EUSIPCO 2013 1569744623

DEREVERBERATION IN NOISY ENVIRONMENTS USING REFERENCE SIGNALS AND A MAXIMUM LIKELIHOOD ESTIMATOR

Sebastian Braun and Emanuël A.P. Habets

International Audio Laboratories Erlangen†, Am Wolfsmantel 33, 91058 Erlangen, Germany

ABSTRACT

In speech communication systems the received microphone signals are commonly degraded by reverberation and ambient noise that can decrease the fidelity and intelligibility of a desired speaker. Reverberation can be modeled as non-stationary diffuse sound which is not directly observable. In this work, we derive an optimal signal-dependent informed spatial filter to reduce jointly reverberation and noise. In addition, we derive a novel maximum likelihood estimator for the power spectral density (PSD) of the diffuse sound, that makes use of a set of linearly independent reference signals. Experimental results show, that the proposed method provides an accurate estimate of the diffuse PSD, even with a small number of microphones. The proposed method outperforms two existing methods and is able to provide an optimal tradeoff between noise reduction and dereverberation.

Index Terms— Speech enhancement, dereverberation, spatial filtering

1. INTRODUCTION

There is a growing market for hands-free communication devices for cars, smart TVs and mobile device. The received microphone signals of these devices are commonly degraded by reverberation and ambient noise that can decrease the fidelity and intelligibility of a desired speaker. The effect of early reflections mainly results in coloration while late reverberation increases the duration of speech phonemes. Whereas noise can be observed during periods in which the desired speakers are inactive, reverberation is not directly observable. In addition, the reverberation is highly time-varying while the noise can often be assumed to be time-invariant or slowly time-varying.

Both single- and multi-microphone techniques have been proposed to reduce reverberation (see [1] and the references therein). Many of these techniques require information about the sound scene, which in practice is difficult to gather, e.g. the acoustic impulse responses (AIRs) or the reverberation time. Recently, an informed spatial filter was proposed for dereverberation using a spherical microphone array [2]. The filter incorporates instantaneous information about the diffuseness of the sound field into the design of the filter and does not require an estimate of the AIRs or the reverberation time. For linear arrays, there are some differences to consider, but the principle of the spatial filter in the spherical harmonic domain can be also applied to other array types. In [3,4] a reference signal was used to estimate the PSD of the late reverberation. In [3], a statistical model is employed, whereas [4] uses an blind source separation technique to directly estimate the late reverberant signal.

In this work, an informed spatial filter is derived in a similar manner to [2], but generalized for linear arrays. Moreover, the PSD of the late reverberation is estimated from a set of linearly independent reference signals in a maximum likelihood sense. The proposed estimator was inspired by the work presented in [5], were an maximum likelihood estimator for time-varying ambient noise with a fixed coherence matrix was proposed. The advantage of the proposed method is that only information of the direction of arrival (DOA) of the desired source is required. In contrast to [3,4], we employ an optimal spatial filter and explicitly take into account ambient noise.

The paper is organized as follows. In Section 2, the problem is formulated. In Section 3, the optimal spatial filter is derived. The reference signals and the maximum likelihood estimator for the diffuse sound PSD using multiple reference signals is derived are derived in Section 4. Section 5 presents the simulation setup, evaluates the performance of the proposed method and compare it to two existing methods. Finally, conclusions are drawn and the work is summarized.

2. PROBLEM FORMULATION

We consider an array of M microphones capturing the sound field. In the short-time Fourier transform (STFT) domain, the microphone signals are written into vectors of length M so that $\mathbf{y}(k,n) = [Y_1(k,n), \ldots, Y_M(k,n)]^T$, where k denotes the STFT coefficent index and n the time frame index. Our model assumes a single non-moving speaker in a room with additive noise. The microphone signals can be described by

$$\mathbf{y}(k,n) = \mathbf{d}(k)S(k,n) + \mathbf{x}_{\mathrm{R}}(k,n) + \mathbf{v}(k,n), \qquad (1)$$

*This research was partly supported by a Grant from the GIF, the German-Israeli Foundation for Scientific Research and Development.

