ADAPTIVE DEREVERBERATION USING MULTI-CHANNEL LINEAR PREDICTION WITH DEFICIENT LENGTH FILTER

Guanjun Li[1,2], *Shan Liang*[1], *Shuai Nie*[1], *Wenju Liu*[1]

[1]National Laboratory of Patten Recognition, Institute of Automation, Chinese Academy of Sciences, China
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences, China
{guanjun.li, sliang, shuai.nie, lwj}@nlpr.ia.ac.cn

ABSTRACT

In almost all adaptive dereverberation algorithms based on the multi-channel linear prediction (MCLP) model, it is assumed that the filter length can cover the reverberation time. However, in many practical situations, a deficient length filter, whose length is less than the reverberation time, is employed in consideration of computational cost. A deficient length filter fails to fully model the late reverberation, resulting in degraded performance. In this paper, we present a new MCLP-based adaptive dereverberation algorithm to improve the dereverberation performance when using a deficient length filter. We introduce a gain and use the filter coefficients estimated from the previous frame to track the MCLP modeling errors of the current frame. The gain and the filter coefficients are jointly optimized and solved by using an alternating minimization technique. Experimental results show the superiority of the proposed algorithm. The shorter the filter length is, the more advantageous the proposed algorithm is.

Index Terms— Dereverberation, multi-channel linear prediction, deficient length filter, adaptive processing

1. INTRODUCTION

Speech signals captured by distant microphones within an enclosure are usually contaminated by reverberation. To alleviate the negative effect of reverberation, various dereverberation techniques have been proposed [1]. At present, multichannel linear prediction (MCLP) [2] has been considered to be one of the most appealing frameworks for dereverberation. MCLP utilizes an autoregressive filter to predict late reverberation using past microphone signals, and it is able to preserve speech quality while suppressing reverberation effectively [3].

To cope with adaptive processing, Yoshioka [4, 5] applies a recursive least squares (RLS) algorithm to update the filter coefficients. Jukić [6, 7] exploits a constrained objective function to deal with the signal cancellation problem when using a small forgetting factor in the RLS algorithm. Braun [8] further estimates the filter coefficients using a Kalman filter. In [4, 9, 10], it is recommended that the filter length should cover the reverberation time. Unfortunately, both the RLS algorithm and the Kalman filter suffer from high computational complexity, e.g., the computational complexity of the RLS algorithm is $\mathcal{O}(N^2)$. Thus, in order to minimize the processing delay in practice, a deficient length filter is preferred, i.e., the actual filter length is typically less than the reverberation time. A deficient length filter is unable to fully model the late reverberation, resulting in degraded dereverberation performance. To the best of our knowledge, little attention has been paid to adaptive dereverberation using a deficient length filter.

In this paper, a new MCLP-based adaptive dereverberation algorithm is proposed to improve the dereverberation performance in the case of using a deficient length filter. Assuming that the filter coefficients are slowly time-varying, we use a gain and the filter coefficients obtained from the previous frame to track the MCLP modeling errors of the current frame. Furthermore, the gain and the filter coefficients are jointly optimized and they are solved by an alternating minimization technique [11]. The proposed algorithm has the same computational complexity as the RLS algorithm, but achieves better dereverberation performance with a deficient length filter. Moreover, the proposed algorithm becomes identical to the RLS algorithm if the gain is forced to be zero.

2. MCLP MODEL

We consider a single speech source captured by M microphones in a noiseless reverberant environment. Let $x_m(n,k)$ denote the short-time Fourier transform (STFT) representation of the m-th microphone signal at time index n and frequency index k. The first microphone signal can be decomposed as $x_1(n,k) = d(n,k) + u(n,k)$, where $d(n,k)$ is the desired signal and $u(n,k)$ is the late reverberation. In the following, we omit the frequency index k, since each subband can be processed independently. The MCLP model [2] can be

This work was supported in part by the National Key R&D Plan of China (No. 2016YFB1001404) and the China National Nature Science Foundation (No. 61573357, No. 61503382, No. 61403370, No. 61273267, No. 91120303). This work was sponsored by CCF-Tencent Open Fund (No. RAGR20180106) and Xueersi Cooperation Fund.

