Online Dereverberation for Dynamic Scenarios Using a Kalman Filter With an Autoregressive Model

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

***Abstract*—Reverberant signals can be modeled in the short-time Fourier transform domain using a frequency-dependent autoregressive (AR) model. In state-of-the-art, these AR coefficients have been considered stationary, which does not hold in time-varying environments. We propose to model these AR coefficients using a first-order Markov process, whereas the reverberant microphone signal observations are considered deterministic. This leads to a framework where the AR coefficients can be optimally estimated using a Kalman filter per subband. As a consequence, we can dereverberate the observed signals by applying the estimated AR coefficients as an adaptive linear filter per subband. Estimators for the required statistical parameters in the Kalman filter are derived. Due to the adaptive solution, the algorithm is suitable for real-time applications. It is shown that the proposed method outperforms an existing recursive least-squares solution in terms of reverberation reduction, convergence time, and tracking changes in the acoustic scene.**

***Index Terms*—Dereverberation, Kalman filter, Markov model, multichannel autoregressive (AR) model.**

I. Introduction

REVERBERATION reduction is still a challenging task in many audio processing applications like speech communication, audio feature extraction, source separation, and speech recognition.

Multichannel dereverberation methods can be divided into the following classes: methods using spatiospectral filters [1], methods computing the inverse equalizer using the multiple-input/output inverse theorem (MINT) [2], direct estimation of the clean speech using a moving average model [3], and methods based on an autoregressive (AR) model [4] estimating directly the acoustic channel equalizer. To the latter related methods in the time-domain using linear prediction-based methods [5]–[7] require very long filters and can potentially whiten the desired signal. It has been shown in several contributions [4], [8]–[11] that the reverberant speech in the short-time Fourier transform (STFT) domain can be described using an AR model per frequency band, which allows the use of relatively short filters per subband. Advantages of multichannel algorithms based on this model are that they require no prior knowledge of the source and microphone positions, and do not change the inter-microphone correlation of the direct sound, which makes them suitable as a preprocessing technique for other multichannel algorithms. In contrast to spatiospectral filtering techniques [1] that depend on the direct-to-reverberation ratio, the filtering techniques based on the AR model are effective even far beyond the critical distance. Unfortunately, most proposed solutions are batch algorithms and assume a stationary sound scene, which often does not hold in practice. However, in [12] and [13], a recursive least-squares (RLS) algorithm to estimate the AR coefficients is proposed.

In the present contribution, we formulate the multichannel autoregressive (MAR) signal model with time-varying coefficients in the STFT domain. A Kalman filter is used to estimate the MAR coefficients, which has been applied similarly in the context of acoustic echo cancellation [14], [15] to estimate the finite echo response filter. In the field of speech enhancement, Kalman filters have been applied for noise reduction [16]–[18] and for dereverberation using a moving average reverberation model [3], [19]. The proposed approach can be seen as a generalization of the RLS [13] algorithm, or as an online solution to the blind acoustic impulse response shortening task [9], although the underlying reverberation model is substantially different. In contrast to [9] and [13], we use a time-varying statistical model for the MAR coefficients and also take the intermicrophone correlations of the desired signal into account. Furthermore, we derive a maximum likelihood (ML) estimator for the desired signal covariance matrix, without requiring additional information about the reverberation time as in [13]. Finally, we establish a connection between the proposed algorithm and [13], and compare them in simulations. The proposed adaptive algorithm yields faster convergence, better stability, and higher dereverberation performance, but exhibits a higher computational cost.

