CONSTRAINED MULTI-CHANNEL LINEAR PREDICTION FOR ADAPTIVE SPEECH DEREVERBERATION

Ante Jukić[1], Zichao Wang[2], Toon van Waterschoot[3], Timo Gerkmann[4], Simon Doclo[1]

[1]University of Oldenburg, Department of Medical Physics and Acoustics and the Cluster of Excellence Hearing4All, Oldenburg, Germany
[2]Rice University, Houston, TX, USA
[3]KU Leuven, Department of Electrical Engineering (ESAT-STADIUS/ETC), Leuven, Belgium
[4]Technicolor Research and Innovation, Hanover, Germany
ante.jukic@uni-oldenburg.de

ABSTRACT

This paper presents a speech dereverberation algorithm combining adaptive multi-channel linear prediction (MCLP) with a statistical model for the undesired reverberation. More specifically, we propose to constrain the power of the MCLP-based late reverberation estimate with the late reverberant power estimated using the exponential decay model, thereby preventing excessive cancellation of the speech signal. Simulation results show that incorporating the constraint improves the performance of the adaptive dereverberation method when the prediction filters need to adapt quickly.

***Index Terms*—** speech dereverberation, multi-channel linear prediction, constrained linear prediction, adaptive filtering

1. INTRODUCTION

Microphone recordings of a distant speech source in an enclosure are typically corrupted by reverberation, caused by reflections against boundaries and objects in the enclosure. Although a small amount of reverberation can be beneficial, many speech communication applications suffer in highly reverberant conditions, resulting in a degraded speech intelligibility and speech recognition performance [1]. Hence, effective speech dereverberation is required for many applications, and several speech dereverberation methods have been proposed in the literature [2].

An important class of blind multi-channel (MC) speech dereverberation methods is based on multi-channel linear prediction (MCLP) [3, 4]. MCLP-based methods aim to predict the undesired reverberant component by filtering and summing the delayed microphone signals, and dereverberation is performed by subtracting the predicted reverberant component from the microphone signals. The prediction filters are typically obtained by maximizing sparsity of the output signal in the time-frequency domain [5, 6], and the delay is introduced in the prediction to ensure that the short-time speech correlation and early reflections are preserved [3]. Adaptive versions of MCLP-based dereverberation methods have been proposed in [7, 8], where the filter adaptation is based on a variant of a recursive least squares (RLS) algorithm [9]. Unfortunately, in some cases these methods may lead to a significant over-estimation of the undesired component and hence high distortions of the desired speech component [8].

In this paper, we propose a constrained MCLP optimization problem for adaptive speech dereverberation. The goal is to prevent over-estimation of the undesired component by incorporating prior knowledge about the undesired late reverberation into the adaptive MCLP-based method. More specifically, we use the late reverberant power spectral density (PSD) estimated using the exponential decay model [10] to constrain the power of the undesired component estimated using the MCLP. Simulation results show that incorporating this constraint improves the performance of the adaptive method when the prediction filters need to adapt quickly, e.g., for a moving source.

2. SIGNAL MODEL

We consider a scenario with a single speech source and M microphones in a reverberant room. Let $s(k,n)$ denote the clean speech signal in the short-time Fourier transform (STFT) domain, with k the frequency bin index and n the frame index. We assume that the m-th reverberant microphone signal $x_m(k,n)$ can be decomposed into a desired component $d_m(k,n)$ and an undesired component $u_m(k,n)$ as

$$x_m(k,n) = d_m(k,n) + u_m(k,n), \tag{1}$$

where the desired component contains the direct speech and early reflections, while the undesired component contains the late reflections. In the following, we omit the index k and use the model in (1) in each frequency bin independently. When multiple microphones are available, the undesired component $u_m(k,n)$ can be modeled using multi-channel linear prediction [4] as the sum of filtered (delayed) microphone signals, i.e.,

$$u_m(n) = \mathbf{g}_m^{\mathsf{H}}(n)\tilde{\mathbf{x}}(n-\tau), \tag{2}$$

where τ is the prediction delay, $\mathbf{g}_m(n) \in \mathbb{C}^{ML_g}$ is the MC prediction filter, and $\tilde{\mathbf{x}}(n) \in \mathbb{C}^{ML_g}$ is a vector of L_g coefficients for all microphones, defined as

$$\begin{aligned}\tilde{\mathbf{x}}(n) = [&x_1(n), \ldots, x_1(n-L_g+1), \ldots \\ &\ldots, x_M(n), \ldots, x_M(n-L_g+1)]^{\mathsf{T}}. \end{aligned}\tag{3}$$

