A Blind Algorithm for Joint Noise Suppression and Dereverberation

Heinrich W. Löllmann and Peter Vary [1]

Institute of Communication Systems and Data Processing (ind), RWTH Aachen University, 52056 Aachen
E-Mail: {loellmann,vary}@ind.rwth-aachen.de Web: www.ind.rwth-aachen.de

Abstract

This contribution presents a novel speech enhancement algorithm for the suppression of background noise and late reverberation without a priori knowledge. The speech enhancement is performed by a generalized spectral weighting rule based on spectral power estimations for the late reverberant speech and background noise. By this, speech distortions due to late room reflections can be suppressed without knowledge about the complete room impulse response. In contrast to existing methods, all needed quantities are estimated entire blindly from the reverberant and noisy speech signal. A possible application of the proposed algorithm are speech enhancement systems for hands-free devices, mobile phones or hearing aids.

1 Introduction

Algorithms for the enhancement of distorted speech signals are used for a variety of applications including hands-free communications systems, mobile phones or hearing aids. Commonly employed systems for (single-channel) speech enhancement aim mostly at a suppression of disturbing background noise, e.g., [1] but do not treat additional signal distortions due to room reverberation. Such impairments are caused through the multiple reflections and diffraction of the sound on walls and objects of a room. These multiple echoes add to the direct sound at the receiver and blur its temporal and spectral characteristic. As a consequence, reverberation and background noise reduce listening comfort and speech intelligibility, especially for hearing impaired persons [2].

Numerous proposals for speech dereverberation have been made in recent years (see, e.g., [3,4] for an overview). Of special interest for practical applications are dereverberation algorithms which are based on spectral speech estimators, e.g., [4–6]. They are derived by a simple statistical model for the *room impulse response* (RIR) and perform speech enhancement by spectral weighting. This suppresses the effects of late reverberation, which cause detrimental overlap-masking effects where a phoneme is smeared over time and overlaps with following phonemes, which impairs speech intelligibility. Spectral speech enhancement algorithms have a manageable computational complexity and do not require knowledge about the complete RIR, but only some characteristic parameters such as the *reverberation time* (RT).

However, the reverberation time estimation (RTE) is particularly difficult as a *blind* estimation is required out of the reverberant and noisy speech signal. Hence, many proposals for speech dereverberation exclude this difficulty by assuming that the RT is known or can be estimated reliably, e.g., [4, 6]. Lebart et al. [5] propose a partially blind estimation of the RT by detecting suitable 'gaps' in the reverberant (noiseless) speech signal which contain, e.g., a sound decay after a sharp speech drop-off, but they do not describe their estimation method in detail.

In this paper, a new algorithm for joint noise suppression and dereverberation is proposed, which estimates all needed quantities blindly so that no a priori knowledge is required. The RT is estimated by a maximum likelihood approach without the need for a heuristic gap detection. The new enhancement algorithm has a low signal delay and manageable computational complexity, which enables its use for speech enhancement systems in mobile communication devices or hearing aids.

[1] This work was supported by GN ReSound, Eindhoven, The Netherlands.

2 Speech Enhancement System

2.1 Signal Model

The distorted signal $x(k)$ is given by the superposition of the reverberant speech signal $z(k)$ and additive noise $v(k)$, where k marks the discrete sampling index. The received signal $x(k)$ and the original (undistorted) speech signal $s(k)$ are related by

$$x(k) = z(k) + v(k) = \sum_{n=0}^{L_\mathrm{R}-1} s(k-n)\, h_\mathrm{R}(n,k) + v(k) \quad (1)$$

with $h_\mathrm{R}(n,k)$ denoting the time-varying RIR of length L_R between source and receiver. The reverberant signal can be decomposed into

$$z(k) = \underbrace{\sum_{n=0}^{L_\mathrm{e}-1} s(k-n)\, h_\mathrm{R}(n,k)}_{=\, z_\mathrm{e}(k)} + \underbrace{\sum_{n=L_\mathrm{e}}^{L_\mathrm{R}-1} s(k-n)\, h_\mathrm{R}(n,k)}_{=\, z_\mathrm{l}(k)} . \quad (2)$$