II. Signal Model

A. Autoregressive Reverberation Signal Model

We assume M microphones with arbitrary directivity positioned arbitrarily in a reverberant environment with an unknown number of sound sources. We denote the observed microphone signals in the STFT domain by the vector $\mathbf{y}(k,n) = [Y_1(k,n), \ldots, Y_M(k,n)]^{\mathrm{T}}$, where $Y_m(k,n)$ is the STFT representation of the mth microphone signal at frequency index k and time index n. We assume that the reverberant signals are generated by an MAR process per frequency subband [7], [8], [10], i.e.

$$\mathbf{y}(k,n) = \underbrace{\sum_{\ell=D}^{L} \mathbf{C}_\ell(k,n)\mathbf{y}(k,n-\ell)}_{\mathbf{r}(k,n)} \; + \; \mathbf{x}(k,n) \qquad (1)$$

where the vector $\mathbf{x}(k,n) = [X_1(k,n), \ldots, X_M(k,n)]^{\mathrm{T}}$ containing the desired dereverberated signals $X_m(k,n)$ is driving the MAR process, and the $M \times M$ matrices $\mathbf{C}_\ell(k,n)$ contain

Manuscript received September 21, 2016; accepted October 9, 2016. Date of publication October 12, 2016; date of current version October 28, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Rita Singh.

The authors are with the International Audio Laboratories Erlangen, Erlangen 91058, Germany (e-mail: sebastian.braun@audiolabs-erlangen.de; emanuel.habets@audiolabs-erlangen.de).

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

Digital Object Identifier 10.1109/LSP.2016.2616888

the MAR coefficients for $\ell = [D, D+1, \ldots, L]$, which are also known as *room regression coefficients* [8]. Correlation between frequency subbands introduced by the STFT filterbank is neglected in this model. The delay $D > 1$ is chosen depending on the STFT overlap such that there is negligible correlation between $\mathbf{x}(k,n)$ and the term $\mathbf{r}(k,n)$ in (1). As a consequence, $\mathbf{r}(k,n)$ models the late reverberation, and $\mathbf{x}(k,n)$ models the direct sound and early reflections. Note that there is no assumption on the number of sound sources such that $\mathbf{x}(k,n)$ can contain multiple dereverberated sound sources. In contrast to previous work [8]–[10], we assume the MAR coefficients $\mathbf{C}_\ell(k,n)$ to be time varying. If we are able to estimate the MAR coefficients, we can recover the desired signal vector $\mathbf{x}(k,n)$ by a multiple-input-multiple-output finite impulse response filter that is described by the coefficients $\mathbf{C}_\ell(k,n)$ in (1).

Since all variables are frequency dependent, we omit the frequency index k in the following to improve the readability, and we define

$$\mathbf{Y}(n) = \mathbf{I}_M \otimes [\mathbf{y}^{\mathrm{T}}(n), \ldots, \mathbf{y}^{\mathrm{T}}(n-L+D)] \quad (2)$$

$$\mathbf{c}(n) = \mathrm{Vec}\left\{[\mathbf{C}_D(n), \ldots, \mathbf{C}_L(n)]^{\mathrm{T}}\right\} \quad (3)$$

where $\mathbf{I}_M$ is the $M \times M$ identity matrix, $\otimes$ denotes the Kronecker product, and $\mathrm{Vec}\{\cdot\}$ is the matrix column stacking operator. Note that the vector $\mathbf{c}(n)$ is of length $L_{\mathrm{c}} = M^2(L-D+1)$ and $\mathbf{Y}(n)$ is a sparse matrix of size $M \times L_{\mathrm{c}}$. By rewriting (1) using (2) and (3), we obtain the fully vectorized notation of the observed reverberant signal vector

$$\mathbf{y}(n) = \mathbf{Y}(n-D)\mathbf{c}(n) + \mathbf{x}(n). \quad (4)$$

We assume that the vector $\mathbf{x}(n)$ is drawn from a zero-mean complex normal distribution

$$\mathbf{x}(n) \sim \mathcal{N}\left(\mathbf{0}_{M\times 1}, \mathbf{\Phi}_{\mathbf{x}}(n)\right) \quad (5)$$

where $\mathbf{\Phi}_{\mathbf{x}}(n) = E\{\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n)\}$ is the respective covariance matrix. Furthermore, we assume that the desired signals are mutually uncorrelated between different time frames. These are commonly accepted and widely used assumptions for nonreverberant speech STFT coefficients [3], [17], [20].