written as

$$x_1(n) = d(n) + \underbrace{\mathbf{c}^H(n)\widetilde{\mathbf{x}}_{\tau,L_c}(n)}_{u(n)}, \tag{1}$$

where $\mathbf{c}(n) \in \mathbb{C}^{ML_c\times 1}$ is a prediction filter containing L_c taps per channel, $(.)^H$ denotes the conjugate transposition operator and $\widetilde{\mathbf{x}}_{\tau,L_c}(n) \in \mathbb{C}^{ML_c\times 1}$ is a microphone signal buffer defined as

$$\begin{aligned}\widetilde{\mathbf{x}}_{\tau,L_c}(n) = \; & [x_1(n-\tau), \ldots, x_M(n-\tau), \ldots, \\ & x_1(n-\tau-L_c+1), \ldots, x_M(n-\tau-L_c+1)]^T, \end{aligned} \tag{2}$$

where $(.)^T$ denotes the transposition operator and τ is a prediction delay corresponding to the duration of the early reverberation.

However, as mentioned in the introduction, in order to reduce the processing delay in practice, a deficient length filter is often used, i.e., the actual filter length is less than the reverberation time. A deficient length filter can not model the late reverberation well, and modeling errors can result in reduced dereverberation performance. The goal of this work is to design an adaptive dereverberation algorithm that takes into account MCLP modeling errors so that it can suppress reverberation more accurately with a deficient length filter.

3. PROPOSED DEREVERBERATION ALGORITHM

Suppose that in practice we use a filter $\mathbf{g}(n)$ of length L_g ($L_g < L_c$) to predict the late reverberation[1], we write $\mathbf{c}(n)$ and $\widetilde{\mathbf{x}}_{\tau,L_c}(n)$, respectively, as

$$\mathbf{c}(n) = \begin{bmatrix}\mathbf{g}(n)\\ \bar{\mathbf{c}}(n)\end{bmatrix}, \quad \widetilde{\mathbf{x}}_{\tau,L_c}(n) = \begin{bmatrix}\widetilde{\mathbf{x}}_{\tau,L_g}(n)\\ \widetilde{\mathbf{x}}_{\tau+L_g,L_c-L_g}(n)\end{bmatrix}, \tag{3}$$

where $\bar{\mathbf{c}}(n)$ is the prediction filter corresponding to the signal buffer $\widetilde{\mathbf{x}}_{\tau+L_g,L_c-L_g}(n)$. Then Eq. (1) can be rewritten as

$$x_1(n) = d(n) + \mathbf{g}^H(n)\widetilde{\mathbf{x}}_{\tau,L_g}(n) + e(n), \tag{4}$$

where $e(n) = \bar{\mathbf{c}}^H(n)\widetilde{\mathbf{x}}_{\tau+L_g,L_c-L_g}(n)$ represents the modeling error term. We use $d_{\text{mclp}}(n) = x_1(n) - \mathbf{g}^H(n)\widetilde{\mathbf{x}}_{\tau,L_g}(n)$ to represent the output of the MCLP, then Eq. (4) can be rewritten after some manipulations as

$$d_{\text{mclp}}(n) = d(n) + e(n). \tag{5}$$

Just like the single-channel speech enhancement method [12], we estimate $e(n)$ by applying a gain $w(n)$ to $d_{\text{mclp}}(n)$, that is, $\hat{e}(n) = w(n)d_{\text{mclp}}(n)$, where superscript ˆ denotes an estimated value. If we use the definition of $d_{\text{mclp}}(n)$ directly, a good estimate of $\mathbf{g}(n)$ is required to obtain $\hat{e}(n)$. However, this is not easy to achieve since the best available estimate at time index n and before estimating $e(n)$ is $\mathbf{g}(n-1)$. To solve this issue, we assume that the reverberation environment is stationary or the filter coefficients are slowly time-varying. Under the assumptions made above, we can get

$$\hat{e}(n) = w(n)\left[x_1(n) - \mathbf{g}^H(n-1)\widetilde{\mathbf{x}}_{\tau,L_g}(n)\right]. \tag{6}$$

