LATE REVERBERANT SPECTRAL VARIANCE ESTIMATION USING ACOUSTIC CHANNEL EQUALIZATION

Benjamin Cauchi[1,4], Patrick A. Naylor[2], Timo Gerkmann[3,4], Simon Doclo[1,3,4], Stefan Goetze[1,4]

[1]Fraunhofer IDMT, Project Group Hearing, Speech and Audio Technology, Oldenburg, Germany
[2]Imperial College London, Dept. of Electrical Engineering, London, United Kingdom
[3]University of Oldenburg, Dept. of Medical Physics and Acoustics, Oldenburg, Germany
[4]Cluster of Excellence Hearing4all, Oldenburg, Germany

ABSTRACT

In many single- and multi-channel speech dereverberation methods an estimate of the late reverberant spectral variance (LRSV) is required. Contrary to LRSV estimators based on room acoustical properties, such as reverberation time, or based on isotropic models of the reverberant sound field, in this paper we propose to use acoustic channel equalization with estimated room impulse responses (RIRs) for LRSV estimation. Unlike the typical application of acoustic channel equalization, where the objective is to estimate the anechoic or the early reverberant speech component, here the late reverberant part of the estimated RIR is set as the target response. The combination of the proposed LRSV estimator with a beamformer and a spectral gain aims at a tradeoff between the performance of acoustic channel equalization and the robustness of methods based on models of the reverberant sound field. The performance, evaluated for different levels of RIR estimation error, is compared to the results obtained using a maximum likelihood estimator (MLE) of the LRSV, based on an isotropic model of the reverberant sound field, and to a state-of-the-art acoustic channel equalization method. Experimental results for different acoustic scenarios show that for medium levels of RIR estimation errors the proposed method outperforms acoustic channel equalization as well as the maximum-likelihood LRSV estimator in terms of instrumental speech quality measures.

Index Terms— Dereverberation, spectral suppression, blocking matrix, channel equalization

1. INTRODUCTION

In many speech communication applications distant microphones are used to record the signal of a target speaker. In an enclosed space the microphone signals are, thus, corrupted by reverberation, which can be characterized by the RIRs between the speaker and the microphones. These RIRs depend on the characteristics of the room as well as on the positions of the speaker and the microphones [1]. As the late part of the RIR is known to be the major source of degradation in terms of speech quality and intelligibility [2, 3], several methods aim to suppress the late reverberation while preserving the direct path and the early reflections [4–8]. A common approach consists in applying a beamformer to the microphone signals, aiming to preserve the direct speech while suppressing the reflections whose directions of arrival (DOA) differ from the direction of the target speaker. However, the output of the beamformer generally still contains a large amount of reverberation. In order to suppress the residual reverberation, a spectral gain is typically applied to the beamformer output [9]. The computation of this spectral gain requires an estimate of the LRSV, whose estimation is the focus of this paper.

Several methods have been proposed for LRSV estimation. A popular class of methods is based on a statistical model of the RIR and the acoustical properties of the room, such as the reverberation time (T_{60}) or the direct to reverberant ratio (DRR) [4, 5, 10]. Other methods instead proposed to estimate the LRSV using the output signal(s) of a blocking matrix, suppressing the signal to be preserved, and from which the LRSV at a reference position can be estimated. This blocking matrix can be designed as a delay-and-subtract beamformer cancelling the direct speech component [11, 12] or as a blind source separation (BSS) scheme cancelling both the direct speech and the early reflections [13, 14]. Alternatively, the LRSV at a reference position can be obtained using a MLE and a model of the reverberant sound field [15]. The LRSV to be used in the computation of the spectral gain is then obtained by correcting the LRSV at the reference position. This correction can be done using an adaptive filter [12], back-projection [13, 14] or the relative transfer functions between the target speaker and the microphones [15].

