EFFICIENT BLIND ESTIMATION OF SUBBAND REVERBERATION TIME FROM SPEECH IN NON-DIFFUSE ENVIRONMENTS

Salomon Diether[1], Lukas Bruderer[2], Andreas Streich[3], and Hans-Andrea Loeliger[4]

[1,2,4]Department of Information Technology and Electrical Engineering, ETH Zurich
[3]Phonak AG, Staefa, Switzerland
{sdiether, bruderer, loeliger}@isi.ee.ethz.ch, andreas.streich@ieee.org

ABSTRACT

To respond to reverberation effectively, modern speech processing tasks like signal enhancement in digital hearing aids or distant-talking speech recognition often require precise knowledge of the acoustic situation. A common measure is the (frequency-dependent) reverberation time T_{60}. Explicit measurement using sound excitations is not possible for most applications, therefore blind estimation of subband T_{60} from speech signals is employed. Previous approaches are either limited to T_{60} between 0...1.5 s, or require long speech signals and costly computation. We propose an efficient algorithm that estimates T_{60} up to 6 s from short speech signals of no more than 20 s. In experiments with real room impulse responses (RIR) the algorithm exhibits state-of-the-art performance for T_{60} up to 1.5 s and superior performance for longer T_{60}.

Index Terms— reverberation time, blind estimation, non-diffuse, low complexity.

1. INTRODUCTION

Late reverberation degrades speech intelligibility by interfering the clean signal with diffuse sound reflections. This poses a great challenge to modern speech processing tasks such as signal enhancement in digital hearing aids or (distant-talking) automatic speech recognition (ASR). On the other hand, early sound reflections do not affect speech intelligibility, both for human hearing [1, 2] and ASR tasks [3]. A common measure for *late reverberation* is the reverberation time T_{60}, which describes the time span until a sound impulse, peaking at 0 decibel (dB), decays from -5 dB to -65 dB. T_{60} is used e.g. to steer dereverberation algorithms [4, 5, 6, 7] or to select/adapt acoustic models in the context of distant-talking ASR [8, 9].

Despite considerable efforts in the past, blind estimation of T_{60} from speech signals remains challenging, especially in highly reverberant, non-diffuse environments with T_{60} greater than 1.5 s. Most related work assumes diffuse single-slope reverberation models [10, 11, 12, 13, 14, 15, 16] to ensure robustness and computational efficiency. Such an approach is effective for estimation of $T_{60} \sim 0...1$ s or even 1.5 s, but is not applicable in highly non-diffuse environments. Estimation of long T_{60} is performed accurately in [17] by employing a two-slope sound decay model rather than a single-slope model and by restricting estimation to long speech pauses with many dB of sound decay. The approach delivers accurate estimation results for $T_{60} \sim 0...5$ s. However, it requires speech signals in the order of minutes. Another issue is the high computational complexity of the method, rendering it impractical for (real-time) applications on mobile devices. In the context of ASR, acoustic models incorporating two-slope reverberation were used before: [8, 18, 19] adapt the acoustic model according to T_{60}. However, [8] uses actual T_{60} *measurements* due to the challenge of estimating T_{60} accurately. [18, 19] estimate *late reverberation*, however indirectly by inference from single-slope T_{60} estimates and Early Decay Time (EDT) estimates. As EDT describes the time needed by a sound to decay from 0 to -10 dB, its robust estimation is even more challenging than T_{60}, due to non-diffuse early sound reflections (see [17]).

1.1. Contributions

In this paper, we present a low-complexity and recursive method for blind estimation of subband T_{60} from single-channel speech signals. We base our algorithm on an efficient technique for *late reverberation* estimation from a room impulse response (RIR) or a speech pause (Section 2). Its core is a model fitting approach between a *late reverberation* model and a sound decay. Second, we describe a free decay region (FDR) detection method, which improves the one in [17, 20], to identify sound decays suitable for subband T_{60} estimation from speech (Section 3). *Fit likelihoods* are used to examine the data structure and to determine, together with other criteria, whether a sound decay is sufficiently similar to one of the predefined *FDR profiles* to be accepted for estimation. This allows fast, flexible and robust identification of useful data. Third, we show that the challenging T_{60} estimation in low frequency bands [12, 17] benefits from *Frequency band boosting* (Section 4). This technique essentially smooths signal fluctuations within a subband, but preserves its characteristic decay rate required for estimation. Experiments show significant gains in estimation accuracy for low frequency bands.