An estimate of the desired signal vector is obtained by linear filtering using the estimated MAR coefficients $\widehat{\mathbf{c}}(n)$, i.e.

$$\widehat{\mathbf{x}}(n) = \mathbf{y}(n) - \mathbf{Y}(n-D)\widehat{\mathbf{c}}(n). \quad (6)$$

B. Stochastic State-Space Modeling of AR Coefficients

In previous works [7], [8], [10], the MAR coefficients are assumed to be stationary, which does not hold in time-varying acoustic scenarios. Furthermore, our experiments have shown that the MAR coefficients $\mathbf{c}(n)$ can be time varying even in spatially stationary scenarios due to model errors of the STFT domain model (1).

To model the uncertainty of the MAR coefficient vector, we assume that $\mathbf{c}(n)$ consists of independent complex random variables. We describe the MAR coefficient state vector by a first-order Markov model [14]

$$\mathbf{c}(n) = \mathbf{A}(n)\,\mathbf{c}(n-1) + \mathbf{w}(n) \quad (7)$$

where the matrix $\mathbf{A}(n)$ models the propagation of the state vector over time and $\mathbf{w}(n) \sim \mathcal{N}\left(\mathbf{0}_{L_{\mathrm{c}}\times 1}, \mathbf{\Phi}_{\mathbf{w}}(n)\right)$ is a zero-mean complex Gaussian perturbation process with covariance matrix $\mathbf{\Phi}_{\mathbf{w}}(n)$. Further, we assume that $\mathbf{w}(n)$ and $\mathbf{x}(n)$ are mutually uncorrelated.

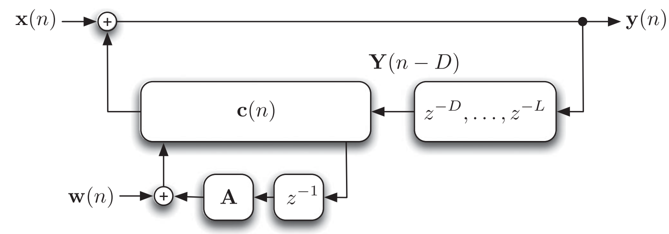

Fig. 1. Multichannel autoregressive signal model with Markov modeling of the regression coefficients.

The signal model is depicted in Fig. 1. In the next section, we use the Kalman filter to estimate the MAR coefficients.

III. Kalman Filter to Estimate MAR Coefficients

An optimal estimate of the MAR coefficient vector is obtained by minimizing the mean-squared error

$$E\left\{\|\mathbf{c}(n) - \widehat{\mathbf{c}}(n)\|_2^2\right\}. \quad (8)$$

We denote the Kronecker delta function by $\delta(n)$, and define the state vector error covariance matrix as

$$\mathbf{\Phi}_\epsilon(n) = E\left\{[\mathbf{c}(n) - \widehat{\mathbf{c}}(n)][\mathbf{c}(n) - \widehat{\mathbf{c}}(n)]^{\mathrm{H}}\right\}. \quad (9)$$

Given that we have a dynamic system described by the state (7), the observation (4), and that assumptions

$$\mathbf{w}(n) \sim \mathcal{N}\left(\mathbf{0}_{L_{\mathrm{c}}\times 1}, \mathbf{\Phi}_{\mathbf{w}}(n)\right) \quad (10)$$

$$\mathbf{x}(n) \sim \mathcal{N}\left(\mathbf{0}_{M\times 1}, \mathbf{\Phi}_{\mathbf{x}}(n)\right) \quad (11)$$

$$E\left\{\mathbf{w}(n)\mathbf{w}^{\mathrm{H}}(n-j)\right\} = \mathbf{\Phi}_{\mathbf{w}}(n)\delta(n-j) \quad (12)$$

$$E\left\{\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n-j)\right\} = \mathbf{\Phi}_{\mathbf{x}}(n)\delta(n-j) \quad (13)$$