In principle, channel equalization can achieve perfect reconstruction of the target signal if the RIRs are perfectly es-

The research leading to these results has received funding from the EU Seventh Framework Programme project DREAMS under grant agreement ITN-GA-2012-316969 as well as by the DFG-Cluster of Excellence EXC 1077/1, Hearing4all.

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

timated [7]. However, robustness against RIRs estimation error is still a major issue [7, 8, 16]. In this paper, a reverberant variance estimator based on channel equalization (REVECE) is proposed. It consists in applying partial channel equalization as a blocking matrix to obtain the late reverberant speech component present in the output of the beamformer, as depicted in Fig. 1. Combining spectral enhancement and channel equalization, the proposed method aims at achieving a trade-off between the performance of channel equalization and the robustness of the estimators based on models of the reverberant sound field.

The remainder of this paper is structured as follows. First, the notations and the use of the LRSV estimate for spectral enhancement is decribed in Section 2. Channel equalization is described in Section 3 and the proposed LRSV estimation method based on channel equalization is decribed in Section 4. Experimental results are presented in Section 5, in which the proposed method is compared to both channel equalization and a MLE, in terms of instrumental speech quality measures.

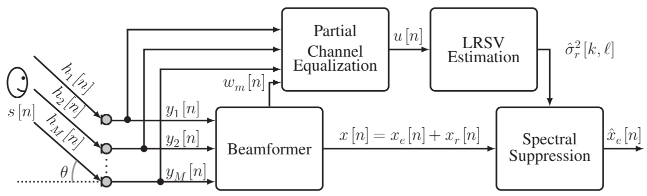

Fig. 1. Overview of the proposed method using channel equalization as a blocking matrix.

2. PROBLEM STATEMENT

Consider an acoustic system with a single speech source and M microphones. In the absence of noise, the signal $y_m[n]$ received by the m-th microphone can be expressed as

$$y_m[n] = s[n] * h_m[n] = s[n] * h_m^e[n] + s[n] * h_m^r[n], \quad (1)$$

with n denoting the sample index, $s[n]$ denoting the clean speech signal, $h_m[n]$ denoting the RIR of length L_h between the speech source and the m-th microphone and

$$h_m^e[n] = \begin{cases} h_m[n] \text{ if } n \leq L_e, \\ 0 \text{ otherwise,} \end{cases} \quad (2)$$

$$h_m^r[n] = \begin{cases} h_m[n] \text{ if } n > L_e, \\ 0 \text{ otherwise,} \end{cases} \quad (3)$$

where L_e is set so that $h_m^e[n]$ contains the direct path and a few early reflections while $h_m^r[n]$ contains the late reflections, i.e, the reverberant tail.

In order to preserve the clean speech signal while suppressing the reflections whose DOAs differ from the DOA θ of the target speaker, a beamformer can be applied to the microphone signals. The beamformer output $x[n]$ is equal to

$$x[n] = \sum_{m=1}^{M} y_m[n] * w_m[n] = x_e[n] + x_r[n], \quad (4)$$

with $w_m[n]$ denoting the beamformer filter, of length L_w, for the m-th microphone signal and with $x_e[n]$ and $x_r[n]$ denoting the spatially filtered early and late reverberant speech components of the beamformer output, respectively.

In the following, the short-time Fourier transforms (STFTs) of $x[n]$, $y_m[n]$, $x_e[n]$ and $x_r[n]$ are denoted by $x[k,\ell]$, $y_m[k,\ell]$, $x_e[k,\ell]$ and $x_r[k,\ell]$ respectively, with k and ℓ denoting the frequency bin and frame indices, respectively. Therefore, in the STFT domain, (4) becomes

$$x[k,\ell] = x_e[k,\ell] + x_r[k,\ell]. \quad (5)$$

In the remainder of this paper, $\sigma_e^2[k,\ell] = \mathcal{E}\left\{|x_e[k,\ell]|^2\right\}$ and $\sigma_r^2[k,\ell] = \mathcal{E}\left\{|x_r[k,\ell]|^2\right\}$ denote the early and late reverberant spectral variances, respectively, with $\mathcal{E}\{\cdot\}$ denoting the expectation operator.

