Linear Prediction-Based Online Dereverberation and Noise Reduction Using Alternating Kalman Filters

Sebastian Braun, *Student Member, IEEE* and Emanuël A. P. Habets, *Senior Member, IEEE*

***Abstract*—Multichannel linear prediction-based dereverberation in the short-time Fourier transform (STFT) domain has been shown to be highly effective. Using this framework, the desired dereverberated multichannel signal is obtained by filtering the noise-free reverberant signals using the estimated multichannel autoregressive (MAR) coefficients. To use such methods in the presence of noise, especially in the case of online processing, remains a challenging problem. Existing sequential enhancement structures, which first remove the noise and then estimate the MAR coefficients, suffer from a causality problem as both the optimal noise reduction and dereverberation stages depend on the current output of each other. To address this problem, an algorithm that consists of two alternating Kalman filters to estimate the noise-free reverberant signals and the (MAR) coefficients is proposed. The causality of the estimation procedure is important when dealing with time-variant acoustic scenarios, where the MAR coefficients are time-varying. The proposed method is evaluated using simulated and measured acoustic impulse responses and is compared to a method based on the same signal model. In addition, a method to control the reverberation reduction and noise reduction independently is derived.**

***Index Terms*—Dereverberation, multichannel linear prediction, autoregressive model, Kalman filter, alternating minimization.**

I. Introduction

In DISTANT speech communication scenarios, where the desired speech source is far from the capturing device, the speech quality and intelligibility is typically degraded due to high levels of reverberation and noise compared to the desired speech level [1]. Also the performance of speech recognizers degrades drastically in distant talking scenarios [2], [3]. Therefore, dereverberation in noisy environments for real-time frame-by-frame processing with high perceptual quality remains a challenging and partly unsolved problem.

State-of-the-art multichannel dereverberation algorithms are based on spatio-spectral filtering [4], [5], system identification [6], [7], acoustic channel inversion [8], [9] or linear prediction using an autoregressive reverberation model [10]–[12]. Successful application of the linear prediction-based approaches was achieved by using a multichannel autoregressive (MAR) model for each short-time Fourier transform (STFT) domain frequency band. Advantages of methods based on the MAR model are that they are valid for multiple sources, they directly estimate a dereverberation filter of finite length, the required filters are relatively short, and they are suitable as pre-processing techniques for beamforming algorithms. A great challenge of the MAR signal model is the integration of additive noise, which has to be removed in advance [11], [13] without destroying the relation between successive frames of the reverberant signal. In [14], a generalized framework for the multichannel linear prediction methods called *blind impulse response shortening* was presented, which aims at shortening the reverberant tail in each microphone signal and results in the same number of output as input channels, while preserving the inter-microphone correlation of the desired signal.

As early solutions based on the multichannel linear prediction framework were batch algorithms, further efforts have been made to develop online algorithms, which are suitable for real-time processing [15]–[19]. However, the reduction of additive noise using an online solution has been considered only in [16] to the best of our knowledge.

In this paper, we propose a method based on the MAR reverberation model to reduce reverberation and noise using an online algorithm as an extension of the noise-free solution presented in [20], where the MAR coefficients are modeled by a time-varying first-order Markov model. To obtain the desired dereverberated multichannel speech signal, we have to estimate the MAR coefficients and the multichannel noise-free reverberant speech signal.

The proposed solution has several advantages when compared to state-of-the-art solutions: Firstly, in contrast to the sequential signal and autoregressive (AR) parameter estimation methods used for noise reduction presented in [21], [22], we propose a parallel estimation structure and use an alternating minimization algorithm which consists of two interacting Kalman filters to estimate the MAR coefficients and the noise-free reverberant multichannel signal. This parallel structure allows a fully causal estimation chain as opposed to a sequential structure, where the noise reduction stage would use outdated MAR coefficients. Secondly, in the proposed method we assume the MAR coefficients can be modeled using a time-varying stochastic process, instead of a time-varying deterministic process as in the expectation-maximization (EM) algorithm proposed in [16]. Thirdly, our proposed algorithm does not require multiple iterations per time frame but is an adaptive algorithm that converges

Manuscript received September 29, 2017; revised January 16, 2018; accepted February 11, 2018. Date of publication March 7, 2018; date of current version April 11, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Tan Lee. *(Corresponding author: Sebastian Braun.)*

The authors are with the International Audio Laboratories Erlangen, a joint institution of the Fraunhofer IIS and the Friedrich-Alexander University Erlangen-Nürnberg, Erlangen 91054, Germany (e-mail: sebastian.braun@audiolabs-erlangen.de; emanuel.habets@audiolabs-erlangen.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASLP.2018.2811247

over time. Finally, we propose a method to control the amount of reverberation and noise reduction independently.

The remainder of the paper is organized as follows. In Section II, the signal models for the reverberant signal, the noisy observation, and the MAR coefficients are presented, and the problem is formulated. In Section III, two alternating Kalman filters are derived as part of an alternating minimization problem to estimate the MAR coefficients and the noise-free multichannel signal. An optional method to control the reverberation and noise reduction is presented in Section IV. In Section V, the proposed method is evaluated and compared to state-of-the-art methods. The paper is finally concluded in Section VI.

Notation: Vectors are denoted as lower case bold symbols, e.g., $\mathbf{a}$, matrices as upper case bold symbols, e.g., $\mathbf{A}$ and scalars in normal font, e.g., A. Estimated quantities are denoted by $\widehat{\cdot}$, e.g., $\widehat{A}$.

II. Signal Model and Problem Formulation

We assume an array of M microphones with arbitrary directivity and arbitrary geometry. The microphone signals are given in the STFT domain by $Y_m(k,n)$ for $m \in \{1,\dots,M\}$, where k and n denote the frequency and time indices, respectively. In vector notation, the microphone signals can be written as $\mathbf{y}(k,n) = [Y_1(k,n),\dots,Y_M(k,n)]^{\mathrm{T}}$. We assume that the multichannel microphone signal vector is composed as

$$\mathbf{y}(k,n) = \mathbf{x}(k,n) + \mathbf{v}(k,n), \tag{1}$$

where the vectors $\mathbf{x}(k,n)$ and $\mathbf{v}(k,n)$ contain the reverberant speech at each microphone and additive noise, respectively.

A. Multichannel Autoregressive Reverberation Model

As proposed in [10], [11], [14], we model the reverberant speech signal vector $\mathbf{x}(k,n)$ as an MAR process

$$\mathbf{x}(k,n) = \underbrace{\sum_{\ell=D}^{L} \mathbf{C}_\ell(k,n)\mathbf{x}(k,n-\ell)}_{\mathbf{r}(k,n)} + \mathbf{s}(k,n), \tag{2}$$

where the vector $\mathbf{s}(k,n) = [S_1(k,n),\dots,S_M(k,n)]^{\mathrm{T}}$ contains the desired early speech at each microphone $S_m(k,n)$, and the $M \times M$ matrices $\mathbf{C}_\ell(k,n)$, $\ell \in \{D, D+1,\dots,L\}$ contain the MAR coefficients predicting the late reverberation component $\mathbf{r}(k,n)$ from past frames of $\mathbf{x}(k,n)$. The desired early speech $\mathbf{s}(k,n)$ is the innovation in this autoregressive process (also known as the prediction error in the linear prediction terminology). The choice of the delay $D \geq 1$ determines the amount of early reflections preserved in the desired signal, and should be chosen depending on the amount of overlap between STFT frames, such that there is little to no correlation between the direct sound contained in $\mathbf{s}(k,n)$ and the late reverberation $\mathbf{r}(k,n)$. The length $L > D$ determines the number of past frames that are used to predict the reverberant signal in each frequency band.

We assume that the desired early speech vector $\mathbf{s}(k,n) \sim \mathcal{N}(\mathbf{0}_{M\times 1}, \mathbf{\Phi}_{\mathbf{s}}(k,n))$ and the noise vector $\mathbf{v}(k,n) \sim \mathcal{N}(\mathbf{0}_{M\times 1}, \mathbf{\Phi}_{\mathbf{v}}(k,n))$ are circularly complex zero-mean Gaussian random variables with the respective covariance matrices $\mathbf{\Phi}_{\mathbf{s}}(k,n) = E\{\mathbf{s}(k,n)\mathbf{s}^{\mathrm{H}}(k,n)\}$ and $\mathbf{\Phi}_{\mathbf{v}}(k,n) = E\{\mathbf{v}(k,n)\mathbf{v}^{\mathrm{H}}(k,n)\}$. Furthermore we assume that $\mathbf{s}(k,n)$ and $\mathbf{v}(k,n)$ are uncorrelated across time and both variables are mutually uncorrelated. These assumptions hold well for the STFT coefficients of non-reverberant speech and a wide variety of noise types that typically have short to moderate temporal correlation in the time domain, and are widely used in speech processing methods [6], [23], [24].