$$E\left\{\mathbf{w}(n)\mathbf{x}^{\mathrm{H}}(n-j)\right\} = \mathbf{0}_{L_{\mathrm{c}}\times M} \quad (14)$$

are fulfilled, the well-known Kalman filter [21] is a suitable recursive estimator of $\mathbf{c}(n)$ that minimizes (8) given all currently available observations $\{\mathbf{y}(0), \mathbf{y}(1), \ldots \mathbf{y}(n)\}$ for $n \to N$, where $N \gg 0$. The Kalman filter prediction and update equations are given by

$$\mathbf{\Phi}_\epsilon(n|n-1) = \mathbf{A}(n)\widehat{\mathbf{\Phi}}_\epsilon(n-1)\mathbf{A}^{\mathrm{H}}(n) + \mathbf{\Phi}_{\mathbf{w}}(n) \quad (15)$$

$$\mathbf{c}(n|n-1) = \mathbf{A}(n)\widehat{\mathbf{c}}(n-1) \quad (16)$$

$$\mathbf{e}(n) = \mathbf{y}(n) - \mathbf{Y}(n-D)\mathbf{c}(n|n-1) \quad (17)$$

$$\mathbf{K}(n) = \mathbf{\Phi}_\epsilon(n|n-1)\mathbf{Y}^{\mathrm{H}}(n-D) \quad (18)$$
$$\left[\mathbf{Y}(n-D)\mathbf{\Phi}_\epsilon(n|n-1)\mathbf{Y}^{\mathrm{H}}(n-D) + \mathbf{\Phi}_{\mathbf{x}}(n)\right]^{-1}$$

$$\widehat{\mathbf{\Phi}}_\epsilon(n) = [\mathbf{I}_{L_{\mathrm{c}}} - \mathbf{K}(n)\mathbf{Y}(n-D)]\;\mathbf{\Phi}_\epsilon(n|n-1) \quad (19)$$

$$\widehat{\mathbf{c}}(n) = \mathbf{c}(n|n-1) + \mathbf{K}(n)\mathbf{e}(n) \quad (20)$$

where the vector $\mathbf{e}(n)$ is called the prediction error and the matrix $\mathbf{K}(n)$ is called the Kalman gain, which minimizes the trace of the error matrix $\mathbf{\Phi}_\epsilon(n)$. If the state of the Kalman filter at $n = 0$ is unknown, we propose to initialize with the values $\widehat{\mathbf{\Phi}}_\epsilon(0) = \mathbf{I}_{L_{\mathrm{c}}}$ and $\widehat{\mathbf{c}}(0) = \mathbf{0}_{L_{\mathrm{c}}\times 1}$.

By reviewing (6) and (17), we observe that the prediction error $\mathbf{e}(n)$ is an estimate of the desired signals $\mathbf{x}(n)$ given the predicted MAR coefficients $\mathbf{c}(n|n-1)$.

IV. Parameter Estimation for Kalman Filter

The Kalman filter requires the covariance of the state vector perturbation process $\mathbf{\Phi}_\mathbf{w}(n)$, the state propagation matrix $\mathbf{A}(n)$ and the covariance matrix $\mathbf{\Phi}_\mathbf{x}(n)$. In this section, we propose suitable estimators for these generally time-varying parameters.

A. Dynamic State Modeling

Similarly as in [15], we want to model slowly time-varying acoustic conditions by the first-order Markov model given by (7). As the transitions of the MAR coefficients over time are unknown, we choose $\mathbf{A}(n) = \mathbf{I}_{L_c}$. By assuming the elements of $\mathbf{w}(n)$ uncorrelated and identically distributed, the uncertainty of the MAR coefficients is determined by the scalar process noise variance $\phi_w(n)$, such that $\mathbf{\Phi}_\mathbf{w}(n) = \phi_w(n)\mathbf{I}_{L_c}$. We propose to estimate the variance $\phi_w(n)$ depending on the change of the MAR coefficients between subsequent frames using

$$\widehat{\phi}_w(n) = \frac{1}{L_\mathrm{c}} E\left\{\|\widehat{\mathbf{c}}(n) - \widehat{\mathbf{c}}(n-1)\|_2^2\right\} + \eta \tag{21}$$

