Blind Dereverberation for Hearing Aids with Binaural Link

Marco Jeub, Heinrich W. Löllmann, and Peter Vary

Institute of Communication Systems and Data Processing (IND), RWTH Aachen University, Germany
E-Mail: {jeub,loellmann,vary}@ind.rwth-aachen.de
Web: www.ind.rwth-aachen.de

Abstract

A new two-stage algorithm for binaural dereverberation is proposed which achieves a joint suppression of early and late reverberant speech. All needed quantities are estimated blindly from the reverberant speech and no information about the acoustical environment such as the reverberation time (RT) is required.

The first stage of the algorithm is based on a spectral subtraction rule which depends on the spectral variance of the late reverberant speech. The calculation of the spectral variances of the late reverberant speech requires an estimate of the reverberation time. This is accomplished by an efficient algorithm which is based on a maximum likelihood (ML) estimation. In a second stage, the output is further enhanced by a multi-channel Wiener filter. This is derived by a coherence model which takes the shadowing effects of the head into account. The overall binaural input-output processing does not affect the most important binaural cues, i.e., the interaural time difference (ITD) and interaural level difference (ILD). This is important especially for speech enhancement in hearing aids to preserve the ability for source localization in the azimuth plane.

Experiments have shown that the new system achieves a significant reduction of early and late reverberation.

1 Introduction

Room reverberation can lead to a degradation of speech quality and intelligibility especially for hearing impaired people. Therefore, it is desirable that modern hearing aids can reduce the detrimental effects of speech reverberation.

Since a joint suppression of both early *and* late reverberation is quite challenging, several (single- and multi-channel) two-stage algorithms are proposed in the literature. The authors in [1] present an inverse filtering algorithm which maximizes the kurtosis of the residual signal obtained by linear prediction (LP) for the reduction of early reverberation, followed by a spectral subtraction rule that reduces long-term reverberation. A similar approach is described in [2] where spatiotemporal averaging is combined with a spectral subtraction algorithm.

The major drawback is that most of these techniques were developed for systems with a single output channel given one or possibly multiple input channels. Therefore, they are only suitable for independent processing, termed as *bilateral*. Several studies have shown that such processing degrades the ability for sound localization and that hearing impaired persons localize sounds better without their independent bilateral hearing aids than with them, cf., [3]. This can be explained by the fact that the binaural cues, which are the basis for human sound localization, are not preserved. This comprises mostly the interaural level difference (ILD) and interaural time difference (ITD).

Therefore, it is advantageous to perform *binaural* instead of bilateral processing, especially as an appropriate data link between both sides of the hearing aid can be assumed in the future, cf., [4]. In order to preserve the binaural cues and to allow for a "real" binaural processing, several approaches have been proposed in the past. A comprehensive study how binaural noise reduction algorithms can preserve binaural cues can be found in [3]. A binaural blind source separation (BSS) strategy is proposed, e.g., in [5]. The problem in terms of binaural dereverberation is addressed, e.g., in [6].

In this contribution, a novel *blind* two-stage binaural dereverberation system is proposed. All needed quantities are estimated blindly from the reverberant speech such that no a priori information about acoustical parameters is required.