†A joint institution of the University Erlangen-Nuremberg and Fraunhofer IIS, Germany.

where $S(k, n)$ denotes the direct sound of the desired speaker as received by a reference microphone, $\mathbf{d}(k)$ is the relative complex propagation vector of the direct sound from the reference microphone to all microphones. The signal vector $\mathbf{x}_\mathrm{R}(k, n)$ represents the reverberant signal and $\mathbf{v}(k, n)$ additive noise. In the following, the time frame index n is omitted where possible for the sake of brevity.

The PSD matrix of the late reverberation signal $\mathbf{x}_\mathrm{R}(k)$ can be modeled as a scaled diffuse sound field, which holds statistically for the late reverberation tail and frequencies above the Schroeder frequency [6]. We assume all components to be mutually uncorrelated such that the PSD matrix of the microphone signals can be expressed as

$$\begin{aligned}\mathbf{\Phi}_\mathbf{y}(k) &= E\left\{\mathbf{y}(k)\mathbf{y}^H(k)\right\} \\ &= \phi_S(k)\,\mathbf{d}(k)\mathbf{d}^H(k) + \underbrace{\phi_\mathrm{R}(k)\,\mathbf{\Gamma}_\mathrm{diff}(k) + \mathbf{\Phi}_\mathbf{v}(k)} \\ &= \phi_S(k)\,\mathbf{d}(k)\mathbf{d}^H(k) \qquad + \qquad \mathbf{\Phi}_\mathrm{in}(k), \end{aligned} \tag{2}$$

where $\phi_S(k)$ is the PSD of the desired speech signal, $\phi_\mathrm{R}(k)$ the PSD of the reverberation, $\mathbf{\Gamma}_\mathrm{diff}(k)$ denotes the diffuse coherence matrix, $\mathbf{\Phi}_\mathbf{v}(k)$ is the PSD matrix of $\mathbf{v}(k)$ and $\mathbf{\Phi}_\mathrm{in}(k)$ denotes the interference matrix. In an ideal diffuse sound field, $\mathbf{\Gamma}_\mathrm{diff}(k)$ is given by

$$\mathbf{\Gamma}_\mathrm{diff}(k) = \mathrm{sinc}\left(2\pi \frac{k f_\mathrm{s}}{N_\mathrm{FFT}} \frac{\mathbf{D}_\mathrm{mic}}{c}\right), \tag{3}$$

where $\mathrm{sinc}(\cdot) = \frac{\sin(\cdot)}{(\cdot)}$, N_FFT is the STFT length, f_S is the sampling frequency, $\mathbf{D}_\mathrm{mic}$ is a $M \times M$ matrix containing the distances between the microphones and c is the speed of sound.

Our objective is to obtain an estimate of the desired speech signal $S(k, n)$. A spatial filter is applied to the microphone signals such that

$$\begin{aligned} Z(k) &= \mathbf{h}^H(k)\mathbf{y}(k) \\ &= \mathbf{h}^H(k)\mathbf{d}(k)S(k) + \mathbf{h}^H(k)\mathbf{x}_\mathrm{R}(k) + \mathbf{h}^H(k)\mathbf{v}(k). \end{aligned} \tag{4}$$

In the following section, an spatial filter that is optimal in the MMSE sense is derived.