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

B. Estimation of the Spatial Covariance Matrix

The estimation accuracy of the spatial covariance matrix $\mathbf{\Phi}_\mathbf{x}(n)$ greatly effects the quality of the estimated desired signals $\widehat{\mathbf{x}}(n)$. Therefore, we derive an ML estimator and an approximation thereof, that can be computed given the currently available data.

According to the signal model (1), only the last L observations are required to model the current observed signal. While the probability distribution of $\mathbf{y}(n)$ is dependent on the previous time frames, the conditional probability given the previous L time frames is independent. The conditional probability at frame n is $f\{\mathbf{y}(n) \,|\, \mathbf{\Theta}(n)\}$, where we define the parameter set $\mathbf{\Theta}(n) = \{\mathbf{y}(n), \mathbf{y}(n-1), \ldots, \mathbf{y}(n-L), \mathbf{c}(n)\}$. The conditional mean and covariance of $\mathbf{y}(n)$ given $\mathbf{\Theta}(n)$ are

$$\boldsymbol{\mu}_{\mathbf{y}|\mathbf{\Theta}} = E\{\mathbf{y}(n) \,|\, \mathbf{\Theta}(n)\} = \mathbf{Y}(n-D)\mathbf{c}(n) \tag{22}$$

and

$$E\left\{[\mathbf{y}(n) - \boldsymbol{\mu}_{\mathbf{y}|\mathbf{\Theta}}]\,[\mathbf{y}(n) - \boldsymbol{\mu}_{\mathbf{y}|\mathbf{\Theta}}]^\mathrm{H} \,|\, \mathbf{\Theta}(n)\right\} = \mathbf{\Phi}_\mathbf{x}(n) \tag{23}$$

respectively. As the desired signal vector $\mathbf{x}(n)$ is Gaussian distributed, the PDF of $\mathbf{y}(n)$ is a Gaussian conditional likelihood function with mean (22) and covariance (23) and is given by

$$f\{\mathbf{y}(n) \,|\, \mathbf{\Theta}(n)\} = \frac{1}{\pi^M |\mathbf{\Phi}_\mathbf{x}(n)|} e^{-[\mathbf{y}(n)-\mathbf{Y}(n-D)\mathbf{c}(n)]^\mathrm{H} \mathbf{\Phi}_\mathbf{x}{}^{-1}(n)\,[\mathbf{y}(n)-\mathbf{Y}(n-D)\mathbf{c}(n)]}. \tag{24}$$

By assuming short-time stationarity of the PDF of $\mathbf{y}(n)$ within N frames, the joint PDF at frame n is obtained by

$$f\{\mathbf{y}(n-N+1), \ldots, \mathbf{y}(n)\} = \prod_{\ell=n-N+1}^{n} f\{\mathbf{y}(\ell) \,|\, \mathbf{\Theta}(\ell)\}. \tag{25}$$

By maximizing the log-likelihood function of (25) w.r.t. $\mathbf{\Phi}_\mathbf{x}(n)$, the covariance at frame n is given by

$$\mathbf{\Phi}_\mathbf{x}^{(\mathrm{ML})}(n) = \frac{1}{N} \sum_{\ell=n-N+1}^{n} [\mathbf{y}(\ell) - \mathbf{Y}(\ell-D)\mathbf{c}(\ell)] [\mathbf{y}(\ell) - \mathbf{Y}(\ell-D)\mathbf{c}(\ell)]^\mathrm{H}. \tag{26}$$

We can compute (26) using the estimates $\{\widehat{\mathbf{c}}(n-N+1), \ldots, \widehat{\mathbf{c}}(n-1)\}$. As we do not have yet the estimate $\widehat{\mathbf{c}}(n)$ at frame n available when we need to compute (26), we approximate it by the predicted estimate $\widehat{\mathbf{c}}(n) \approx \mathbf{c}(n|n-1)$ in (26). Under this assumption and by using (6) and (17), an estimate of (26) is given by

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

For practical reasons, we replace the arithmetic average in (27) by an exponentially weighted average, such that we can compute the estimate recursively, i.e.