1.2. Proposed algorithm

The proposed algorithm aims at estimating subband T_{60} blindly from sound measurements. The first stage decomposes a speech signal with discrete length N into F octave band signals $\mathbf{x}_1[1...N], ..., \mathbf{x}_F[1...N]$, for subband T_{60} estimation [17]. *Frequency band boosting* is applied to improve performance at low frequencies. The second stage converts each subband signal $\mathbf{x}_f[\cdot]$ into an envelope signal $\mathbf{s}_f[\cdot]$ and fits a *late reverberation* model to all sound decays that are potential FDRs. If a fitted model resembles a FDR closely enough, its T_{60} estimate is extracted. The final subband T_{60} estimate is given by the median of extracted T_{60} estimates.

2. LATE REVERBERATION ESTIMATION

We first address the need for an efficient estimation technique that handles $T_{60} \sim 0.3...6\,\mathrm{s}$. Single-slope reverberation models (e.g. [11, 14]) are only applicable for T_{60} less than 1-1.5 s. Two-slope models capture highly reverberant scenarios accurately, but are expensive to compute [17]. Our proposed technique estimates *late reverberation* efficiently from impulse responses or speech pauses and captures even long T_{60} accurately.

Prior to T_{60} estimation, we apply recursive smoothing (RS) to the signal energy of each subband, in contrast to e.g. [17, 21, 22] who use Hilbert envelopes or rectangular smoothing for signal pre-processing. The RS envelope $\tilde{\mathbf{s}}[\cdot]$ of a subband speech signal $\tilde{\mathbf{x}}[\cdot]$[1] at discrete time k is defined as

$$\tilde{\mathbf{s}}[k] = \alpha\ \tilde{\mathbf{x}}^2[k-1] + (1-\alpha)\ \tilde{\mathbf{x}}^2[k] \tag{1}$$

with smoothness parameter $\alpha \in\ ]0...1[$. We found pre-processing using RS especially beneficial for estimation of long T_{60}. RS smooths data asymmetrically, only according to past data, and this resembles a constant, minimum amount of reverberation, that is introduced into the signal. The resulting smooth envelope allows robust T_{60} estimation, as all decay rates above the minimum level of reverberation, determined by α, are preserved. The RS envelope of a subband sound decay is modeled accurately using a two-slope exponential decay multiplied with a *log-normally* distributed Gaussian noise term (see Figure 1). We define the two-slope model of a sound energy decay with $k = 1...K$ as

$$\tilde{\mathbf{E}}_d[k] \triangleq \left(\tilde{c}_1^2 e^{-2\tilde{\rho}_1 T_s k} + \tilde{c}_2^2 e^{-2\tilde{\rho}_2 T_s k}\right)\tilde{\mathbf{w}}[k] \tag{2}$$

where $\tilde{c}_1, \tilde{c}_2 > 0$ are amplitude constants, $\tilde{\rho}_1$ is the EDT decay rate stemming from early sound reflections, and $\tilde{\rho}_2$ is the T_{60} decay rate from *late reverberation* ($\tilde{\rho}_1 > \tilde{\rho}_2 > 0$). $T_s = 1/f_s$ marks the sampling period and f_s the sampling rate of the *subband* signal. The *multiplicative, log-normally* distributed noise term $\tilde{\mathbf{w}}[k] \sim \ln \mathcal{N}\left(0, \tilde{\sigma}_{\mathbf{w}}^2\right)$ models the signal deflections of the RS envelope from the ideal model (see Figure 1b). Its variance $\tilde{\sigma}_{\mathbf{w}}^2$ is chosen according to the smoothness parameter α of the RS envelope. We transform the two-slope model into the logarithmic dB domain and obtain