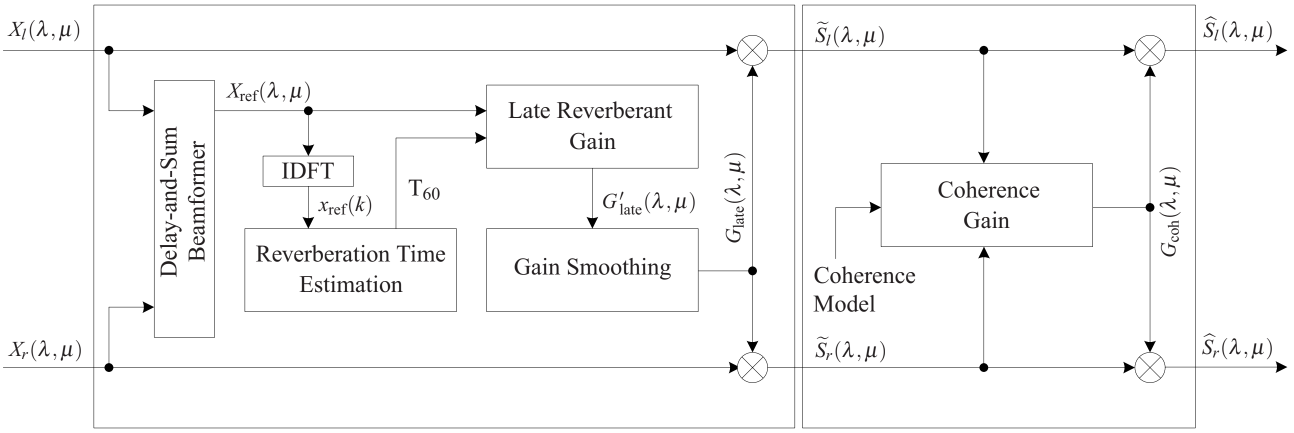
2 Dereverberation System

The proposed dereverberation system is based on [6, 7, 8] and consists of two independent stages as depicted in Fig. 1. The considered algorithms are realized by short-term spectral weighting using the weighted overlap-add method. For the transformation into the frequency domain, the disturbed input signals $x_{l|r}(k)$ at sampling frequency f_s are first segmented into overlapping frames (Hann window, 50% overlap) of length L. After windowing, these frames are transformed via Fast Fourier Transform (FFT) of length M into the short-term spectral domain. At discrete frequency bins μ and frame λ, the distorted signals for right and left channel are $X_{l|r}(\lambda,\mu)$. The enhanced spectra $\widehat{S}_{l|r}(\lambda,\mu)$ can be obtained by multiplying the coefficients $X_{l|r}(\lambda,\mu)$ with the weighting gains $G_{\text{late}}(\lambda,\mu)$ and $G_{\text{coh}}(\lambda,\mu)$ of the two stages. The same weighting gains are applied to the left and right channel such that the ILD is unaffected. In order to ensure an unaffected interaural phase difference (which is used by the human auditory system for the determination of the ITD), the phase of the disturbed input signals is kept. Additionally, each channel shows the same algorithmic delay. This concept is also used in binaural noise reduction algorithms, cf., [9]. The enhanced time domain signals $\widehat{s}_{l|r}(k)$ are obtained by an Inverse Fast Fourier Transform (IFFT) followed by an overlap-add operation. In the following, the calculation of the spectral weighting gains is described.

2.1 Stage I: Dereverberation Based on a Statistical Model of Late Reverberation

The first stage of our binaural dereverberation system is based on the model-based algorithm proposed in [10]. The basic idea is to estimate the variance of the late reverberant speech components and to formulate a weighting rule that aims to suppress late reverberant components while leaving the direct path and early reflections unaltered. In the presented binaural system, a delay-and-sum beamformer (two microphones, left and right) is used to generate the reference signal $X_{\text{ref}}(\lambda,\mu)$ from which the binaural spectral weights are calculated. One crucial aspect of this stage

ISBN 978-3-8007-3300-2 © VDE VERLAG GMBH · Berlin · Offenbach


$X_l(\lambda,\mu)$
$X_{\mathrm{ref}}(\lambda,\mu)$
Late Reverberant Gain
Delay-and-Sum Beamformer
IDFT
$x_{\mathrm{ref}}(k)$
T_{60}
$G'_{\mathrm{late}}(\lambda,\mu)$
Reverberation Time Estimation
Gain Smoothing
$G_{\mathrm{late}}(\lambda,\mu)$
$X_r(\lambda,\mu)$
$\widetilde{S}_l(\lambda,\mu)$
$\widehat{S}_l(\lambda,\mu)$
Coherence Gain
Coherence Model
$G_{\mathrm{coh}}(\lambda,\mu)$
$\widetilde{S}_r(\lambda,\mu)$
$\widehat{S}_r(\lambda,\mu)$

Stage I. Binaural Spectral Subtraction.

Stage II. Binaural Coherence Filter.

Figure 1: Schematic diagram of the proposed two-stage binaural cue preserving dereverberation algorithm.

is a robust estimation of the reverberation time (RT) T_{60} which is described in Section 2.3.