The role of the delay τ is to include only the late reflections in the undesired component, thereby preserving the early reflections and the short-time speech correlation in the desired component [3,4]. By combining the models in (1) and (2) for all channels, the multiple-input multiple-output (MIMO) signal model can be written as

$$\mathbf{x}(n) = \mathbf{d}(n) + \underbrace{\mathbf{G}^{\mathsf{H}}(n)\tilde{\mathbf{x}}(n-\tau)}_{\mathbf{u}(n)}, \tag{4}$$

This research was supported by the Marie Curie Initial Training Network DREAMS (Grant agreement no. ITN-GA-2012-316969), and in part by the Cluster of Excellence 1077 "Hearing4All", funded by the German Research Foundation (DFG).

978-1-5090-2007-2/16/$31.00 ©2016 IEEE

where $\mathbf{x}(n) = [x_1(n), \ldots, x_M(n)]^{\mathsf{T}}$ is the MC reverberant signal, $\mathbf{G}(n) = [\mathbf{g}_1(n), \ldots, \mathbf{g}_M(n)] \in \mathbb{C}^{ML_g \times M}$ is the MIMO prediction filter, and $\mathbf{d}(n)$ and $\mathbf{u}(n)$ are the MC desired and undesired undesired component, respectively. The goal of speech dereverberation is then to recover the desired component $\mathbf{d}(n)$, which can be achieved by estimating the prediction filter $\mathbf{G}(n)$ and consequently subtracting the estimated undesired component from $\mathbf{x}(n)$ in (4).

3. MCLP FOR SPEECH DEREVERBERATION

In this section we give a brief overview of two MCLP-based speech dereverberation methods: the batch generalized weighted prediction error (GWPE) method and its adaptive variant (A-GWPE).

3.1. Batch processing (GWPE)

The batch GWPE method [11] assumes that the prediction filter $\mathbf{G}(n)$ does not change over time, i.e., $\mathbf{G}(n) = \mathbf{G}$ for all n. Assuming that a batch of N time-frames is available, the prediction filter $\mathbf{G}$ can be estimated by minimizing the temporal correlation of the desired component [11], which is equivalent to maximizing sparsity across time [12]. The corresponding optimization problem can be formulated as [11]

$$\hat{\mathbf{G}} = \arg\min_{\mathbf{G}} \sum_{n=1}^{N} \log \|\mathbf{d}(n)\|_2^2, \tag{5}$$

where $\|.\|_2$ denotes the ℓ_2-norm of a vector, and the logarithmic function promotes sparsity across time. This optimization problem can be solved using the iteratively reweighted least squares (IRLS) algorithm [13], leading to the following iterative updates [11,12]

$$\hat{w}^i(n) = \left(\frac{1}{M}\|\hat{\mathbf{d}}^{i-1}(n)\|_2^2 + \varepsilon\right)^{-1}, \ \forall n \in \{1, \ldots, N\}, \tag{6}$$

$$\hat{\mathbf{G}}^i = \arg\min_{\mathbf{G}} \sum_{n=1}^{N} \hat{w}^i(n) \|\mathbf{x}(n) - \mathbf{G}^{\mathsf{H}}\tilde{\mathbf{x}}(n-\tau)\|_2^2, \tag{7}$$

$$\hat{\mathbf{d}}^i(n) = \mathbf{x}(n) - \left(\hat{\mathbf{G}}^i\right)^{\mathsf{H}} \tilde{\mathbf{x}}(n-\tau), \ \forall n \in \{1, \ldots, N\}, \tag{8}$$

with i the iteration index, and ε a small constant for regularization. Intuitively, the weights $\hat{w}(n)$ put more emphasis on time frames where the desired component is expected to have small power, corresponding to the sparsity-promoting behavior of the logarithmic function in (5) [5,8].

