_{\text{euc}}(\underline{\mathbf{w}}_\tau, \underline{\mathbf{w}}_\kappa^{\text{tr}}) = ||\underline{\mathbf{w}}_\tau - \underline{\mathbf{w}}_\kappa^{\text{tr}}||^2 \tag{22}$$

between the optimum Q-dimensional MISO AIR vector $\underline{\mathbf{w}}_\tau = (\mathbf{I}_B \otimes \mathbf{Q}_3^{\text{T}})\underline{\mathbf{h}}_\tau$ and the training data samples $\underline{\mathbf{w}}_\kappa^{\text{tr}}$ is used to select the NN. The respective NN approximation attains an average system mismatch of $\bar{\Upsilon} = -6.7$ dB (cf. *KNN Proj.* at subspace dimension $D_i = 0$ in Fig. 4). We suspect that the NN model generalizes poorly to AIRs in between the training samples as the respective subspace is zero-dimensional, i.e., condensed to a single element. To remedy this limitation we suggest to remove the condition of non-overlapping local training data sets $\mathcal{U}_i$ as used in [21]. In particular we propose to estimate the affine subspace parameters from the K_τ NN training samples $\underline{\mathbf{w}}_\kappa^{\text{tr}}$ to the optimum AIR vector $\underline{\mathbf{w}}_\tau$. Note that we index the subspace-related parameters by τ in the following to stress the dependence on the ground-truth AIR vector. The corresponding projection suggests a reconstruction of $\underline{\mathbf{w}}_\tau$ from its surrounding K_τ training samples. Note that if the subspace dimension is chosen as $D_\tau = K_\tau - 1$, with K_τ being the number of neighbors, the projection can be computed directly from the training samples (*KNN Proj.*) as shown in the sequel. Note that one degree of freedom is required for computing the local offset vector $\underline{\bar{\mathbf{w}}}_\tau$ (cf. Eq. (17)). The projection matrix $\underline{\mathbf{L}}_\tau$ is obtained without eigenvalue decomposition of the local covariance matrix by choosing

$$\underline{\mathbf{V}}_\tau = \left[\underline{\mathbf{w}}_1^{\text{tr}} - \underline{\bar{\mathbf{w}}}_\tau \;\cdots\; \underline{\mathbf{w}}_{K_\tau - 1}^{\text{tr}} - \underline{\bar{\mathbf{w}}}_\tau\right] \tag{23}$$

in Eq. (17). The corresponding *KNN Proj.* algorithm is evaluated in Fig. 4. The respective K_τ training AIRs are computed based on the squared Euclidean distance (22) w.r.t. the optimum AIR vector $\underline{\mathbf{w}}_\tau$. We observe that, for an equivalent subspace dimension $D_i = D_\tau$, the proposed method achieves a much lower average system mismatch $\bar{\Upsilon}$ in comparison to both the global and the mixture approach. Furthermore, due to removing the condition of disjoint local training data sets, the KNN-based projection achieves the benchmark performance (cf. *Oracle GT*) at a much lower subspace dimension in comparison to the mixture approach (cf. *Mixture Proj.*).

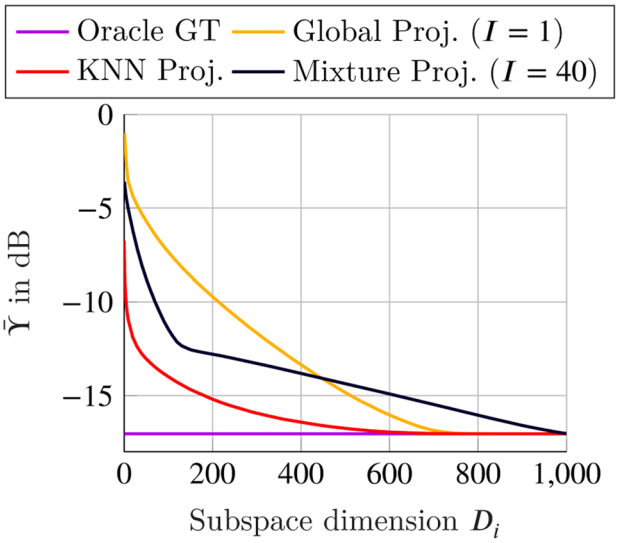

Figure 4 Average system mismatch $\bar{\Upsilon}$ of projecting 500 ground-truth MISO AIR vectors $\underline{\mathbf{h}}_\tau \in \mathbb{R}^{BW}$ onto various affine subspace models ($B = 2$, $W = 4096$, $L = 512$). The subspace dimension $D_i = 0$ corresponds to using only the offset vector of the model as projected AIR vector.

This property is beneficial as the orthogonal space to the affine subspace is treated as noisy part of an AF estimate in the following (cf. Sec. 4.3).

4 Acoustic Impulse Response Denoising

In this section, we introduce the proposed OSASI algorithm which fuses a state-of-the-art KF adaptation with an adaptive AIR subspace model. The subspace parameters are estimated from the NN training samples (cf. Sec. 3.4). For computing the closest neighbors we propose a novel distance which takes the state uncertainty of the KF into account. Finally, we describe a probabilistically motivated frequency-dependent convex combination of the KF estimate and its projection onto the affine subspace which improves the performance of the baseline KF.

4.1 Kalman Filter-based Acoustic Impulse Response Estimation

The DFT-domain KF [12, 28] approach to OSASI suggests a probabilistic inference of the latent ATF vector $\mathbf{w}_\tau$. For this, the conditional PDF of $\mathbf{w}_\tau$, given the current and the preceding observations $\mathbf{Y}_{1:\tau} = [\mathbf{y}_1, \ldots, \mathbf{y}_\tau]$, is modeled by [28]

$$p(\mathbf{w}_\tau|\mathbf{Y}_{1:\tau}) = \mathcal{N}_c(\mathbf{w}_\tau|\hat{\mathbf{w}}_\tau^{\text{kf}}, \mathbf{P}_\tau) \tag{24}$$

with the ATF mean vector

$$\hat{\mathbf{w}}_\tau^{\text{kf}} = \left[\left(\hat{\mathbf{w}}_{1,\tau}^{\text{kf}}\right)^{\text{T}} \;\cdots\; \left(\hat{\mathbf{w}}_{B,\tau}^{\text{kf}}\right)^{\text{T}}\right]^{\text{T}} \in \mathbb{C}^{MB} \tag{25}$$

and the state uncertainty matrix

$$\mathbf{P}_\tau = \begin{bmatrix} \mathbf{P}_{11,\tau} & \cdots & \mathbf{P}_{1B,\tau} \\ \vdots & \ddots & \vdots \\ \mathbf{P}_{B1,\tau} & \cdots & \mathbf{P}_{BB,\tau} \end{bmatrix} \in \mathbb{C}^{BM \times BM}. \tag{26}$$