B. Signal Model Formulated Using Two Compact Notations

To formulate a cost-function, which is decomposed into two sub-cost-functions in Section III, we first introduce two equivalently usable matrix notations to describe the observed signal vector (1). For the sake of a more compact notation, the frequency indices k are omitted in the remainder of the paper. Let us first define the quantities

$$\mathbf{X}(n) = \mathbf{I}_M \otimes \begin{bmatrix} \mathbf{x}^{\mathrm{T}}(n-L+D) & \dots & \mathbf{x}^{\mathrm{T}}(n) \end{bmatrix} \tag{3}$$

$$\mathbf{c}(n) = \mathrm{Vec}\left\{ \begin{bmatrix} \mathbf{C}_L(n) & \dots & \mathbf{C}_D(n) \end{bmatrix}^{\mathrm{T}} \right\}, \tag{4}$$

where $\mathbf{I}_M$ is the $M \times M$ identity matrix, $\otimes$ denotes the Kronecker product, and the operator $\mathrm{Vec}\{\cdot\}$ stacks the columns of a matrix sequentially into a vector. Consequently, $\mathbf{c}(n)$ is column vector of length $L_{\mathrm{c}} = M^2(L-D+1)$ and $\mathbf{X}(n)$ is a sparse matrix of size $M \times L_{\mathrm{c}}$. Using the definitions (3) and (4) with the signal model (1) and (2), the observed signal vector is given by

$$\mathbf{y}(n) = \underbrace{\mathbf{X}(n-D)\mathbf{c}(n)}_{\mathbf{r}(n)} + \underbrace{\mathbf{s}(n) + \mathbf{v}(n)}_{\mathbf{u}(n)}, \tag{5}$$

where the vector $\mathbf{u}(n)$ contains the early speech plus noise signals that consequently have the covariance matrix $\mathbf{\Phi}_{\mathbf{u}}(k,n) = E\{\mathbf{u}(k,n)\mathbf{u}^{\mathrm{H}}(k,n)\}$, and $\mathbf{u}(k,n) \sim \mathcal{N}(\mathbf{0}_{M\times 1}, \mathbf{\Phi}_{\mathbf{u}}(k,n))$.

The second compact notation uses the stacked vectors

$$\underline{\mathbf{x}}(n) = \begin{bmatrix} \mathbf{x}^{\mathrm{T}}(n-L+1) & \dots & \mathbf{x}^{\mathrm{T}}(n) \end{bmatrix}^{\mathrm{T}} \tag{6}$$

$$\underline{\mathbf{s}}(n) = \begin{bmatrix} \mathbf{0}_{1\times M(L-1)} & \mathbf{s}^{\mathrm{T}}(n) \end{bmatrix}^{\mathrm{T}}, \tag{7}$$

indicated as underlined variables, which are column vectors of length ML, and the propagation and observation matrices

$$\mathbf{F}(n) = \begin{bmatrix} \mathbf{0}_{M(L-1)\times M} & & \mathbf{I}_{M(L-1)} \\ \mathbf{C}_L(n) & \dots & \mathbf{C}_D(n) \quad \mathbf{0}_{M\times M(D-1)} \end{bmatrix} \tag{8}$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{0}_{M\times M(L-1)} & \mathbf{I}_M \end{bmatrix}, \tag{9}$$

respectively, where the $ML \times ML$ propagation matrix $\mathbf{F}(n)$ contains the MAR coefficients $\mathbf{C}_\ell(n)$ in the bottom M rows, and $\mathbf{H}$ is a $M \times ML$ selection matrix. Using (8) and (9), we can alternatively recast (2) and (1) to

$$\underline{\mathbf{x}}(n) = \mathbf{F}(n)\underline{\mathbf{x}}(n-1) + \underline{\mathbf{s}}(n) \tag{10}$$

$$\mathbf{y}(n) = \mathbf{H}\underline{\mathbf{x}}(n) + \mathbf{v}(n). \tag{11}$$

Note that (5) and (11) are equivalent using different notations.

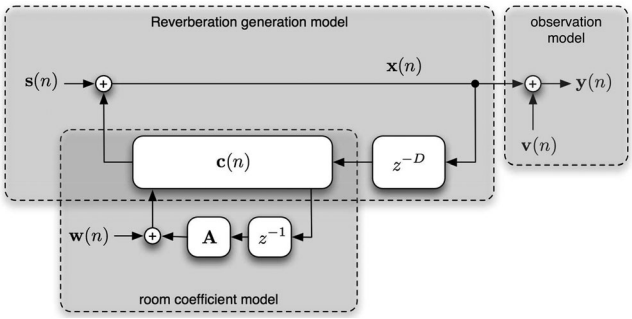

Fig. 1. Generative model of the reverberant signals, multichannel autoregressive coefficients and noisy observation.

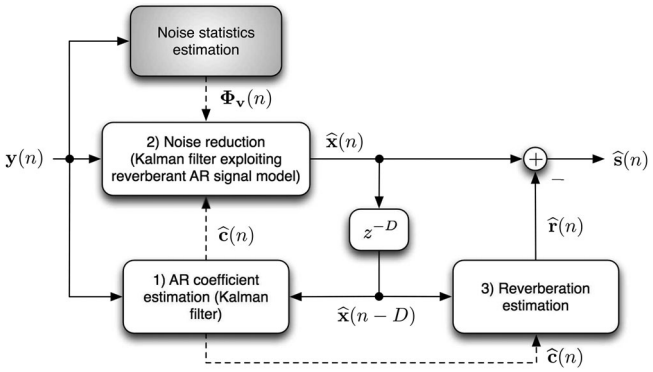

Fig. 2. Proposed parallel dual Kalman filter structure. The three-step procedure ensures that all blocks receive current parameter estimates without delay at each time step n. For the grey noise estimation block, there exist several suitable solutions, which are beyond the scope of this paper.

C. Stochastic State-Space Modeling of MAR Coefficients

To model possibly time-varying acoustic environments and the non-stationarity of the MAR coefficients due to model errors of the STFT domain model [20], we use a first-order Markov model to describe the MAR coefficient vector [25]

$$\mathbf{c}(n) = \mathbf{A}\,\mathbf{c}(n-1) + \mathbf{w}(n). \tag{12}$$

We assume that the transition matrix $\mathbf{A} = \mathbf{I}_{L_\mathrm{c}}$ is an identity matrix, while the process noise $\mathbf{w}(n)$ models the uncertainty of $\mathbf{c}(n)$ over time. We assume that $\mathbf{w}(n) \sim \mathcal{N}\left(\mathbf{0}_{M\times 1}, \mathbf{\Phi}_{\mathbf{w}}(n)\right)$ is a circularly complex zero-mean Gaussian random variable with covariance $\mathbf{\Phi}_{\mathbf{w}}(n)$, and that $\mathbf{w}(n)$ is uncorrelated across time and uncorrelated with $\mathbf{u}(n)$.

Fig. 1 shows the generation process of the observed signals and the underlying (hidden) processes of the reverberant signals and the MAR coefficients.

D. Problem Formulation

Our goal is to obtain an estimate of the multichannel early speech signal $\mathbf{s}(n)$. Instead of directly estimating $\mathbf{s}(n)$, we propose to first estimate the noise-free reverberant signals $\mathbf{x}(n)$ and the MAR coefficients $\mathbf{c}(n)$, denoted by $\widehat{\mathbf{x}}(n)$ and $\widehat{\mathbf{c}}(n)$. Then we can obtain an estimate of the desired signals by applying the MAR coefficients in the manner of a finite multiple-input multiple-output (MIMO) filter to the reverberant signals, i.e.,

$$\widehat{\mathbf{s}}(n) = \widehat{\mathbf{x}}(n) - \underbrace{\widehat{\mathbf{X}}(n-D)\widehat{\mathbf{c}}(n)}_{\widehat{\mathbf{r}}(n)}, \tag{13}$$

where $\widehat{\mathbf{X}}(n)$ is constructed using (3) with $\widehat{\mathbf{x}}(n)$, and $\widehat{\mathbf{r}}(n)$ is considered as the estimated late reverberation. In the following section we show how we can jointly estimate $\mathbf{x}(n)$ and $\mathbf{c}(n)$.

III. MMSE Estimation by Alternating Minimization

The stacked reverberant speech signal vector $\underline{\mathbf{x}}(n)$ and the MAR coefficient vector $\mathbf{c}(n)$ (which is encapsulated in $\mathbf{F}(n)$) can be estimated in the minimum mean-square error (MMSE) sense by minimizing the cost function

$$J(\underline{\mathbf{x}}, \mathbf{c}) = E\left\{\left\|\underline{\mathbf{x}}(n) - \underbrace{\widehat{\mathbf{F}}(n)\underline{\widehat{\mathbf{x}}}(n-1) + \underline{\widehat{\mathbf{s}}}(n)}_{\underline{\widehat{\mathbf{x}}}(n)}\right\|_2^2\right\}. \tag{14}$$

To simplify the estimation problem (14) to obtain a closed-form solution, we resort to an *alternating minimization* technique [26], which minimizes the cost function for each variable separately, while keeping the other variable fixed and using the available estimated value. The two sub-cost-functions, where the respective other variable is assumed as fixed, are given by

$$J_c\left(\mathbf{c}(n)|\underline{\mathbf{x}}(n)\right) = E\left\{\|\mathbf{c}(n) - \widehat{\mathbf{c}}(n)\|_2^2\right\} \tag{15}$$

$$J_x\left(\underline{\mathbf{x}}(n)|\mathbf{c}(n)\right) = E\left\{\|\underline{\mathbf{x}}(n) - \underline{\widehat{\mathbf{x}}}(n)\|_2^2\right\}. \tag{16}$$

Note that to solve (15) at frame n, it is sufficient to know the delayed stacked vector $\underline{\mathbf{x}}(n-D)$ to construct $\mathbf{X}(n-D)$, since the signal model (5) at time frame n depends only on past values of $\mathbf{x}(n)$ with $D \geq 1$. Therefore we can state for the given signal model $J_c\left(\mathbf{c}(n)|\underline{\mathbf{x}}(n)\right) = J_c\left(\mathbf{c}(n)|\underline{\mathbf{x}}(n-D)\right)$.

By now replacing the deterministic dependencies of the cost functions (15) and (16) on $\underline{\mathbf{x}}(n)$ and $\mathbf{c}(n)$ by the available estimates, we naturally arrive at the alternating minimization procedure for each time step n:

$$1)\quad \widehat{\mathbf{c}}(n) = \arg\min_{\mathbf{c}} J_c\left(\mathbf{c}(n)|\underline{\widehat{\mathbf{x}}}(n-D)\right) \tag{17}$$

$$2)\quad \underline{\widehat{\mathbf{x}}}(n) = \arg\min_{\underline{\mathbf{x}}} J_x\left(\underline{\mathbf{x}}(n)|\widehat{\mathbf{c}}(n)\right). \tag{18}$$

The ordering of solving (17) before (18) is especially important if the coefficients $\mathbf{c}(n)$ are time-varying. Although convergence of the global cost function (14) to the global minimum is not guaranteed, it converges to local minima if (15) and (16) decrease individually. For the given signal model, (15) and (16) can be solved using the Kalman filter [27].