3. OPTIMAL SPATIAL FILTER

In this work, we aim at estimating the desired speech component in the MMSE sense. The MMSE cost function is given by

$$\begin{aligned} J(\mathbf{h}) &= E\left\{|Z(k) - S(k)|^2\right\} \\ &= E\left\{|\mathbf{h}^H(k)\left[S(k)\mathbf{d}(k) + \mathbf{x}_\mathrm{R}(k) + \mathbf{v}(k)\right] - S(k)|^2\right\}. \end{aligned} \tag{5}$$

The solution is the well-known multichannel Wiener filter that can be split into an MVDR beamformer, $\mathbf{h}_\mathrm{MVDR}(k)$, and a single-channel Wiener filter, $H_\mathrm{WF}(k)$, i.e.

$$\mathbf{h}_\mathrm{MWF}(k) = \frac{\phi_S(k)\,{\mathbf{\Phi}_\mathrm{in}}^{-1}(k)\,\mathbf{d}(k)}{\phi_S(k)\,\mathbf{d}^H(k)\,{\mathbf{\Phi}_\mathrm{in}}^{-1}(k)\,\mathbf{d}(k) + 1} \tag{6a}$$

$$= \underbrace{\frac{{\mathbf{\Phi}_\mathrm{in}}^{-1}(k)\,\mathbf{d}(k)}{\mathbf{d}^H(k)\,{\mathbf{\Phi}_\mathrm{in}}^{-1}(k)\,\mathbf{d}(k)}}_{\mathbf{h}_\mathrm{MVDR}(k)} \cdot \underbrace{\frac{\phi_S(k)}{\phi_S(k) + \phi_\mathrm{ri}(k)}}_{H_\mathrm{WF}(k)}. \tag{6b}$$

The form in (6b) has the advantage over (6a) that it provides direct control over the Wiener filter. The PSD of the desired signal at the MVDR output is given by

$$\phi_S(k) = \mathbf{h}_\mathrm{MVDR}^H(k)\left[\mathbf{\Phi}_\mathbf{y}(k) - \mathbf{\Phi}_\mathrm{in}(k)\right]\mathbf{h}_\mathrm{MVDR}(k) \tag{7}$$

and the PSD of the residual interference by

$$\phi_\mathrm{ri}(k) = \mathbf{h}_\mathrm{MVDR}^H(k)\mathbf{\Phi}_\mathrm{in}(k)\mathbf{h}_\mathrm{MVDR}(k). \tag{8}$$

It should be noted that by considering also the diffuse sound field as interference, the PSD matrix is highly time-varying. At those time and frequencies where the diffuse sound is much stronger than the noise, the MVDR filter is approximately equal to a super-directive beamformer. When the noise is spatially incoherent and is much stronger than the diffuse sound, the MVDR is approximately equal to a delay-and-sum beamformer.

To compute the weights of the MVDR filter, we also require an estimate of the propagation vector $\mathbf{d}(k)$. In this work, we assume a far-field model such that $\mathbf{d}(k)$ can be computed given the DOA of the source and the geometry of the microphone array. The main challenge is to get a proper estimate of the interference matrix $\mathbf{\Phi}_\mathrm{in}(k)$, which is in this case highly time-varying. In this work we assume that $\mathbf{\Gamma}_\mathrm{diff}(k)$ and $\mathbf{\Phi}_\mathbf{v}(k)$ are known such that only $\phi_\mathrm{R}(k)$ needs to be estimated. In the next section, an estimator for the PSD of the late reverberation $\phi_\mathrm{R}(k)$ is derived.

4. ESTIMATION OF LATE REVERBERATION PSD

In this section, we derive an estimator for the diffuse sound PSD $\phi_R(k)$ based on $M-1$ reference signals. In Section 4.1, the computation of the reference signals is described. In Section 4.2, a maximum likelihood estimator (MLE) for the diffuse sound PSD is derived based on the computed reference signals.

