Open access • Proceedings Article • DOI:10.1109/ICASSP.2014.6854589

Speech dereverberation using weighted prediction error with laplacian model of the desired signal — Source link

Ante Jukic, Simon Doclo

Institutions: University of Oldenburg

Published on: 04 May 2014 - International Conference on Acoustics, Speech, and Signal Processing

Topics: Linear predictive coding, Intelligibility (communication), Linear prediction and Reverberation

SPEECH DEREVERBERATION USING WEIGHTED PREDICTION ERROR WITH LAPLACIAN MODEL OF THE DESIRED SIGNAL

Ante Jukić, Simon Doclo

University of Oldenburg, Department of Medical Physics and Acoustics, Oldenburg, Germany
{ante.jukic,simon.doclo}@uni-oldenburg.de

ABSTRACT

Reverberation has a considerable impact on the quality and intelligibility of captured speech signals. In this paper we present an approach for blind multi-microphone speech dereverberation based on the weighted prediction error method, where the reverberant observations are modeled using multi-channel linear prediction in the short-time Fourier transform domain. Instead of using the commonly employed Gaussian distribution for the desired speech signal, the proposed approach uses a Laplacian distribution which is known to be more accurate in modeling speech signals. Maximum-likelihood estimation is used for estimating the model parameters, leading to a linear programming optimization problem. Experimental results, obtained using measured impulse responses, indicate that the proposed approach could be used to improve the dereverberation performance compared to the classical technique.

Index Terms— Dereverberation, speech enhancement, model-based signal processing

1. INTRODUCTION

Capturing speech in an enclosed space with microphones placed at a distance from the speaker typically results in microphone recordings corrupted by reverberation, caused by reflections against the walls and objects in the enclosure. It is well known that reverberation often results in a decrease of speech intelligibility and reduces the performance of automatic speech recognition systems [1]. Many speech communication applications, such as hands-free telephony, teleconferencing and voice-controlled systems, therefore benefit from effective dereverberation. Recently, various techniques have been proposed that aim to reduce the reverberant components. Some techniques are based on first blindly estimating the room impulse responses (RIR) [2] followed by multichannel equalization [3]. Although in theory these approaches can perform perfect dereverberation, the achieved performance is limited by the accuracy of the RIR estimation step and robust equalization techniques are needed [4]. More robust speech dereverberation approaches are based on spectral enhancement, which however typically introduce a trade-off between reverberation suppression and speech distortion [5]. In addition, several blind speech dereverberation techniques, which do not employ any knowledge of the room acoustics properties, were proposed recently [6, 7, 8, 9, 10].

An approach for blind speech dereverberation based on multi-channel linear prediction (MCLP) was proposed in [6, 7]. The efficient implementation in the short-time Fourier transform (STFT) domain, named weighted prediction error (WPE) [6], uses an autoregressive generative model for the acoustic transfer functions (ATFs), and models the spectral coefficients of the desired (clean) speech signal using a Gaussian distribution. Dereverberation is then performed by maximum likelihood (ML) estimation of all unknown model parameters.

While the assumption of Gaussianity often leads to closed-form expressions, it is well known that the speech signals can be modeled more accurately using Laplacian or Gamma distribution, both in terms of time-domain samples as well as spectral coefficients [11, 12, 13, 14]. Motivated by these facts, in this paper we propose a speech dereverberation method based on MCLP with a Laplacian distribution as a local model of the desired speech signal coefficients. Since maximal likelihood estimation of the regression parameters however no longer results in a closed-form expression, we have to resort to numerical optimization for solving the corresponding linear programming (LP) problem. The results presented in the experimental section show that for different acoustic scenarios the proposed method achieves better performance, in terms of cepstral distance and PESQ score, when compared to the classical WPE based on the Gaussian distribution.

The paper is organized as follows. In Section 2 we introduce notation and formulate the problem of speech dereverberation. In Section 3 we give an overview of the WPE based on Gaussian distribution and present the proposed method based on a Laplacian distribution. The experimental results are presented in Section 4.