The resulting procedure to estimate the desired signal vector $\mathbf{s}(n)$ by (13) results in the following three steps, which are also outlined in Fig. 2:

1) Estimate the MAR coefficients $\mathbf{c}(n)$ from the noisy observed signals and delayed noise-free signals $\mathbf{x}(n')$ for $n' \in \{1, \ldots, n-D\}$, which are assumed to be deterministic and known. In practice, these signals are replaced by the estimates $\widehat{\mathbf{x}}(n')$ obtained from the second Kalman filter in Step 2.

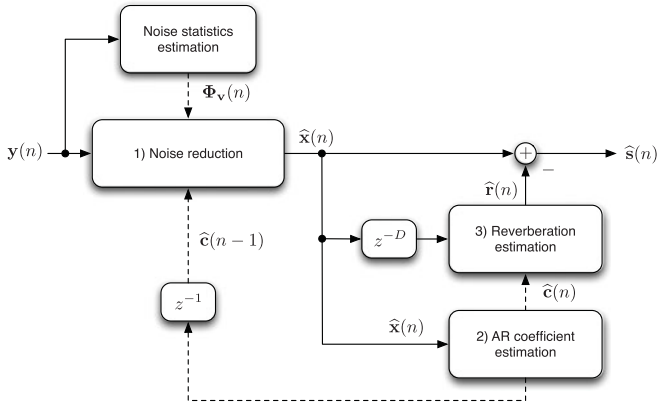

Fig. 3. State-of-the-art sequential noise reduction and dereverberation structure [16]. As the noise reduction receives delayed AR coefficients, they have to be assumed stationary or slowly time-varying.

2) Estimate the set of reverberant microphone signals $\underline{\mathbf{x}}(n)$ by exploiting the autoregressive model. This step is considered as noise reduction stage. Here, the MAR coefficients $\mathbf{c}(n)$ are assumed to be deterministic and known. In practice, the MAR coefficients are given by the estimate $\widehat{\mathbf{c}}(n)$ from Step 1. The obtained Kalman filter is similar to the Kalman smoother used in [13].
3) From the estimated MAR coefficients $\widehat{\mathbf{c}}(n)$ and from delayed versions of the noise-free signals $\widehat{\mathbf{x}}(n)$, an estimate of the late reverberation $\mathbf{r}(n)$ can be obtained. The desired signal is obtained by subtracting the estimated reverberation from the noise-free signal using (13).

The noise reduction stage requires the second-order noise statistics $\mathbf{\Phi}_{\mathbf{v}}(n)$ as indicated by the grey estimation block in Fig. 2. As there exist sophisticated methods to estimate second-order noise statistics, e.g., [28]–[30], further investigation of the noise statistics estimation is beyond the scope of this paper, and we assume the noise statistics to be known.

The proposed structure overcomes the causality problem of commonly used sequential structures for AR signal and parameter estimation [16], [21], where each estimation step requires a current estimate from each other. Such state-of-the-art sequential structures are illustrated in Fig. 3 for the given signal model, where in this case the noise reduction stage would receive delayed MAR coefficients. This would be suboptimal in the case of time-varying coefficients $\mathbf{c}(n)$.

In contrast to related state-parameter estimation methods [21], [22], our desired signal is not the state variable but a signal obtained from both state estimates (13).

A. Optimal Sequential Estimation of MAR Coefficients

Given knowledge of the delayed reverberant signals $\mathbf{x}(n)$ that are estimated as shown in Fig. 2, we derive a Kalman filter to estimate the MAR coefficients in this section.

1) Kalman filter for MAR coefficient estimation: Let us assume, we have knowledge of the past reverberant signals contained in the matrix $\mathbf{X}(n-D)$. In the following, we consider (12) and (5) as state and observation equations, respectively. Given that $\mathbf{w}(n)$ and $\mathbf{u}(n)$ are zero-mean Gaussian noise processes, which are mutually uncorrelated, we can obtain an optimal sequential estimate of the MAR coefficient vector by minimizing the trace of the error matrix

$$\mathbf{\Phi}_{\Delta\mathbf{c}}(n) = E\left\{[\mathbf{c}(n) - \widehat{\mathbf{c}}(n)][\mathbf{c}(n) - \widehat{\mathbf{c}}(n)]^{\mathrm{H}}\right\}. \tag{19}$$

The solution is obtained using the well-known Kalman filter equations [20], [27]

$$\widehat{\mathbf{\Phi}}_{\Delta\mathbf{c}}(n|n-1) = \mathbf{A}\widehat{\mathbf{\Phi}}_{\Delta\mathbf{c}}(n-1)\mathbf{A}^{\mathrm{H}} + \mathbf{\Phi}_{\mathbf{w}}(n) \tag{20}$$

$$\widehat{\mathbf{c}}(n|n-1) = \mathbf{A}\widehat{\mathbf{c}}(n-1) \tag{21}$$

$$\mathbf{e}(n) = \mathbf{y}(n) - \mathbf{X}(n-D)\widehat{\mathbf{c}}(n|n-1) \tag{22}$$

$$\mathbf{K}(n) = \widehat{\mathbf{\Phi}}_{\Delta\mathbf{c}}(n|n-1)\mathbf{X}^{\mathrm{H}}(n-D) \tag{23}$$

$$\left[\mathbf{X}(n-D)\widehat{\mathbf{\Phi}}_{\Delta\mathbf{c}}(n|n-1)\mathbf{X}^{\mathrm{H}}(n-D) + \mathbf{\Phi}_{\mathbf{u}}(n)\right]^{-1}$$

$$\widehat{\mathbf{\Phi}}_{\Delta\mathbf{c}}(n) = \left[\mathbf{I}_{L_{\mathrm{c}}} - \mathbf{K}(n)\mathbf{X}(n-D)\right]\widehat{\mathbf{\Phi}}_{\Delta\mathbf{c}}(n|n-1) \tag{24}$$

$$\widehat{\mathbf{c}}(n) = \widehat{\mathbf{c}}(n|n-1) + \mathbf{K}(n)\mathbf{e}(n), \tag{25}$$

where $\mathbf{K}(n)$ is called the Kalman gain and $\mathbf{e}(n)$ is the prediction error. Note that the prediction error is an estimate of the early speech plus noise vector $\mathbf{u}(n)$ using the predicted MAR coefficients, i.e., $\mathbf{e}(n) = \mathbf{u}(n|n-1)$.

2) Parameter estimation: The matrix $\mathbf{X}(n-D)$ containing only delayed frames of the reverberant signals $\mathbf{x}(n)$ is estimated using the second Kalman filter described in Section III-B.

We assume $\mathbf{A} = \mathbf{I}_{L_{\mathrm{c}}}$ and the covariance of the uncertainty noise $\mathbf{\Phi}_{\mathbf{w}}(n) = \phi_w(n)\mathbf{I}_{L_{\mathrm{c}}}$, where we propose to estimate the scalar variance $\phi_w(n)$ by [25]

$$\widehat{\phi}_w(n) = \frac{1}{L_{\mathrm{c}}}\|\widehat{\mathbf{c}}(n) - \widehat{\mathbf{c}}(n-1)\|_2^2 + \eta, \tag{26}$$

and η is a small positive number to model the continuous variability of the MAR coefficients if the difference between subsequent estimated coefficients is zero.

The covariance $\mathbf{\Phi}_{\mathbf{u}}(n)$ can be estimated in the maximum likelihood (ML) sense as proposed in [20] given the p.d.f. $f(\mathbf{y}(n) \,|\, \widehat{\mathbf{\Theta}}(n))$, where $\widehat{\mathbf{\Theta}}(n) = \{\widehat{\mathbf{x}}(n-L), \ldots, \widehat{\mathbf{x}}(n-1), \widehat{\mathbf{c}}(n)\}$ are the currently available parameter estimates at frame n. By assuming stationarity of $\mathbf{\Phi}_{\mathbf{u}}(n)$ within N frames, the ML estimate given the currently available information is obtained by

$$\widehat{\mathbf{\Phi}}_{\mathbf{u}}^{\mathrm{ML}}(n) = \frac{1}{N}\left(\sum_{\ell=n-N+1}^{n-1} \widehat{\mathbf{u}}(n-\ell)\widehat{\mathbf{u}}^{\mathrm{H}}(n-\ell) + \mathbf{e}(n)\mathbf{e}^{\mathrm{H}}(n)\right), \tag{27}$$

where $\widehat{\mathbf{u}}(n) = \mathbf{y}(n) - \widehat{\mathbf{X}}(n-D)\widehat{\mathbf{c}}(n)$ and $\mathbf{e}(n) = \mathbf{u}(n|n-1)$ is the predicted speech plus noise signal, since $\widehat{\mathbf{c}}(n)$ is not yet available.

In practice, the arithmetic average in (27) can be replaced by a recursive average, yielding the recursive estimate

$$\widehat{\mathbf{\Phi}}_{\mathbf{u}}(n) = \alpha\widehat{\mathbf{\Phi}}_{\mathbf{u}}^{\mathrm{pos}}(n-1) + (1-\alpha)\mathbf{e}(n)\mathbf{e}^{\mathrm{H}}(n), \tag{28}$$

where the recursive a posteriori covariance estimate, which can be computed only for the previous frame, is given by

$$\widehat{\mathbf{\Phi}}_{\mathbf{u}}^{\text{pos}}(n) = \alpha \widehat{\mathbf{\Phi}}_{\mathbf{u}}^{\text{pos}}(n-1) + (1-\alpha)\widehat{\mathbf{u}}(n)\widehat{\mathbf{u}}^{\text{H}}(n). \quad (29)$$

The recursive averaging factor $\alpha = e^{-\frac{\Delta t}{\tau}}$ depends on the exponential smoothing constant τ given in seconds, and the frame shift Δt in seconds. Since $\mathbf{u}(n)$ can be assumed stationary only within a short time period of a few frames, the recursive estimator given by (28) is preferred over the ML estimator. Furthermore, we can adjust the time constant with continuous values, whereas the arithmetic averaging length in (27) can be adjusted only in discrete time steps as N multiples of Δt.