4.1. Computation of the reference signals

The reference signal vector $\mathbf{u}(k)$ is obtained as the output of a blocking matrix $\mathbf{B}(k) \in \mathbb{C}_{M \times M-1}$

$$\mathbf{u}(k) = \mathbf{B}^H(k)\mathbf{y}(k). \tag{9}$$

Our objective is to design the blocking matrix $\mathbf{B}(k)$ such that it generates signals that contain only the undesired signal components, i.e. the reverberant signal component plus

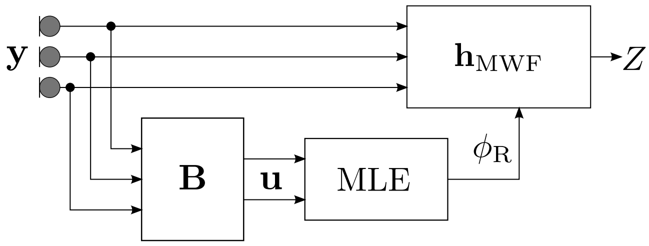

Fig. 1. Proposed system for $M = 3$ using a maximum likelihood estimator for the late reverberant PSD and a multichannel Wiener filter.

residual noise. Given the propagation vector $\mathbf{d}(k)$, the blocking matrix should satisfy

$$\mathbf{B}^H(k)\mathbf{d}(k) = \mathbf{0}_{M-1\times 1}. \qquad (10)$$

There exist many blocking matrices that satisfy (10). In this work, we use the solution proposed in [5], where $\mathbf{A}(k)$ is a matrix of size $M \times M$ and the blocking matrix consists of its first $M-1$ columns, i.e.,

$$\mathbf{A}(k) = \mathbf{I} - \frac{\mathbf{d}(k)\,\mathbf{d}^H(k)}{\|\mathbf{d}(k)\|_2^2}$$
$$[\mathbf{B}(k)\ \mathbf{b}_M(k)] = \mathbf{A}(k).$$

According to (2), the PSD matrix of the blocking matrix output consists of the residual diffuse and residual noise PSD matrices, i.e.

$$\begin{aligned}\mathbf{\Phi}_{\mathbf{u}}(k) &= \mathbf{B}^H(k)\left[\phi_{\mathrm{R}}(k)\,\mathbf{\Gamma}_{\text{diff}}(k) + \mathbf{\Phi}_{\mathbf{v}}(k)\right]\mathbf{B}(k)\\ &= \phi_{\mathrm{R}}(k)\,\widetilde{\mathbf{\Gamma}}_{\text{diff}}(k) + \widetilde{\mathbf{\Phi}}_{\mathbf{v}}(k),\end{aligned} \qquad (11)$$

where the matrices $\widetilde{\mathbf{\Gamma}}_{\text{diff}}(k) = \mathbf{B}^H(k)\,\mathbf{\Gamma}_{\text{diff}}(k)\,\mathbf{B}(k)$ and $\widetilde{\mathbf{\Phi}}_{\mathbf{v}}(k) = \mathbf{B}^H(k)\,\mathbf{\Phi}_{\mathbf{v}}(k)\,\mathbf{B}(k)$ denote the diffuse coherence matrix and the noise PSD matrix at the output of the blocking matrix, respectively.

4.2. Maximum likelihood estimator

As proposed in [7], we define an error matrix

$$\begin{aligned}\mathbf{\Phi}_{\mathbf{e}}(k) &= \underbrace{\mathbf{\Phi}_{\mathbf{u}}(k) - \widetilde{\mathbf{\Phi}}_{\mathbf{v}}(k)} - \phi_{\mathrm{R}}(k)\,\widetilde{\mathbf{\Gamma}}_{\text{diff}}(k)\\ &= \qquad \hat{\mathbf{\Phi}}_{\mathrm{R}}(k) \qquad - \phi_{\mathrm{R}}(k)\,\widetilde{\mathbf{\Gamma}}_{\text{diff}}(k).\end{aligned} \qquad (12)$$