The signal $z_\mathrm{e}(k)$ is termed as *early speech component* and constitutes the target signal of our speech enhancement algorithm. The suppression of the *late reverberant speech* $z_\mathrm{l}(k)$ and additive noise $v(k)$ is accomplished by modeling them both as uncorrelated, random noise processes so that spectral enhancement techniques can be applied. This concept, which has been introduced by Lebart et al. [5] and further improved by Habets [4, 6], forms the basis for our speech enhancement algorithm.

2.2 Low Delay Filtering

The speech enhancement is performed by spectral weighting in the frequency-domain

$$\widehat{S}(i,\lambda) = X(i,\lambda) \cdot W_i(\lambda)\,; \quad i \in \{0,1,\ldots,M-1\} \quad (3)$$

with i denoting the frequency (channel) index and λ marking the subsampled time index $\lambda = \lfloor \frac{k}{R} \rfloor$. (The operation $\lfloor . \rfloor$ returns the greatest integer value which is lower than or equal to the argument.) For block-wise processing, the downsampling rate $R \in \mathbb{N}\backslash\{0\}$ corresponds to the frame shift and λ to the frame index.

Eq. (3) can be either implemented by the overlap-add method [7], or the *filter-bank equalizer* of [8] which is considered here. This filter-bank concept allows for a low algorithmic signal delay and is beneficial if the vantages of a non-uniform (Bark-scaled) frequency resolution should be exploited, e.g., for speech enhancement in hearing aids [9] or near end listening enhancement [10].

A general representation of the overall system is provided by Fig. 1. The subband signals $X(i,\lambda)$ are calcu-

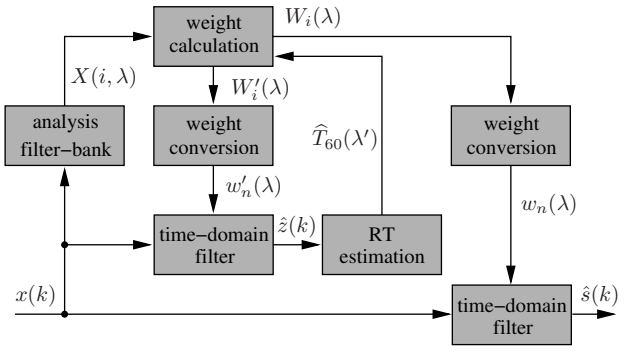

Figure 1: System for low delay noise reduction and dereverberation with blind RT estimation.

lated by a (uniform or warped) DFT analysis filter-bank with downsampling by R, which can be efficiently implemented by a polyphase network. The spectral coefficients $X(i,\lambda)$ are needed to calculate the spectral weights for speech enhancement $W_i(\lambda)$ as well as the weights $W_i'(\lambda)$ for speech denoising prior to the RTE. These spectral weights are converted to time-domain filter coefficients $w_n(\lambda)$ and $w_n'(\lambda)$ by means of the generalized discrete Fourier transform (GDFT)

$$w_n(\lambda) = \frac{h(n)}{M} \sum_{i=0}^{M-1} W_i(\lambda)\, e^{-j\frac{2\pi}{M}\, i\,(n-L/2)} \qquad n \in \{\,0,1,\ldots L-1\,\} \tag{4}$$

where $h(n)$ denotes the real impulse response of the prototype lowpass filter for the analysis filter-bank. It is also possible to approximate the (uniform or warped) time-domain filter by an FIR or IIR filter of lower degree to further reduce the signal delay. A more detailed description of the outlined filter (bank) concept can be found in [8, 9].