Due to the linear Gaussian DFT-domain state transition model (12) and observation model (6), a closed form recursive update of the likelihood (24) is given by the KF equations [33]. In particular, by assuming decorrelated frequency components, i.e., diagonal forms for the submatrices $\mathbf{P}_{ij,\tau}\,(i,j \in \{1,\dots,B\})$ of the state uncertainty matrix $\mathbf{P}_\tau$, the process noise covariance matrix $\mathbf{\Psi}_\tau^{\Delta\mathrm{W}}$ and the observation noise covariance matrix $\mathbf{\Psi}_\tau^{\mathrm{N}}$, computationally efficient update rules are obtained [28]:

$$\mathbf{e}_\tau^+ = \mathbf{y}_\tau - \mathbf{C}_\tau \hat{\mathbf{w}}_{\tau-1}^{\mathrm{kf}} \approx \mathbf{y}_\tau - A\,\mathbf{C}_\tau \hat{\mathbf{w}}_{\tau-1}^{\mathrm{kf}} \tag{27}$$

$$\mathbf{P}_{ij,\tau-1}^+ = A^2\,\mathbf{P}_{ij,\tau-1} + \mathbf{\Psi}_{ij,\tau}^{\Delta\mathrm{W}} \tag{28}$$

$$\mathbf{D}_\tau = \sum_{i,j=1}^{B} \mathbf{X}_{i,\tau}\mathbf{P}_{ij,\tau-1}^+\mathbf{X}_{j,\tau}^{\mathrm{H}} + \frac{M}{R}\mathbf{\Psi}_\tau^{\mathrm{N}} \tag{29}$$

$$\mathbf{\Lambda}_{i,\tau} = \sum_{j=1}^{B} \left(\mathbf{P}_{ij,\tau-1}^+\mathbf{X}_{j,\tau}^{\mathrm{H}}\right)\mathbf{D}_\tau^{-1} \tag{30}$$

$$\hat{\mathbf{w}}_{i,\tau}^{\mathrm{kf}} = \hat{\mathbf{w}}_{i,\tau-1}^{\mathrm{kf}} + \mathbf{G}\mathbf{\Lambda}_{i,\tau}\mathbf{e}_\tau^+ \tag{31}$$

$$\mathbf{P}_{ij,\tau} = \mathbf{P}_{ij,\tau-1}^+ - \frac{R}{M}\mathbf{K}_{i,\tau}\sum_{l=1}^{B}\mathbf{X}_{l,\tau}\mathbf{P}_{lj,\tau-1}^+. \tag{32}$$

In contrast to [28], we introduced a gradient constraint matrix $\mathbf{G} = \mathbf{F}_M\mathbf{Q}_2\mathbf{Q}_2^{\mathrm{T}}\mathbf{F}_M^{-1}$ to the filter update (31) to ensure zero-padded time-domain AIR estimates [6, 34]. Note that this could also have been motivated by the introduction of a constrained state transition model (12) as discussed in [35] which we avoided for brevity. Furthermore, we set the state transition factor A to one in the prior error computation (27) (cf. [6]) which allows to interpret the update rules (27) - (32) as a noise-aware multichannel extension of the classical Frequency-Domain Adaptive Filter (FDAF) [7]. Note that the approximation in (27) has only marginal effects on the computation of the prior error $\mathbf{e}_\tau^+$ as the state transition factor A is usually chosen slightly smaller than one [12] (cf. also Sec. 5).

The diagonal entries of the required process and observation noise covariance matrices

$$\left[\mathbf{\Psi}_\tau^{\Delta\mathrm{W}}\right]_{ll} = (1-A^2)\,\left[\mathbf{\Psi}_\tau^{\mathrm{W}}\right]_{ll} \tag{33}$$

$$\left[\mathbf{\Psi}_\tau^{\mathrm{N}}\right]_{mm} = \lambda_{\mathrm{N}}\left[\mathbf{\Psi}_{\tau-1}^{\mathrm{N}}\right]_{mm} + (1-\lambda_{\mathrm{N}})\left|\left[\mathbf{e}_\tau^+\right]_m\right|^2 \tag{34}$$

are estimated from the observed microphone signals using the estimated ATF power

$$\left[\mathbf{\Psi}_\tau^{\mathrm{W}}\right]_{ll} = \lambda_{\mathrm{W}}\left[\mathbf{\Psi}_{\tau-1}^{\mathrm{W}}\right]_{ll} + (1-\lambda_{\mathrm{W}})\left|\left[\hat{\mathbf{w}}_{\tau-1}^{\mathrm{kf}}\right]_l\right|^2 \tag{35}$$

and the recursive averaging factors λ_{W} and λ_{N} [6, 15, 36].

4.2 Adaptive Subspace Tracking

In Sec. 3.4 we described the idea of learning an affine subspace $\mathcal{M}_\tau$ for the current test sample $\underline{\mathbf{w}}_\tau$ from the surrounding K_τ NNs in the training data. This approach is straightforwardly extended to the OSASI application by computing the affine subspace $\mathcal{M}_\tau$ based on the NNs w.r.t to the current AF estimate $\hat{\mathbf{w}}_\tau^{\mathrm{kf}}$ (cf. Eq. (24)). We now discuss the question if there are better choices than the simple squared Euclidean distance (22) for computing the closest neighbors in the training data set. For this we exploit the probabilistic KF model (24) which renders an uncertainty measure $\mathbf{P}_\tau$ of the current mean estimate $\hat{\mathbf{w}}_\tau^{\mathrm{kf}}$. We suggest the negated likelihood of the training data samples given the KF estimate (cf. Eq. 24) as squared distance measure

$$-\log\, p(\mathbf{w}_\kappa^{\mathrm{tr}}|\mathbf{Y}_{1:\tau}) \tag{36}$$

$$= -\log \mathcal{N}_c(\mathbf{w}_\kappa^{\mathrm{tr}}|\hat{\mathbf{w}}_\tau^{\mathrm{kf}}, \mathbf{P}_\tau) \tag{37}$$

$$\overset{\mathrm{c}}{=} \left(\mathbf{w}_\kappa^{\mathrm{tr}} - \hat{\mathbf{w}}_\tau^{\mathrm{kf}}\right)^{\mathrm{H}}\mathbf{P}_\tau^{-1}\left(\mathbf{w}_\kappa^{\mathrm{tr}} - \hat{\mathbf{w}}_\tau^{\mathrm{kf}}\right) \tag{38}$$

$$= d_{\mathrm{kf}}(\mathbf{w}_\kappa^{\mathrm{tr}}, \hat{\mathbf{w}}_\tau^{\mathrm{kf}}). \tag{39}$$

Eq. (39) describes a frequency-dependent weighted squared Euclidean distance for which more reliable estimates, i.e., those having a lower state uncertainty, are more important, i.e., have a higher weight. Note that if the state uncertainty matrix $\mathbf{P}_\tau$ is chosen as identity matrix, as usually done in classical AF algorithms, the probabilistic distance measure (39) simplifies to the squared Euclidean distance (22).