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

where

$$\widehat{\mathbf{\Phi}}_\mathbf{x}(n) = \alpha\widehat{\mathbf{\Phi}}_\mathbf{x}(n-1) + (1-\alpha)\widehat{\mathbf{x}}(n)\widehat{\mathbf{x}}^\mathrm{H}(n) \tag{29}$$

and α is an exponential recursive smoothing factor with a time constant corresponding to N frames.

Although it is proposed in some solutions in [9] that the spatial covariance matrix $\mathbf{\Phi}_\mathbf{x}(n)$ is diagonal, this is typically not the case due to strong inter-microphone coherence of the desired signals. If only a single sound source is present, this matrix is of rank one.

C. Relation to the RLS Algorithm

It is well known that the RLS algorithm has a strong similarity to the Kalman filter [22], although they are significantly different from the theoretical point of view. The equations of the proposed Kalman filter become identical to the RLS algorithm proposed in [13], if $\mathbf{\Phi}_\mathbf{x}(n)$ is assumed to be a scaled identity matrix, $\mathbf{A}(n) = \mathbf{I}_{L_c}$, and (15) changes to

$$\mathbf{\Phi}_\epsilon^\mathrm{RLS}(n|n-1) = \frac{1}{\gamma}\widehat{\mathbf{\Phi}}_\epsilon(n-1) \tag{30}$$

where γ is a forgetting factor that controls the convergence. The RLS algorithm is then described by (30) and (17)–(20), where $\mathbf{\Phi}_\epsilon(n|n-1)$ is replaced by $\mathbf{\Phi}_\epsilon^\mathrm{RLS}(n|n-1)$. Note that in the RLS algorithm, the covariance matrix $\mathbf{\Phi}_\epsilon(n)$ denotes the covariance of the MAR coefficients $\mathbf{c}(n)$ instead of the error covariance matrix.

The major advantages of the proposed Kalman filter over the RLS algorithm are as follows:

1) time varying stochastic modeling of MAR coefficients;
2) modeling of the spatial covariance matrix $\mathbf{\Phi}_\mathbf{x}(n)$;
3) adaptively controlled updating of $\mathbf{\Phi}_\epsilon(n|n-1)$;
4) guaranteed stability and controllability of the Kalman filter under the given assumptions;
5) ML-based estimation of $\widehat{\mathbf{\Phi}}_\mathbf{x}(n)$.

In contrast, by exploiting the diagonality of $\mathbf{\Phi}_\mathbf{x}(n)$, the RLS has a lower computational complexity by a factor of M^2 and does not require an $M \times M$ matrix inversion.