2. PROBLEM FORMULATION

We consider a scenario where a single speech source in an enclosure is captured by M microphones. Let $s_{n,k}$ denote the clean speech signal in the STFT domain with time frame index $n \in \{1, \dots, N\}$, and frequency bin index $k \in \{1, \dots, K\}$. The reverberant speech signal observed at the m-th microphone, $m \in \{1, \dots, M\}$, is typically modeled in the STFT domain as [7]

$$x_{n,k}^m = \sum_{l=0}^{L_h-1} (h_{l,k}^m)^* s_{n-l,k} + e_{n,k}^m, \tag{1}$$

where $h_{l,k}^m$ models the ATF between the speech source and m-th microphone in the STFT domain, the length of ATF equals L_h, and $(.)^*$ denotes the complex conjugate operator. The additive term $e_{n,k}^m$ jointly represents modeling errors and the additive noise signal. The convolutive model in (1) is often rewritten as

$$x_{n,k}^m = d_{n,k}^m + \sum_{l=D}^{L_h-1} (h_{l,k}^m)^* s_{n-l,k} + e_{n,k}^m, \tag{2}$$

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).

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

where the signal

$$d_{n,k}^{m} = \sum_{l=0}^{D-1} (h_{l,k}^{m})^{*} s_{n-l,k} \tag{3}$$

is composed of the anechoic speech signal and early reflections at the m-th microphone, and D corresponds to the duration of the early reflections. Dereverberation methods often aim to recover the anechoic signal together with the early reflections, since the early reflections tend to improve speech inteligibility [1].

In several methods it was proposed to replace the convolutive model in (1) and (2) with an autoregressive model [6, 7, 9]. In [7, 15], the observation model was further simplified by assuming $e_{n,k}^{m} = 0$, $\forall n, k, m$. Under these assumptions, the signal observed at the first microphone ($m = 1$) can be written in the well-known multi-channel linear prediction form, i.e.,

$$x_{n,k}^{1} = d_{n,k} + \sum_{m=1}^{M} (\mathbf{g}_{k}^{m})^{H} \mathbf{x}_{n-D,k}^{m} \tag{4}$$

where $d_{n,k} \equiv d_{n,k}^{1}$ is the desired signal, and $(.)^{H}$ denotes the conjugate transposition operator. The vector $\mathbf{g}_{k}^{m} \in \mathbb{C}^{L_k}$ is the regression vector of order L_k for the m-th channel and $\mathbf{x}_{n,k}^{m}$ is defined as

$$\mathbf{x}_{n,k}^{m} = \left[x_{n,k}^{m}, \ldots, x_{n-L_k+1,k}^{m}\right]^{T}, \tag{5}$$

with $(.)^{T}$ denoting the transposition operator. The MCLP model (4) can be written in a compact form using the multi-channel regression vector $\mathbf{g}_{k} \in \mathbb{C}^{ML_k}$ as

$$\boxed{x_{n,k}^{1} = d_{n,k} + \mathbf{g}_{k}^{H} \mathbf{x}_{n-D,k}} \tag{6}$$

with the following notation

$$\mathbf{g}_{k} = \left[(\mathbf{g}_{k}^{1})^{T}, \ldots, (\mathbf{g}_{k}^{M})^{T}\right]^{T}, \tag{7}$$

$$\mathbf{x}_{n,k} = \left[(\mathbf{x}_{n,k}^{1})^{T}, \ldots, (\mathbf{x}_{n,k}^{M})^{T}\right]^{T}. \tag{8}$$

In the presented scenario the problem of speech dereverberation is formulated as a blind estimation of the desired signal $d_{n,k}$, consisting of the direct speech signal and early reflections, from the reverberant observations $x_{n,k}^{m}, \forall m, n, k$. From (4) it follows that the desired signal can be estimated as

$$\boxed{\hat{d}_{n,k} = x_{n,k}^{1} - \hat{\mathbf{g}}_{k}^{H} \mathbf{x}_{n-D,k}} \tag{9}$$

where $\hat{}$ denotes an estimated value. Therefore, dereverberation can be performed by estimating the regression vectors $\hat{\mathbf{g}}_{k}$, and calculating an estimate of the desired speech signal $\hat{d}_{n,k}$ as in (9).