Assuming that the early and late components are uncorrelated, i.e. that the spectral variance $\sigma_x^2[k,\ell]$ of the beamformer output can be expressed as

$$\sigma_x^2[k,\ell] = \sigma_e^2[k,\ell] + \sigma_r^2[k,\ell], \quad (6)$$

an estimate $\hat{x}_e[k,\ell]$ of the direct and early speech component in the beamformer output, $x_e[k,\ell]$, can be obtained using a spectral gain $g[k,\ell]$ as

$$\hat{x}_e[k,\ell] = g[k,\ell]\,x[k,\ell]. \quad (7)$$

The spectral gain $g[k,\ell]$ can be computed using e.g. a Wiener filter [17]

$$g[k,\ell] = \frac{\hat{\sigma}_e^2[k,\ell]}{\hat{\sigma}_e^2[k,\ell] + \hat{\sigma}_r^2[k,\ell]}, \quad (8)$$

with $\hat{\sigma}_e^2[k,\ell]$ and $\hat{\sigma}_r^2[k,\ell]$ denoting estimates of $\sigma_e^2[k,\ell]$ and $\sigma_r^2[k,\ell]$, respectively. This paper focuses on estimating the LRSV, i.e. $\hat{\sigma}_r^2[k,\ell]$.

3. CHANNEL EQUALIZATION

Channel equalization consists in applying a set of filters $c_m[n]$ of length L_c to the microphone signals to obtain the equalized signal $u[n]$ such that

$$u[n] = \sum_{m=1}^{M} y_m[n] * c_m[n] = s[n] * h_t[n], \quad (9)$$

where $h_t[n]$ is the target RIR of length L_t. As most of the applications of channel equalization aim at estimating either the anechoic or the early reverberant speech, the target RIR is usually set to a Dirac delta function or to the early part of the RIR between the source and a reference microphone [7, 16]. The computation of the set of equalization filters $c_m[n]$ requires knowledge of the RIRs $h_m[n]$ which in practice are not available. Therefore, an estimate $\hat{h}_m[n]$ of length $L_{\hat{h}}$ has to be used instead and $c_m[n]$ is the set of filters aiming to

$$h_t[n] = \sum_{m=1}^{M} \hat{h}_m[n] * c_m[n]. \tag{10}$$

In vector notation, (10) can be expressed as

$$\boldsymbol{h}_t = \hat{\boldsymbol{H}}\boldsymbol{c}, \tag{11}$$

with

$$\boldsymbol{h}_t = [\, h_t[0], h_t[1], \cdots, h_t[L_t - 1]\,]^{\mathrm{T}}, \tag{12}$$

and $\boldsymbol{c}$ denoting the ML_c-dimensional stacked filter vector, i.e.

$$\boldsymbol{c} = \left[\, \boldsymbol{c}_1^{\mathrm{T}}\ \boldsymbol{c}_2^{\mathrm{T}}\ \cdots\ \boldsymbol{c}_M^{\mathrm{T}}\,\right]^{\mathrm{T}}, \tag{13}$$

and $\hat{\boldsymbol{H}}$ denoting the $L_t \times ML_c$-dimensional multichannel convolution matrix using the estimated RIRs. Assuming that $\hat{\boldsymbol{H}}$ is full row-rank, the solution to (11) can be computed as

$$\boldsymbol{c} = \hat{\boldsymbol{H}}^{\mathrm{T}}(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{\mathrm{T}})^{-1}\boldsymbol{h}_t, \tag{14}$$

which is refered to as the multiple-input/output inverse theorem (MINT) solution [16] if $\boldsymbol{h}_t$ is set as Dirac delta function and as the partial MINT (PMINT) solution [7] if $\boldsymbol{h}_t$ is set as the early part of the RIR between the source and a reference microphone. However, since the solution in (14) is known to be highly sensitive to RIRs estimation errors, it has been proposed to reduce the norm of the regularization filters [7, 16], using