4.3 Soft Subspace Projection

In Sec. 3 the local affine subspace approximation of an AIR manifold has been evaluated by the average system mismatch which results from a projection of an optimum AIR vector $\underline{\mathbf{w}}_\tau$ onto a local affine subspace. Here, the projective mapping (16) sets the coordinates with little influence on the AIR vector to zero, i.e., yields a compressed description of the AIR vectors. This observation can now be exploited to denoise an AF estimate $\underline{\hat{\mathbf{w}}}_\tau^{\mathrm{kf}}$ by

$$\hat{\mathbf{w}}_\tau^{\mathrm{kf,p}} = \left(\mathbf{I} \otimes \left(\mathbf{F}_M\mathbf{Q}_2\right)\right) f_{\mathcal{M}_\tau}\left(\underline{\hat{\mathbf{w}}}_\tau^{\mathrm{kf}}\right) \tag{40}$$

with the time-domain AF estimate

$$\underline{\hat{\mathbf{w}}}_{\tau}^{\mathrm{kf}} = \left(\mathbf{I} \otimes \left(\mathbf{Q}_2^{\mathrm{T}} \mathbf{F}_M^{-1}\right)\right) \hat{\mathbf{w}}_{\tau}^{\mathrm{kf}} \in \mathbb{R}^{Q}. \tag{41}$$

This projection removes all components of the KF-based AIR estimate $\underline{\mathbf{w}}_{\tau}^{\mathrm{kf}}$ which are not supported by the surrounding training data samples spanning the affine subspace $\mathcal{M}_{\tau}$. This is visualized in Fig. 5.

However, the experiments in Sec. 3 showed also the imperfection of the local affine subspace models as optimum AIRs could not be perfectly reconstructed by the projection. This motivates the idea of modelling an uncertainty measure around the manifold, i.e., diluting the deterministic model. We suggest to model this uncertainty by a proper complex Gaussian PDF

$$p(\mathbf{w}_{\tau}|\mathcal{M}_{\tau}) = \mathcal{N}_c(\mathbf{w}_{\tau}|\hat{\mathbf{w}}_{\tau}^{\mathrm{kf,p}}, \mathbf{\Psi}_{\mathcal{M}_{\tau}}) \tag{42}$$

with the mean vector $\hat{\mathbf{w}}_{\tau}^{\mathrm{kf,p}}$ being the projection of $\hat{\mathbf{w}}_{\tau}^{\mathrm{kf}}$ onto the manifold and the covariance matrix $\mathbf{\Psi}_{\mathcal{M}_{\tau}} \in \mathbb{C}^{MB \times MB}$ expressing the uncertainty of the projection. Subsequently, the probability of the ATF vector $\mathbf{w}_{\tau}$ given the KF likelihood (24) and the probabilistic manifold model (42) is given by

$$p(\mathbf{w}_{\tau}|\mathbf{Y}_{1:\tau}, \mathcal{M}_{\tau}) = \tag{43}$$

$$\mathcal{N}_c(\mathbf{w}_{\tau}|\hat{\mathbf{w}}_{\tau}^{\mathrm{kf}}, \mathbf{P}_{\tau})\, \mathcal{N}_c(\mathbf{w}_{\tau}|\hat{\mathbf{w}}_{\tau}^{\mathrm{kf,p}}, \mathbf{\Psi}_{\mathcal{M}_{\tau}}) \tag{44}$$

assuming $p(\mathbf{w}_{\tau}|\mathbf{Y}_{1:\tau}, \mathcal{M}_{\tau}) = p(\mathbf{w}_{\tau}|\mathbf{Y}_{1:\tau})\, p(\mathbf{w}_{\tau}|\mathcal{M}_{\tau})$. The Maximum-Likelihood (ML) estimate of $\mathbf{w}_{\tau}$ based on (44) is given by [37]:

$$\hat{\mathbf{w}}_{\tau}^{\mathrm{den}} = \left(\mathbf{P}_{\tau}^{-1} + \mathbf{\Psi}_{\mathcal{M}_{\tau}}^{-1}\right)^{-1} \left(\mathbf{P}_{\tau}^{-1} \hat{\mathbf{w}}_{\tau}^{\mathrm{kf}} + \mathbf{\Psi}_{\mathcal{M}_{\tau}}^{-1} \hat{\mathbf{w}}_{\tau}^{\mathrm{kf,p}}\right). \tag{45}$$

If the KF state uncertainty matrix $\mathbf{P}_{\tau}$ and the prior covariance matrix $\mathbf{\Psi}_{\mathcal{M}_{\tau}}$ are assumed to be diagonal, Eq. (45) simplifies to

$$\left[\hat{\mathbf{w}}_{b,\tau}^{\mathrm{den}}\right]_m = (1 - \alpha_{b,\tau,m}) \left[\hat{\mathbf{w}}_{b,\tau}^{\mathrm{kf}}\right]_m + \alpha_{b,\tau,m} \left[\hat{\mathbf{w}}_{b,\tau}^{\mathrm{kf,p}}\right]_m \tag{46}$$

with the ATF index b, the frequency bin index m and the convex combination weight

$$\alpha_{b,\tau,m} = \frac{\left[\mathbf{P}_{bb,\tau}\right]_{mm}}{\left[\mathbf{P}_{bb,\tau}\right]_{mm} + \left[\mathbf{\Psi}_{\mathcal{M}_{\tau},bb}\right]_{mm}}. \tag{47}$$

Here, the bth block diagonal element of $\mathbf{\Psi}_{\mathcal{M}_{\tau}}$ is denoted by $\mathbf{\Psi}_{\mathcal{M}_{\tau},bb} \in \mathbb{C}^{M \times M}$. Eq. (46) represents a frequency-dependent convex combination of the KF estimate $\hat{\mathbf{w}}_{\tau}^{\mathrm{kf}}$ and its projection $\hat{\mathbf{w}}_{\tau}^{\mathrm{kf,p}}$ onto the manifold which is visualized in Fig. 5. The proposed approach favours the KF estimate $\hat{\mathbf{w}}_{\tau}^{\mathrm{kf}}$ whenever the KF state uncertainty $\mathbf{P}_{\tau}$ is smaller than the prior model uncertainty $\mathbf{\Psi}_{\mathcal{M}_{\tau}}$ and vice versa.

We will conclude this section by proposing a model for the prior uncertainty $\mathbf{\Psi}_{\mathcal{M}_{\tau}}$. It is assumed that the model uncertainty

$$\mathbf{\Psi}_{\mathcal{M}_{\tau}} = \frac{\beta_{\mathrm{pr}}}{1 - A^2} \mathbf{\Psi}_{\tau}^{\Delta \mathrm{W}} \tag{48}$$

is a scaled version of the process noise covariance matrix $\mathbf{\Psi}_{\tau}^{\Delta \mathrm{W}}$ in the Markov model (12) with $\beta_{\mathrm{pr}} > 0$ being a hyperparameter. This is motivated by the intrinsic ATF variability which is assumed to be proportional to the process noise power.