B. Optimal Sequential Noise Reduction

Given knowledge of the current MAR coefficients $\mathbf{c}(n)$ that are estimated as shown in Fig. 2, we derive a second Kalman filter to estimate the noise-free reverberant signal vector $\underline{\mathbf{x}}(n)$ in this section.

1) Kalman filter for noise reduction: By assuming the MAR coefficients $\mathbf{c}(n)$, respectively the matrix $\mathbf{F}(n)$, as given, and by considering the stacked reverberant signal vector $\underline{\mathbf{x}}(n)$ containing the latest L frames of $\mathbf{x}(n)$ as state variable, we consider (10) and (11) as state and observation equations. Due to the assumptions on $\mathbf{s}(n)$ and (7), $\underline{\mathbf{s}}(n)$ is also a zero-mean Gaussian random variable and its covariance matrix $\mathbf{\Phi}_{\underline{\mathbf{s}}}(n) = E\{\underline{\mathbf{s}}(n)\underline{\mathbf{s}}^{\text{H}}(n)\}$ contains $\mathbf{\Phi}_{\mathbf{s}}(n)$ in the lower right corner and is zero elsewhere.

Given that $\underline{\mathbf{s}}(n)$ and $\mathbf{v}(n)$ are zero-mean Gaussian noise processes, which are mutually uncorrelated, we can obtain an optimal sequential estimate of $\underline{\mathbf{x}}(n)$ by minimizing the trace of the error matrix

$$\mathbf{\Phi}_{\Delta\mathbf{x}}(n) = E\left\{[\underline{\mathbf{x}}(n) - \widehat{\underline{\mathbf{x}}}(n)][\underline{\mathbf{x}}(n) - \widehat{\underline{\mathbf{x}}}(n)]^{\text{H}}\right\}. \quad (30)$$

The standard Kalman filtering equations to estimate the state vector $\underline{\mathbf{x}}(n)$ are given by the predictions

$$\widehat{\mathbf{\Phi}}_{\Delta\mathbf{x}}(n|n-1) = \mathbf{F}(n)\widehat{\mathbf{\Phi}}_{\Delta\mathbf{x}}(n-1)\mathbf{F}^{\text{H}}(n) + \mathbf{\Phi}_{\underline{\mathbf{s}}}(n) \quad (31)$$

$$\widehat{\underline{\mathbf{x}}}(n|n-1) = \mathbf{F}(n)\widehat{\underline{\mathbf{x}}}(n-1) \quad (32)$$

and updates

$$\mathbf{K}_{\mathbf{x}}(n) = \widehat{\mathbf{\Phi}}_{\Delta\mathbf{x}}(n|n-1)\mathbf{H}^{\text{H}} \times \left[\mathbf{H}\widehat{\mathbf{\Phi}}_{\Delta\mathbf{x}}(n|n-1)\mathbf{H}^{\text{H}} + \mathbf{\Phi}_{\mathbf{v}}(n)\right]^{-1} \quad (33)$$

$$\mathbf{e}_{\mathbf{x}}(n) = \mathbf{y}(n) - \mathbf{H}\widehat{\underline{\mathbf{x}}}(n|n-1) \quad (34)$$

$$\widehat{\mathbf{\Phi}}_{\Delta\mathbf{x}}(n) = [\mathbf{I}_{ML} - \mathbf{K}_{\mathbf{x}}(n)\mathbf{H}]\,\widehat{\mathbf{\Phi}}_{\Delta\mathbf{x}}(n|n-1), \quad (35)$$

$$\widehat{\underline{\mathbf{x}}}(n) = \widehat{\underline{\mathbf{x}}}(n|n-1) + \mathbf{K}_{\mathbf{x}}(n)\mathbf{e}_{\mathbf{x}}(n) \quad (36)$$

where $\mathbf{K}_{\mathbf{x}}(n)$ and $\mathbf{e}_{\mathbf{x}}(n)$ are the Kalman gain and the prediction error of the noise reduction Kalman filter.

The estimated noise-free reverberant signal vector at frame n is contained in the state vector and given by $\widehat{\mathbf{x}}(n) = \mathbf{H}\widehat{\underline{\mathbf{x}}}(n)$.

2) Parameter estimation: The noise covariance matrix $\mathbf{\Phi}_{\mathbf{v}}(n)$ is assumed to be known in advance in this paper. For stationary noise, it can be estimated from the microphone signals during speech absence e.g., using the methods proposed in [28]–[32].

Further, we have to estimate $\mathbf{\Phi}_{\underline{\mathbf{s}}}(n)$, i.e., the desired speech covariance matrix $\mathbf{\Phi}_{\mathbf{s}}(n)$. To reduce musical tones arising from the noise reduction procedure performed by the Kalman filter, we use a decision-directed approach [33] to estimate the current speech covariance matrix $\mathbf{\Phi}_{\mathbf{s}}(n)$, which is in this case a weighting between the a posteriori estimate $\widehat{\mathbf{\Phi}}_{\mathbf{s}}^{\text{pos}}(n) = E\{\mathbf{\Phi}_{\mathbf{s}}(n)|\widehat{\mathbf{s}}(n)\}$ at the previous frame and the a priori estimate $\widehat{\mathbf{\Phi}}_{\mathbf{s}}^{\text{pri}}(n) = E\{\mathbf{\Phi}_{\mathbf{s}}(n)|\mathbf{y}(n), \widehat{\mathbf{r}}(n)\}$ at the current frame. The decision-directed estimate is given by

$$\widehat{\mathbf{\Phi}}_{\mathbf{s}}(n) = \gamma\widehat{\mathbf{\Phi}}_{\mathbf{s}}^{\text{pos}}(n-1) + (1-\gamma)\widehat{\mathbf{\Phi}}_{\mathbf{s}}^{\text{pri}}(n), \quad (37)$$

where γ is the decision-directed weighting parameter. To reduce musical tones, the parameter is typically chosen to put more weight on the previous a posteriori estimate.

The recursive a posteriori ML estimate is obtained by

$$\widehat{\mathbf{\Phi}}_{\mathbf{s}}^{\text{pos}}(n) = \alpha\widehat{\mathbf{\Phi}}_{\mathbf{s}}^{\text{pos}}(n-1) + (1-\alpha)\widehat{\mathbf{s}}(n)\widehat{\mathbf{s}}^{\text{H}}(n), \quad (38)$$

where $\alpha = e^{-\frac{\Delta t}{\tau}}$ is a recursive averaging factor.

To obtain the a priori estimate $\widehat{\mathbf{\Phi}}_{\mathbf{s}}^{\text{pri}}(n)$, we derive a multichannel Wiener filter (MWF), i.e.,

$$\mathbf{W}_{\text{MWF}}(n) = \arg\min_{\mathbf{W}} E\{\left\|\mathbf{s}(n) - \mathbf{W}^{\text{H}}\mathbf{y}(n)\right\|_2^2\}. \quad (39)$$

By inserting (10) in (11), we can rewrite the observed signal vector as

$$\mathbf{y}(n) = \mathbf{s}(n) + \underbrace{\mathbf{H}\mathbf{F}(n)\underline{\mathbf{x}}(n-1)}_{\mathbf{r}(n)} + \mathbf{v}(n), \quad (40)$$

where all three components are mutually uncorrelated. Note that estimates of all components of the late reverberation $\mathbf{r}(n)$ are already available at this point. An instantaneous estimate of $\mathbf{\Phi}_{\mathbf{s}}(n)$ using an MMSE estimator given the currently available information is then obtained by

$$\widehat{\mathbf{\Phi}}_{\mathbf{s}}^{\text{pri}}(n) = \mathbf{W}_{\text{MWF}}^{\text{H}}(n)\,\mathbf{y}(n)\mathbf{y}^{\text{H}}(n)\mathbf{W}_{\text{MWF}}(n). \quad (41)$$

The MWF filter matrix is given by

$$\mathbf{W}_{\text{MWF}}(n) = \mathbf{\Phi}_{\mathbf{y}}^{-1}(n)\,[\mathbf{\Phi}_{\mathbf{y}}(n) - \mathbf{\Phi}_{\mathbf{r}}(n) - \mathbf{\Phi}_{\mathbf{v}}(n)], \quad (42)$$

where $\mathbf{\Phi}_{\mathbf{y}}(n)$ and $\mathbf{\Phi}_{\mathbf{r}}(n)$ are estimated using recursive averaging from the signals $\mathbf{y}(n)$ and $\widehat{\mathbf{r}}(n)$, similar to (38).

C. Algorithm Overview

The complete algorithm is outlined in Algorithm 1. The initialization of the Kalman filters was found to be uncritical. Although the initial convergence phase could be improved by using better initial estimates of the state variables, the algorithm converged within a few seconds and stayed stable in practice when using the proposed initialization.

The proposed algorithm is suitable for real-time processing applications requiring low algorithmic delay. As a matter of fact, the delay depends only on the time-frequency analysis and synthesis stages. However, the computational complexity, which depends on the number of microphones M, the filter length L per frequency, and the number of frequency bands, can be high. The complexity of the first and second Kalman filters rises quadratically with the length of the state vectors,

Algorithm 1: Proposed algorithm per frequency band k.