2.3 Spectral Weight Calculation

The weights are calculated by the spectral subtraction rule

$$W_i^{(\mathrm{ss})}(\lambda) = 1 - \frac{1}{\sqrt{\hat{\gamma}(i,\lambda)}}\,; \quad i \in \{0,1,\ldots,M-1\}\,. \tag{5}$$

This method achieves a good speech quality with low computational complexity, but other, more sophisticated estimators such as [11] can be employed as well, cf. [4].

The spectral weights of Eq. (5) depend on an estimation of the *a posteriori signal-to-interference ratio* (SIR)

$$\gamma(i,\lambda) = \frac{E\left\{|X(i,\lambda)|^2\right\}}{\sigma_{z_\mathrm{l}}^2(i,\lambda) + \sigma_v^2(i,\lambda)} \tag{6}$$

which is related to the *a priori* SIR according to

$$\xi(i,\lambda) = \frac{E\left\{|Z_\mathrm{e}(i,\lambda)|^2\right\}}{\sigma_{z_\mathrm{l}}^2(i,\lambda) + \sigma_v^2(i,\lambda)} = \gamma(i,\lambda) - 1 \tag{7}$$

due to the model of Eq. (2). If no reverberation is present, i.e., $z(k) = s(k)$, Eq. (6) and Eq. (7) reduce to the well-known *a posteriori* and *a priori signal-to-noise ratio* (SNR). The *a priori* SIR can be estimated by using the recursive decision-directed approach of [11]

$$\hat{\xi}(i,\lambda) = \eta \cdot \frac{|\hat{Z}_e(i,\lambda-1)|^2}{\hat{\sigma}_{z_\mathrm{l}}^2(i,\lambda-1) + \hat{\sigma}_v^2(i,\lambda-1)} + (1-\eta)\cdot \max\left\{\hat{\gamma}(i,\lambda) - 1, 0\right\}\,;\; 0.8 < \eta < 1 \tag{8}$$

which is beneficial to avoid audible artifacts such as musical tones. The spectral weights are confined by a lower threshold[1]

$$W_i(\lambda) = \max\left\{W_i^{(\mathrm{ss})}(\lambda), \epsilon_{\mathrm{thr}}(i,\lambda)\right\}\,. \tag{9}$$

This allows to balance the trade-off between the amount of interference suppression on one hand, and musical tones on the other hand.

2.4 Interference Power Estimation

A crucial issue is the estimation of the power spectral densities (PSDs) of the interfering sources to obtain the *a priori* SIR of Eq. (6). The noise PSD $\sigma_v^2(i,\lambda)$ can be estimated by common techniques such as minimum statistics [12].

The estimator for the variance of the late reverberant speech $\sigma_{z_\mathrm{l}}^2(i,\lambda)$ can be obtained by means of a statistical model for the RIR [4]

$$h_\mathrm{M}(t) = \begin{cases} b_\mathrm{d}(t)\, e^{-j\rho\, t} & \text{for } \; 0 \le t < T_\mathrm{r} \\ b_\mathrm{r}(t)\, e^{-j\rho\, t} & \text{for } \; T_\mathrm{r} \le t \\ 0 & \text{for } \; 0 > t \end{cases} \tag{10}$$

with $b_\mathrm{d}(t)$ and $b_\mathrm{r}(t)$ representing uncorrelated, stationary, white Gaussian noise processes with zero mean. The value for T_r distinguishes between the part of the RIR modeling the direct path ($t < T_\mathrm{r}$), and the part modeling all later reflections ($t > T_\mathrm{r}$). If this distinction is not made (i.e., $T_\mathrm{r} = 0$), the RIR model reduces to

$$h_\mathrm{M}(t) = \begin{cases} b_\mathrm{r}(t)\, e^{-j\rho\, t} & \text{for } \; 0 \le t \\ 0 & \text{for } \; 0 > t, \end{cases} \tag{11}$$

which is considered in [5, 6].