The matrix $\hat{\mathbf{\Phi}}_{\mathrm{R}}(k)$ can be estimated using an estimate of the PSD matrix $\mathbf{\Phi}_{\mathbf{u}}(k) = E\{\mathbf{u}(k)\mathbf{u}^H(k)\}$, where $\mathbf{u}(k)$ is computed using (9), and an estimate of the residual noise PSD matrix $\widetilde{\mathbf{\Phi}}_{\mathbf{v}}(k)$. The real and imaginary elements of $\mathbf{\Phi}_{\mathbf{e}}(k)$ are modeled as independent zero-mean Gaussian distributions with a certain standard deviation σ, which is assumed to be equal for all random variables. The joint probability density function of $\mathbf{\Phi}_{\mathbf{e}}(k)$ is given as

$$\begin{aligned}f\left(\mathbf{\Phi}_{\mathbf{e}}\right) = &\frac{1}{(\sigma\sqrt{2\pi})^{(M-1)^2}}\\ &\times \prod_{p,q=1}^{M-1} \exp\left(-\frac{\left(\Re\{\hat{\Phi}_{\mathrm{R}}^{p,q}\} - \phi_{\mathrm{R}}\,\Re\{\widetilde{\Gamma}_{\text{diff}}^{p,q}\}\right)^2}{2\sigma^2}\right)\\ &\times \prod_{p,q=1}^{M-1} \exp\left(-\frac{\left(\Im\left\{\hat{\Phi}_{\mathrm{R}}^{p,q}\right\} - \phi_{\mathrm{R}}\,\Im\left\{\widetilde{\Gamma}_{\text{diff}}^{p,q}\right\}\right)^2}{2\sigma^2}\right),\end{aligned} \qquad (13)$$

where $\hat{\Phi}_{\mathrm{R}}^{p,q}(k)$ and $\widetilde{\Gamma}_{\text{diff}}^{p,q}(k)$ denote the elements of $\hat{\mathbf{\Phi}}_{\mathrm{R}}(k)$ and $\widetilde{\mathbf{\Gamma}}_{\text{diff}}(k)$, respectively, $\Re\{\cdot\}$ and $\Im\{\cdot\}$ the real and imaginary part operators. The frequency index k is omitted in (13) for better readability. The log-likelihood function is given by [7]

$$\begin{aligned}\ln f\left(\mathbf{\Phi}_{\mathbf{e}}\right) = &-(M-1)^2\ln(\sigma\sqrt{2\pi})\\ &- \frac{1}{2\sigma^2}\|\hat{\mathbf{\Phi}}_{\mathrm{R}}(k) - \phi_{\mathrm{R}}(k)\cdot\widetilde{\mathbf{\Gamma}}_{\text{diff}}(k)\|_F^2,\end{aligned} \qquad (14)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. The maximum likelihood estimate of the diffuse PSD is obtained by minimizing the log-likelihood function (14) with respect to $\phi_{\mathrm{R}}(k)$, i.e.,

$$\hat{\phi}_{\mathrm{R}}(k) = \underset{\phi_{\mathrm{R}}(k)}{\arg\min}\ \|\hat{\mathbf{\Phi}}_{\mathrm{R}}(k) - \phi_{\mathrm{R}}(k)\cdot\widetilde{\mathbf{\Gamma}}_{\text{diff}}(k)\|_F^2. \qquad (15)$$

The solution of the minimization problem in (15) is given by

$$\hat{\phi}_{\mathrm{R}}(k) = \frac{\operatorname{tr}\left\{\widetilde{\mathbf{\Gamma}}_{\text{diff}}^H(k)\,\hat{\mathbf{\Phi}}_{\mathrm{R}}(k)\right\}}{\operatorname{tr}\left\{\widetilde{\mathbf{\Gamma}}_{\text{diff}}^H(k)\,\widetilde{\mathbf{\Gamma}}_{\text{diff}}(k)\right\}}, \qquad (16)$$

where $\operatorname{tr}\{\cdot\}$ denotes the trace operator.