The least-squares (LS) optimization problem for estimating the prediction filter $\mathbf{G}$ in (7) can be written as

$$\boxed{\min_{\mathbf{G}} \operatorname{tr}\left[\mathbf{G}^{\mathsf{H}}\hat{\mathbf{Q}}^i\mathbf{G}\right] - 2\Re\left\{\operatorname{tr}\left[\mathbf{G}^{\mathsf{H}}\hat{\mathbf{R}}^i\right]\right\}, \qquad (9)}$$

with the matrices $\hat{\mathbf{Q}}^i$ and $\hat{\mathbf{R}}^i$ defined as

$$\hat{\mathbf{Q}}^i = \sum_{n=1}^{N} \hat{w}^i(n)\tilde{\mathbf{x}}(n-\tau)\tilde{\mathbf{x}}^{\mathsf{H}}(n-\tau), \tag{10}$$

$$\hat{\mathbf{R}}^i = \sum_{n=1}^{N} \hat{w}^i(n)\tilde{\mathbf{x}}(n-\tau)\mathbf{x}^{\mathsf{H}}(n). \tag{11}$$

The closed-form solution for the prediction filter is given with

$$\hat{\mathbf{G}}^i = \left(\hat{\mathbf{Q}}^i\right)^{-1}\hat{\mathbf{R}}^i. \tag{12}$$

The GWPE method is typically initialized with the reverberant signals, i.e., $\hat{\mathbf{d}}^0(n) = \mathbf{x}(n)$, or equivalently with $\hat{\mathbf{G}}^0 = \mathbf{0}$, after which a number of reweighting iterations in (6)-(8) are performed.

3.2. Adaptive processing (A-GWPE)

An adaptive version of the GWPE method (A-GWPE) has been proposed in [8]. The A-GWPE method estimates the prediction filter $\mathbf{G}(n)$ at each frame n by applying the RLS algorithm on the LS problem in (9) [9]. Assuming that the weights are fixed, the prediction filter $\mathbf{G}(n)$ can be estimated, similarly as in (9), by solving the following LS problem

$$\boxed{\min_{\mathbf{G}(n)} \operatorname{tr}\left[\mathbf{G}^{\mathsf{H}}(n)\hat{\mathbf{Q}}(n)\mathbf{G}(n)\right] - 2\Re\left\{\operatorname{tr}\left[\mathbf{G}^{\mathsf{H}}(n)\hat{\mathbf{R}}(n)\right]\right\}, \ (13)}$$

with the matrices $\hat{\mathbf{Q}}(n)$ and $\hat{\mathbf{R}}(n)$ defined as

$$\hat{\mathbf{Q}}(n) = \sum_{t=1}^{n} \gamma^{n-t}\hat{w}(t)\tilde{\mathbf{x}}(t-\tau)\tilde{\mathbf{x}}^{\mathsf{H}}(t-\tau), \tag{14}$$

$$\hat{\mathbf{R}}(n) = \sum_{t=1}^{n} \gamma^{n-t}\hat{w}(t)\tilde{\mathbf{x}}(t-\tau)\mathbf{x}^{\mathsf{H}}(t), \tag{15}$$

where $\gamma \in (0,1)$ is a forgetting factor typically used in RLS algorithms. The closed-form solution for the prediction filter in (13) is given by

$$\hat{\mathbf{G}}(n) = \hat{\mathbf{Q}}^{-1}(n)\hat{\mathbf{R}}(n), \tag{16}$$

and dereverberation is performed by subtracting the estimated undesired component $\hat{\mathbf{u}}(n) = \hat{\mathbf{G}}^{\mathsf{H}}(n)\tilde{\mathbf{x}}(n-\tau)$ from $\mathbf{x}(n)$.

By observing that the matrices $\hat{\mathbf{Q}}(n)$ and $\hat{\mathbf{R}}(n)$ in (14)-(15) are obtained by adding rank-1 perturbations to $\gamma\hat{\mathbf{Q}}(n-1)$ and $\gamma\hat{\mathbf{R}}(n-1)$, the Woodbury matrix inversion lemma can be used to compute the inverse $\hat{\mathbf{Q}}^{-1}(n)$ recursively as

$$\hat{\mathbf{Q}}^{-1}(n) = \frac{1}{\gamma}\left[\hat{\mathbf{Q}}^{-1}(n-1) - \hat{\mathbf{k}}(n)\tilde{\mathbf{x}}^{\mathsf{H}}(n-\tau)\hat{\mathbf{Q}}^{-1}(n-1)\right], \tag{17}$$