3. WPE METHOD

3.1. Original approach - Gaussian model

The original weighted prediction error (WPE) method proposed in [7] is based on a time-varying power spectrum model (TVPS) of the desired signal, assuming locally Gaussian distribution for the desired speech coefficients. More specifically, the desired signal in each time-frequency bin is modeled as a zero-mean random variable with a circular complex Gaussian distribution and a time- and frequency-dependent variance $\lambda_{n,k}$. The probability density function $p\left(d_{n,k}\right)$ of the desired signal is then given by

$$\boxed{p\left(d_{n,k}\right) = \frac{1}{\pi \lambda_{n,k}} e^{-\frac{|d_{n,k}|^{2}}{\lambda_{n,k}}}} \tag{10}$$

Additionally, it is assumed that d_{n_1,k_1} and d_{n_2,k_2} are independent for $(n_1, k_1) \neq (n_2, k_2)$.

The unknown parameters to be estimated from the reverberant observations $x_{n,k}^{m}$ are the (time- and frequency-dependent) speech variances $\lambda_{n,k}$ and the (frequency-dependent) regression vectors $\mathbf{g}_{k}$, modeling the ATFs. Note that we can work in each frequency bin independently, since the observation model (4) and the TVPS model do not assume any dependency across different frequency bins. In [7] a ML estimation of the parameters has been proposed, by maximizing the likelihood function

$$\mathcal{L}\left(\Theta_{k}\right) = \prod_{n=1}^{N} p\left(d_{n,k}\right), \tag{11}$$

where $\Theta_{k} = \{\mathbf{g}_{k}, \lambda_{1,k}, \ldots, \lambda_{N,k}\}$ is the set of unknown parameters for the k-th frequency bin. This estimation is equivalent to minimization of the cost function

$$\boxed{\ell\left(\Theta_{k}\right) = \sum_{n=1}^{N} \left(\log \lambda_{n,k} + \frac{|x_{n,k}^{1} - \mathbf{g}_{k}^{H} \mathbf{x}_{n-D,k}|^{2}}{\lambda_{n,k}}\right)} \tag{12}$$

which is obtained by taking the negative logarithm of (11) and ignoring the constant terms. However, minimizing (12) with respect to the parameters Θ_{k} can not be performed analytically, so an alternating two-step optimization scheme was proposed. In the first step, (12) is minimized with respect to $\mathbf{g}_{k}$, while the other parameters (variances $\lambda_{n,k}$) are kept fixed. In this case, the cost function (12) can be rewritten as a function of $\mathbf{g}_{k}$ as follows

$$\ell\left(\mathbf{g}_{k}\right) = \sum_{n=1}^{N} \left|\frac{1}{\sqrt{\lambda_{n,k}}} x_{n,k}^{1} - \frac{1}{\sqrt{\lambda_{n,k}}} \mathbf{g}_{k}^{H} \mathbf{x}_{n-D,k}\right|^{2} + r_{k}, \tag{13}$$

where $r_{k} = \sum_{n=1}^{N} \log \lambda_{n,k}$ does not depend on $\mathbf{g}_{k}$. Minimization of (13) is a linear least squares problem in variable $\mathbf{g}_{k}$ with a closed-form solution (cf. Table 1). In the second step parameters swap roles, i.e., (12) is minimized with respect to the unknown variances $\lambda_{n,k}$ while $\mathbf{g}_{k}$ is kept fixed. In this case, each variance is obtained as

$$\hat{\lambda}_{n,k} = \underset{\lambda_{n,k}>0}{\arg\min} \left(\log \lambda_{n,k} + \frac{|d_{n,k}|^{2}}{\lambda_{n,k}}\right) = |d_{n,k}|^{2}. \tag{14}$$

This two-step procedure is repeated until some convergence criterion is satisfied or a maximum number of iterations is exceeded. Additionally, a small positive constant ε_{k} is included as a lower bound for the estimated variance, to prevent division by zero. The complete procedure is outlined in Table 1.