4.4 Algorithmic Description

Alg. 1 provides a detailed algorithmic description of the proposed KF with an Adaptive Subspace Projection (KS-ASP) algorithm for OSASI. For each block of microphone observations $\underline{\mathbf{y}}_{\tau}$ and loudspeaker excitations $\underline{\mathbf{x}}_{b,\tau}$, the prior error $\mathbf{e}_{\tau}^{+}$ is computed by using the ATF estimate of the previous time step $\hat{\mathbf{w}}_{\tau-1}^{\mathrm{kf}}$ (cf. Eq. (27)). Subsequently, the diagonal process noise and observation noise covariance matrices $\mathbf{\Psi}_{\tau}^{\Delta \mathrm{W}}$ and $\mathbf{\Psi}_{\tau}^{\mathrm{N}}$ are updated by Eqs. (33) - (35). Afterwards, the posterior mean vector $\hat{\mathbf{w}}_{\tau}^{\mathrm{kf}}$ and the state uncertainty matrix $\mathbf{P}_{\tau}$ are updated by the KF Eqs. (28)-(32). The K_{τ} NNs to the posterior mean vector $\hat{\mathbf{w}}_{\tau}^{\mathrm{kf}}$ are computed by the respective distance metric, i.e., $d_{\mathrm{euc}}(\cdot,\cdot)$ or $d_{\mathrm{kf}}(\cdot,\cdot)$ (cf. Eqs. (22), (39)). For an efficient computation of the inverse in Eq. (38), we only use the elements of the main diagonal of the state uncertainty matrix $\mathbf{P}_{\tau}$. This reduces the computation of a full matrix inverse to a

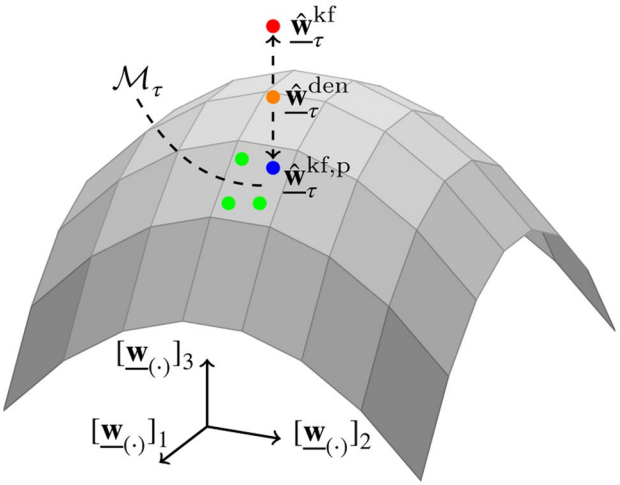

Figure 5 The denoised estimate $\underline{\hat{\mathbf{w}}}_{\tau}^{\mathrm{den}} \in \mathbb{R}^3$ is a convex combination of the KF-based AIR vector estimate $\underline{\hat{\mathbf{w}}}_{\tau}^{\mathrm{kf}}$ and its projection $\underline{\hat{\mathbf{w}}}_{\tau}^{\mathrm{kf,p}}$ onto the affine subspace $\mathcal{M}_{\tau}$. The NNs which span the subspace are visualized as green dots.

element-wise scalar inversion. The corresponding K_τ closest samples w.r.t. d_{euc} or d_{kf} are used as local training data set $\mathcal{U}_\tau$ to compute the associated subspace offset $\underline{\bar{\mathbf{w}}}_\tau$ and projection matrix $\underline{\mathbf{L}}_\tau$ by Eqs. (19) and (17), respectively. Because of the chosen subspace dimension $D_\tau = K_\tau - 1$, the basis matrix $\underline{\mathbf{V}}_\tau$ Eq. (23) can be used in Eq. (17). The KF estimate $\hat{\mathbf{w}}_\tau^{\text{kf}}$ is then projected onto the affine subspace $\mathcal{M}_\tau$ by Eq. (16). Subsequently, the convex combination weights $\alpha_{b,\tau,f}$ (cf. Eq. (47)) are used to fuse the KF estimate $\hat{\mathbf{w}}_\tau^{\text{kf}}$ and its projection $\hat{\mathbf{w}}_\tau^{\text{kf,p}}$. Finally, the respective denoised estimated filter vector $\hat{\mathbf{w}}_\tau^{\text{den}}$ is used as posterior mean vector $\hat{\mathbf{w}}_\tau^{\text{kf}}$ of the KF.

Algorithm 1 Proposed KF-ASP-based AIR estimation for one signal block $\underline{\mathbf{y}}_\tau$.

Compute prior error $\mathbf{e}_\tau^+$ by Eq. (27)
Estimate $\mathbf{\Psi}_\tau^{\Delta\text{W}}$ and $\mathbf{\Psi}_\tau^{\text{N}}$ by Eqs. (33)-(35)
Update $\hat{\mathbf{w}}_\tau^{\text{kf}}$ and $\mathbf{P}_\tau$ by Eqs. (28)-(32)
Compute NN training samples by distance metric $d_{(\cdot)}(\cdot,\cdot)$
Compute subspace offset $\underline{\bar{\mathbf{w}}}_\tau$ by Eq. (19)
Compute projection matrix $\underline{\mathbf{L}}_\tau$ by Eq. (17)
Project KF estimate onto affine subspace $\mathcal{M}_\tau$ by Eq. (16)
Compute convex combination weights $\alpha_{b,\tau,f}$ by Eq. (47)
Compute denoised KF estimate $\hat{\mathbf{w}}_\tau^{\text{den}}$ by Eq. (46)
Assign denoised estimate to KF: $\hat{\mathbf{w}}_\tau^{\text{kf}} \leftarrow \hat{\mathbf{w}}_\tau^{\text{den}}$

5 Experiments

We now evaluate the proposed algorithm for a typical MISO acoustic system identification scenario. The acoustic scene is characterized by a loudspeaker array comprising $B = 2$ elements with a spacing of 10 cm and a single microphone. Loudspeakers and microphone are located in a room of dimensions $[6\text{ m},\ 5\text{ m},\ 3.5\text{ m}]$ and a reverberation time of $T_{60} = 0.3$ s. While the center position and orientation of the loudspeaker array are kept fixed at $\mathbf{r}_{\text{ls}} = [3.0\text{ m},\ 2.0\text{ m},\ 1.2\text{ m}]$ and parallel to the x-axis of the room, respectively, the microphone position

$$\mathbf{r}_{\text{mic}} = \mathbf{r}_{\text{ls}} + \begin{bmatrix} r\ \cos(\phi)\ \cos(\theta) \\ r\ \cos(\phi)\ \sin(\theta) \\ r\ \sin(\phi) \end{bmatrix} \tag{49}$$

is constrained to a volumetric segment of a sphere with radius $r \in [1.2\text{ m},\ 1.4\text{ m}]$, azimuth angle $\theta \in [45°,\ 135°]$ and elevation angle $\phi \in [-5°,\ 40°]$, respectively. All ground-truth AIRs $\mathbf{h}_{b,\tau}$ of length $W = 4096$ have been simulated using the image method [38, 39] with a maximum reflection order and a sampling frequency of $f_s = 8$ kHz. For each experiment the noise-free observation $\underline{\mathbf{d}}_\tau$ is simulated by convolving the ground-truth AIRs $\underline{\mathbf{h}}_{b,\tau}$, corresponding to a random radius r, azimuth angle θ and elevation angle ϕ, with randomly chosen excitation signals $\underline{\mathbf{x}}_{b,\tau}$ and subsequently adding up the convolution products. Here, we consider two types of excitation signals: spatially uncorrelated stationary White Gaussian Noise (WGN) and speech. The speech signals are taken from a subset of the UWNU database [40] which comprises 15 different speakers. For the speech-excited scenarios we consider in addition to playing independent speech signals at each loudspeaker also a teleconferencing setup with strong correlation between the loudspeaker signals. The additive noise signal $\underline{\mathbf{n}}_\tau$ is composed of an interfering WGN component $\underline{\mathbf{n}}_{\text{wgn},\tau}$ and a non-stationary speech component $\underline{\mathbf{n}}_{\text{sp},\tau}$, i.e., $\underline{\mathbf{n}}_\tau = \underline{\mathbf{n}}_{\text{wgn},\tau} + \underline{\mathbf{n}}_{\text{sp},\tau}$. The variances of the noise signals $\underline{\mathbf{n}}_{\text{wgn},\tau}$ and $\underline{\mathbf{n}}_{\text{sp},\tau}$ are prescribed by SNR_{wgn} and SNR_{sp}, respectively. The set of interfering speech signals consists of 15 additional talkers from [40].