with the gain vector defined as

$$\hat{\mathbf{k}}(n) = \frac{\hat{\mathbf{Q}}^{-1}(n-1)\tilde{\mathbf{x}}(n-\tau)}{\frac{\gamma}{\hat{w}(n)} + \tilde{\mathbf{x}}^{\mathsf{H}}(n-\tau)\hat{\mathbf{Q}}^{-1}(n-1)\tilde{\mathbf{x}}(n-\tau)}, \tag{18}$$

consequently leading to a recursive update for the prediction filter $\hat{\mathbf{G}}(n)$ as

$$\hat{\mathbf{G}}(n) = \hat{\mathbf{G}}(n-1) + \hat{\mathbf{k}}(n)\left[\mathbf{x}(n) - \hat{\mathbf{G}}^{\mathsf{H}}(n-1)\tilde{\mathbf{x}}(n-\tau)\right]^{\mathsf{H}}. \tag{19}$$

The effective forgetting factor in (18) is equal to $\gamma/\hat{w}(n)$, and therefore related to the expected power of the desired component in the n-th frame through the weight $\hat{w}(n)$.

As opposed to the batch method, where the weights are iteratively updated, in the adaptive method it is assumed that the weights are fixed at each frame. Since the weights are related to the expected power of the desired component, cf. (6), in [11] it has been proposed to compute them using a statistical model of the late reverberation [10]. The PSD of the desired component in the m-th microphone $\hat{\sigma}^2_{d,m}(n)$ can be estimated using recursive smoothing and assuming the exponential decay model [10] for the late reverberant PSD, i.e.,

$$\hat{\sigma}^2_{x,m}(n) = \alpha\,\hat{\sigma}^2_{x,m}(n-1) + (1-\alpha)|x_m(n)|^2, \tag{20}$$

$$\hat{\sigma}^2_{u,m}(n) = e^{-2\Delta}\,\hat{\sigma}^2_{x,m}(n-n_d), \tag{21}$$

$$\begin{aligned}\hat{\sigma}^2_{d,m}(n) &= \alpha\,\hat{\sigma}^2_{d,m}(n-1) \\ &\quad + (1-\alpha)\ \max\left\{|x_m(n)|^2 - \hat{\sigma}^2_{u,m}(n), 0\right\},\end{aligned} \tag{22}$$

where α is a smoothing parameter, $\Delta = \frac{3 \ln 10}{T_{60}/T_d}$ is the decay parameter, T_{60} is the reverberation time, T_d is the duration of the direct path and early reflections (typically around 50 ms), and n_d is the number of frames corresponding to T_d. The weight $\hat{w}(n)$ is then computed, similarly as in (6), based on the average estimated PSD of the desired component, as

$$\hat{w}(n) = \left(\frac{1}{M}\|\hat{\boldsymbol{\sigma}}_d(n)\|_2^2 + \varepsilon\right)^{-1} \tag{23}$$

where $\hat{\boldsymbol{\sigma}}_d(n) = [\hat{\sigma}_{d,1}(n), \ldots, \hat{\sigma}_{d,M}(n)]^{\mathsf{T}}$.

4. CONSTRAINED MCLP FOR ADAPTIVE SPEECH DEREVERBERATION

As previously mentioned, in some cases the presented batch and adaptive speech dereverberation methods may result in high distortions of the desired speech component [8]. This can, for example, be illustrated using the cost function for the batch GWPE method in (7). Assuming that $N = ML_g$, i.e., having either a relatively short utterance or relatively long filters, the minimum of the cost function in (7) is equal to zero, resulting in the estimated desired component equal to zero. For the adaptive version a similar situation occurs when the forgetting factor is relatively small, reducing the effective window length in (14)-(15). In both cases the undesired component is significantly over-estimated due to the available data being small compared to the number of free parameters ML_g.