$$\boldsymbol{c} = (\hat{\boldsymbol{H}}^{\mathrm{T}}\hat{\boldsymbol{H}} + \delta\boldsymbol{I})^{-1}\hat{\boldsymbol{H}}^{\mathrm{T}}\boldsymbol{h}_t, \tag{15}$$

with $\boldsymbol{I}$ denoting the identity matrix and δ denoting a regularization parameter. In [7], a procedure has been presented for automatically determining the parameter δ, leading to the so-called automatically regularized PMINT (ARPMINT) method.

4. LRSV ESTIMATION USING ACOUSTIC CHANNEL EQUALIZATION

The proposed method aims at estimating the LRSV $\sigma_r^2[k,\ell]$ using REVECE, in order to compute a spectral posfilter applied to the output of a beamformer. This estimation consists in constructing an intermediate signal $u[n]$ from which $\hat{\sigma}_r^2[k,\ell]$ is computed. The signal $u[n]$ is the interference signal to be suppressed, i.e. the late reverberation, and can be interpreted as the output of a blocking matrix, as in the context of a generalized sidelobe canceller (GSC).

In this paper, partial channel equalization, requiring an estimate of the RIRs, is used to compute $u[n]$ aiming to obtain

$$u[n] \approx x_r[n] = s[n] * \sum_{m=1}^{M} h_m^r[n] * w_m[n]. \tag{16}$$

It can be seen from (16) that the computation of $u[n]$ is similar to the channel equalization described in (9). However, unlike in the direct application of channel equalization, which aims at estimating either the anechoic or the early reverberant speech as described in Section 3, the target RIR $h_t[n]$ is here set as

$$h_t[n] = \sum_{m=1}^{M} \hat{h}_m^r[n] * w_m[n]. \tag{17}$$

Therefore, using estimates of the RIRs, the signal $u[n]$ can be obtained as

$$u[n] = \sum_{m=1}^{M} y_m[n] * c_m[n] \approx x_r[n], \tag{18}$$

where $c_m[n]$ is the equalization filter which can be computed as described in Section 3. The ARPMINT method (15) is used for this purpose in the remainder of this paper. It can be noted from the convolution of $\hat{h}_m^r[n]$ with the beamformer filter coefficients $w_m[n]$ in (16) that $u[n]$ estimates the late reverberant speech component in the beamformer output. Therefore, the LRSV estimate $\hat{\sigma}_r^2[k,\ell]$ can then be estimated from $|u[k,\ell]|^2$, e.g. using recursive smoothing

$$\hat{\sigma}_r^2[k,\ell] = \beta\hat{\sigma}_r^2[k,\ell-1] + (1-\beta)|u[k,\ell]|^2, \tag{19}$$

where β denotes a smoothing parameter.

5. SIMULATIONS

5.1. Implementation

The simulations have been carried out using speech samples from the TIMIT database [18] convolved with measured RIRs. The RIRs have been measured in two different rooms, an office room with $\mathrm{T}_{60} \approx 500$ ms and a meeting room with $\mathrm{T}_{60} \approx 750$ ms, using a circular array with 10 cm of diameter [19], with M =4 equidistant microphones located 2 m away from the source. All signals were sampled at 16 kHz and the RIRs had a length $L_h = 16000$. The STFTs have been computed using a 32 ms Hann window with 50 % overlap. The smoothing parameter β from (19) has been set to $\beta = 0.67$, corresponding to 40 ms.