In [7] it was proposed to use just a single iteration of the algorithm, since it was observed that subsequent iterations did not always increase the quality of the recovered signal and could even lead to degradations. These degradations typically occurred for short observed signals, and it was recently proposed to introduce additional information, through spectral priors, to mitigate the degradation [15].

input: $x^m_{n,k}, \forall n, m\ ; L_k$
initialization: $\hat{\lambda}_{n,k} \leftarrow |x^1_{n,k}|^2$
repeat
$\mathbf{A}_k \leftarrow \sum_{n=1}^{N} \frac{\mathbf{x}_{n-D,k}\mathbf{x}^H_{n-D,k}}{\hat{\lambda}_{n,k}}$
$\mathbf{b}_k \leftarrow \sum_{n=1}^{N} \frac{\mathbf{x}_{n-D,k}(x^1_{n,k})^*}{\hat{\lambda}_{n,k}}$
$\hat{\mathbf{g}}_k \leftarrow \mathbf{A}_k^{-1}\mathbf{b}_k$
$\hat{d}_{n,k} \leftarrow x^1_{n,k} - \hat{\mathbf{g}}_k^H \mathbf{x}_{n-D,k}$
$\hat{\lambda}_{n,k} \leftarrow \max\{|\hat{d}_{n,k}|^2, \varepsilon_k\}$
until condition satisfied

Table 1. Outline of the WPE method based on the Gaussian distribution

3.2. Proposed approach - Laplacian model

The assumption that the STFT coefficients of the desired signal can be modeled using a Gaussian distribution results in a closed-form solution for estimating the regression vector $\mathbf{g}_k$ in WPE. However, it is often stated that the STFT coefficients of speech signals can be more accurately modeled using Laplacian or Gamma distributions [11, 12, 13]. Motivated by these facts, we propose to model the STFT coefficients of the desired signal locally, in each time-frequency bin, using a Laplacian distribution. We assume that the real and imaginary parts of $d_{n,k}$ have a Laplacian distribution with equal variance $\lambda_{n,k}/2$, and that they are independent. The probability density function of the desired signal can then be written as

$$p\left(d_{n,k}\right) = \frac{1}{\lambda_{n,k}} e^{-2\frac{|\Re(d_{n,k})|+|\Im(d_{n,k})|}{\sqrt{\lambda_{n,k}}}} \tag{15}$$

where $\Re(.)$ and $\Im(.)$ denote the real and imaginary part of a complex number. Similarly as in the original WPE method, the ML estimate of the parameters Θ_k can now be obtained by minimizing the cost function

$$\tilde{\ell}\left(\Theta_k\right) = \sum_{n=1}^{N}\left(\log\lambda_{n,k} + 2\frac{\left|\Re\left(d_{n,k}\right)\right| + \left|\Im\left(d_{n,k}\right)\right|}{\sqrt{\lambda_{n,k}}}\right) \tag{16}$$

which is obtained by taking the negative logarithm of the likelihood function and ignoring the constant terms. Again, we resort to a two-step alternating scheme to obtain estimates for the parameters Θ_k.

Step 1 - estimation of $\mathbf{g}_k$: Assuming that the variances $\lambda_{n,k}$ are fixed, the regression vector can be estimated by minimizing the cost function (16) with respect to $\mathbf{g}_k$. In this case, the cost function (16) can be rewritten as a function of $\mathbf{g}_k$ as follows

$$\begin{aligned}\tilde{\ell}\left(\mathbf{g}_k\right) = \sum_{n=1}^{N}\frac{2}{\sqrt{\lambda_{n,k}}}\Bigg(\left|\Re\left(x^1_{n,k} - \mathbf{g}_k^H\mathbf{x}_{n-D,k}\right)\right| + \\ + \left|\Im\left(x^1_{n,k} - \mathbf{g}_k^H\mathbf{x}_{n-D,k}\right)\right|\Bigg) + r_k,\end{aligned} \tag{17}$$

with r_k does not depend on $\mathbf{g}_k$. The first term within the brackets in (17) can be rewritten as