Initialize: $\widehat{\mathbf{c}}(0)=\mathbf{0}$, $\underline{\widehat{\mathbf{x}}}(0)=\mathbf{0}$, $\widehat{\mathbf{\Phi}}_{\Delta\mathbf{c}}(n)=\mathbf{I}_{L_\mathrm{c}}$, $\widehat{\mathbf{\Phi}}_{\Delta\mathbf{x}}(n)=\mathbf{I}_{ML}$
for each n **do**
  Estimate the noise covariance $\mathbf{\Phi}_{\mathbf{v}}(n)$, e.g. using [29]
  $\mathbf{X}(n-D)\leftarrow\underline{\widehat{\mathbf{x}}}(n-1)$
  Compute $\widehat{\mathbf{\Phi}}_{\mathbf{w}}(n)=\phi_w(n)\mathbf{I}_{L_\mathrm{c}}$ using (26)
  Obtain $\widehat{\mathbf{c}}(n)$ by calculating (20)-(22), (27), (23)-(25)
  $\mathbf{F}(n)\leftarrow\widehat{\mathbf{c}}(n)$
  $\mathbf{\Phi}_{\underline{\mathbf{s}}}(n)\leftarrow\widehat{\mathbf{\Phi}}_{\mathbf{s}}(n)$ using (37)
  Obtain $\underline{\widehat{\mathbf{x}}}(n)$ by calculating (32)-(35)
  Estimate the desired signal by (13)
end for

$M^2(L-D+1)$ and ML respectively. However, complexity can be reduced by exploiting the sparse or block-diagonal structure of some matrices [34], and some matrix multiplications are simple index shifts operations.

IV. Reduction Control

In some applications it is beneficial to have independent control over the reduction of the undesired sound components such as reverberation and noise. In many cases, the subjective sound quality can be significantly improved by controlling the amount of reduction to mask artifacts and mitigate speech distortion [35]–[37]. In communication scenarios, it is often preferred to maintain a small amount of residual noise; otherwise the listener might have the impression that the connection is lost (also known as *comfort noise*) [38]. For dereverberation, it might be subjectively preferable to maintain some residual late reverberation, as it can sound unnatural if the early reflections are preserved while the late reverberation is strongly reduced. In this section, we show how to compute an alternative output signal $\mathbf{z}(n)$, where we have control over the reduction of reverberation and noise.

The desired controlled output signal is given by

$$\mathbf{z}(n)=\mathbf{s}(n)+\beta_r\mathbf{r}(n)+\beta_v\mathbf{v}(n), \tag{43}$$

where β_r and β_v are attenuation factors of the reverberation and noise. By re-arranging (43) using (5) and replacing unknown variables by the available estimates, we can compute the desired controlled output signal vector by

$$\widehat{\mathbf{z}}(n)=\beta_v\mathbf{y}(n)+(1-\beta_v)\widehat{\mathbf{x}}(n)-(1-\beta_r)\widehat{\mathbf{r}}(n). \tag{44}$$

Note that for $\beta_v=\beta_r=0$, the output $\widehat{\mathbf{z}}(n)$ is identical to the early speech estimate $\widehat{\mathbf{s}}(n)$, and for $\beta_v=\beta_r=1$, the output $\widehat{\mathbf{z}}(n)$ is equal to $\mathbf{y}(n)$.

Typically, speech enhancement algorithms have a trade-off between the amount of interference reduction and artifacts such as speech distortion or musical tones. To reduce audible artifacts in periods where the MAR coefficient estimation Kalman filter is adapting fast and exhibits a high prediction error, we use the estimated error covariance matrix $\widehat{\mathbf{\Phi}}_{\Delta\mathbf{c}}(n)$ given by (24) to adaptively control the reverberation attenuation factor β_r. If the error of the Kalman filter is high, we like the attenuation factor β_r to be close to one. We propose to compute the reverberation attenuation factor at time frame n by the heuristically chosen mapping function

$$\beta_r(n)=\max\left(\frac{1}{1+\mu_r L_\mathrm{c}\,\mathrm{tr}\left\{\widehat{\mathbf{\Phi}}_{\Delta\mathbf{c}}(n)\right\}^{-1}},\ \beta_{r,\min}\right), \tag{45}$$

where the fixed lower bound $\beta_{r,\min}$ limits the allowed reverberation attenuation, and the factor μ_r controls the attenuation depending on the Kalman error.

The structure of the proposed system with reduction control is illustrated in Fig. 4. The noise estimation block is omitted here as it can be also integrated in the noise reduction block.

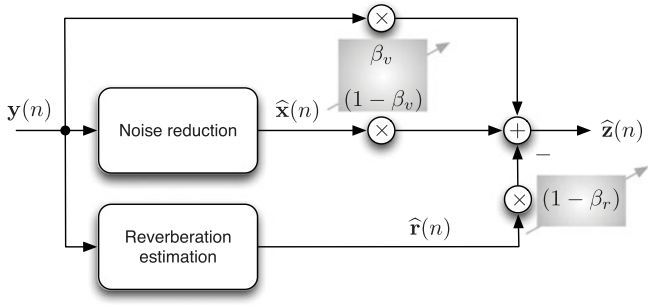

Fig. 4. Proposed structure to control the amount of noise reduction β_v and reverberation reduction β_r.

V. Evaluation

In this section, we evaluate the proposed system using the experimental setup described in Section V-A by comparing to the two reference methods reviewed in Section V-B. The results are shown in Section V-C.

A. Experimental Setup

The reverberant signals were generated by convolving room impulse responses (RIRs) with anechoic speech signals from [39]. We used two different kinds of RIRs: measured RIRs in an acoustic lab with variable acoustics at Bar-Ilan University, Israel, or simulated RIRs using the image method [40] for moving sources. In the case of moving sources, the simulated RIRs facilitate the evaluation, as in this case, it is possible to additionally generate RIRs containing only direct sound and early reflections to obtain the target signal for evaluation.

In simulated and measured cases, we used a linear microphone array with up to $M=4$ omnidirectional microphones with inter-microphone spacings $\{11, 7, 14\}$ cm. Note that in all experiments except in Section V-C1, only 2 microphones with spacing 11 cm are used. Either stationary pink noise or babble noise, a recording in a cafeteria from [41], was added to the reverberant signals with a certain input signal-to-noise ratio (iSNR). We used a sampling frequency of 16 kHz and the STFT parameters were a square-root Hann window of 32 ms length, 50% overlap and an FFT length of

1024 samples. The delay preserving early reflections was set to $D = 2$. The recursive averaging factor was $\alpha = e^{-\frac{\Delta t}{\tau}}$ with a time constant of $\tau = 25$ ms, where $\Delta t = 16$ ms is the frame shift. The decision-directed weighting factor was $\gamma = 0.98$ and we chose $\eta = 10^{-4}$. We present results without reduction control (RC), i.e., $\beta_v = \beta_r = 0$, and with RC using different settings for β_v and $\beta_{r,\min}$, where we chose $\mu_r = -10$ dB in (45). The noise covariance matrix was computed as long-term average over non-speech segments to exclude effects of noise estimation errors. In practice, similar noise covariance estimates can be obtained using online estimation methods [30], [31].

For evaluation, the target signals were generated as the direct speech signal with early reflections up to 32 ms after the direct sound peak (corresponds to a delay of $D = 2$ frames). The processed signals are evaluated in terms of the *cepstral distance* (CD) [42], the *perceptual evaluation of speech quality* (PESQ) [43], the *frequency-weighted segmental signal-to-interference ratio* (fwSSIR) [44], where reverberation and noise are considered as interference, and the normalized *speech-to-reverberation modulation ratio* (SRMR) [45]. These measures have been shown to yield reasonable correlation with the perceived amount of reverberation and overall quality in the context of dereverberation [3], [46]. The CD reflects more the overall quality and is sensitive to speech distortion, while PESQ, fwSSIR, and SRMR are more sensitive to reverberation/interference reduction. Note that for the CD, lower values are better, while for PESQ, fwSSIR, and SRMR higher values are better. We present only results for the first microphone as all other microphones behave similarly.

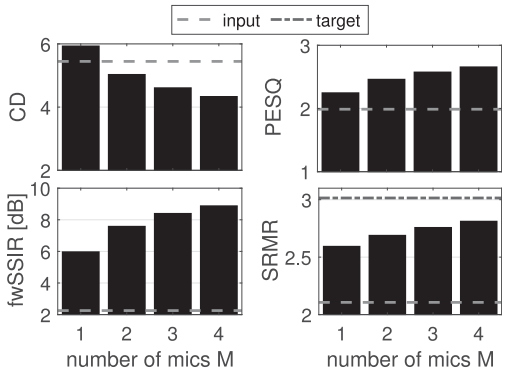

Fig. 5. Objective measures for varying microphone number using measured RIRs. iSNR $= 10$ dB, $L = 15$, no reduction control ($\beta_v = \beta_r = 0$).

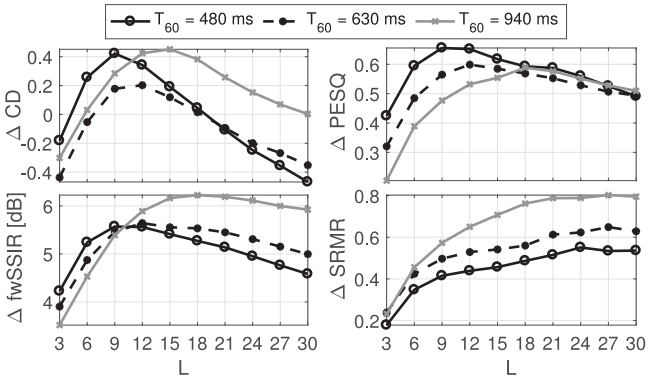

Fig. 6. Objective measures for varying filter length L. Parameters: iSNR $=$ 15 dB, $M = 2$, no reduction control ($\beta_v = \beta_r = 0$).