The simulations aim at evaluating the performance of REVECE in terms of speech quality. This performance has been assessed using two instrumental speech quality measures, namely the perceptual evaluation of speech quality (PESQ) [20] and the frequency-weighted segmental signal

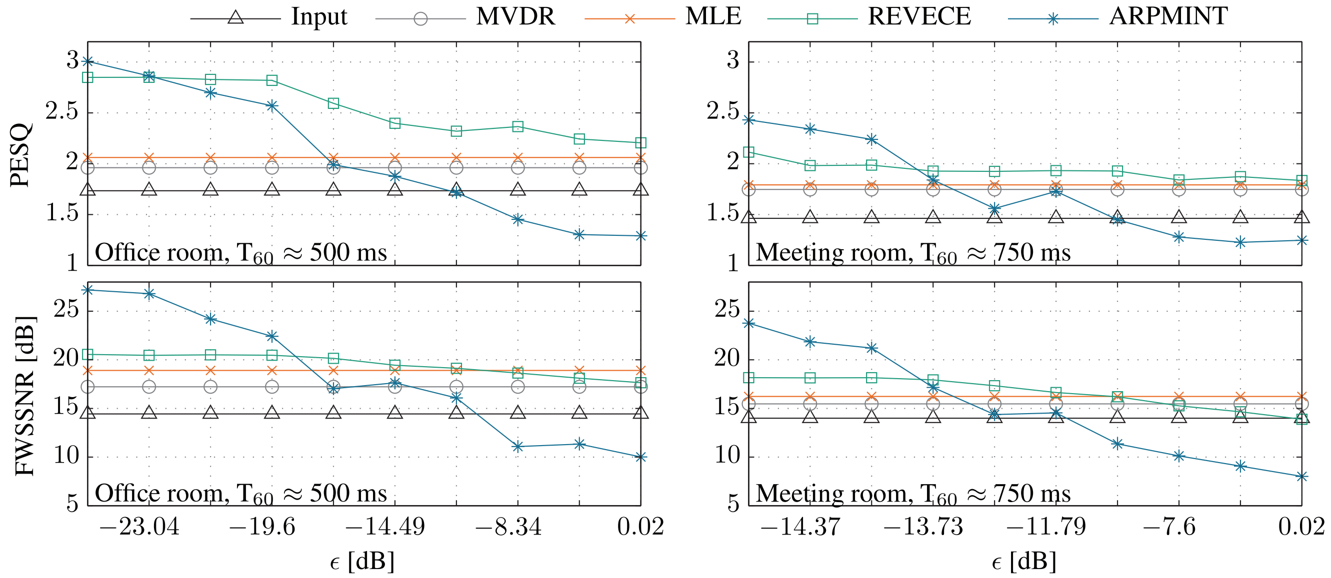

Fig. 2. Speech enhancement achieved in terms of PESQ (top row) and FWSSNR (bottom) using RIRs recorded in an office room (left column) and a meeting room (right).

to noise ratio (FWSSNR) [21] using the implementations provided by the REVERB challenge [19]. As reference signal, the early reverberant speech, i.e. $s[n] * h_1^e[n]$, has been used with L_e corresponding to the first 50 ms of the RIR. The proposed method, REVECE, is compared with the MLE of the LRSV presented in [15], the output of the beamformer (4) and the output obtained by direct application of channel equalization to the microphone signals, ARPMINT [7].

The LRSV estimate $\hat{\sigma}_r^2[k,\ell]$ using REVECE and MLE is used to compute the spectral gain in (8), which is applied to the beamformer output. For the beamformer, we have used a fixed minimum variance distortionless response (MVDR) beamformer in which the steering vector has been computed using far-field assumption and the true DOA and the coherence matrix has been computed assuming a spherically isotropic noisefield. For both REVECE and MLE, the early reverberant spectral variance, $\hat{\sigma}_e^2[k,\ell]$, used in (8), has been estimated using the decision directed approach [22], i.e.

$$\begin{aligned}\hat{\sigma}_e^2[k,\ell] =& \alpha|g[k,\ell-1]x[k,\ell-1]|^2 \\ &+(1-\alpha)\max\left\{0, |x[k,\ell]|^2 - \hat{\sigma}_r^2[k,\ell]\right\},\end{aligned} \tag{20}$$

with the forgetting factor $\alpha = 0.98$, corresponding to 800 ms.