$$\Re\left(x^1_{n,k} - \mathbf{g}_k^H\mathbf{x}_{n-D,k}\right) = \Re(x^1_{n,k}) - \bar{\mathbf{g}}_k^T\bar{\mathbf{x}}_{n-D,k}, \tag{18}$$

with

$$\bar{\mathbf{x}}_{n,k} = \begin{bmatrix}\Re(\mathbf{x}_{n,k})\\ \Im(\mathbf{x}_{n,k})\end{bmatrix}, \bar{\mathbf{g}}_k = \begin{bmatrix}\Re(\mathbf{g}_k)\\ \Im(\mathbf{g}_k)\end{bmatrix}. \tag{19}$$

Similarly, we have

$$\Im\left(x^1_{n,k} - \mathbf{g}_k^H\mathbf{x}_{n-D,k}\right) = \Im(x^1_{n,k}) - \bar{\mathbf{g}}_k^T\tilde{\mathbf{x}}_{n-D,k}, \tag{20}$$

with $\tilde{\mathbf{x}}_{n,k} = \left[\Im(\mathbf{x}_{n,k})^T\ -\Re(\mathbf{x}_{n,k})^T\right]^T$. The cost function (17) can hence be rewritten as

$$\begin{aligned}\tilde{\ell}\left(\mathbf{g}_k\right) = \sum_{n=1}^{N}\frac{2}{\sqrt{\lambda_{n,k}}}\Bigg(\left|\Re(x^1_{n,k}) - \bar{\mathbf{g}}_k^T\bar{\mathbf{x}}_{n-D,k}\right| + \\ + \left|\Im(x^1_{n,k}) - \bar{\mathbf{g}}_k^T\tilde{\mathbf{x}}_{n-D,k}\right|\Bigg) + r_k.\end{aligned} \tag{21}$$

Minimizing (21) can be formulated as the following linear programming (LP) problem [16]

$$\begin{aligned}&\min_{\mathbf{t},\bar{\mathbf{g}}_k} && \|\mathbf{t}\|_1\\ &\text{subject to} && \mathbf{t} \geq 0\\ &&& |\Re(x^1_{n,k}) - \bar{\mathbf{g}}_k^T\bar{\mathbf{x}}_{n-D,k}| \leq \frac{\sqrt{\lambda_{n,k}}}{2}t_{2n-1}\\ &&& |\Im(x^1_{n,k}) - \bar{\mathbf{g}}_k^T\tilde{\mathbf{x}}_{n-D,k}| \leq \frac{\sqrt{\lambda_{n,k}}}{2}t_{2n}\end{aligned} \tag{22}$$

with variables $\mathbf{t} \in \mathbb{R}^{2N}$, $\bar{\mathbf{g}}_k \in \mathbb{R}^{2ML_k}$, where t_n denotes the n-th element of the vector $\mathbf{t}$ and $\|.\|_1$ is the ℓ_1-norm.

Step 2 - estimation of $\lambda_{n,k}$: Assuming that the regression vector $\mathbf{g}_k$ is fixed, the variances can be estimated by minimizing the cost function in (16) with respect to $\lambda_{n,k}$. In this step, each variance is obtained by solving

$$\min_{\lambda_{n,k}>0}\left(\log\lambda_{n,k} + 2\frac{|\Re(d_{n,k})| + |\Im(d_{n,k})|}{\sqrt{\lambda_{n,k}}}\right), \tag{23}$$

yielding a closed-form solution

$$\lambda_{n,k} = \left(|\Re(d_{n,k})| + |\Im(d_{n,k})|\right)^2. \tag{24}$$

This two-step procedure is repeated in an alternating fashion, until some convergence criterion is satisfied or a maximum number of iterations is exceeded. Again, a small positive constant ε_k is included as a lower bound for the estimated variance. The complete procedure is outlined in Table 2.