A reverberant signal $x(k)$ can be decomposed into its early and late reverberant speech components $x_{\mathrm{early}}(k)$ and $x_{\mathrm{late}}(k)$ according to

$$x(k) = \underbrace{\sum_{n=0}^{\mathrm{T}_{\mathrm{late}} f_s - 1} s(k-n)h(n)}_{x_{\mathrm{early}}(k)} + \underbrace{\sum_{n=\mathrm{T}_{\mathrm{late}} f_s}^{\mathrm{T} f_s - 1} s(k-n)h(n)}_{x_{\mathrm{late}}(k)}, \quad (1)$$

where $\mathrm{T}_{\mathrm{late}}$ marks the time span after which the late reverberation begins and T the total length of the RIR. The corresponding DFT spectra are marked by $X_{\mathrm{early}}(\lambda,\mu)$ and $X_{\mathrm{late}}(\lambda,\mu)$, respectively.

Based on a statistical model of the RIR proposed by Polack, it can be shown that the late reverberant component $x_{\mathrm{late}}(k)$ can be modeled as an uncorrelated noise process [10].

An estimator for the variance of the late reverberant speech is given by

$$\sigma^2_{x_{\mathrm{late}}}(\lambda,\mu) = e^{-2\rho \mathrm{T}_{\mathrm{late}}} \cdot \sigma^2_x(\lambda - \mathrm{N}_{\mathrm{late}},\mu), \quad (2)$$

with spectral variance of the reverberant speech denoted by $\sigma^2_x(\lambda,\mu)$ and $\mathrm{N}_{\mathrm{late}}$ marks the number of frames corresponding to the time span $\mathrm{T}_{\mathrm{late}}$. Decay rate ρ and RT T_{60} are related by

$$\rho = \frac{3 \cdot \ln 10}{\mathrm{T}_{60}}. \quad (3)$$

In order to estimate the *a posteriori* signal-to-interference ratio (SIR)

$$\eta(\lambda,\mu) = \frac{|X(\lambda,\mu)|^2}{\sigma^2_{x_{\mathrm{late}}}(\lambda,\mu)}, \quad (4)$$

the spectral variance of the reverberant speech is calculated by recursive averaging

$$\sigma^2_x(\lambda,\mu) = \alpha_1 \cdot \sigma^2_x(\lambda-1,\mu) + (1-\alpha_1) \cdot |X(\lambda,\mu)|^2, \quad (5)$$

with a smoothing factor $0 \le \alpha_1 \le 1$. The weights for the suppression of the late reverberant components are calculated by the spectral magnitude subtraction rule

$$G'_{\mathrm{late}}(\lambda,\mu) = 1 - \frac{1}{\sqrt{\eta(\lambda,\mu)}}. \quad (6)$$

Additionally, a lower bound $G^{\mathrm{late}}_{\mathrm{min}}$ is applied to all weighting gains to counter an overestimation of $\sigma^2_{x_{\mathrm{late}}}(\lambda,\mu)$.

For reducing the amount of musical tones of the spectral subtraction approach of (6), spectral smoothing of the magnitudes $G'_{\mathrm{late}}(\lambda,\mu)$ is performed as proposed in [11]. The main idea is to reduce the annoying musical tones especially in low signal-to-interference ratio (SIR) regions. Within the smoothing procedure, the weighting gain magnitudes are convoluted over frequency μ by a lowpass filter $H_s(\lambda,\mu)$ in every frame λ:

$$G_{\mathrm{late}}(\lambda,\mu) = G'_{\mathrm{late}}(\lambda,\mu) * H_s(\lambda,\mu). \quad (7)$$

Finally, the smoothed weighting gains $G_{\mathrm{late}}(\lambda,\mu)$ are applied to the disturbed input spectra by

$$\widetilde{S}_l(\lambda,\mu) = X_l(\lambda,\mu) \cdot G_{\mathrm{late}}(\lambda,\mu) \quad (8a)$$

$$\widetilde{S}_r(\lambda,\mu) = X_r(\lambda,\mu) \cdot G_{\mathrm{late}}(\lambda,\mu). \quad (8b)$$

2.2 Stage II: Reduction of Early Reverberation Exploiting Sound Field Coherence

The motivation for a second processing step is that the spectral subtraction rule described in the previous subsection aims at reducing late reverberation only and hence, residual reverberation remains. The subsequent coherence-based dereverberation algorithm exploits the low coherence of the reverberant sound field between different microphones to estimate the (direct) power spectral density and to remove all non-coherent signal parts while keeping the coherent parts unaffected. Since only the direct speech shows a high coherence between both microphones, this approach also reduces early reverberation. The stage is based on a Wiener filter where the coefficients are calculated based on the two input signals of the left and right channel.