B. Reference Methods

To show the effectiveness and performance of the proposed method (*dual-Kalman*), we compare it to the following two methods:

- *single-Kalman*: A single Kalman filter to estimate the MAR coefficients without noise reduction as proposed in [20]. The original algorithm assumes no additive noise. However, it can be still used to estimate the MAR coefficients from the noisy signal and then obtain a dereverberated, but still noisy filtered signal as output.
- *MAP-EM*: In the method proposed in [16], the MAR coefficients are estimated using a Bayesian approach based on maximum a posteriori (MAP) estimation and the noise-free desired signal is then estimated using an EM algorithm. The algorithm is online, but the EM procedure requires about 20 iterations per frame to converge.

C. Results

1) Dependence on number of microphones: We investigated the performance of the proposed algorithm depending on the number of microphones M. The desired signal with a total length of 34 s consisted of two non-concurrent speakers at different positions: During the first 15 s the first speaker was active, while after 15 s, the second speaker was active. Each speaker signal was convolved with measured RIRs at different positions with a $T_{60} = 630$ ms. Stationary pink noise was added to the reverberant signals with iSNR $= 15$ dB.

Fig. 5 shows CD, PESQ, fwSSIR and SRMR for a varying number of microphones M. The measures for the noisy reverberant input signal are indicated as light grey dashed line, and the SRMR of the target signal, i.e., the early speech, is indicated as dark grey dash-dotted line. For $M = 1$, the CD is larger than for the input signal, which indicates an overall quality deterioration, whereas PESQ, fwSSIR and SRMR still improve over the input, i.e., reverberation and noise are reduced. The performance in terms of all measures increases by increasing the number of microphones.

2) Dependence on filter length: The effect of the filter length L was investigated using measured RIRs with different reverberation times. As in the first experiment, two non-concurrent speakers were active at different positions, and stationary pink noise was added with iSNR $= 15$ dB. Fig. 6 shows the improvement of the objective measures compared to the unprocessed microphone signal. Positive values indicate an improvement for all relative measures, where Δ denotes the improvement. Considering the given STFT parameters, the reverberation times $T_{60} = \{480, 630, 940\}$ ms correspond to filter lengths $L = \{30, 39, 58\}$ frames. We can observe that the best CD, PESQ and fwSSIR values depend on the reverberation time, but the optimal values are obtained at around 25% of the

TABLE I
OBJECTIVE MEASURES FOR VARYING iSNRs (STATIONARY NOISE) USING MEASURED RIRs

iSNR [dB]	0	5	15	25	35
	Δ CD				
single-Kalman [20]	−0.03	0.04	0.39	0.76	1.01
MAP-EM [16]	0.10	0.20	0.34	0.07	−1.31
dual-Kalman	0.00	0.02	0.41	0.83	1.02
dual-Kalman RC	**0.42**	**0.56**	**0.77**	**0.96**	**1.04**
	Δ PESQ				
single-Kalman [20]	0.05	0.12	0.20	0.33	0.40
MAP-EM [16]	0.15	0.21	0.20	0.24	0.25
dual-Kalman	**0.54**	**0.53**	**0.47**	**0.51**	**0.50**
dual-Kalman RC	0.34	0.40	0.34	0.39	0.40
	Δ fwSSIR [dB]				
single-Kalman [20]	0.63	1.13	2.78	4.35	5.48
MAP-EM [16]	1.52	2.32	3.31	3.75	3.73
dual-Kalman	**6.70**	**6.62**	**5.65**	**5.79**	**6.20**
dual-Kalman RC	2.85	3.95	4.56	5.02	5.30
	Δ SRMR				
single-Kalman [20]	0.21	0.29	0.38	0.42	0.44
MAP-EM [16]	0.46	0.44	0.31	0.28	0.27
dual-Kalman	**1.37**	**0.98**	**0.58**	**0.50**	**0.47**
dual-Kalman RC	0.89	0.75	0.47	0.45	0.41

$M = 2$, $L = 12$, $\beta_v = -10$ dB, $\beta_{r,\min} = -15$ dB

TABLE II
OBJECTIVE MEASURES FOR VARYING iSNRs (BABBLE NOISE) USING MEASURED RIRs

iSNR [dB]	0	5	15	25	35
	Δ CD				
single-Kalman [20]	−0.04	0.08	0.45	0.78	0.93
MAP-EM [16]	−0.18	−0.01	0.38	0.41	−1.01
dual-Kalman	−1.61	−1.20	−0.26	0.56	**0.94**
dual-Kalman RC	**−0.03**	**0.10**	**0.49**	**0.79**	**0.94**
	Δ PESQ				
single-Kalman [20]	0.03	0.16	0.21	0.29	0.37
MAP-EM [16]	−0.04	0.05	0.19	0.23	0.24
dual-Kalman	−0.13	0.14	**0.31**	**0.43**	**0.46**
dual-Kalman RC	**0.18**	**0.20**	0.28	0.34	0.37
	Δ fwSSIR [dB]				
single-Kalman [20]	0.71	1.35	3.01	4.41	5.29
MAP-EM [16]	1.10	1.93	3.36	3.83	3.36
dual-Kalman	**3.45**	**4.04**	**4.34**	**5.06**	**5.83**
dual-Kalman RC	2.03	3.03	3.95	4.55	5.03
	Δ SRMR				
single-Kalman [20]	0.22	0.31	0.39	0.40	0.42
MAP-EM [16]	0.30	0.38	0.32	0.27	0.25
dual-Kalman	**1.21**	**1.08**	**0.55**	**0.45**	**0.46**
dual-Kalman RC	0.72	0.77	0.48	0.41	0.41

$M = 2$, $L = 12$, $\beta_v = -10$ dB, $\beta_{r,\min} = -15$ dB

corresponding length of the reverberation time. In contrast, the SRMR monotonously grows with increasing L. It is worthwhile to note that the reverberation reduction becomes more aggressive with increasing L. If the reduction is too aggressive by choosing L too large, the desired speech is distorted as the Δ CD indicates with negative values.

3) Comparison with state-of-the-art methods: The proposed algorithm and the two reference algorithms were evaluated for two noise types in varying iSNRs. As in the first two experiments, the desired signal consisted of two non-concurrent speakers at different positions with a total length of 34 s using measured RIRs with $T_{60} = 630$ ms. Either stationary pink noise or recorded babble noise was added with varying iSNRs. Tables I and II show the improvement of the objective measures compared to the unprocessed microphone signal in stationary pink noise and in babble noise, respectively. Note that although the babble noise is not short-term stationary, we used a stationary long-term estimate of the noise covariance matrix, which is realistic to obtain as an estimate in practice.

It can be observed that the proposed algorithm either without or with RC outperforms both competing algorithms in all conditions. The RC provides a trade-off between interference reduction and desired signal distortion. The CD as an indicator for speech distortion is consistently better with RC, whereas the other measures, which majorly reflect the amount of interference reduction, consistently achieve slightly higher results without RC in stationary noise. In babble noise, the dual-Kalman with RC yields higher PESQ at low iSNRs than without RC. This indicates that the RC can help to improve the quality by masking artifacts in challenging iSNR conditions and in the presence of noise covariance estimation errors. In high iSNR conditions, the performance of the dual-Kalman becomes similar to the performance of the single-Kalman as expected.

4) Tracking of moving speakers: A moving source was simulated using simulated RIRs in a shoebox room with $T_{60} = 500$ ms based on the image method [40], [47]: The desired source was first at position A, and during the time interval [8,13] s it moved continuously from position A to B, where it stayed then for the rest of the time. Position A and B were 2 m apart. Fig. 7 shows the segmental improvement of CD, PESQ, SIR and SRMR for this dynamic scenario. The segmental measures were computed from 50% overlapping segments of 2 s. In this experiment, the target signal for evaluation is generated by simulating the wall reflections only up to the second order.

We observe that all measures decrease during the movement, while after the speaker has reached position B, the measures reach high improvements again. The convergence of all methods behaves similar, while the dual-Kalman without and with RC perform best. During the moving time period, the MAP-EM yields sometimes higher fwSSIR and SRMR, but at the price of much worse CD and PESQ. The reduction control improves the CD, such that the CD improvement always stays positive, which indicates that the RC can reduce speech distortion and artifacts. It is worthwhile to note that even if the reverberation reduction can become less effective during movement of the speech source, the dual-Kalman algorithm did not become unstable, the improvements of PESQ, fwSSIR and SRMR were always positive, and the Δ CD was always positive by using the

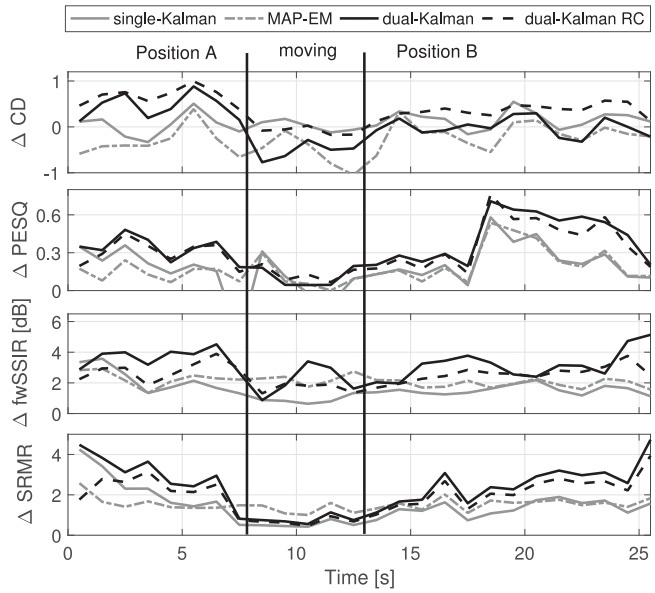

Fig. 7. Short-term measures for a moving source between 8–13 s in a simulated shoebox room with $T_{60} = 500$ ms. iSNR = 15 dB, $M = 2$, $L = 15$, $\beta_v = -10$ dB, $\beta_{r,\text{min}} = -15$ dB.