To evaluate the performance of the proposed algorithm we use the system mismatch Υ_τ (cf. Eq. (21)) and the Echo Return Loss Enhancement (ERLE)

$$\mathcal{E}_\tau = 10\log_{10}\frac{\mathbb{E}\left[||\underline{\mathbf{d}}_\tau||^2\right]}{\mathbb{E}\left[||\underline{\mathbf{d}}_\tau - \underline{\hat{\mathbf{d}}}_\tau||^2\right]} \tag{50}$$

with the noise-free observation estimate $\hat{\mathbf{d}}_\tau = \mathbf{C}_\tau \hat{\mathbf{w}}_{\tau-1}^{\text{kf}}$ (cf. Eq. (27)). Note that, in contrast to the system mismatch Υ_τ, the ERLE $\mathcal{E}_\tau$ represents a signal-dependent performance measure and, thus, is of particular interest for signal cancellation applications. The expectation operator $\mathbb{E}[\cdot]$ in Eq. (50) is here approximated by recursive averaging over time. To allow for more general conclusions, the system mismatch Υ_τ and the ERLE $\mathcal{E}_\tau$ are averaged over 50 trials of the random experiment with randomly chosen excitation signals, random microphone positions and random interfering noise signals. The respective averaged performance measures are denoted by an overbar $\bar{(\cdot)}$.

In the following experiments we evaluate the proposed KF-ASP algorithm (cf. Alg. 1) for the previously described OSASI scenario. As baseline we use the state-of-the-art KF [28] (cf. Sec. 4.1) with a state transition factor $A = 0.9999$, a frame shift $R = 512$, a filter length $L = 512$ and recursive averaging factors $\lambda_{\text{W}} = 0.9$ and $\lambda_{\text{N}} = 0.5$. The state uncertainty matrix $\mathbf{P}_\tau$ (cf. Eq. (26)) was initialized with an identity matrix scaled by $P_0 = 0.01$. The training data set is composed of $K = 5000$ simulated AIRs from the acoustic scene which correspond to microphone positions in the volumetric sphere segment with random radius r, azimuth angle θ and elevation angle ϕ (cf. Eq. (49)). Note that in real applications this training data set can straightforwardly be obtained by collecting previous estimates that were deemed reliable, indicated by a small state uncertainty $\mathbf{P}_\tau$ (cf. Eq. (26)), of the OSASI algorithm. For the proposed algorithm we investigate three variants: KF-ASP (d_{kf}, Proj.), KF-ASP (d_{kf}, Comb.) and KF-ASP (d_{euc}, Proj.). Here, the variants including d_{kf} use the probabilistic quadratic distance (39) whereas the last one, labeled by d_{euc}, uses the quadratic

Euclidean distance (22) to compute the $K_\tau = 80$ NN training samples for the update in time step τ (cf. Subsec. 4.2). Furthermore, we compare the effect of the convex combination-based enhancement (cf. Subsec. 4.3), labeled by *Comb.*, in contrast to a hard projection, i.e., choosing $\alpha_{b,\tau,m} = 1$ in Eq. (46), which is labeled by *Proj.* For the model prior (cf. Eq. (48)) $\beta_{\text{pr}} = 5$ is chosen. The state uncertainty matrices of the ASP-based algorithms were initialized with a scaling factor of $P_0 = 0.1$. Note that this higher initial state uncertainty could not be used in the baseline KF as it led to divergence in our experiments. In addition, we evaluated two oracle baselines, i.e., Oracle GT and Oracle NN. Here, Oracle GT uses the optimum Q-dimensional MISO AIR vector $\underline{\mathbf{w}}_\tau = \left(\mathbf{I}_B \otimes \mathbf{Q}_3^{\text{T}}\right)\underline{\mathbf{h}}_\tau$, i.e., the first L taps of each ground-truth AIR $\underline{\mathbf{h}}_{b,\tau}$ with $b = 1, \ldots, B$, as estimate and and Oracle NN the training sample $\underline{\mathbf{w}}_\kappa^{\text{tr}}$ with the smallest squared Euclidean distance (22) to $\underline{\mathbf{w}}_\tau$.

Figure 6 shows the average ERLE $\bar{\mathcal{E}}_\tau$ and system mismatch $\bar{\Upsilon}_\tau$ for an excitation with spatially uncorrelated stationary WGN input signal ($\text{SNR}_{\text{wgn}} = 0\,\text{dB}, \text{SNR}_{\text{sp}} = \infty\,\text{dB}$). We conclude from Fig. 6 that the proposed KF-ASP algorithms significantly increase the convergence rate of the baseline KF. However, the steady-state performance of the variants KF-ASP (d_{euc}, Proj.) and KF-ASP (d_{kf}, Proj.) which rely entirely on the hard projection, i.e., the choice $\alpha_{b,\tau,m} = 1$ in the convex combination (46), are significantly worse than the baseline KF. This is due to the imperfections of the affine subspace model whose modeling capability depends on the training sample size K and the choice of the subspace dimension D_τ (cf. Fig. 4). In contrast, the soft projection-based KF-ASP approach (cf. Subsec. 4.3) achieves an increased steady-state performance in comparison to the hard projection-based variants due to the convex combination (cf. Eq. (46)), which takes the uncertainty of the model into account. Yet, the baseline KF still shows a slightly improved steady-state performance which motivates an adaptive control of the prior covariance matrix (48). Furthermore, we observe that the probabilistic distance d_{kf} performs similarly to the Euclidean distance d_{euc}. This is explained by the stochastic properties of the excitation and noise signals which do not result in a nonuniform frequency-dependent weighting in Eq. (38). Finally, by inspecting the performance of the Oracle NN algorithm we conclude that the affine subspace models generalize better to AIRs in between the training data samples.