Aiming to obtain a more robust method, we therefore propose to directly incorporate knowledge about the expected undesired component in the adaptive MCLP-based method. More specifically, we propose to add a constraint to (13), forcing the power of the estimated undesired component $\mathbf{u}(n)$ not to exceed the late reverberant PSD estimate $\hat{\boldsymbol{\sigma}}_u(n)$ based on the exponential decay model in (21), leading to the following optimization problem

$$\boxed{\begin{aligned} &\min_{\mathbf{G}(n)} \mathrm{tr}\left[\mathbf{G}^{\mathsf{H}}(n)\hat{\mathbf{Q}}(n)\mathbf{G}(n)\right] - 2\Re\left\{\mathrm{tr}\left[\mathbf{G}^{\mathsf{H}}(n)\hat{\mathbf{R}}(n)\right]\right\} \\ &\qquad\qquad \text{subject to} \quad |\mathbf{G}^{\mathsf{H}}(n)\tilde{\mathbf{x}}(n-\tau)|^2 \leq \hat{\boldsymbol{\sigma}}_u^2(n), \quad (24) \end{aligned}}$$

with $\hat{\boldsymbol{\sigma}}_u(n) = [\hat{\sigma}_{u,1}(n), \ldots, \hat{\sigma}_{u,M}(n)]^{\mathsf{T}}$. It is expected that this will reduce the undesired speech cancellation for small values of the forgetting factor γ, while not deteriorating the performance for large values of the forgetting factor γ.

The optimization problem in (24) can be efficiently solved using the alternating direction method of multipliers (ADMM) [14]. The problem in (24) can be rewritten in a form that is suitable for ADMM by introducing a splitting variable $\mathbf{z}$ as

$$\begin{aligned} \min_{\mathbf{G}(n)} \mathrm{tr}\left[\mathbf{G}^{\mathsf{H}}(n)\hat{\mathbf{Q}}(n)\mathbf{G}(n)\right] - 2\Re\left\{\mathrm{tr}\left[\mathbf{G}^{\mathsf{H}}(n)\hat{\mathbf{R}}(n)\right]\right\} + c\left(\mathbf{z}\right) \\ \text{subject to} \quad \mathbf{z} = \mathbf{G}^{\mathsf{H}}(n)\tilde{\mathbf{x}}(n-\tau), \end{aligned} \tag{25}$$

where $c : \mathbb{C}^M \to \mathbb{R}$ is a convex function enforcing the constraint, i.e.,

$$c\left(\mathbf{z}\right) = \begin{cases} 0, & \text{if } |z_m| \leq \hat{\sigma}_{u,m}(n), \forall m, \\ +\infty, & \text{otherwise} \end{cases}. \tag{26}$$

The augmented Lagrangian for the problem in (25) can be written as

$$\begin{aligned} \mathcal{L}\left(\mathbf{G}(n), \mathbf{z}, \boldsymbol{\mu}\right) = \mathrm{tr}\left[\mathbf{G}(n)^{\mathsf{H}}\check{\mathbf{Q}}(n)\mathbf{G}(n)\right] - \\ - 2\Re\left\{\mathrm{tr}\left[\mathbf{G}^{\mathsf{H}}(n)\check{\mathbf{R}}(n)\right]\right\} + c\left(\mathbf{z}\right) - \frac{\rho}{2}\|\boldsymbol{\mu}\|_2^2, \end{aligned} \tag{27}$$

with the matrices $\check{\mathbf{Q}}(n)$ and $\check{\mathbf{R}}(n)$ defined as

$$\check{\mathbf{Q}}(n) = \hat{\mathbf{Q}}(n) + \frac{\rho}{2}\tilde{\mathbf{x}}(n-\tau)\tilde{\mathbf{x}}^{\mathsf{H}}(n-\tau), \tag{28}$$

$$\check{\mathbf{R}}(n) = \hat{\mathbf{R}}(n) + \frac{\rho}{2}\tilde{\mathbf{x}}(n-\tau)\left(\mathbf{z} + \boldsymbol{\mu}\right)^{\mathsf{H}}, \tag{29}$$

where ρ is a penalty parameter and $\boldsymbol{\mu}$ is the dual variable. Following the ADMM algorithm [14], alternating minimization of $\mathcal{L}$ in (27) with respect to the prediction filter and the splitting variable followed by a dual ascent results in the following iterative updates

$$\check{\mathbf{G}}^i(n) \leftarrow \hat{\mathbf{G}}(n) + \check{\mathbf{k}}(n)\left[\mathbf{z}^{i-1} + \boldsymbol{\mu}^{i-1} - \hat{\mathbf{u}}(n)\right]^{\mathsf{H}}, \tag{30}$$