Now the gain $w(n)$ and the filter $\mathbf{g}(n)$ can be estimated in the minimum mean-square error (MMSE) sense by minimizing the cost function

$$J\left(w(n), \mathbf{g}(n)\right) = E\left\{|d_{\text{mclp}}(n) - \hat{e}(n) - d(n)|^2\right\}. \tag{7}$$

Jointly minimizing (7) with respect to $w(n)$ and $\mathbf{g}(n)$ is not straightforward. Here, we resort to an alternating minimization technique [11], which minimizes the cost function for one variable while keeping the other one fixed. To be more precise, we iteratively perform the following update rules for each time index n:

$$1)\quad \hat{w}(n) = \arg\min_{w(n)} J\left(w(n)|\hat{\mathbf{g}}(n-1)\right), \tag{8}$$

$$2)\quad \hat{\mathbf{g}}(n) = \arg\min_{\mathbf{g}(n)} J\left(\mathbf{g}(n)|\hat{w}(n)\right), \tag{9}$$

In the following subsections, we describe the procedures for accomplishing (8) and (9).

3.1. Gain update

Given knowledge of the filter coefficients estimated from the previous frame, the cost function $J(w(n)|\hat{\mathbf{g}}(n-1))$ can be written as

$$J(w(n)|\hat{\mathbf{g}}(n-1)) = E\left\{\left|(1-w(n))d_{\text{mclp}}(n) - d(n)\right|^2\right\}, \tag{10}$$

where $\hat{\mathbf{g}}(n-1)$ is contained in $d_{\text{mclp}}(n)$. $w(n)$ that minimizes (10) is easily derived by setting $\frac{\partial J(w(n)|\hat{\mathbf{g}}(n-1))}{\partial w(n)} = 0$, leading to

$$\hat{w}(n) = 1 - \frac{E\{d^*_{\text{mclp}}(n)d(n)\} + E\{d_{\text{mclp}}(n)d^*(n)\}}{2E\{|d_{\text{mclp}}(n)|^2\}}, \tag{11}$$

where superscript $*$ denotes the complex conjugate. By assuming that $d(n)$ and $e(n)$ are independent and $d(n)$ has zero mean, i.e., $E\{d^*(n)e(n)\} = 0$, using Eq. (5), Eq. (11) can be rewritten as

$$\hat{w}(n) = 1 - \frac{\sigma_d^2(n)}{E\{|d_{\text{mclp}}(n)|^2\}}, \tag{12}$$

where $\sigma_d^2(n) = E\{d(n)d^*(n)\}$ is the power spectral density (PSD) of the desired signal.

In order to estimate $\sigma_d^2(n)$, the exponential decay model [13, 14] is employed in calculating the late reverberant PSD

[1] Although $\mathbf{g}(n) \in \mathbb{C}^{ML_g\times 1}$, for clarity of presentation, we use L_g to represent the filter length when given a microphone array with M microphones in advance.

and $\hat{\sigma}_d^2(n)$ is then obtained by using the spectral subtractive method[5, 6], i.e.

$$\hat{\sigma}_x^2(n) = \alpha\hat{\sigma}_x^2(n-1) + (1-\alpha)|x_1(n)|^2 \tag{13}$$

$$\hat{\sigma}_u^2(n) = e^{\frac{-6R\tau \ln 10}{T_{60} f_s}} \hat{\sigma}_x^2(n-\tau) \tag{14}$$

$$\hat{\sigma}_{\tilde{d}}^2(n) = \max\{|x_1(n)|^2 - \hat{\sigma}_u^2(n), 0\} \tag{15}$$

$$\hat{\sigma}_d^2(n) = \alpha\hat{\sigma}_d^2(n-1) + (1-\alpha)\hat{\sigma}_{\tilde{d}}^2(n) \tag{16}$$

where $\hat{\sigma}_u^2(n)$ is the estimate of the late reverberant PSD, α is a smoothing parameter, R is the STFT time shift in samples, T_{60} denotes the reverberation time and f_s indicates the sampling frequency in Hz.