For calculating the Wiener filter coefficients in multi-channel systems, several approaches have been presented in the past. They all have in common that the estimation procedure is optimized for a specific model of the sound field coherence $\Gamma_{x_l x_r}(\Omega)$. A well-known technique by Zelinski assumes a perfectly incoherent sound field and hence, uncorrelated noise at different sensors. Since this assumption does not hold in real sound fields, an improved approach was presented by McCowan [12] who proposed to use a model for the coherence of a spherically isotropic (diffuse) sound field. Since the head-shadowing has a severe impact on the coherence between the microphone signals, we use the coherence model for a binaural isotropic sound field as discussed in [6, 7] in the following.

An estimate of the original (undistorted) speech auto-power spectral density (APSD) is calculated by the rule of [12]

$$\widehat{\Phi}_{ss}(\lambda,\mu) = \frac{\mathrm{Re}\{\widehat{\Phi}_{\widetilde{s}_l\widetilde{s}_r}(\lambda,\mu)\} - \frac{1}{2}\mathrm{Re}\{\Gamma_{\widetilde{s}_l\widetilde{s}_r}(\Omega)\}\Big(\widehat{\Phi}_{\widetilde{s}_l\widetilde{s}_l}(\lambda,\mu) + \widehat{\Phi}_{\widetilde{s}_r\widetilde{s}_r}(\lambda,\mu)\Big)}{1 - \mathrm{Re}\{\Gamma_{\widetilde{s}_l\widetilde{s}_r}(\Omega)\}}, \tag{9}$$

where the hat-operator $\{\widehat{\cdot}\}$ indicates an estimate as shown later. The function $\mathrm{Re}\{\cdot\}$ returns the real part of its argument. Since the estimate of the signal APSD may not be negative or singular, a maximum threshold $\Gamma_{\max}$ for the coherence function has to be applied to ensure that $1 - \mathrm{Re}\{\Gamma_{\widetilde{s}_l\widetilde{s}_r}(\Omega)\} > 0$. The coherence function $\Gamma_{\widetilde{s}_l\widetilde{s}_r}(\Omega)$ is derived by an improved model as proposed in [6, 7].

The resulting spectral weights of the Wiener filter can now be calculated by

$$G_{\mathrm{coh}}(\lambda,\mu) = \frac{\widehat{\Phi}_{ss}(\lambda,\mu)}{\frac{1}{2}\cdot\Big(\widehat{\Phi}_{\widetilde{s}_l\widetilde{s}_l}(\lambda,\mu) + \widehat{\Phi}_{\widetilde{s}_r\widetilde{s}_r}(\lambda,\mu)\Big)}. \tag{10}$$

The spectral weights are further confined by a lower threshold $G_{\min}^{\mathrm{coh}}$ for robustness against overestimation errors and to control the amount by which reverberation is attenuated. The spectral weights are applied to each of the two channels by

$$\widehat{S}_l(\lambda,\mu) = \widetilde{S}_l(\lambda,\mu)\cdot G_{\mathrm{coh}}(\lambda,\mu) \tag{11a}$$

$$\widehat{S}_r(\lambda,\mu) = \widetilde{S}_r(\lambda,\mu)\cdot G_{\mathrm{coh}}(\lambda,\mu). \tag{11b}$$

The required auto- and cross-power spectral densities are calculated by a recursive periodogram approach with smoothing factor α_2.

2.3 Blind Reverberation Time Estimation

The estimation of the RT T_{60} or decay rate ρ, respectively, is important for the calculation of the spectral variance of the late reverberant speech according to Eq. (2). This estimation is performed by the algorithm proposed in [8]. In contrast to previous approaches for a blind RT estimation based on a maximum likelihood (ML) estimation [13, 14], this improved algorithm exhibits a significantly reduced computational complexity and is more suitable to track time-varying RTs. Such properties are of special importance for an application within hearing aids where only a very limited computational power is available. In the following, only the main steps and properties of this algorithm are outlined where a more detailed description is provided in [8].