$$\check{\mathbf{u}}^i(n) \leftarrow \left(\check{\mathbf{G}}^i(n)\right)^{\mathsf{H}}\tilde{\mathbf{x}}(n-\tau), \tag{31}$$

$$\mathbf{z}^i \leftarrow \arg\min_{\mathbf{z}} c(\mathbf{z}) + \frac{\rho}{2}\|\mathbf{z} - \check{\mathbf{u}}^i(n) + \boldsymbol{\mu}^{i-1}\|_2^2, \tag{32}$$

$$\boldsymbol{\mu}^i \leftarrow \boldsymbol{\mu}^{i-1} + \mathbf{z}^i - \check{\mathbf{u}}^i(n), \tag{33}$$

where $\hat{\mathbf{G}}(n)$ is the unconstrained filter computed in (19), $\hat{\mathbf{u}}(n) = \hat{\mathbf{G}}^{\mathsf{H}}(n)\tilde{\mathbf{x}}(n-\tau)$ is the undesired component estimated using the unconstrained filter $\hat{\mathbf{G}}(n)$, and

$$\check{\mathbf{k}}(n) = \frac{\hat{\mathbf{Q}}^{-1}(n)\tilde{\mathbf{x}}(n-\tau)}{\frac{2}{\rho} + \tilde{\mathbf{x}}^{\mathsf{H}}(n-\tau)\hat{\mathbf{Q}}^{-1}(n)\tilde{\mathbf{x}}(n-\tau)}, \tag{34}$$

is the gain vector for the ADMM iterations. The update for $\mathbf{z}$ in (32) is a projection step, which can be computed element-wise as

$$z_m^i \leftarrow \min\left\{\frac{\hat{\sigma}_{u,m}(n)}{\left|\check{u}_m^i(n) - \mu_m^{i-1}\right|}, 1\right\} \cdot \left(\check{u}_m^i(n) - \mu_m^{i-1}\right). \tag{35}$$

The obtained iterative updates in (30)-(33) can be interpreted as an iterative correction of the unconstrained filter $\hat{\mathbf{G}}(n)$ to obtain the constrained filter $\check{\mathbf{G}}(n)$ which satisfies the constraint in (24).

5. SIMULATIONS

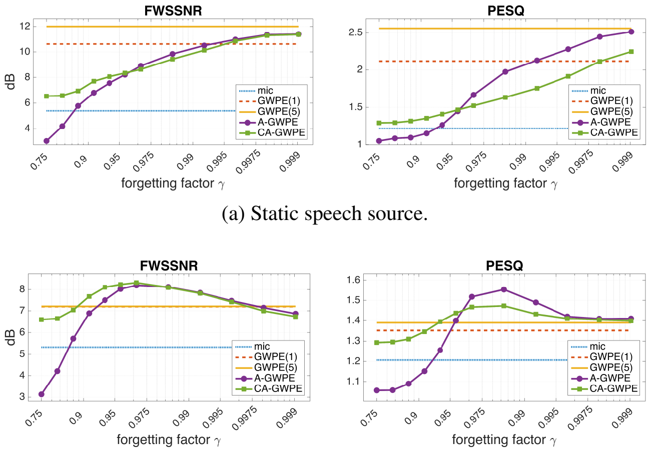
We consider an acoustic scenario with a single speech source and $M = 2$ microphones. The reverberation time was $T_{60} \approx 700$ ms, the distance between the microphones was approximately 14 cm, and the distance between the speech source and the microphones was approximately 2 m. The microphone signals were obtained by convolving the clean speech signal with measured RIRs [15]. The speech signal was constructed by concatenating a set of 4 utterances (2 male and 2 female) with a total length of approximately 11 s, sampled at $f_s = 16$ kHz.