In order to estimate $E\{|d_{\text{mclp}}(n)|^2\}$, the recursive averaging approach can be used. Specifically,

$$E\{|d_{\text{mclp}}(n)|^2\} = \beta E\{|d_{\text{mclp}}(n-1)|^2\} + (1-\beta)|x_1(n) - \mathbf{g}^H(n-1)\widetilde{\mathbf{x}}_{\tau,L_g}(n)|^2, \tag{17}$$

where β is a smoothing parameter.

3.2. Filter coefficients update

Given knowledge of the current gain $\hat{w}(n)$, according to (9), the filter estimate $\hat{\mathbf{g}}(n)$ requires that $\frac{\partial J(\mathbf{g}(n)|\hat{w}(n))}{\partial \mathbf{g}(n)} = 0$. In this operation, it is necessary to estimate $E\{\widetilde{\mathbf{x}}_{\tau,L_g}(n)d(n)\}$, but it is not easy to obtain.

Instead of solving (9), we can alternatively update $\hat{\mathbf{g}}(n)$ using the maximum likelihood estimation (MLE) algorithm. In this case, cost function (7) is not necessarily minimized. Nevertheless, we found experimentally that using MLE also allowed the proposed algorithm to have a better performance than the comparison algorithms.

To begin with, we substitute $\hat{e}(n)$ for $e(n)$ and then insert Eq. (6) into Eq. (4), obtaining

$$x_1(n) = \frac{d(n)}{1-w(n)} + \frac{\left[\mathbf{g}^H(n) - w(n)\mathbf{g}^H(n-1)\right]\widetilde{\mathbf{x}}_{\tau,L_g}(n)}{1-w(n)}. \tag{18}$$

If we assume that the desired signal $d(n)$ follows a zero-mean complex Gaussian distribution with time-varying variance $\sigma_d^2(n)$ [2], i.e., $p(d(n)) = \frac{1}{\pi\sigma_d^2(n)} e^{-\frac{|d(n)|^2}{\sigma_d^2(n)}}$, using Eq. (18), the weighted exponential likelihood function can be defined as

$$\begin{aligned}\mathcal{L}(\mathbf{g}(n)) &= \sum_{t=1}^{n} \gamma^{n-t} \ln p\left(x_1(t)\right) \\ &= -\sum_{t=1}^{n} \gamma^{n-t} \frac{|x_1(t) - \mathbf{g}(n)\widetilde{\mathbf{x}}_{\tau,L_g}(t) - \hat{e}(t)|^2}{\sigma_d^2(t)} + c,\end{aligned} \tag{19}$$

Algorithm 1 Proposed algorithm per subband.

Initialization: $\hat{\mathbf{g}}(0) = \mathbf{0}_{ML_g \times 1}$, $\mathbf{\Phi}(0) = \mathbf{I}_{MLg}$

for $n = 1, 2, 3, \cdots$ **do**
- Compute $\hat{\sigma}_d^2(n)$ using (13), (14), (15), (16)
- $\hat{d}_{\text{mclp}}(n) = x_1(n) - \mathbf{g}^H(n-1)\widetilde{\mathbf{x}}_{\tau,L_g}(n)$
- Compute $w(n)$ using (12)
- $\mathbf{k}(n) = \frac{\mathbf{\Phi}(n-1)\widetilde{\mathbf{x}}_{\tau,L_g}(n)}{\gamma\hat{\sigma}_d^2(n) + \widetilde{\mathbf{x}}_{\tau,L_g}^H(n)\mathbf{\Phi}(n-1)\widetilde{\mathbf{x}}_{\tau,L_g}(n)}$
- $\mathbf{\Phi}(n) = \frac{1}{\gamma}\left[\mathbf{\Phi}(n-1) - \mathbf{k}(n)\widetilde{\mathbf{x}}_{\tau,L_g}^H(n)\mathbf{\Phi}(n-1)\right]$
- $\hat{\mathbf{g}}(n) = \hat{\mathbf{g}}(n-1) + \mathbf{k}(n)\left[1 - w(n)\right]\hat{d}_{\text{mclp}}^*(n)$
- $\hat{d}(n) = x_1(n) - \hat{\mathbf{g}}^H(n)\widetilde{\mathbf{x}}_{\tau,L_g}(n) - w(n)\hat{d}_{\text{mclp}}(n)$
- Output $\hat{d}(n)$ as the desired signal