5. PERFORMANCE EVALUATION

5.1. Simulation setup

The performance of the proposed algorithm is evaluated using the following setup. A room of size $6\times 5\times 4$ m with a reverberation time T_{60} of 500 ms was simulated using the image method [8]. The desired speaker was positioned in the broadside of a microphone array with $M = 4$ microphones at a distance of 2 m. The microphone signals are corrupted by additive spatially uncorrelated white Gaussian noise with different input SNRs. A sampling rate of $f_{\mathrm{s}} = 16$ kHz was used with a STFT length of 1024 points, a Hamming window of length of 32 ms and a hop size of 8 ms. The PSD matrices $\mathbf{\Phi}_{\mathbf{y}}(k,n)$ and $\mathbf{\Phi}_{\mathbf{u}}(k,n)$ were estimated using recursive averaging with a time constant of 50 ms, which was empirically found to provide a good tradeoff between dereverberation performance and audible artifacts. The noise PSD matrix was calculated when the speaker was silent, assuming an ideal voice activity detector.

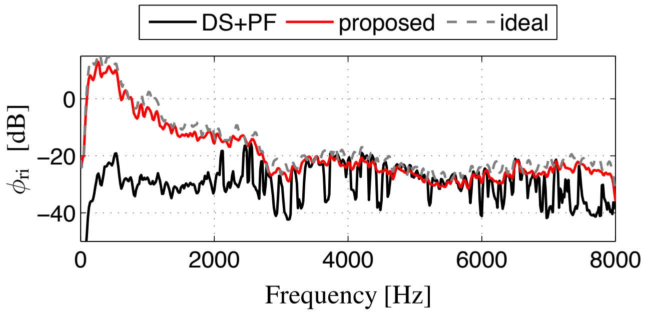

Fig. 2. Estimated residual late reverberation PSD averaged over time.

5.2. Evaluation of the proposed estimator

The dereverberation performance of the proposed algorithm highly depends on the performance of the late reverberant PSD estimator. In this section, we compare the performance of the estimator proposed in [3] with the proposed MLE. In [3], Habets and Gannot proposed a system similar to the one proposed here. As opposed to the optimal spatial filter, a delay-and-sum (DS) beamformer followed by a post-filter was used. Their main contribution was the estimator for the late reverberant PSD at the output of the DS beamformer that was obtained using a simple blocking matrix and an adaptive algorithm. In order for the adaptive algorithm to work properly and to mitigate speech distortion, it was proposed to delay the output of the blocking matrix. The delay was justified using a statistical model for the late reverberant PSD. In the following, we used a delay of 40 ms. Since the algorithm in [3] was proposed for $M = 2$, we extended it for a fair comparison to an arbitrary number of microphones such that the DS beamformer and the blocking matrix use M microphones. The estimation of the late reverberant PSD remains equal and uses only one reference signal.

Fig. 2 shows the estimator proposed in [3] (black) and the estimated late reverberation PSD at the output of the beamformer (red) obtained using (8). The PSDs in Fig. 2 are long-term averaged over a speech segment of 8 s. The input signal-to-noise ratio (SNR), denoted by iSNR, was set to 60 dB because the estimator in [3] does not take additive noise into account. For the evaluation in Fig. 2, hte optimal spatial filter $\mathbf{h}_{\text{MVDR}}(k)$ was replaced by a fixed DS beamformer to ensure a fair comparison. We observe that the estimate of $\phi_{\text{ri}}(k)$ highly correlates with the late reverberation at the output of the DS beamformer and that the proposed MLE matches the ideal PSD closely, especially at lower frequencies.