In the case of both REVECE and ARPMINT, the length of the equalization filter has been set to

$$L_c = \left\lceil \frac{L_{\hat{h}} - 1}{M-1} \right\rceil, \tag{21}$$

with $L_{\hat{h}} = 4400$. The choice of $L_{\hat{h}} < L_h$ is motivated by the fact that in practice an estimated RIR of finite length, typically shorter than L_h, is identified using blind system identification algorithms.

Different models of RIR estimation errors have been proposed [7, 23, 24]. Here, the estimated RIR is modeled as the sum of the first $L_{\hat{h}}$ samples of the true RIR and of an error signal $\rho_m[n]$ such that for each channel m

$$\hat{h}_m[n] = h_m[n](1+\rho_m[n]), \text{ for } 0 \leq n \leq L_{\hat{h}} - 1 \tag{22}$$

where $\rho_m[n]$ is a zero-mean Gaussian noise with its variance normalized in order to obtain a certain normalized channel mismatch (NCM) ϵ_m in dB, defined as

$$\epsilon_m = 10\log_{10}\frac{\sum_{n=0}^{L_{\hat{h}}-1}\rho_m^2[n] + \sum_{n=L_{\hat{h}}}^{L_h-1}h_m^2[n]}{\sum_{n=0}^{L_h-1}h_m^2[n]}. \tag{23}$$

5.2. Results

The obtained results are presented in Fig. 2. First, comparing the performance of ARPMINT and of the proposed method, it can be observed that the performance of both ARPMINT and REVECE decreases for high values of NCM. ARPMINT achieves higher PESQ scores for low values of NCM but lower PESQ scores than REVECE for NCM higher than -23 and -13 dB, in the office room and in the meeting room, respectively (cf. top panels). It appears as well that for high values of NCM, $\epsilon >$ -10 dB, the distortions introduced by ARPMINT reduce the speech quality when compared to the output of the MVDR beamformer. Similar effects are observed in terms of FWSSNR (cf. lower panels). The performance of both REVECE and ARPMINT is lower in the case of the meeting room (cf. right panels), for which the energy of the unestimated part of the RIR, $n > L_{\hat{h}}$, is larger as illustrated by the higher T_{60}.

On the other hand, spectral suppression improves PESQ scores compared to the MVDR beamformer when using either MLE or REVECE, for all tested values of NCM. REVECE achieves higher PESQ scores than MLE though the improvement becomes smaller for large identification errors. In terms of FWSSNR, REVECE achieves lower scores than MLE for high values of error, $\epsilon > -10$ dB, suggesting that the estimate of the LRSV becomes less accurate than the one obtained using MLE.

Considering the comparison of REVECE with both MLE and ARPMINT, it appears that the use of channel equalization as blocking matrix for LRSV estimation provides a tradeoff between the performance of channel equalization at low NCM and the robustness of spectral suppression using MLE at high NCM.

6. CONCLUSION

This paper proposes a LRSV estimation method based on acoustic channel equalization (REVECE) to be used in the suppression of the late reverberation present in the output of a beamformer. This method has been compared in terms of instrumental speech quality measure with a MLE and with a state-of-the art channel equalization method. For low RIRs estimation error, channel equalization achieves the best performance while both equalization and REVECE are outperformed by MLE for high error values. For medium levels of identification error, the proposed method outperforms both channel equalization and MLE. Unfortunately, RIR estimates obtained using state-of-the art blind system identification algorithms are not accurate enough yet for the proposed method to be beneficial in realistic conditions.

REFERENCES

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

[2] J. S. Bradley, H. Sato, and M. Picard, "On the importance of early reflections for speech in rooms," *J. Acoust. Soc. Am.*, vol. 113, no. 6, pp. 3233–3244, June 2003.