RC. This was also verified using real recordings with moving speakers.[1]

5) Evaluation of reduction control: In this section, we evaluate the performance of the RC in terms of the reduction of noise and reverberation by the proposed system. In the appendix, it is shown how the residual noise and reverberation signals after processing with RC $\mathbf{z}_\mathbf{v}(n)$ and $\mathbf{z}_\mathbf{r}(n)$ for the proposed dual-Kalman filter system can be computed. The noise reduction and reverberation reduction measures are then computed by

$$\text{NR}(n) = \frac{\sum_k \|\mathbf{z}_\mathbf{v}(k,n)\|_2^2}{\sum_k \|\mathbf{v}(k,n)\|_2^2} \tag{46}$$

$$\text{RR}(n) = \frac{\sum_k \|\mathbf{z}_\mathbf{r}(k,n)\|_2^2}{\sum_k \|\mathbf{r}(k,n)\|_2^2}. \tag{47}$$

In this experiment, we simulated a scenario with a single speaker at a stationary position using measured RIRs in the acoustic lab with $T_{60} = 630$ ms. In Fig. 8, five different settings for the attenuation factors are shown: No reduction control ($\beta_v = \beta_{r,\text{min}} = 0$), a moderate setting with $\beta_v = \beta_{r,\text{min}} = -7$ dB, reducing either only reverberation or only noise, and a stronger attenuation setting with $\beta_v = \beta_{r,\text{min}} = -15$ dB. We can observe that the noise reduction measure yields the desired reduction levels only during speech pauses. The reverberation reduction measure surprisingly shows that a high reduction is only achieved during speech absence. This does not mean that the residual reverberation is more audible during speech presence, as the direct sound of the speech perceptually masks the residual reverberation. During the first 5 seconds, we can observe the reduced reverberation reduction caused by the adaptive

[1]Examples online available at www.audiolabs-erlangen.de/resources/2017-dualkalman.

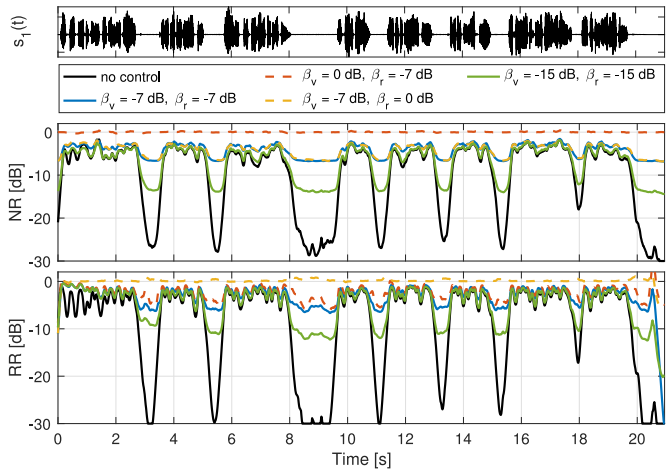

Fig. 8. Noise reduction and reverberation reduction for varying control parameters β_v and $\beta_{r,\text{min}}$. iSNR = 15 dB, $M = 2$, $L = 12$. The desired speech signal at the first microphone $s_1(t)$ indicates the speech activity.

reverberation attenuation factor (45), as the Kalman filter error is high during the initial convergence.

VI. Conclusion

We presented an alternating minimization algorithm based on two interacting Kalman filters to estimate multichannel autoregressive parameters and the reverberant signal to reduce noise and reverberation from each microphone signal. The proposed solution using recursive Kalman filters is suitable for online processing applications. We showed the effectiveness and superior performance to similar online methods in various experiments. In addition, we proposed a method to control the reduction of noise and reverberation independently to mask possible artifacts and to adjust the output signal to perceptual requirements.

Appendix
Computation of Residual Noise and Reverberation

To compute the residual power of noise and reverberation at the output of the proposed system, we need to propagate these signals through the system. By propagating only the noise at the input $\mathbf{v}(n)$ through the dual-Kalman system instead of $\mathbf{y}(n)$ as in Fig. 2, we obtain the output $\widehat{\mathbf{s}}_\mathbf{v}(n)$, which is the residual noise contained in $\widehat{\mathbf{s}}(n)$. By also taking the RC into account, the residual contribution of the noise $\mathbf{v}(n)$ in the output signal $\mathbf{z}(n)$ is $\mathbf{z}_\mathbf{v}(n)$. By inspecting (32), (34) and (36), the noise is fed through the noise reduction Kalman filter by the equation

$$\begin{aligned}\widetilde{\underline{\mathbf{v}}}(n) &= \mathbf{F}(n)\widetilde{\underline{\mathbf{v}}}(n-1) + \mathbf{K}_\mathbf{x}(n)\left[\mathbf{v}(n) - \mathbf{H}\mathbf{F}(n)\widetilde{\underline{\mathbf{v}}}(n-1)\right] \\ &= \mathbf{K}_\mathbf{x}(n)\mathbf{v}(n) + \left[\mathbf{F}(n) - \mathbf{K}_\mathbf{x}(n)\mathbf{H}\mathbf{F}(n)\right]\widetilde{\underline{\mathbf{v}}}(n-1),\end{aligned} \tag{48}$$

where $\widetilde{\underline{\mathbf{v}}}(n)$ is the residual noise vector of length ML, similarly defined as (6), after noise reduction. The output after the dereverberation step is obtained by

$$\widehat{\mathbf{s}}_\mathbf{v}(n) = \underbrace{\mathbf{H}\widetilde{\underline{\mathbf{v}}}(n)}_{\widetilde{\mathbf{v}}(n)} - \underbrace{\mathbf{H}\mathbf{F}(n)\widetilde{\underline{\mathbf{v}}}(n-1)}_{\widetilde{\mathbf{v}}(n|n-1)}. \tag{49}$$

With RC, the residual noise is given in analogy to (44) by

$$\mathbf{z_v}(n) = \beta_v \mathbf{v}(n) + (1-\beta_v)\widetilde{\mathbf{v}}(n) - (1-\beta_r)\widetilde{\mathbf{v}}(n|n-1). \tag{50}$$

The calculation of the residual reverberation $\mathbf{z_r}(n)$ is more difficult. To exclude the noise from this calculation, we first feed the oracle reverberant noise-free signal vector $\mathbf{x}(n)$ through the noise reduction stage:

$$\begin{aligned}\widetilde{\underline{\mathbf{x}}}(n) &= \mathbf{F}(n)\widetilde{\underline{\mathbf{x}}}(n-1) + \mathbf{K_x}(n)\left[\mathbf{x}(n) - \mathbf{HF}(n)\widetilde{\underline{\mathbf{x}}}(n-1)\right] \\ &= \mathbf{K_x}(n)\mathbf{x}(n) + \left[\mathbf{F}(n) - \mathbf{K_x}(n)\mathbf{HF}(n)\right]\widetilde{\underline{\mathbf{x}}}(n-1),\end{aligned} \tag{51}$$

where $\widetilde{\mathbf{x}}(n) = \mathbf{H}\widetilde{\underline{\mathbf{x}}}(n)$ is the output of the noise-free signal vector $\mathbf{x}(n)$ after the noise reduction stage. According to (44) the output of the noise-free signal vector after dereverberation and RC is obtained by

$$\mathbf{z_x}(n) = \beta_v \mathbf{x}(n) + (1-\beta_v)\widetilde{\mathbf{x}}(n) - (1-\beta_r)\widetilde{\mathbf{r}}(n) \tag{52}$$

where $\widetilde{\mathbf{r}}(n) = \widetilde{\mathbf{X}}(n-D)\widehat{\mathbf{c}}(n)$ and the matrix $\widetilde{\mathbf{X}}(n)$ is obtained using $\widetilde{\mathbf{x}}(n)$ in analogy to (3).

Now let us assume that the noise-free signal vector after the noise reduction $\widetilde{\mathbf{x}}(n)$ and the noise-free output signal vector after dereverberation and RC $\mathbf{z_x}(n)$ are composed as

$$\widetilde{\mathbf{x}}(n) \approx \mathbf{s}(n) + \mathbf{r}(n) \tag{53}$$

$$\mathbf{z_x}(n) \approx \mathbf{s}(n) + \mathbf{z_r}(n), \tag{54}$$

where $\mathbf{z_r}(n)$ denotes the residual reverberation in the RC output $\mathbf{z}(n)$. By using (53) and knowledge of the oracle desired signal vector $\mathbf{s}(n)$, we can compute the reverberation signal

$$\mathbf{r}(n) = \widetilde{\mathbf{x}}(n) - \mathbf{s}(n). \tag{55}$$

From the difference of (53) and (54) and using (55), we can obtain the residual reverberation signals as

$$\mathbf{z_r}(n) = \mathbf{r}(n) - \underbrace{\left[\widetilde{\mathbf{x}}(n) - \mathbf{z_x}(n)\right]}_{\mathbf{r}(n)-\mathbf{z_r}(n)}. \tag{56}$$

Now we can analyze the power of residual noise and reverberation at the output and compare it to their respective power at the input.

ACKNOWLEDGMENT

The authors would like to thank Dr. M. Togami for the helpful discussion on the implementation of the MAP-EM method that was used for comparison.

REFERENCES

[1] A. K. Nábělek and D. Mason, "Effect of noise and reverberation on binaural and monaural word identification by subjects with various audiograms," *J. Speech Hearing Res.*, vol. 24, pp. 375–383, 1981.

[2] T. Yoshioka *et al.*, "Making machines understand us in reverberant rooms: Robustness against reverberation for automatic speech recognition," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 114–126, Nov. 2012.

[3] K. Kinoshita *et al.*, "A summary of the REVERB challenge: state-of-the-art and remaining challenges in reverberant speech processing research," *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 1, p. 7, Jan 2016.

[4] O. Schwartz, S. Gannot, and E. Habets, "Multi-microphone speech dereverberation and noise reduction using relative early transfer functions," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 2, pp. 240–251, Jan. 2015.