A scenario with the two loudspeakers playing independent speech excitation signals is shown in Fig. 7 ($\text{SNR}_{\text{wgn}} = 5\,\text{dB}$, $\text{SNR}_{\text{sp}} = 0\,\text{dB}$). Similarly as for the WGN excitation the affine subspace projection greatly improves the convergence rate of the baseline KF. However, in contrast to the WGN excitation, the speech excitation benefits from the proposed probabilistic distance measure d_{kf} (39). This is explained by the nonuniform weighting in Eq. (38) which results from

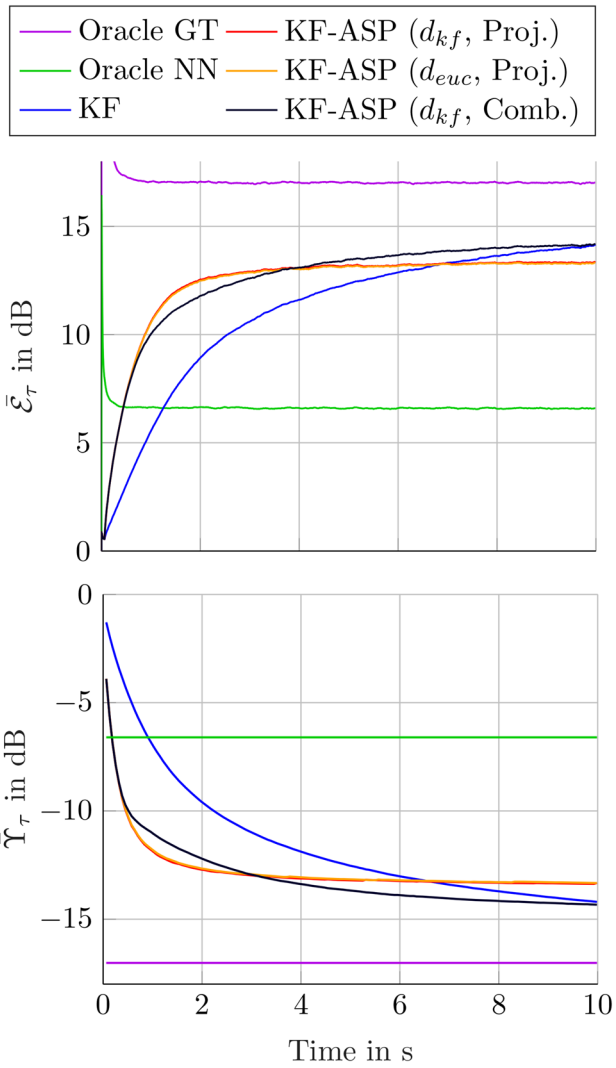

Figure 6 Evaluation of the proposed KF-ASP-based algorithmic variants for a system identification application ($\text{SNR}_{\text{wgn}} = 0$ dB, $\text{SNR}_{\text{sp}} = \infty$ dB) with a spatially uncorrelated WGN excitation signal.

the frequency-dependent power of the speech excitation. Furthermore, the proposed KF-ASP algorithms show here an improved steady-state performance in comparison to the baseline KF and almost achieve the Oracle GT performance.

We now extend the previous system identification scenarios to a typical Acoustic Echo Cancellation (AEC) task in a teleconferencing setup ($\text{SNR}_{\text{wgn}} = 10\,\text{dB}$, $\text{SNR}_{\text{sp}} = \infty\,\text{dB}$). Here, we assume that the loudspeaker array in the near-end room plays two microphone signals which are recorded in a distant far-end room [1]. The far-end microphone signals are composed of the reverberant speech signals of two spatially separated far-end speakers with disjoint activity. Due to the high spatial correlation of the recorded speech signals, the system identification in the near-end room is complicated by the nonuniqueness problem [26, 27]. Here,

the near-end OSASI AEC algorithm can only estimate a nonunique cancellation filter which often leads to severe performance drops whenever the activity of the far-end speakers changes. Figure 8 shows the average ERLE $\bar{\mathcal{E}}_\tau$ and system mismatch $\bar{\Upsilon}_\tau$ computed from 50 trials for the experiment. We conclude that, while the ERLE is high, the baseline KF does not converge towards the true AIRs. Due to the nonunique filter estimate the performance of the baseline KF significantly drops after the source activity switching at $t = 5$ s. In contrast, the KF-ASP-based algorithms achieve a much better average system mismatch $\bar{\Upsilon}_\tau$. This might be explained by the projections keeping the solution closer to the true AIR. Thus, there is no performance drop when the far-end speakers flip.

Finally, we evaluate the effect of the training data size K. To this end, we simulate an additional scenario with spatially uncorrelated speech excitation ($\text{SNR}_{\text{wgn}} = 0$ dB, $\text{SNR}_{\text{sp}} = \infty$ dB). However, in contrast to the previous experiments ($K = 5000$), the affine subspace models are learned from a training data set including only $K = 1000$ samples. The results shown in Fig. 9 suggest a rapidly decreasing performance of the algorithms relying on hard projection onto affine subspaces. This is explained by the reduced number of training data samples K which only allow to learn coarse AIR models. In contrast, the soft projection-based approach, which does not enforce that all AIR estimates are confined to the subspace model, still achieves an excellent steady-state performance in addition to the improved convergence rate.

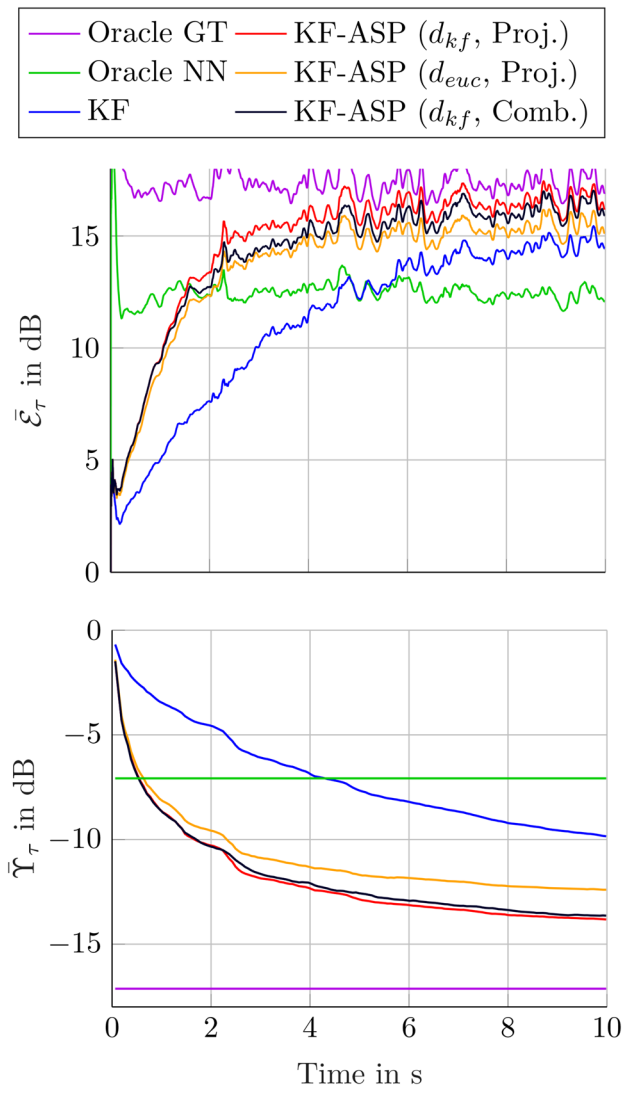

Figure 7 Evaluation of the proposed KF-ASP-based algorithmic variants for a system identification application ($\text{SNR}_{\text{wgn}} = 5$ dB, $\text{SNR}_{\text{sp}} = 0$ dB) with two independent speech excitation signals.