The blind RT estimation is performed on the time domain signal $x_{\mathrm{ref}}(k)$ (see Fig. 1). In a first step, this signal is downsampled by a ratio of five to reduce the computational burden for estimating the RT. Afterwards, a pre-selection is performed to detect segments within the reference signal which possibly contains only sound decay. If such a sound decay is detected, the RT is estimated by a ML estimation and the obtained value is used to update a histogram determined by the most recent ML estimates. The value associated with the maximum of this histogram is taken as estimate for the RT. In order to detect changes of the RT more rapidly, a second histogram with a lower number of ML estimates for the RT is also calculated. If the maximum of this second histogram differs from that of the first histogram by more than 0.2 s for a certain period, the first histogram is replaced by the second histogram. The RT value associated with the maximum of this histogram is taken as RT estimate. A recursive smoothing is finally applied to this RT value to reduce the variance for the estimate. The decay rate ρ can be finally obtained by Eq. (3).

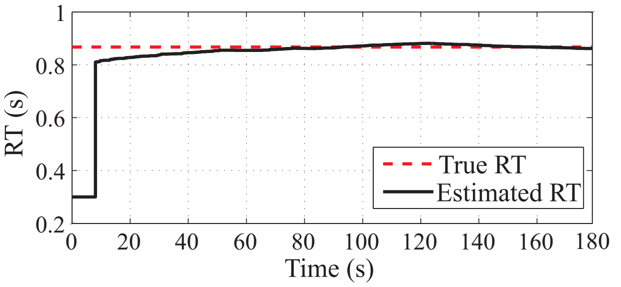


Figure 2: Tracking performance of the proposed reverberation time estimator (Lecture room, $\mathrm{T}_{60} = 0.86\,\mathrm{s}$).

2.4 Experiments

The performance of the blind dereverberation system will be evaluated by some simulation examples. An anechoic speech signal of 180 s was convolved with three different binaural room impulse responses from the AIR database [15]:

- Office : $\mathrm{T}_{60} = 0.37\,\mathrm{s}$,
- Lecture room : $\mathrm{T}_{60} = 0.86\,\mathrm{s}$,
- Stairway : $\mathrm{T}_{60} = 0.69\,\mathrm{s}$.

The single speech file with such long duration was used in order to take the tracking performance and the adaptation speed of the RT estimation procedure into account.

In a first experiment, the tracking performance of the RT estimator is investigated. Figure 2 shows the true and estimated reverberation time exemplary for the lecture room. It can be seen that after a short adaptation period, the estimate converges towards the true RT. Similar results can be obtained with the other reverberant signals. (The performance of this RT estimation for time-varying room impulse responses is analyzed in [8].) In a second experiment, the influence of the blind RT estimator is investigated. The dereverberation is performed with either

- a priori knowledge of the RT or
- blind estimation of the RT.

For an objective evaluation, the non-intrusive measurement based on the Speech to Reverberation Modulation energy Ratio (SRMR) is employed [16]. Furthermore, the Bark Spectral Distortion (BSD) is used as a perceptually motivated spectral distance measure. The reference signal for the BSD is the direct path signal. All signal levels are normalized to −26 dBov using the ITU-T Rec. P.56 speech voltmeter. Further simulation parameters are listed in Table 1.

It can be observed that the blind estimation does not cause a significant difference in comparison to the algorithm with knowledge about the actual RT. The slightly

Table 1: Main simulation parameters.

Parameter	Setting
Sampling frequency	$f_s = 16\,\text{kHz}$
Frame length	$L = 256$
FFT length	$M = 256$
Frame overlap	50% (Hann window)
Smoothing factors	$\alpha_1 = 0.9, \alpha_2 = 0.8$
Coherence threshold	$\Gamma_{\max} = 0.99$
Gain factor thresholds	$G_{\min}^{\text{late}} = G_{\min}^{\text{coh}} = 0.3$
Late reverberant time span	$\text{T}_{\text{late}} = 0.08\,\text{s}$

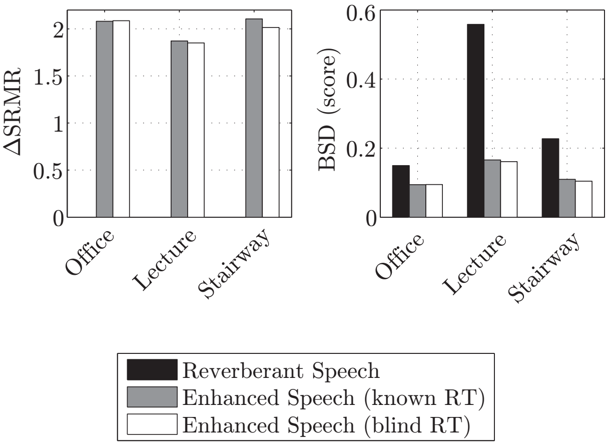


Figure 3: Evaluation of the binaural dereverberation algorithm with known and blindly estimated reverberation time. SRMR entries give the difference to the reverberant speech and BSD indicates the spectral distortions between the direct signal and the reverberant/processed signal.