Although solving the LP problem in (22) is more complicated than calculating the closed-form solution in the original WPE, it should be noted that solvers for LP problems are a mature technology [16]. Recent trends show that modern solvers, even for a wider class of convex problems, are often fast enough to be used in real-time systems [17].

input: $x^m_{n,k}, \forall n, m\ ; L_k$
initialization: $\hat{\lambda}_{n,k} \leftarrow \left(|\Re(x^1_{n,k})| + |\Im(x^1_{n,k})|\right)^2$
repeat
$\hat{\mathbf{g}}_k \leftarrow$ solve LP (22)
$\hat{d}_{n,k} \leftarrow x^1_{n,k} - \hat{\mathbf{g}}_k^H\mathbf{x}_{n-D,k}$
$\hat{\lambda}_{n,k} \leftarrow \max\left\{\left(|\Re(\hat{d}_{n,k})| + |\Im(\hat{d}_{n,k})|\right)^2, \varepsilon_k\right\}$
until condition satisfied

Table 2. Outline of the proposed method based on the Laplacian distribution

4. EXPERIMENTS

To evaluate the performance of the proposed method, we performed two experiments using sound samples of 10 different (5 male and 5 female) speakers, with one utterance for each of the speakers. The average length of the utterances was 3.6s, and the sampling frequency was equal to 8 kHz. The reverberant observations were generated by convolving each utterance with measured RIRs for $M = 2$ omni-directional microphones. In the first experiment we used two RIRs from the MARDY database [18], with reverberation time $T_{60} \approx 450$ms, and the loudspeaker positioned centrally at a distance of 1 m from the array. In the second experiment we considered an acoustic scenario in a room with reverberation time $T_{60} \approx 550$ms, where the RIRs were measured using the swept-sine technique.

In both experiments the STFT was calculated using a 32ms Hamming window with 50% overlap. The order of the multichannel regression vector was set to $L_k = 15$ in the first experiment, and to $L_k = 20$ in the second experiment. In both experiments the prediction delay was set to $D = 3$, and the small positive constant was fixed to $\varepsilon_k = 10^{-6}$. The linear programming problem (22) was solved using the CVX software package [19].

The dereverberation performance was evaluated in terms of cepstral distance (CD) and the objective speech quality measure PESQ [20]. The cepstral distance between two signals is defined as

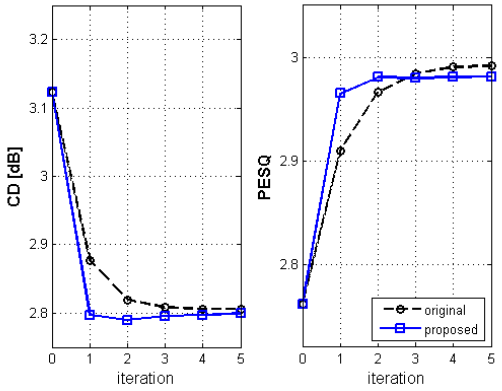
$$CD = \frac{10}{\log 10}\sqrt{(c_0 - \hat{c}_0)^2 + 2\sum_{k=1}^{12}(c_k - \hat{c}_k)^2}, \qquad (25)$$

where c_k and $\hat{c}_k$ are the cepstral coefficients of the anechoic speech signal and the estimated desired signal, respectively. PESQ quantifies the level of similarity between a reference signal and an estimated signal, with the output in the range $1 - 4.5$. The PESQ score was calculated with the anechoic speech signal as the reference signal. For each of the experiments we report the values of CD and PESQ averaged over all of the speakers.