We evaluate the speech dereverberation performance of the following methods: the batch GWPE with 1 and with 5 reweighting iterations (GWPE(1) and GWPE(5)), the adaptive GWPE (A-GWPE), and the proposed constrained adaptive GWPE (CA-GWPE). For all methods, the STFT is computed using a 64 ms Hann window with 16 ms shift. The prediction delay is set to $\tau = 2$ and the length of the prediction filter is set to $L_g = 20$. The value of the forgetting factor γ for the adaptive methods is selected between 0.75 and 0.999, the prediction filters are initialized with zeros, and the inverse matrices are initialized with a scaled identity matrix. The parameters required for the PSD estimation in (20)-(22) are set to $\alpha = 0.3$ and $T_d = 50$ ms. The ADMM iterations in (30)-(33) are performed 20 times. To reduce the initialization effects, we first processed an additional 5 s speech signal before processing the test signal described above. The dereverberation performance is evaluated in terms of frequency-weighted segmental signal-to-noise ratio (FWSSNR) and PESQ [15]. We evaluate the performance using the first output signal (although the methods generate $M = 2$ output channels) and the clean speech signal as the reference.

In the first experiment, we consider a scenario with the speech source positioned at 45° left of the broadside direction of the array. Fig. 1a depicts the obtained instrumental measures for the reverberant first microphone signal and the output signals obtained using the considered dereverberation methods. It can be observed that the batch methods result in significant improvements compared to the microphone signal, with GWPE(5) performing better than GWPE(1). As expected, the performance of the adaptive methods highly depends on the forgetting factor γ. For relatively large values of the forgetting factor γ, the adaptive A-GWPE and CA-GWPE result in a similar performance as the batch method. In terms of PESQ, the performance of the CA-GWPE method is somewhat lower than the A-GWPE method, since the proposed constraint may result in a lower suppression of reverberation. For relatively small values of the forgetting factor γ, the performance of both adaptive methods is significantly decreased. On the one hand, the A-GWPE method results in a significantly worse performance than the microphone signal due to the significant cancellation of the desired signal. On the other hand, the CA-GWPE method still results in some improvements over the microphone signal due to the proposed constraint, which prevents excessive signal cancellation.

In the second experiment, we consider a scenario with the speech source switching between two positions at 45° left and 45° right of the broadside direction of the array. The test signal is obtained by alternating the source position per utterance, without any overlap between successive utterances. Fig. 1b depicts the obtained instrumental measures for the reverberant microphone signal and the output signals obtained using the considered speech dereverberation methods. As expected, it can be observed that the improvements with the batch method are much smaller than for the static case, with GWPE(1) and GWPE(5) resulting in almost the same performance. For relatively large values of the forgetting factor γ, the adaptive methods again achieve a very similar performance as the batch methods. By slightly decreasing the forgetting factor, both the A-GWPE and CA-GWPE methods outperform the batch method. By further decreasing the forgetting factor, the performance of the adaptive methods is in general decreased. On the one hand, the A-GWPE method again results in a significantly worse performance than the microphone signal. On the other hand, the CA-GWPE method still results in some improvements over the microphone signal due to the proposed constraint. To better illustrate the effect of the proposed constraint, Fig. 2 depicts the spectrograms of the clean speech signal, the reverberant microphone signal, and the output signals obtained using the A-GWPE and CA-GWPE for two exemplary values of the forgetting factor γ. Comparing the spectrograms of the output signals obtained using the smaller γ, it can be observed that the A-GWPE method results in almost complete cancellation of the desired speech signal at the output, while the CA-GWPE preserves the desired speech much better. Comparing the spectrograms with the larger γ it can be seen that A-GWPE and CA-GWPE result in a very similar output signal, with CA-GWPE resulting in a slightly reduced dereverberation.

In summary, the simulations confirm that the constrained linear prediction for adaptive MC speech dereverberation results in a significant increase in the performance for small values of the forgetting factor, i.e., when the prediction filters adapt quickly, while at the same time not having a large influence on the performance for large values of the forgetting factor, i.e., when the prediction filters change slowly. Therefore, the proposed constraint improves robust the robustness of the dereverberation method with respect to the selection of the forgetting factor.



(a) Static speech source.

(b) Switching between two speech sources.

Fig. 1. Instrumental measures vs. forgetting factor γ for the considered experimental scenarios.