lower value for the SRMR can be explained by the fact that the RT estimator tends to underestimate the RT which has no noticeable effect for the speech quality. Informal listening tests revealed no audible difference. A comparison of the algorithm with known RT to related binaural dereverberation algorithms can be found in [6].

2.5 Conclusions

This paper proposes a blind two-stage speech enhancement algorithm for binaural dereverberation which is based on a model of the room impulse response (RIR) and a model of the sound field coherence. The required reverberation time (RT) is estimated blindly by means of a Maximum Likelihood (ML) based approach which has a rather low computational complexity. The overall binaural input-output structure does not affect the most important binaural cues, i.e., interaural time difference (ITD) and interaural level difference (ILD), and hence, keeps the localization ability. In simulations with measured binaural room impulse responses, the proposed system achieves a significant reduction of early and late reverberation, which was confirmed by informal listening tests. An audible difference between speech signals processed by the system with known RT and estimated RT was not noticeable. In a further step, the algorithm can be extended by a noise reduction module in order to allow for a joint noise reduction and dereverberation. Additionally, an adaptive coherence model based on a measure of the "diffusiveness" can be employed.

References

[1] M. Wu and D. Wang, "A two-stage algorithm for one-microphone reverberant speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 3, pp. 774–784, 2006.

[2] N.D. Gaubitch, E.A.P. Habets, and P.A. Naylor, "Multimicrophone speech dereverberation using spatiotemporal and spectral processing," in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS)*, 2008, pp. 3222–3225.

[3] T. van den Bogeart, *Preserving binaural cues in noise reduction algorithms for hearing aids*, Ph.D. thesis, Katholieke Universiteit Leuven, Leuven, Belgium, 2008.

[4] V. Hamacher, J. Chalupper, J. Eggers, E. Fischer, U. Kornagel, H. Puder, and U. Rass, "Signal processing in high-end hearing aids: State of the art, challenges, and future trends," *EURASIP Journal on Applied Signal Processing*, vol. 18, pp. 2915—-2929, 2005.

[5] K. Reindl, Y. Zheng, and W. Kellermann, "Speech enhancement for binaural hearing aids based on blind source separation," in *Proc. 4th Int. Symp.on Communications, Control, and Signal Proc. (ISCCSP)*, Limassol, Cyprus, 2010.

[6] M. Jeub, M. Schäfer, T. Esch, and P. Vary, "Model-based dereverberation preserving binaural cues," *IEEE Transactions on Audio, Speech, and Language Processing*, to appear.

[7] M. Jeub and P. Vary, "Binaural dereverberation based on a dual-channel wiener filter with optimized noise field coherence," in *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, TX, USA, 2010.

[8] H.W. Löllmann, E. Yilmaz, M. Jeub, and P. Vary, "An improved algorithm for blind reverberation time estimation," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, 2010.

[9] T. Lotter and P. Vary, "Dual-channel speech enhancement by superdirective beamforming," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1–14, 2006.

[10] 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.

[11] T. Esch and P. Vary, "Efficient musical noise suppression for speech enhancement systems," in *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Taipei, Taiwan, 2009, pp. 4409–4412.

[12] I.A. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence," *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 6, pp. 709–716, 2003.

[13] R. Ratnam, D. L. Jones, B. C. Wheeler, W. D. O'Brien, C. R Lansing, and A. S. Feng, "Blind Estimation of Reverberation Time," *Journal of the Acoustical Society of America*, vol. 114, no. 5, pp. 2877–2892, 2003.

[14] H.W. Löllmann and P. Vary, "Estimation of the reverberation time in noisy environments," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, USA, 2008.

[15] M. Jeub, M. Schäfer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *Proc. Int. Conference on Digital Signal Processing (DSP)*, Santorini, Greece, 2009.

[16] T.H. Falk and W.-Y. Chan, "A non-intrusive quality measure of dereverberated speech," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, USA, 2008.