The RT is defined as time span in which the energy of a steady-state sound field in a room decays 60 dB below its initial level after switching-off the excitation source. It is linked to the *decay rate* ρ of Eq. (11) according to

$$T_{60} = \frac{3}{\rho\, \log_{10}(e)} \approx \frac{6.908}{\rho}\,. \tag{12}$$

Due to this relation, the terms decay rate and reverberation time are used interchangeably in the following. The RIR model of Eq. (10) is rather coarse, but allows to derive a relation between the variance of the late reverberant speech $\sigma_{z_\mathrm{l}}^2$ and the variance of the reverberant speech σ_z^2 [4]

$$\sigma_{z_\mathrm{l}}^2(i,\lambda) = e^{-2\,\nu(i,\lambda)\left(T_\mathrm{l} - \frac{R}{f_\mathrm{s}}\right)} \sigma_{z_\mathrm{r}}^2(i,\lambda - N_\mathrm{l} + 1) \tag{13a}$$

$$\sigma_{z_\mathrm{r}}^2(i,\lambda) = \big(1 - \kappa(i,\lambda)\big)\cdot e^{-2\,\nu(i,\lambda)\frac{R}{f_\mathrm{s}}} \cdot \sigma_{z_\mathrm{r}}^2(i,\lambda-1) + \kappa(i,\lambda)\cdot e^{-2\,\nu(i,\lambda)\frac{R}{f_\mathrm{s}}} \cdot \sigma_z^2(i,\lambda-1). \tag{13b}$$

The value for $\kappa(i,\lambda)$ depends on the direct-to-reverberation ratio of the (unknown) RIR and should prevent an over-estimation for $\hat{\sigma}_{z_\mathrm{r}}^2$, if the source-receiver distance is lower than the critical distance (where the direct path energy is higher than the energy of all reflections). It might be estimated by a heuristic rule devised in [4]. A less complex and more robust approach is to take the fixed value $\kappa \equiv 1$ which implies the RIR model of Eq. (11) so that Eq. (13) simplifies to

$$\sigma_{z_\mathrm{l}}^2(i,\lambda) = e^{-2\,\nu(i,\lambda)\,T_\mathrm{l}} \cdot \sigma_z^2(i,\lambda - N_\mathrm{l})\,. \tag{14}$$

[1] The spectral weights can also be bounded implicitly by imposing a lower threshold to the *a priori* SIR of Eq. (7).

The integer value $N_\mathrm{l} = \lfloor T_\mathrm{l}\, f_\mathrm{s}/R \rfloor$ marks the number of frames corresponding to the chosen time span T_l, where f_s denotes the sampling frequency. The value for T_l is typically in a range of 20 to 100 ms and controls the time span after which the late reverberation (presumably) begins.

The PSD of the reverberant speech σ_z^2 can be estimated from the spectral coefficients $\widehat{Z}(i,\lambda)$ by recursive averaging

$$\hat{\sigma}_z^2(i,\lambda) = \beta \cdot \hat{\sigma}_z^2(i,\lambda-1) + (1-\beta)\cdot |\widehat{Z}(i,\lambda)|^2 \qquad (15)$$

with $0 < \beta < 1$. The spectral coefficients of the reverberant speech are obtained by spectral weighting

$$\hat{Z}(i,\lambda) = X(i,\lambda)\cdot W_i'(\lambda) \qquad (16)$$

using, for instance, the spectral subtraction rule of Eq. (5) based on an estimation of the *a posteriori* SNR.

2.5 Decay Rate Estimation

The value $\nu(i,\lambda)$ in Eq. (13) and Eq. (14) marks the frequency and time dependent decay rate of the RIR in the subband-domain. Non-subsampled subband signals are needed for its estimation which causes a high computational complexity. To avoid this, we estimate the decay rate in the time-domain at decimated time instants $\lambda' = \lfloor k/R' \rfloor$ out of the denoised, reverberant speech signal $\hat{z}(k)$ as sketched by Fig. 1. The prime indicates that the update rate for this estimation R' is not necessarily identical to that for the adaptation of the spectral weights $W_i(\lambda)$ and $W_i'(\lambda)$. The coefficients for the denoising are obtained from the spectral weights $W_i'(\lambda)$, cf. Eq. (16). The frequency dependent decay rate $\nu(i,\lambda')$, needed to evaluate Eq. (13) or Eq. (14), is obtained by the time-domain estimate for the decay rate $\hat{\rho}(\lambda')$ according to

$$\hat{\nu}(i,\lambda') \approx \hat{\rho}(\lambda')\ \forall\ i \in \{\,0,1,\ldots,M-1\,\}\,. \qquad (17)$$