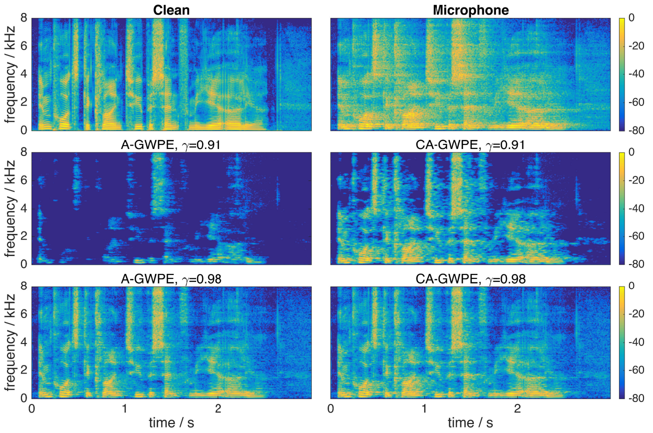


Fig. 2. Spectrograms of the clean speech signal and the microphone signal (top), and the output signals obtained using $\gamma = 0.91$ (middle) and $\gamma = 0.98$ (bottom).

6. CONCLUSION

In this paper we have presented a multi-channel speech dereverberation method based on constrained linear prediction. We have proposed to use a statistical model of the undesired reverberation to constrain the power of the estimated undesired component, aiming to increase the robustness of adaptive MCLP-based speech dereverberation with respect to the forgetting factor, making it more usable in scenarios when the prediction filters need to quickly adapt. The constrained prediction filter has been iteratively computed using the alternating direction method of multipliers. Simulation results have shown that the proposed constrained method significantly improves the performance for small values of the forgetting factor, while not having a large influence on the performance for large values of the forgetting factor, and therefore improves robustness with respect to the selection of the forgetting factor.

7. REFERENCES

[1] T. Yoshioka, A. Sehr, M. Delcroix, K. Kinoshita, R. Maas, T. Nakatani, and W. Kellermann, "Making machines understand us in reverberant rooms: Robustness against reverberation for automatic speech recognition," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 114–126, Nov. 2012.

[2] P. A. Naylor and N. D. Gaubitch, *Speech Dereverberation*, Springer, 2010.

[3] K. Kinoshita, M. Delcroix, T. Nakatani, and M. Miyoshi, "Suppression of late reverberation effect on speech signal using long-term multiple-step linear prediction," *IEEE Trans. Audio Speech Lang. Process.*, vol. 17, no. 4, pp. 534–545, May 2009.

[4] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B. H. Juang, "Speech dereverberation based on variance-normalized delayed linear prediction," *IEEE Trans. Audio Speech Lang. Process.*, vol. 18, no. 7, pp. 1717–1731, Sept. 2010.

[5] A. Jukić, T. van Waterschoot, T. Gerkmann, and S. Doclo, "Multi-channel linear prediction-based speech dereverberation with sparse priors," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 9, pp. 1509–1520, Sept. 2015.

[6] A. Jukić, T. van Waterschoot, T. Gerkmann, and S. Doclo, "A general framework for multi-channel speech dereverberation exploiting sparsity," in *Proc. AES 60th Int. Conf.*, Leuven, Belgium, Feb. 2016.

[7] 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. (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 3733–3736.

[8] T. Yoshioka and T. Nakatani, "Dereverberation for reverberation-robust microphone arrays," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Marrakech, Morocco, Sept. 2013, pp. 1–5.

[9] S. Haykin, *Adaptive Filter Theory,*, Prentice Hall, 3 edition, 2013.

[10] K. Lebart, J. M. Boucher, and P. N. Denbigh, "A new method based on spectral subtraction for speech dereverberation," *Acta Acustica*, vol. 87, no. 3, pp. 359–366, May-Jun 2001.

[11] 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.

[12] A. Jukić, T. van Waterschoot, T. Gerkmann, and S. Doclo, "Group sparsity for MIMO speech dereverberation," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, New Paltz, NY, USA, Oct. 2015, pp. 1–5.

[13] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Las Vegas, NV, USA, March 2008, pp. 3869–3872.

[14] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[15] K. Kinoshita, M. Delcroix, S. Gannot, E. A. P. Habets, R. Haeb-Umbach, W. Kellermann, V. Leutnant, R. Maas, T. Nakatani, B. Raj, A. Sehr, and T. Yoshioka, "A summary of the REVERB challenge: state-of-the-art and remaining challenges in reverberant speech processing research," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, pp. 1–19, 2016.