[5] S. Braun and E. A. P. Habets, "A multichannel diffuse power estimator for dereverberation in the presence of multiple sources," *EURASIP J. Audio, Speech, Music Process.*, vol. 2015, no. 1, pp. 1–14, Dec. 2015.

[6] B. Schwartz, S. Gannot, and E. Habets, "Online speech dereverberation using Kalman filter and EM algorithm," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 2, pp. 394–406, Feb. 2015.

[7] D. Schmid, G. Enzner, S. Malik, D. Kolossa, and R. Martin, "Variational Bayesian inference for multichannel dereverberation and noise reduction," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 8, pp. 1320–1335, Aug. 2014.

[8] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*. London, U.K.: Springer, 2010.

[9] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 2, pp. 145–152, Feb. 1988.

[10] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and J. Biing-Hwang, "Speech dereverberation based on variance-normalized delayed linear prediction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 7, pp. 1717–1731, Sep. 2010.

[11] T. Yoshioka, T. Nakatani, and M. Miyoshi, "Integrated speech enhancement method using noise suppression and dereverberation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 2, pp. 231–246, Feb. 2009.

[12] M. Togami, Y. Kawaguchi, R. Takeda, Y. Obuchi, and N. Nukaga, "Optimized speech dereverberation from probabilistic perspective for time varying acoustic transfer function," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 7, pp. 1369–1380, Jul. 2013.

[13] M. Togami and Y. Kawaguchi, "Noise robust speech dereverberation with Kalman smoother," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2013, pp. 7447–7451.

[14] T. Yoshioka and T. Nakatani, "Generalization of multi-channel linear prediction methods for blind MIMO impulse response shortening," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 10, pp. 2707–2720, Dec. 2012.

[15] T. Yoshioka and T. Nakatani, "Dereverberation for reverberation-robust microphone arrays," in *Proc. Eur. Signal Process. Conf.*, Sep. 2013, pp. 1–5.

[16] M. Togami, "Multichannel online speech dereverberation under noisy environments," in *Proc. Eur. Signal Process. Conf.*, Nice, France, Sep. 2015, pp. 1078–1082.

[17] A. Jukic, Z. Wang, T. van Waterschoot, T. Gerkmann, and S. Doclo, "Constrained multi-channel linear prediction for adaptive speech dereverberation," in *Proc. Int. Workshop Acoust. Signal Enhancement*, Xi'an, China, Sep. 2016, pp. 1–5.

[18] T. Dietzen, A. Spriet, W. Tirry, S. Doclo, M. Moonen, and T. van Waterschoot, "Partitioned block frequency domain Kalman filter for multi-channel linear prediction based blind speech dereverberation," in *Proc. Int. Workshop Acoust. Signal Enhancement*, Xi'an, China, Sep. 2016, pp. 1–5.

[19] A. Jukic, T. van Waterschoot, and S. Doclo, "Adaptive speech dereverberation using constrained sparse multichannel linear prediction," *IEEE Signal Process. Lett.*, vol. 24, no. 1, pp. 101–105, Jan. 2017.

[20] S. Braun and E. A. P. Habets, "Online dereverberation for dynamic scenarios using a Kalman filter with an autoregressive models," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1741–1745, Dec. 2016.

[21] S. Gannot, D. Burshtein, and E. Weinstein, "Iterative and sequential Kalman filter-based speech enhancement algorithms," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 4, pp. 373–385, Jul. 1998.

[22] D. Labarre, E. Grivel, Y. Berthoumieu, E. Todini, and M. Najim, "Consistent estimation of autoregressive parameters from noisy observations based on two interacting Kalman filters," *Signal Process.*, vol. 86, no. 10, pp. 2863–2876, 2006.

[23] T. Esch and P. Vary, "Speech enhancement using a modified Kalman filter based on complex linear prediction and supergaussian priors," in *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Process.*, Mar. 2008, pp. 4877–4880.

[24] J. Erkelens and R. Heusdens, "Correlation-based and model-based blind single-channel late-reverberation suppression in noisy time-varying acoustical environments," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 7, pp. 1746–1765, Sep. 2010.

[25] G. Enzner and P. Vary, "Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones," *Signal Process.*, vol. 86, no. 6, pp. 1140–1156, 2006.

[26] U. Niesen, D. Shah, and G. W. Wornell, "Adaptive alternating minimization algorithms," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1423–1429, Mar. 2009.

[27] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME J. Basic Eng.*, vol. 82, no. Series D, pp. 35–45, 1960.
[28] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, Jul. 2001.
[29] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 4, pp. 1383–1393, May 2012.
[30] M. Taseska and E. A. P. Habets, "MMSE-based blind source extraction in diffuse noise fields using a complex coherence-based *a priori* SAP estimator," in *Proc. Int. Workshop Acoust. Signal Enhancement*, Sep. 2012, pp. 1–4.
[31] M. Souden, J. Chen, J. Benesty, and S. Affes, "An integrated solution for online multichannel noise tracking and reduction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 7, pp. 2159–2169, Sep. 2011.
[32] R. C. Hendriks and T. Gerkmann, "Noise correlation matrix estimation for multi-microphone speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 1, pp. 223–233, Jan. 2012.
[33] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.
[34] T. Dietzen, S. Doclo, A. Spriet, W. Tirry, M. Moonen, and T. van Waterschoot, "Low complexity Kalman filter for multi-channel linear prediction based blind speech dereverberation," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, New Paltz, NY, USA, Oct. 2017, pp. 1–5.
[35] Y. H. J. Chen, J. Benesty, and S. Doclo, "New insights into the noise reduction Wiener filters," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1218–1234, Jul. 2006.
[36] T. J. Klasen, T. V. den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1579–1585, Apr. 2007.
[37] S. Braun, K. Kowalczyk, and E. A. P. Habets, "Residual noise control using a parametric multichannel Wiener filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Brisbane, Australia, Apr. 2015, pp. 360–364.
[38] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*. Hoboken, NJ, USA: Wiley, 2004.
[39] E. B. Union, "Sound quality assessment material recordings for subjective tests," 1998. [Online]. Available: http://tech.ebu.ch/publications/sqamcd
[40] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.
[41] J. Thiemann, N. Ito, and E. Vincent, "Diverse Environments Multichannel Acoustic Noise Database (DEMAND)," Jun. 2013. [Online]. Available: http://parole.loria.fr/DEMAND/
[42] N. Kitawaki, H. Nagabuchi, and K. Itoh, "Objective quality evaluation for low bit-rate speech coding systems," *IEEE J. Sel. Areas Commun.*, vol. 6, no. 2, pp. 262–273, Feb. 1988.
[43] ITU-T, *Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs*, International Telecommunications Union (ITU-T) Recommendation P.862, Feb. 2001.
[44] P. C. Loizou, *Speech Enhancement Theory and Practice*. New York, NY, USA: Taylor & Francis, 2007.
[45] J. F. Santos, M. Senoussaoui, and T. H. Falk, "An updated objective intelligibility estimation metric for normal hearing listeners under noise and reverberation," in *Proc. Int. Workshop Acoust. Signal Enhancement*, Antibes, France, Sep. 2014, pp. 55–59.
[46] S. Goetze *et al.*, "A study on speech quality and speech intelligibility measures for quality assessment of single-channel dereverberation algorithms," in *Proc. Int. Workshop Acoust. Signal Enhancement*, Sep. 2014, pp. 233–237.
[47] 2017. [Online]. Available: https://github.com/ehabets/Signal-Generator

Sebastian Braun received the M.Sc. degree in electrical engineering and sound engineering from the University of Music and Dramatic Arts Graz, Graz, Austria, and the Technical University Graz, Graz, Austria, in 2012. He then joined the International Audio Laboratories Erlangen (a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg and Fraunhofer IIS) as a Ph.D. candidate in the field of acoustic signal processing. His current research interests include spatial audio processing, spatial filtering, speech enhancement (dereverberation, noise reduction, echo cancellation, feedback cancellation, automatic gain control), adaptive filtering, and binaural processing techniques.

Emanuël A. P. Habets (S'02–M'07–SM'11) received the B.Sc. degree in electrical engineering from the Hogeschool Limburg, Limburg, The Netherlands, in 1999, and the M.Sc. and Ph.D. degrees in electrical engineering from the Technische Universiteit Eindhoven, Eindhoven, The Netherlands, in 2002 and 2007, respectively. He is an Associate Professor with the International Audio Laboratories Erlangen (a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg and Fraunhofer IIS), and the Head of the Spatial Audio Research Group, Fraunhofer IIS, Germany.

From 2007 to 2009, he was a Postdoctoral Fellow at the Technion—Israel Institute of Technology and at the Bar-Ilan University, Israel. From 2009 to 2010, he was a Research Fellow in the Communication and Signal Processing Group, Imperial College London, U.K. His research activities center around audio and acoustic signal processing, and include spatial audio signal processing, spatial sound recording and reproduction, speech enhancement (dereverberation, noise reduction, echo reduction), and sound localization and tracking.

Dr. Habets was a member of the organization committee of the 2005 International Workshop on Acoustic Echo and Noise Control, Eindhoven, The Netherlands, a general co-chair of the 2013 International Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York, and a general co-chair of the 2014 International Conference on Spatial Audio, Erlangen, Germany. He was a member of the IEEE Signal Processing Society Standing Committee on Industry Digital Signal Processing Technology (2013–2015), a Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING and the *EURASIP Journal on Advances in Signal Processing*, and an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS (2013–2017). He is the recipient, with S. Gannot and I. Cohen, of the 2014 IEEE Signal Processing Letters Best Paper Award. He is currently a member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing, the Vice Chair of the E*URASIP Special Area Team on Acoustic, Sound and Music Signal Processing*, and the Editor-in-Chief of the *EURASIP Journal on Audio, Speech, and Music Processing*.