5.3. Performance measures

For evaluation purposes, auxiliary signals of the direct-plus-early signal component and the late reverberant signal component were generated. We considered 30 ms as the transition between early reflections and late reverberation. These

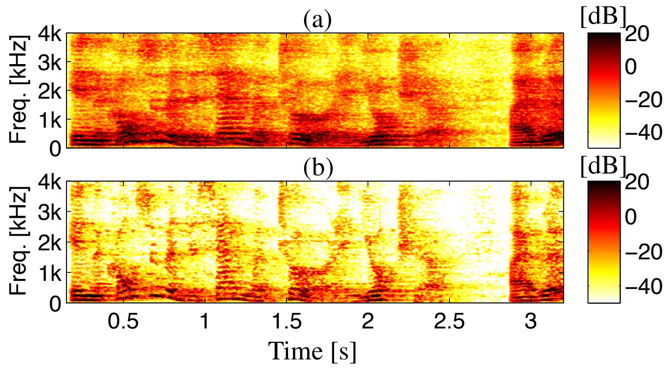

Fig. 3. Spectrograms of reverberant and noisy input signal (a) with iSNR = 40 dB and processed output signal (b).

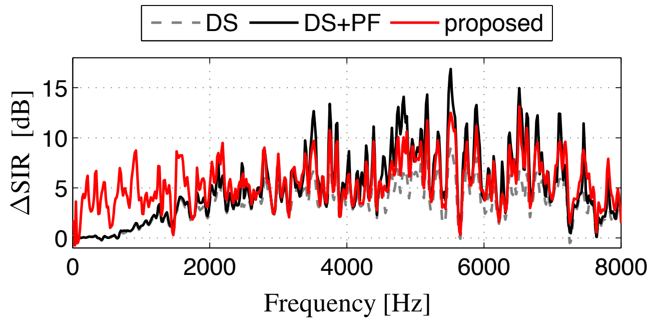

Fig. 4. Average subband SIR gain for iSNR = 40 dB.

auxiliary signals were used to calculate the segmental signal-to-interference ratio (SIR), where the direct-plus-early signal component are considered as desired signal and the late reverberation plus noise as interference. Fullband segmental SIR values are calculated from the time domain signals.

As a second measure, we employed the speech to reverberation energy modulation ratio (SRMR) [9], which intrusively measures the amount of reverberation within a signal. Higher SRMR values correspond to lower reverberation levels. The SRMR values were calculated from noise-free signals, since the measure is biased by additive noise.

As a third and final measure, we employed the log-spectral distance (LSD) [10], where the direct-plus-early signal component was used as a reference signal.

5.4. Results

The effect of the proposed filter can be observed in the input (a) and output (b) spectrograms shown in Fig. 3. The sensor noise as well as late reverberation are reduced, so that the harmonic and temporal structure of the speech is recovered.

Fig. 4 shows the subband SIR gain, denoted by ΔSIR, of a conventional DS beamformer, the DS+PF [3], and the proposed method for an input SNR of 40 dB. At frequencies above 3 kHz, the SIR gain of the DS+PF and the proposed method is very similar and a little superior to the DS beamformer. At frequencies below 3 kHz, where most of

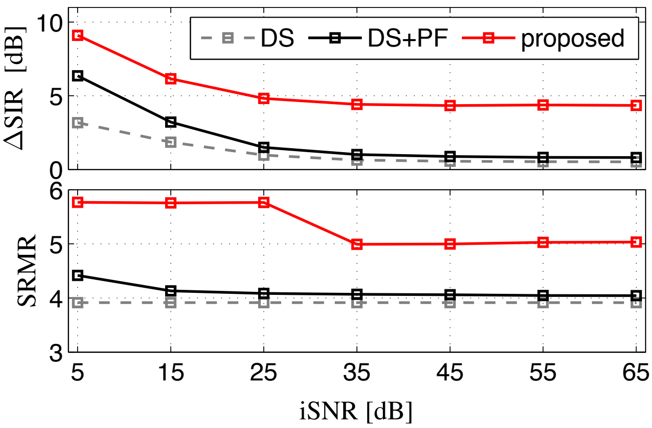

Fig. 5. Dependence on the input noise level: Fullband SIR gain (top) and SRMR (bottom).

the speech energy lies, the proposed method achieves a much higher SIR gain.