end for

where $0 < \gamma < 1$ is a forgetting factor and c is a constant term with respect to $\mathbf{g}(n)$. The maximization of (19) requires that $\frac{\partial \mathcal{L}(\mathbf{g}(n))}{\partial \mathbf{g}(n)} = 0$, leading to

$$\hat{\mathbf{g}}(n) = \mathbf{\Phi}(n)\mathbf{R}(n), \tag{20}$$

where

$$\mathbf{\Phi}(n) = \left(\sum_{t=1}^{n} \gamma^{n-t} \frac{\widetilde{\mathbf{x}}_{\tau,L_g}(t)\widetilde{\mathbf{x}}_{\tau,L_g}^H(t)}{\hat{\sigma}_d^2(t)}\right)^{-1}, \tag{21}$$

$$\mathbf{R}(n) = \sum_{t=1}^{n} \gamma^{n-t} \frac{\widetilde{\mathbf{x}}_{\tau,L_g}(t)\left[x_1^*(t) - \hat{e}^*(t)\right]}{\hat{\sigma}_d^2(t)}. \tag{22}$$

By applying the matrix inversion lemma[15], $\hat{\mathbf{g}}(n)$ can be recursively computed, yielding the proposed dereverberation algorithm listed in Algorithm 1.

The proposed algorithm has the same computational complexity as the RLS algorithm [4, 5], and degenerates into the RLS algorithm when the gain $w(n)$ is forced to be zero.

4. EXPERIMENTAL RESULTS

Two experiments were performed using 20 test utterances with an average length of 25 s. Each test utterance contained concatenated 8 utterances from the TIMIT database [16] and the sampling frequency was 16 kHz. The reverberant observations were generated by convolving each test utterance with measured room impulse responses (RIRs) from REVERB challenge [3] and the source-microphone distance was about 2 m. The STFT was computed using a 32 ms Hann window with 50% overlap and the prediction delay in (2) was set to $\tau = 2$. As for the smoothing parameter, we took α and β to be 0.5 and 0.3, respectively. The forgetting factor was set at $\gamma = 0.99$ as proposed in [4, 5]. The performance was evaluated using the five measures suggested in [3], i.e., the perceptual evaluation of speech quality (PESQ) [17], the

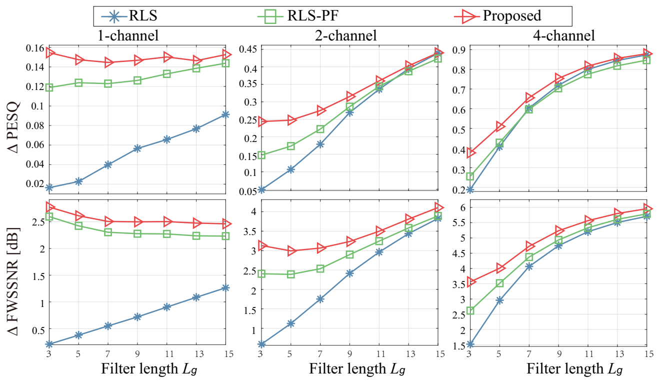

Fig. 1. Improvements of performance measures versus filter length L_g in 1-channel (left), 2-channel (center) and 4-channel (right) scenarios.

frequency-weighted segmental SNR (FWSSNR) [18], the cepstral distance (CD) [18], the speech to reverberation modulation energy ratio (SRMR) [19] and the log likelihood ratio (LLR) [18]. We reported the performance measures averaged over all the test utterances.

For reference, we implemented the RLS algorithm [4, 5] and the RLS algorithm followed by a post-filter (RLS-PF), which simply treats the gain $[1 - w(n)]$ as a post-filter to further suppress residual reverberation [20]. Unlike the proposed algorithm, the gain in RLS-PF does not affect the update of the prediction filter coefficients.