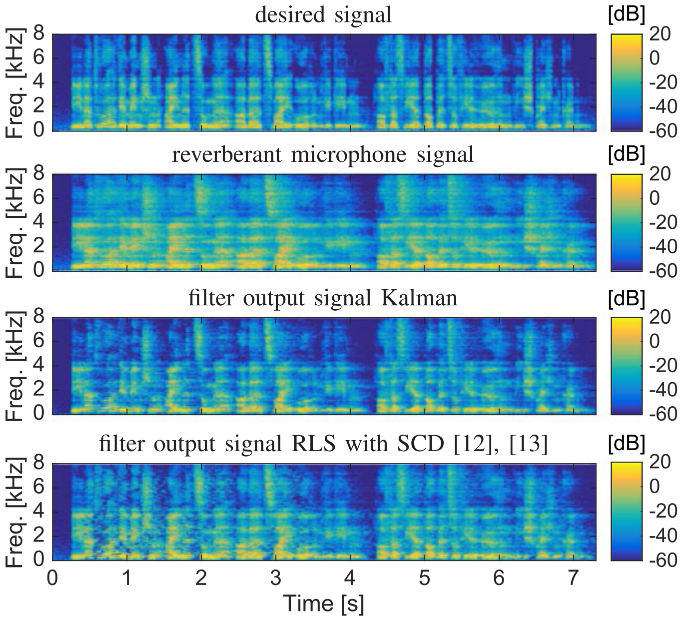

Fig. 2. Spectrograms of desired signal including early reflections, observed signal at the first microphone and filtered signal at first microphone.

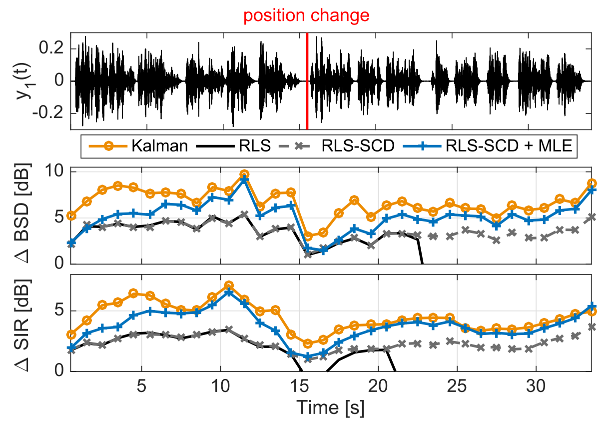

Fig. 3. Improvement of objective measures over the unprocessed microphone signal in 1-s analysis segments.

V. Experimental Results

A. Setup

We created test signals by convolving two-channel room impulse responses measured in an acoustic lab with $T_{60} \approx 0.63$ s. As indicated in Fig. 3, the speaker changes to a different position after 15 s. The speaker-microphone distances are 3.8 m and 2.1 m, respectively. Although the signal model does not include additive noise, a small amount of pink noise was added with an SNR $= 50$ dB to simulate background noise. We used a sampling frequency of 16 kHz, an FFT size of 1024 points, and a square-root Hann window of 32 ms length with 50% overlap. The filter length and delay in each frequency band were set to $L = 15$ and $D = 2$. The smoothing constant in (28) and (29) was $\alpha = 0.4$ and $\eta = -35$ dB.

For reference, we implemented the RLS algorithm [13] described in Section IV-C in following three variants: 1) using the originally proposed estimator for $\mathbf{\Phi}_{\mathbf{x}}(n)$ based on spectral subtraction requiring the reverberation time, without speaker change detection (SCD), 2) the same implementation with SCD [12], and 3) the RLS using the ML spatial covariance estimator (28) derived in Section IV-B. As the matrix $\mathbf{\Phi}_{\mathbf{x}}(n)$ is required to be diagonal in the RLS, we used an identity matrix scaled by the mean of the main diagonal of $\widehat{\mathbf{\Phi}}_{\mathbf{x}}^{(\mathrm{RML})}(n)$. The RLS forgetting factor was set to $\gamma = 0.99$ as proposed in [13].

The desired signal for evaluation was defined as the speech signals convolved with the first part of the room impulse responses containing early reflections until 40 ms after the direct sound.

B. Results

Fig. 2 shows spectrogram parts of the signals at the first microphone: the desired signal $X_1(k, n)$, the reverberant microphone signal $Y_1(k, n)$, the filter output signals obtained using the proposed Kalman filter $\widehat{X}_1(k, n)$, and using the RLS algorithm with SCD using the ML covariance estimator (28). It can be observed that the correlation over time introduced by reverberation is reduced in both filter output signals.

We evaluate the algorithms by applying objective measures to 1 s long segments of the unprocessed and filtered signals at the first microphone. The used objective measures are the bark spectral distortion (BSD) [23] and the signal-to-interference ratio (SIR), where the interference is the residual reverberation plus artifacts.