This simple approximation is motivated by the observation that the estimation error for the decay rate is of similar amount than the approximation error due to Eq. (17). This approach yields good results in practice with low complexity.

The decay rate (or RT) is estimated blindly by the *maximum likelihood* (ML) approach proposed in [13]. A ML estimation for the decay rate ρ is performed at decimated time instants λ' on a frame with N samples according to

$$\hat{\rho}(\lambda') = \arg\left\{ \max_{\rho} \left\{ \mathcal{L}(\lambda') \right\} \right\} \qquad (18)$$

with log-likelihood function given by

$$\mathcal{L}(\lambda') = -\frac{N}{2}\left((N-1)\,\ln(a) + \ln\left(\frac{2\,\pi}{N} \sum_{i=0}^{N-1} a^{-2\,i}\, \hat{z}^2(\lambda' R' - N + 1 + i) \right) + 1 \right) \qquad (19)$$

and damping factor $a = \exp\{-\rho/f_\mathrm{s}\}$. The corresponding RT is obtained by Eq. (12).

A correct RT estimate can be expected, if the current frame captures a free decay period following the sharp offset of a speech sound. Otherwise, an incorrect RT is obtained, e.g., for segments with ongoing speech, speech onsets or gradually declining speech offsets. Such estimates can be expected to overestimate the RT, since the damping of sound cannot occur at a rate faster than the free decay. However, taking the minimum of the last K_l ML estimates is likely to underestimate the RT since the estimation procedure is also a stochastic process. A more robust strategy is to apply an order-statistics filtering approach as known from image processing [14]. The histogram of the K_l most recent ML estimates is built and its first local maximum is taken as RT estimate $\widehat{T}_{60}^{(\mathrm{peak})}(\lambda')$. The effects of outliers can be efficiently reduced by a strong recursive smoothing

$$\widehat{T}_{60}(\lambda') = \beta \cdot \widehat{T}_{60}(\lambda'-1) + (1-\beta)\cdot \widehat{T}_{60}^{(\mathrm{peak})}(\lambda') \qquad (20)$$

with $0.9 < \beta < 1$. The devised RT estimation exploits the fact that speech signals always contain small pauses of some hundred milliseconds, but it does not require an explicit gap detection as, for instance, [5].

3 Simulation Example

The performance of the new algorithm is evaluated by means of instrumental quality measures. The distorted signals are generated according to Eq. (1). A speech signal with $f_\mathrm{s} = 16$ kHz is convolved with a real measured RIR having a RT of $T_{60} = 0.64$ s (measured by the modified Schroeder method as described in [13]). The reverberant speech signal $z(k)$ is distorted by additive car noise from the NOISEX-92 database with varying global SNRs for anechoic speech $s(k)$ and additive noise $n(k)$.

The processing is done by means of the uniform filter-bank equalizer [13] as sketched by Fig. 1. A uniform DFT analysis filter-bank with Hann prototype filter, $M = 256$ frequency channels and downsampling rate of $R = 128$ is used. The time-domain filters are of linear-phase and have a length of $L = M$ taps.[2] Thus, this system has an algorithmic signal delay of $M/2$ samples instead of $M-1$ samples as for the corresponding DFT analysis-synthesis filter-bank (or overlap-add method).