First, we evaluated the effect of filter length on the performance of the algorithms in 1-channel, 2-channel and 4-channel scenarios with $T_{60} \approx 500$ ms. In [2, 6, 8], the filter length L_g is typically set as $L_g \geq 15$. Thus, to simulate the use of deficient length filter, we set $L_g \leq 15$ in this experiment. Fig. 1 shows the positive improvements of PESQ and FWSSNR compared to the unprocessed microphone signal (i.e., the delta values). As can be seen from Fig. 1, the shorter the filter length is, the more advantageous the proposed algorithm is than the RLS algorithm. This is due to the fact that as the filter length decreases, the MCLP modeling errors become larger. Compared with the proposed algorithm, the RLS algorithm can not deal with the modeling errors. We can also find that the proposed algorithm is superior to RLS-PF at different filter lengths, because the separate optimization of RLS-PF is suboptimal with respect to the joint optimization of the proposed algorithm. Fig. 2 shows spectrogram parts of the clean signal, the reverberant microphone signal and the processed signals obtained using the RLS algorithm and using the proposed algorithm when $L_g = 3$ and $M = 2$. It can be observed that the proposed algorithm achieves more reverberation suppression.

Next, we systematically evaluated the performance of the algorithms in different reverberation scenarios. To simplify the experiment, we consider two 2-channel scenarios with $T_{60} \approx 500$ ms and $T_{60} \approx 700$ ms, respectively, which can be regarded as the most common reverberation scenarios in real-world environments. In order to simulate the use of deficient length filter, we set the filter length as $L_g = 9$. Table 1 reports the comparative results in terms of PESQ, FWSSNR, CD, SRMR and LLR. As observed, the proposed algorithm outperforms other algorithms in all the measures.

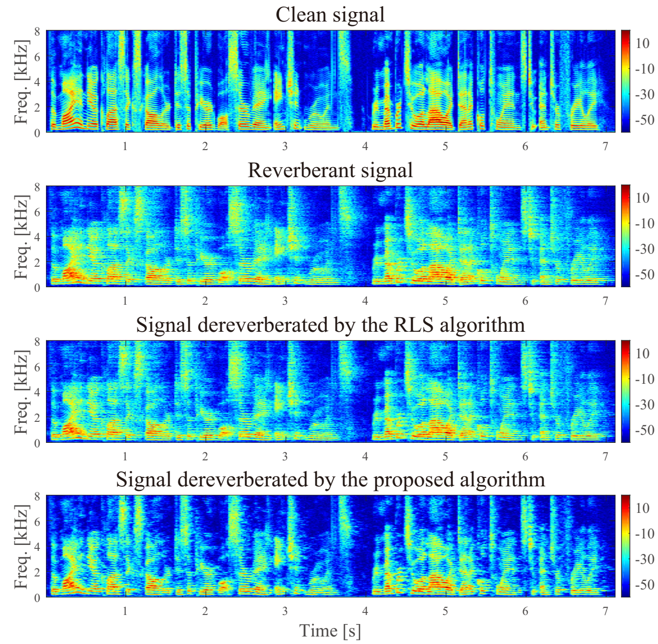

Fig. 2. Spectrograms of clean signal, observed signal at the first microphone and processed signal.

Table 1. Dereverberation results

Method	PESQ	FWSSNR [dB]	CD [dB]	SRMR [dB]	LLR
	Reverberation time $T_{60} \approx 500$ ms				
Unprocessed	1.692	4.068	4.278	2.740	0.711
RLS	1.962	6.481	3.525	3.848	0.531
RLS-PF	1.979	6.964	3.436	4.256	0.520
Proposed	**2.008**	**7.301**	**3.353**	**4.490**	**0.500**
	Reverberation time $T_{60} \approx 700$ ms				
Unprocessed	1.658	3.816	4.351	2.382	0.742
RLS	1.899	5.731	3.552	3.546	0.573
RLS-PF	1.894	6.009	3.505	3.922	0.549
Proposed	**1.918**	**6.292**	**3.445**	**4.039**	**0.532**

5. CONCLUSIONS

In this paper, we proposed a new adaptive dereverberation algorithm considering the MCLP modeling errors in the case of using a deficient length filter. The filter coefficients and the gain used to track the modeling errors are joint optimized. Experimental results show the superiority of the proposed algorithm when using a deficient length filter. Since computational cost should be paid attention to in real applications, the proposed algorithm shows more practical potentialities.