Fig. 3 shows the positive improvement of those objective measures compared to the unprocessed microphone signal. It can be observed that all methods greatly improve the objective measures. Only the RLS without SCD (no markers) diverges after the reverberation path change and deteriorates the objective measures even below the unprocessed reference, yielding a negative improvement. This problem is mitigated by the RLS with SCD ("$\times$" markers) proposed in [12]. By using the proposed ML covariance estimator (28) instead of the original spectral subtraction method, the performance is improved ("$+$" markers). The proposed Kalman based algorithm ("$\circ$" markers) yields faster convergence and is robust against the reverberation path change. It can be noticed that the proposed Kalman filter algorithm converges within just a few seconds and achieves higher reverberation cancellation with low audible artifacts.[1]

VI. Conclusion

We proposed an adaptive algorithm to estimate multichannel regression coefficients per subband to achieve dereverberation. The underlying dynamic model for the regression coefficients is suitable for real-life applications with slowly moving speakers or microphones and changing acoustic conditions. It was shown that the proposed algorithm achieves dereverberation and yields fast convergence in changing acoustic conditions. An existing RLS algorithm is outperformed in terms of convergence speed and dereverberation performance. The proposed ML estimator for the covariance of the desired signal improves also the RLS algorithm.

[1] Audio examples are available at www.audiolabs-erlangen.de/resources/2016-SPL-MAR-KALMAN

REFERENCES

[1] 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, 2015.

[2] 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.

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

[4] 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, Language Process.*, vol. 18, no. 7, pp. 1717–1731, Sep. 2010.

[5] M. Triki and D. T. M. Slock, "Iterated delay and predict equalization for blind speech dereverberation," in *Proc. Intl. Workshop Acoust. Echo Noise Control*, Paris, France, Sep. 2006, pp. 1–4.

[6] M. Delcroix, T. Hikichi, and M. Miyoshi, "Dereverberation and denoising using multichannel linear prediction," *IEEE Trans. Audio, Speech, Language Process.*, vol. 15, no. 6, pp. 1791–1801, Aug. 2007.

[7] T. Nakatani, B.-H. Juang, T. Yoshioka, K. Kinoshita, M. Delcroix, and M. Miyoshi, "Speech dereverberation based on maximum-likelihood estimation with time-varying Gaussian source model," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 8, pp. 1512–1527, Nov. 2008.

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

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

[10] 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, Language Process.*, vol. 21, no. 7, pp. 1369–1380, Jul. 2013.

[11] A. Jukic, T. van Waterschoot, T. Gerkmann, and S. Doclo, "Multi-channel linear prediction-based speech dereverberation with sparse priors," *IEEE Trans. Audio, Speech, Language Process.*, vol. 23, no. 9, pp. 1509–1520, Sep. 2015.

[12] T. Yoshioka, H. Tachibana, T. Nakatani, and M. Miyoshi, "Adaptive dereverberation of speech signals with speaker-position change detection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2009, pp. 3733–3736.

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

[14] 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.

[15] C. Paleologu, J. Benesty, and S. Ciochina, "Study of the general Kalman filter for echo cancellation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 8, pp. 1539–1549, Aug. 2013.

[16] S. Gannot, "Speech processing utilizing the Kalman filters," *IEEE Instrum. Meas. Mag.*, vol. 15, no. 3, pp. 10–14, Jun. 2012.

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

[18] 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.

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

[20] 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, Language Process.*, vol. 18, no. 7, pp. 1746–1765, Sep. 2010.

[21] R. E. Kalman, "A new approach to linear filtering and prediction problems," *ASME J. Basic Eng.*, vol. 82, pp. 35–45, 1960.

[22] A. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Process. Mag.*, vol. 11, no. 3, pp. 18–60, Jul. 1994.

[23] S. Wang, A. Sekey, and A. Gersho, "An objective measure for predicting subjective quality of speech coders," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 5, pp. 819–829, Jun. 1992.