The spectral weights are calculated by the spectral subtraction rule of Eq. (5) using the thresholding of Eq. (9) with $\epsilon_\mathrm{thr} = 0.15$ for the weights $W_i(\lambda)$ and $\epsilon_\mathrm{thr} = 0.1$ for $W_i'(\lambda)$. The noise energy is estimated by minimum statistics [12] and the late reverberant energy by Eq. (14). For the blind estimation of the RT according to Sec. 2.5, a histogram size of $K_\mathrm{l} = 400$ values and adaptation rate $R' = 256$ are used. A smoothing factor of $\beta = 0.995$ is taken for Eq. (20).

The quality of the processed speech is evaluated in the frequency-domain by means of the *cepstral distance* (CD) (cf. [1, 13]) between the (anechoic) speech signal of the direct path $s_\mathrm{d}(k)$ and the enhanced speech $\hat{s}(k)$ (after delay compensation). The *segmental SIR* is used as time-domain measure for the speech quality according to (cf. [3])

$$\frac{\mathrm{SIR}_\mathrm{seg}}{\mathrm{dB}} = \frac{10}{\mathcal{C}(\mathbb{F}_s)} \sum_{l\in\mathbb{F}_s} \log_{10}\left(\frac{\sum_{n=0}^{N_\mathrm{f}-1} s_\mathrm{d}^2(l-n)}{\sum_{n=0}^{N_\mathrm{f}-1} \left(s_\mathrm{d}(l-n) - \hat{s}(l-n) \right)^2} \right). \qquad (21)$$

The set $\mathbb{F}_s$ contains all frame indices corresponding to frames with speech activity and $\mathcal{C}(\mathbb{F}_s)$ represents its total number of elements. Non-overlapping frames with $N_\mathrm{f} = 256$ samples are used and 40 coefficients are considered for the CD measure. The curves for different measures are plotted in Fig. 2.

The joint suppression of noise and later reverberation leads to a significantly better speech quality, in terms of a

[2] This filter-bank configuration is used as the employed instrumental measures are sensitive towards filtering with nonlinear phase, even if this has no audible effect, cf. [9].

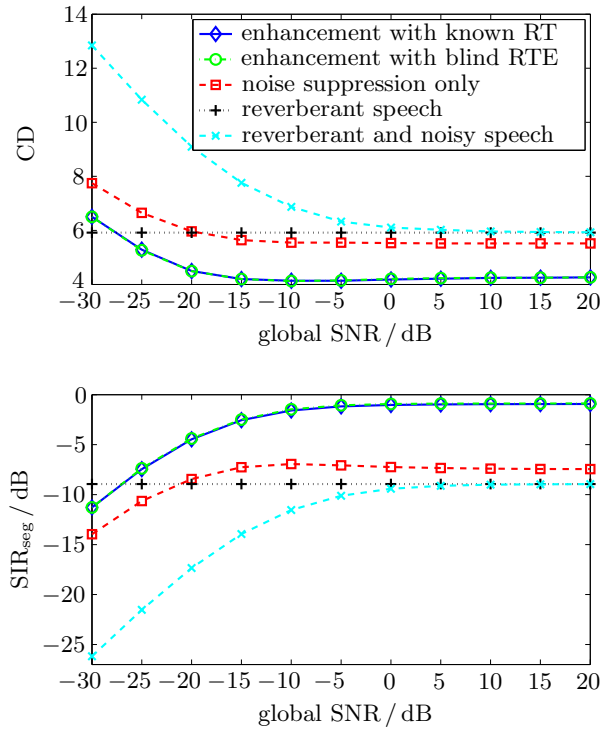

Figure 2: Cepstral distance (CD) and segmental signal-to-interference ratio (SIR) for varying global SNRs and different speech enhancement configurations.

lower CD and higher SIR, in comparison to the noise reduction without dereverberation. For low SNRs, the dereverberation effect becomes less significant due to the high noise energy (see also Eq. (7)). This is a desirable effect as the impact of reverberation is masked by the noise in such cases. For high SNRs, the noise reduction still achieves a slight improvement as the noise power estimation does not yield zero values. The curves for speech enhancement with blind RTE are almost identical to those obtained by using the actual RT. The blind RTE is done with an estimation error of less than $\pm 0.2\,\mathrm{s}$ [13], which is still sufficient for this application and does not cause noteworthy impairments.