[3] A. Warzybok, I. Kodrasi, J.O. Jungmann, E. Habets, T. Gerkmann, A. Mertins, S. Doclo, B. Kollmeier, and S. Goetze, "Subjective speech quality and speech intelligibility evaluation of single-channel dereverberation algorithms," in *Proc. Int. Workshop on Acoustic Signal Enhancement (IWAENC 2014)*, Antibes, France, Sep. 2014, pp. 333–337.

[4] K. Lebart, J. M. Boucher, and P. N. Denbigh, "A new method based on spectral subtraction for speech de-reverberation," *Acta Acoustica*, vol. 87, pp. 359–366, 2001.

[5] E. A. P. Habets, S. Gannot, and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Process. Lett.*, vol. 16, no. 9, pp. 770–774, Sept. 2009.

[6] A. Mertins, T. Mei, and M. Kallinger, "Room impulse response shortening/reshaping with infinity- and p-norm optimization," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 2, pp. 249–259, Feb. 2010.

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

[8] F. Lim, W. Zhang, E.AP. Habets, and P.A Naylor, "Robust multichannel dereverberation using relaxed multichannel least squares," in *IEEE/ACM Trans. Netw.*, Sept 2014, vol. 22, pp. 1379–1390.

[9] C. Marro, Y. Mahieux, and K. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with post-filtering," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 3, pp. 240–259, 1998.

[10] B. Cauchi, I. Kodrasi, R. Rehr, S. Gerlach, A. Jukić, T. Gerkmann, S. Doclo, and S. Goetze, "Joint dereverberation and noise reduction using beamforming and a single-channel speech-enhancement scheme," in *Proc. REVERB challenge workshop*, Florence, Italy, May 2014.

[11] 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.

[12] S. Braun and E.A.P. Habets, "Dereverberation in noisy environments using reference signals and a maximum likelihood estimator," in *Proc. European Signal Processing Conference (EUSIPCO)*, Marrakech, Morocco, Sep 2013.

[13] A. Schwarz, K. Reindl, and W. Kellermann, "On blocking matrix-based dereverberation for automatic speech recognition," in *Proc. Intl. Workshop Acoust. Signal Enhancement (IWAENC)*, Aachen, Germany, Sept 2012, pp. 1–4.

[14] 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)*, May 2012, pp. 113–116.

[15] A. Kuklasiński, S. Doclo, S.H. Jensen, and J. Jensen, "Maximum likelihood based multi-channel isotropic reverberation reduction for hearing aids," in *Proc. European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, Sep 2014.

[16] T. Hikichi, M. Delcroix, and M. Miyoshi, "Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuations," *EURASIP Journal on Applied Signal Processing*, vol. 2007, 2007.

[17] J. Benesty, J. Chen, and E. A. P. Habets, *Speech Enhancement in the STFT Domain*, SpringerBriefs in Electrical and Computer Engineering. Springer-Verlag, 2011.

[18] C. Jankowski, A. Kalyanswamy, S. Basson, and J. Spitz, "NTIMIT: a phonetically balanced, continuous speech, telephone bandwidth speech database," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, White Plains, NY, USA, 1990, pp. 109–112.

[19] K. Kinoshita, M. Delcroix, T. Yoshioka, T. Nakatani, E.A.P. Habets, R. Haeb-Umbach, V. Leutnant, A. Sehr, W. Kellermann, R. Maas, S. Gannot, and B. Raj, "The REVERB challenge: A common evaluation framework for dereverberation and recognition of reverberant speech," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, N.Y., U.S.A., Oct. 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," Feb. 2001.

[21] Y. Hu and P.C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 1, pp. 229–238, 2008.

[22] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[23] F. Lim and P.A. Naylor, "Statistical modelling of multichannel blind system identification errors," in *Proc. Int. Workshop on Acoustic Signal Enhancement (IWAENC 2014)*, Antibes, France, Sep. 2014.

[24] J.O. Jungmann, R. Mazur, and A. Mertins, "Perturbation of room impulse responses and its application in robust listening room compensation," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 433–437.