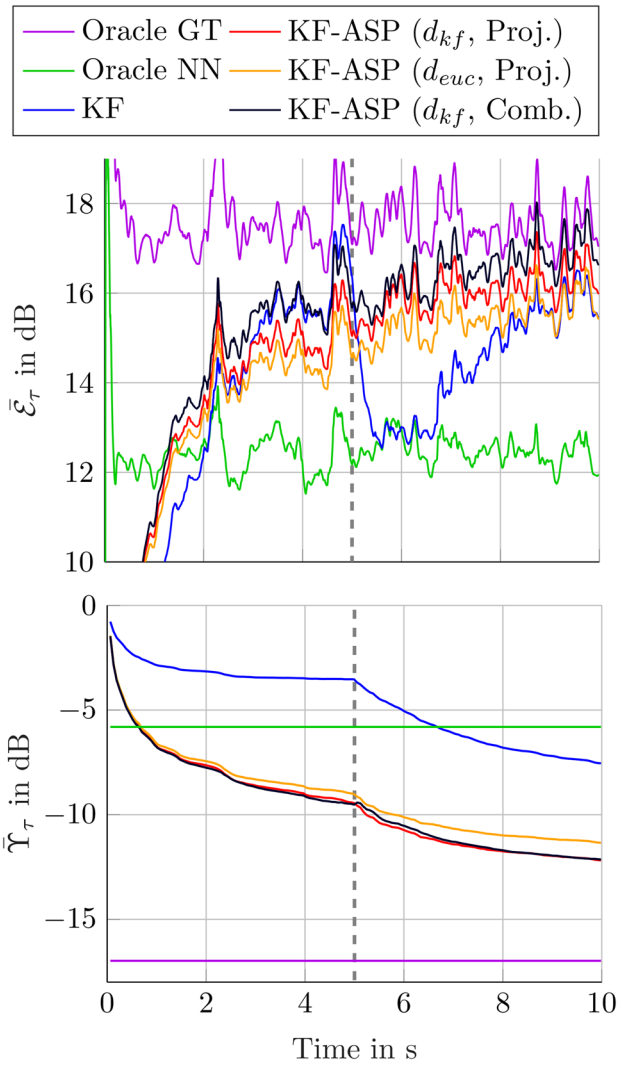

Figure 8 Evaluation of the proposed KF-ASP-based algorithmic variants for a teleconferencing setup ($\text{SNR}_{\text{wgn}} = 10$ dB, $\text{SNR}_{\text{sp}} = \infty$ dB). After 5 s the far-end speakers switch (indicated by a dashed gray line).

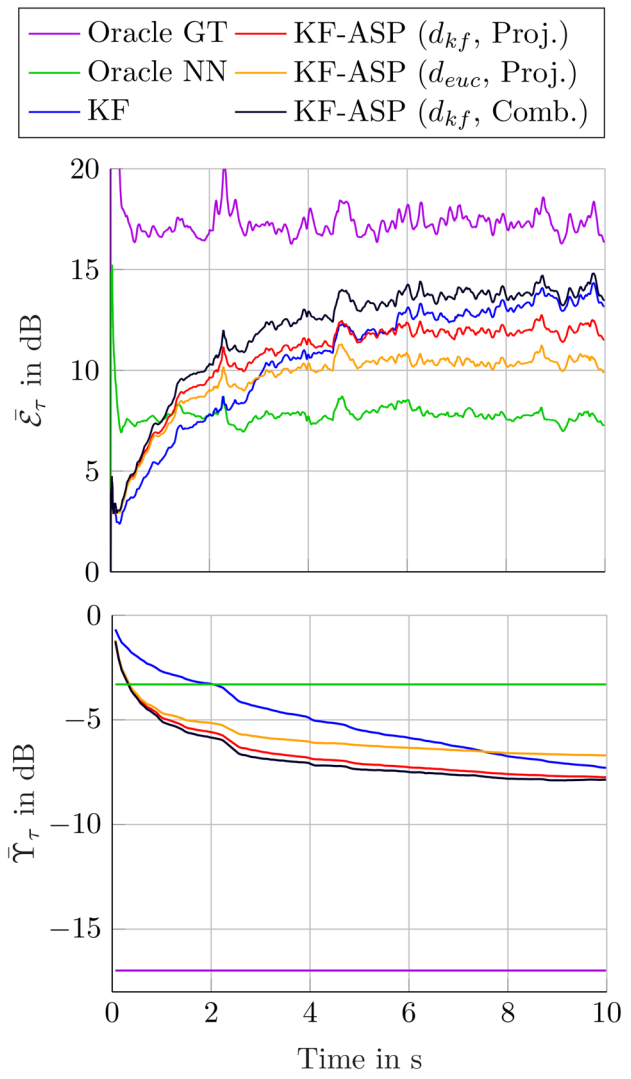

Figure 9 Evaluation of the proposed KF-ASP-based algorithmic variants for a training data size of $K = 1000$ samples. The loudspeakers play two independent speech signals ($\mathrm{SNR}_{\mathrm{wgn}} = 0$ dB, $\mathrm{SNR}_{\mathrm{sp}} = \infty$ dB).

6 Summary and Outlook

In this paper we have introduced a family of novel OSASI algorithms which exhibit significantly faster convergence properties and improved steady-state performance in scenarios suffering from high-level interfering noise and nonuniqueness of optimum filter estimates. The proposed algorithms assume that the variability of AIRs in an acoustic scene is confined to a non-linear manifold which can locally be approximated by an affine subspace. This allows to enhance an AF-based AIR estimate by projecting it onto the learned subspace. As future research we plan to develop improved choices for the prior covariance matrix and evaluate the effect of noisy training data samples on the learned AIR models.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

1. Enzner, G., et al. (2014). Acoustic Echo Control, in *Academic Press Library in Signal Processing*, vol. 4, pp. 807–877. Elsevier.
2. Diniz, P. S. R. (2007). *Adaptive Filtering: Algorithms and Practical Implementation*. Springer: Berlin, Heidelberg.
3. Widrow, B., & Hoff, M. E. (1960). *Adaptive Switching Circuits, in WESCON Convention Record* (pp. 96–104). Los Angeles, CA: USA, Aug.
4. Ferrara, E. (1980). Fast implementations of LMS adaptive filters. *Transactions on Acoustics, Speech, and Signal Processing, 28*(4), 474–475.
5. Benesty, J., et al. (2006). A Nonparametric VSS NLMS Algorithm. *IEEE Signal Processing Letters, 13*(10), 581–584.
6. Kuech, F., et al. (2014). *State-space architecture of the partitioned-block-based acoustic echo controller, in International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1295–1299). Florence: Italy, May.
7. Haykin, S. (2002). *Adaptive Filter Theory* (4th ed.). NJ, USA: Prentice Hall.
8. Hänsler, E., & Schmidt, G. (2004). *Acoustic Echo and Noise Control: A practical Approach*. NJ, USA: Wiley-Interscience.
9. Mansour, D., & Gray, A. (1982). Unconstrained frequency-domain adaptive filter. *Transactions on Acoustics, Speech, and Signal Processing, 30*(5), 726–734.
10. Benesty, J., et al. (2000). A new class of doubletalk detectors based on cross-correlation. *IEEE Transactions on Speech and Audio Processing, 8*(2), 168–172.
11. Nitsch, B. H. (2000). A frequency-selective stepfactor control for an adaptive filter algorithm working in the frequency domain. *Signal Processing, 80*(9), 1733–1745.
12. Enzner, G., & Vary, P. (2006). Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones. *Signal Processing, 86*(6), 1140–1156.
13. Malik, S., & Enzner, G. (2010). *Online maximum-likelihood learning of time-varying dynamical models in block-frequency-domain, in International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Dallas, TX: USA.
14. Huemmer, C., et al. (2015). The NLMS algorithm with time-variant optimum stepsize derived from a bayesian network perspective. *IEEE Signal Processing Letters, 22*(11), 1874–1878.
15. Yang, F., et al. (2017). Frequency-Domain Adaptive Kalman Filter With Fast Recovery of Abrupt Echo-Path Changes. *IEEE Signal Processing Letters, 24*(12), 1778–1782.
16. Haubner, T., et al. (2021). *Noise-robust adaptation control for supervised acoustic system identification exploiting a noise dictionary, in International Conference on Acoustics, Speech*