The results of the instrumental measurements comply with our informal listening tests. The new speech enhancement system achieves a significant reduction of background noise and reverberation, but still preserves a natural sound impression. The speech signals enhanced with blind RTE and known RT have revealed no audible differences. The noise reduction alone achieves only a slightly audible reduction of reverberation.

4 Conclusions

A new speech enhancement algorithm for the joint suppression of background noise and late reverberation is proposed. The enhancement is performed by a spectral subtraction rule based on spectral power estimations for the background noise and late reverberant speech. The variance of the late reverberant speech is calculated by a simple rule in dependence of the RT, which is determined blindly by means of an adaptive maximum likelihood estimation and order-statistics filtering.

In reverberant and noisy environments, the new system achieves a significantly better speech quality than common noise reduction systems. The enhancement with blind RT estimation achieves the same speech quality as a system using the actual RT.

The proposed algorithm has a low signal delay and manageable computational complexity, which is important for speech enhancement systems in hands-free devices, mobile phones or hearing aids. Another interesting application are speech recognition systems where the suppression of noise and reverberation is beneficial to increase the recognition rates, cf. [5].

References

[1] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment*, John Wiley & Sons, Chichester, 2006.

[2] A. K. Nábělek and D. Mason, "Effect of Noise and Reverberation on Binaural and Monaural Word Identification by Subjects with Various Audiograms," *Journal fo Speech and Hearing Research*, vol. 24, pp. 375–383, 1981.

[3] P. A. Naylor and N. D. Gaubitch, "Speech Dereverberation," in *Proc. of Intl. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Eindhoven, The Netherlands, Sept. 2005.

[4] E. A. P. Habets, *Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement*, Ph.D. thesis, Eindhoven University, Eindhoven, The Netherlands, June 2007.

[5] K. Lebart, J. M. Boucher, and P. N. Denbigh, "A New Method Based on Spectral Subtraction for Speech Dereverberation," *acta acoustica - ACOUSTICA*, vol. 87, no. 3, pp. 359–366, 2001.

[6] E. A. Habets, "Multi-Channel Speech Dereverberation Based On A Statistical Model of Late Reverberation," in *Proc. of Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia (Pennsylvania), USA, Mar. 2005, vol. 4, pp. 173–176.

[7] R. E. Crochiere, "A Weighted Overlap-Add Method of Short-Time Fourier Analysis/Synthesis," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-28, no. 10, pp. 99–102, Feb. 1980.

[8] H. W. Löllmann and P. Vary, "Uniform and Warped Low Delay Filter-Banks for Speech Enhancement," *Speech Communication, Elsevier, Special Issue on Speech Enhancement*, vol. 49, pp. 574–587, July 2007.

[9] H. W. Löllmann and P. Vary, "Low Delay Filter-Banks for Speech and Audio Processing," in *Speech and Audio Processing in Adverse Environments*, E. Hänsler and G. Schmidt, Eds., chapter 2, pp. 13–61. Springer, Berlin, New York, 2008.

[10] B. Sauert, H. W. Löllmann, and P. Vary, "Near End Listening Enhancement by Means of Warped Low Delay Filter-Banks," in *Proc. of ITG Conference on Speech Communication*, Aachen, Germany, Oct. 2008.

[11] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.

[12] R. Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," *IEEE Trans. on Speech and Audio Processing*, vol. 9, no. 5, pp. 504–512, July 2001.

[13] H. W. Löllmann and P. Vary, "Estimation of the Reverberation Time in Noisy Environments," in *Proc. of Intl. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle (Washington), USA, Sept. 2008.

[14] I. Pitas and A. N. Venetsanopoulos, "Order Statistics in Digital Image Processing," *Proc. of the IEEE*, vol. 80, no. 12, pp. 1893–1921, Dec. 1992.