In Fig. 5, the SIR gain and SRMR are depicted for different input SNR levels ranging from 5 till 65 dB. The SIR gain is higher for lower input SNR values, since the algorithms have to suppress more noise. The postfilter proposed in [3] suppresses also uncorrelated noise, although it is not designed for this purpose. The SRMR values indicate that the dereverberation performance does not depend much on the input SNR. At a lower SNR, the weak speech components are suppressed more resulting in higher speech distortion. Consequently, the SRMR is higher at low iSNR values.

The results obtained using 2, 4, 6 and 8 microphones are depicted in Table 1. As a reference, the performance measures are also shown for the unprocessed reference microphone. The number of microphones M was increased with a constant microphone spacing of 5 cm. It can be observed that the DS+PF performs better than the DS and that the proposed algorithm performs much better than both. Note that the proposed system yields good values using only 2 or 4 microphones, whereas the DS and DS+PF need at least 8 microphones to reach comparable values.

6. CONCLUSION

A novel method for multi-microphone dereverberation in noisy environments is proposed. We assumed that the reverberant signal component can be modeled as an ideal diffuse sound field. Given an estimate of the DOA of the desired source and the noise PSD matrix, a very accurate estimate of the diffuse sound PSD is obtained using a MLE. In the performance evaluation it was shown that the proposed method outperforms two existing methods. The proposed method was shown to be robust to spatially incoherent noise when an accurate estimate of the noise PSD matrix was used. Investigating the performance in other noise conditions is a topic for future research.

	ΔSIR [dB]			SRMR			LSD		
M	DS	DS+PF	prop.	DS	DS+PF	prop.	DS	DS+PF	prop.
1	0			3.8			3.0		
2	0.3	0.5	3.2	3.9	3.9	5.0	2.9	2.9	2.7
4	0.6	0.9	4.3	3.9	4.1	5.0	2.8	2.7	2.6
6	0.7	1.1	4.4	4.0	4.2	5.2	2.7	2.6	2.6
8	0.8	1.4	4.4	4.1	4.5	5.1	2.6	2.6	2.5

Table 1. Performance for different number of microphones with constant spacing of 5 cm for iSNR $= 40$ dB and $T_{60} =$ 500 ms.

7. REFERENCES

[1] P. A. Naylor, N. D. Gaubitch, and E. A. P. Habets, "Signal-based performance evaluation of dereverberation algorithms," *Journal of Electrical and Computer Engineering*, vol. 2010, pp. 1–5, 2010.

[2] S. Braun, D. P. Jarrett, J. Fischer, and E. A. P. Habets, "An informed spatial filter for dereverberation in the spherical harmonic domain," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013.

[3] E. A. P. Habets and S. Gannot, "Dual-microphone speech dereverberation using a reference signal," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Honolulu, USA, Apr. 2007, vol. IV, pp. 901–904.

[4] A. Schwarz, K. Reindl, and W. Kellermann, "A two-channel reverberation suppression scheme based on blind signal separation and Wiener filtering," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Dec. 2012.

[5] U. Kjems and J. Jensen, "Maximum likelihood based noise covariance matrix estimation for multi-microphone speech enhancement," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Bucharest, Romania, Aug. 2012, pp. 1–5.

[6] M. R. Schroeder, "Statistical parameters of the frequency response curves of large rooms," *Journal Audio Eng. Soc.*, vol. 35, pp. 299–306, 1954.

[7] H. Q. Dam, S. Nordholm, H. H. Dam, and S. Y. Low, "Maximum likelihood estimation and cramer-rao lower bounds for the multichannel spectral evaluation in hands-free communication," in *Asia-Pacific Conference on Communications*, Perth, Australia, Oct. 2005.

[8] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[9] T. Falk, C. Zheng, and W.-Y. Chan, "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 7, pp. 1766–1774, Sept. 2010.

[10] J. Benesty, M. Mohan Sondhi, and Y. Huang, Eds., *Springer Handbook of Speech Processing*, Springer, 2008.