6. REFERENCES

[1] Patrick A Naylor and Nikolay D Gaubitch, *Speech dereverberation*, Springer Science & Business Media, 2010.

[2] Tomohiro Nakatani, Takuya Yoshioka, Keisuke Kinoshita, Masato Miyoshi, and Biing-Hwang Juang, "Speech dereverberation based on variance-normalized delayed linear prediction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1717–1731, 2010.

[3] Keisuke Kinoshita, Marc Delcroix, Sharon Gannot, Emanuël AP Habets, Reinhold Haeb-Umbach, Walter Kellermann, Volker Leutnant, Roland Maas, Tomohiro Nakatani, Bhiksha Raj, et al., "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. 7, 2016.

[4] Takuya Yoshioka, Hideyuki Tachibana, Tomohiro Nakatani, and Masato Miyoshi, "Adaptive dereverberation of speech signals with speaker-position change detection," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 3733–3736.

[5] Takuya Yoshioka and Tomohiro Nakatani, "Dereverberation for reverberation-robust microphone arrays," in *Signal Processing Conference*, 2014, pp. 1–5.

[6] Ante Jukić, Toon van Waterschoot, and Simon Doclo, "Adaptive speech dereverberation using constrained sparse multichannel linear prediction," *IEEE Signal Processing Letters*, vol. 24, no. 1, pp. 101–105, 2017.

[7] Ante Jukić, Zichao Wang, Toon van Waterschoot, Timo Gerkmann, and Simon Doclo, "Constrained multichannel linear prediction for adaptive speech dereverberation," in *Acoustic Signal Enhancement (IWAENC), 2016 IEEE International Workshop on*. IEEE, 2016, pp. 1–5.

[8] Sebastian Braun and Emanuël AP Habets, "Online dereverberation for dynamic scenarios using a kalman filter with an autoregressive model," *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1741–1745, 2016.

[9] Tomohiro Nakatani, Takuya Yoshioka, Keisuke Kinoshita, Masato Miyoshi, and Biing Hwang Juang, "Blind speech dereverberation with multi-channel linear prediction based on short time fourier transform representation," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 85–88.

[10] Takuya Yoshioka, Tomohiro Nakatani, Keisuke Kinoshita, and Masato Miyoshi, "Speech dereverberation and denoising based on time varying speech model and autoregressive reverberation model," in *Speech Processing in Modern Communication*, pp. 151–182. Springer, 2010.

[11] U Niesen, D Shah, and G Wornell, "Adaptive alternating minimization algorithms," in *IEEE International Symposium on Information Theory*, 2007, pp. 1641–1645.

[12] Jacob Benesty, Jingdong Chen, and Emanul A. P. Habets, *Single-Channel Speech Enhancement with a Gain*, Springer Berlin Heidelberg, 2012.

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

[14] Emanuël AP Habets, Sharon Gannot, and Israel Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 770–773, 2009.

[15] Simon Haykin, *Adaptive filter theory (3rd ed.)*, Prentice-Hall, Inc., 1996.

[16] J. S Garofolo, L. F Lamel, W. M Fisher, J. G Fiscus, and D. S Pallett, "Darpa timit acoustic-phonetic continous speech corpus cd-rom. nist speech disc 1-1.1," *Nasa Sti/recon Technical Report N*, vol. 93, 1993.

[17] A. W Rix, J. G Beerends, M. P Hollier, and A. P Hekstra, "Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001. Proceedings*, 2002, pp. 749–752 vol.2.

[18] Yi Hu and P. C. Loizou, *Evaluation of Objective Quality Measures for Speech Enhancement*, IEEE Press, 2008.

[19] Tiago H. Falk, Chenxi Zheng, and Wai Yip Chan, "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech," *IEEE Transactions on Audio Speech & Language Processing*, vol. 18, no. 7, pp. 1766–1774, 2010.

[20] Thomas Dietzen, Simon Doclo, Ann Spriet, Wouter Tirry, Marc Moonen, and Toon van Waterschoot, "Low-complexity kalman filter for multi-channel linear-prediction-based blind speech dereverberation," in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2017 IEEE Workshop on*. IEEE, 2017, pp. 284–288.