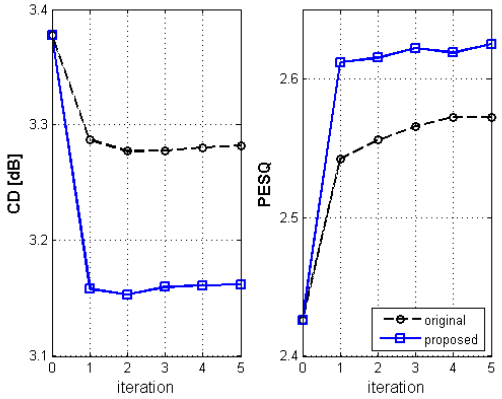
In Figure 1 we compare the original WPE method with the proposed approach in both experimental setups for different number of iterations. The iteration index zero denotes the value of a measure calculated for the observed reverberant signal at the first microphone. It can be seen that the proposed approach outperforms the original WPE both in terms of cepstral distance and perceptual speech quality. The difference is especially visible for a single iteration. Therefore, the proposed approach could be used in a single-iteration mode, similar as was proposed in [7], yielding better performance than the WPE with Gaussian distribution. Additionally, the presented results illustrate that for more iterations both methods exhibit a similar behavior. Hence, it is reasonable to expect that the proposed approach could also benefit by use of additional spectral priors [15], leading to a further increase in the performance.

5. CONCLUSIONS

In this paper we have presented a method for speech dereverberation based on the weighted prediction error method, where the desired signal is modeled as a random variable with a Laplacian distribution. Experimental results in two different acoustic scenarios demonstrate that the proposed approach results in better performance, compared to the original approach that assumes a Gaussian distribution, both in terms of lower speech distortion and higher perceptual speech quality. Incorporation of additional prior knowledge, such as temporal and spectral structure of the desired signal, remains an interesting direction for future research.



(a) Experiment 1



(b) Experiment 2

Fig. 1. Performance for the original WPE and the proposed method (in terms of average CD and PESQ) in two acoustic scenarios for different number of iterations.

6. REFERENCES

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

[2] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. on Signal Processing*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

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

[4] I. Kodrasi, S. Goetze, and S. Doclo, "Regularization for partial multichannel equalization for speech dereverberation," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 21, no. 9, pp. 1879–1890, Sept. 2013.

[5] E. A. P. Habets, S. Gannot, and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 770–773, June 2009.

[6] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B.-H. Juang, "Blind speech dereverberation with multi-channel

linear prediction based on short time Fourier transform representation," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, USA, May 2008, pp. 85–88.

[7] 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 and Language Processing*, vol. 18, no. 7, pp. 1717–1731, Sept. 2010.

[8] D. Schmid, S. Malik, and G. Enzer, "An expectation-maximization algorithm for multichannel adaptive speech dereverberation in the frequency-domain," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, May 2012, pp. 17–20.

[9] M. Togami and Y. Kawaguchi, "Noise robust speech dereverberation with Kalman smoother," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013, pp. 7447–7451.

[10] B. Schwartz, S. Gannot, and E. A. P. Habets, "Multi-microphone speech dereverberation using expectation-maximization and Kalman smoother," in *Proc. European Signal Processing Conference (EUSIPCO)*, Marrakech, Morocco, Sept. 2013.

[11] R. Martin, "Speech enhancement based on minimum mean-square error estimation and supergaussian priors," *IEEE Trans. Speech and Audio Processing*, vol. 13, no. 5, pp. 845–856, Aug. 2005.

[12] J.-H. Chang, "Complex Laplacian probability density function for noisy speech enhancement," *IEICE Electronics Express*, vol. 4, no. 8, pp. 245–250, Apr. 2007.

[13] B. Lee, T. Kalker, and R. W. Schafer, "Maximum-likelihood sound source localization with a multivariate complex Laplacian distribution," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, USA, Sept. 2008.

[14] I. Tashev and A. Acero, "Statistical modeling of the speech signal," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, Sept. 2010.

[15] Y. Iwata and T. Nakatani, "Introduction of speech log-spectral priors into dereverberation based on Itakura-Saito distance minimization," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, May 2012, pp. 245–248.

[16] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.

[17] J. Mattingley and S. Boyd, "Real-time convex optimization in signal processing," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 50–61, May 2010.

[18] J. Y. C. Wen, N. D. Gaubitch, E. A. P. Habets, T. Myatt, and P. A. Naylor, "Evaluation of speech dereverberation algorithms using the MARDY database," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Sept. 2008.

[19] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," http://cvxr.com/cvx, Sept. 2013.

[20] ITU-T, *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs P.862*, International Telecommunications Union (ITU-T) Recommendation, Feb. 2001.