and Signal Processing (ICASSP) (pp. 945–949). Toronto, ON: Canada, June.
17. Haubner, T., et al. (2021). *A Synergistic Kalman- and Deep Postfiltering Approach to Acoustic Echo Cancellation, in European Signal Processing Conference (EUSIPCO)*. Dublin: Ireland.
18. Fozunbal, M., et al. (2008). *Multi-Channel Echo Control by Model Learning, in International Workshop on Acoustic Echo and Noise Control (IWAENC)*. USA: Seattle.
19. Koren, T., et al. (2012). *Supervised system identification based on local PCA models, in International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 541–544). Kyoto: Japan, Mar.
20. Talmon, R., & Gannot, S. (2013). *Relative transfer function identification on manifolds for supervised GSC beamformers, in European Signal Processing Conference (EUSIPCO)*. Marrakech: Morocco.
21. Haubner, T., et al. (2020). *Online supervised acoustic system identification exploiting prelearned local affine subspace models, in International Workshop on Machine Learning for Signal Processing (MLSP)*. Espoo: Finland.
22. Laufer-Goldshtein, B., et al. (2015). *A Study on Manifolds of Acoustic Responses, in Latent Variable Analysis and Signal Separation (LVA/ICA)* (pp. 203–210). Liberec: Czech Republic, Aug.
23. Talmon, R., et al. (2013). Diffusion Maps for Signal Processing: A Deeper Look at Manifold-Learning Techniques Based on Kernels and Graphs. *IEEE Signal Processing Magazine, 30*(4), 75–86.
24. Jolliffe, I. T. (1986). Principal components in regression analysis, in *Principal component analysis*, pp. 129–155. Springer.
25. Hahmann, M., et al. (2019). *Analysis of a sound field in a room using dictionary learning, in 23rd International Congress on Acoustics (ICA)*. Aachen: Germany.
26. Sondhi, M. M., et al. (1995). Stereophonic acoustic echo cancellation-an overview of the fundamental problem. *IEEE Signal Processing Letters, 2*(8), 148–151.
27. Benesty, J., et al. (1998). A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation. *IEEE Transactions on Speech and Audio Processing, 6*(2), 156–165.
28. Malik, S., & Enzner, G. (2011). Recursive Bayesian Control of Multichannel Acoustic Echo Cancellation. *IEEE Signal Processing Letters, 18*(11), 619–622.
29. Tu, L. W. (2010). *An Introduction to Manifolds*. New York: Universitext. Springer.
30. Strang, G. (2006). *Linear Algebra and its Applications*. Brooks/Cole, Belmont, CA: Thomson.
31. Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory, 28*(2), 129–137.
32. Arthur, D., & Vassilvitskii, V. (2007). K-means++: The advantages of careful seeding, in *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms.* New Orleans, LA: USA.
33. Bishop, C. M. (2007). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer: Berlin, Heidelberg.
34. Buchner, H., et al. (2005). Generalized multichannel frequency-domain adaptive filtering: efficient realization and application to hands-free speech communication. *Signal Processing, 85*(3), 549–570.
35. Dietzen, T., et al. (2016). *Partitioned block frequency domain Kalman filter for multi-channel linear prediction based blind speech dereverberation, in 2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*. Xi'an: China
36. Franzen, J., & Fingscheidt, T. (2019). *Improved Measurement Noise Covariance Estimation for N-channel Feedback Cancellation Based on the Frequency Domain Adaptive Kalman Filter, in International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 965–969). Brighton: United Kingdom, May.
37. Petersen, K. B., & Pedersen, M. S. (2012). *The matrix cookbook, Version* 20121115.
38. Allen, J. B., & Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics. *Journal of the Acoustical Society of America, 65*(4), 943–950.
39. Habets, E. (2010). *Room Impulse Response Generator*. Tech. Rep.: Technische Universiteit Eindhoven.
40. Panfili, L. M., et al. (2017). *The UW/NU corpus, version 2.0*, https://depts.washington.edu/phonlab/projects/uwnu.php.

Thomas Haubner studied mechatronics and mechanical engineering at Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Germany, and received his M.Sc. degree (with distinction) in 2017. In the same year, he joined the Chair of Multimedia Communications and Signal Processing, FAU, where he is currently pursuing a Dr.-Ing. degree. He was a visiting scientist at Tsukuba University, Japan (Oct.-Dec. 2017). His research interests include acoustic echo control, blind source separation and adaptive system identification.

Andreas Brendel received the B.Sc. degree and the M.Sc. degree (with distinction) in electrical engineering in 2014 and 2016, respectively. Mr. Brendel is a recipient of the Deutschlandsti-pendium. Since 2016, he is with the Chair of Multimedia Communications and Signal Processing of the Friedrich-Alexander Universität Erlangen-Nürnberg working towards the Dr.-Ing. degree. He was a Visiting Scientist at Tsukuba University, Japan (Sep.–Nov. 2016) and at Bar Ilan University, Israel (Sep.–Oct. 2018). His research interests include localization and tracking of acoustic sources, acoustic sensor networks, blind source separation and acoustic metaparameter estimation.

Walter Kellermann is a professor for communications at the University of Erlangen-Nuremberg (FAU), Germany, since 1999. He received the Dipl.-Ing. (univ.) degree in Electrical Engineering from FAU, in 1983, and the Dr.-Ing. degree from the TH Darmstadt, Germany, in 1988. From 1989 to 1990, he was a Postdoc at AT&T Bell Laboratories, Murray Hill, NJ. In 1990, he joined Philips Kommunikations Industrie, Nuremberg, Germany. From 1993 to 1999, he was a Professor at the Fachhochschule Regensburg. In 1999, he cofounded DSP Solutions, a consulting firm in digital signal processing, and he joined FAU as a tenured Professor. Since then, he has also been a consultant to numerous research divisions in industry and an advisor and reviewer for national and international government and funding organizations. He authored or coauthored 21 book chapters, 300+ refereed papers in journals and conference proceedings, and 70+ patents, and is a corecipient of ten best paper awards. Aside from various roles in the IEEE Signal Processing Society, he served as an Associate Editor and Guest Editor to various IEEE and EURASIP journals. His current research interests include speech signal processing, array signal processing, machine learning, and its applications to acoustic human–machine interfaces and autonomous systems. He was the General Chair of seven mostly IEEE-sponsored workshops and conferences. He was awarded the *Julius von Haast Fellowship* by the Royal Society of New Zealand in 2012 and the *Group Technical Achievement Award* of the European Association for Signal Processing (EURASIP) in 2015. He is an IEEE Fellow since 2008 and was elevated to EURASIP Fellow in 2021.