$$\mathbf{E}_d[k] = 10\log_{10}\left(\tilde{c}_1^2 e^{-2\tilde{\rho}_1 T_s k} + \tilde{c}_2^2 e^{-2\tilde{\rho}_2 T_s k}\right) + \mathbf{w}_k \tag{3}$$

with $\mathbf{w}_k = 10\log_{10}(\tilde{\mathbf{w}}_k) \sim \mathcal{N}\left(0, \sigma_{\mathbf{w}}^2\right)$, and $\sigma_{\mathbf{w}} = 10\log_{10}(\tilde{\sigma}_{\mathbf{w}})$. For large $k \gg 1$, the *late reverberation* tail dominates the sound decay and thus $\mathbf{E}_d[k]$ is approximated well by the linear model

$$\begin{aligned} \mathbf{E}_d[k] &\overset{k\gg 1}{\approx} 10\log_{10}\left(\tilde{c}_2^2 e^{-2\tilde{\rho}_2 T_s k}\right) + \mathbf{w}_k \\ &= c_2 - \rho_2 T_s k + \mathbf{w}_k \end{aligned} \tag{4}$$

where $c_2 = 10\log_{10}(\tilde{c}_2^2)$ and $\rho_2 = \frac{20}{\ln(10)}\tilde{\rho}_2$.

2.1. Estimation using Weighted Least Squares

To estimate ρ_2 from a sound decay in the logarithmic domain $\mathbf{s}[k] = 10\log_{10}(\tilde{\mathbf{s}}[k])$ with $k = 1...K$, we fit the model $\mathbf{E}_d[k]$ from Eq. (4) to the *late* samples of $\mathbf{s}[k]$, i.e. for large k. To this end, we employ *weighted least-squares* (WLS) minimization,

[1] From now on we refer to subband signals without subscript.

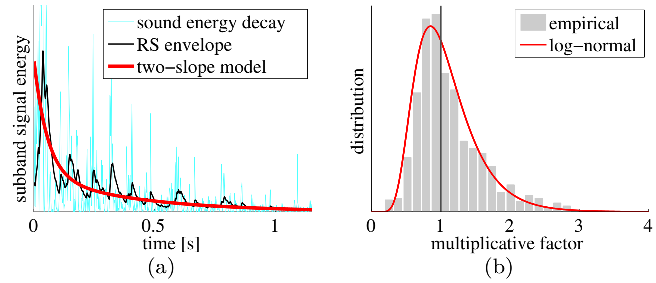


Fig. 1: (a) Two-slope decay model with multiplicative noise $\tilde{w}_k \sim \ln \mathcal{N}(0, \tilde{\sigma}_{\mathbf{w}})$ fitted to the recursive smoothing (RS) envelope of a sound energy decay. (b) The log-normal distribution describes the multiplicative error between the observed RS envelope signal and the fitted two-slope model for data in (a).

with a forgetting factor $\gamma < 1$ that defines the exponential weighting. This avoids the early reflections and focuses on the *late reverberation* tail:

$$(\hat{c}_2, \hat{\rho}_2) = \underset{c_2,\rho_2}{\operatorname{argmin}} \sum_{k=1}^{K} \gamma^{K-k} (\mathbf{s}[k] - \mathbf{E}_d[k])^2 \tag{5}$$

Eq. (5) is solved efficiently in a recursive fashion using Gaussian message passing in factor graphs [23]. The subband T_{60} estimate is given by

$$\hat{T}_{60} \approx \frac{60}{\hat{\rho}_2\ f_s} \tag{6}$$

3. FREE DECAY REGION DETECTION

Accurate estimation of (long) T_{60} from FDRs, i.e. sound decays suitable for estimation, depends on their availability and quality. In order to cope even with short speech signals, one must choose wisely which of the few available decay regions to use. This section describes an effective method to identify FDRs within subbands of speech signals. In contrast to related approaches [11, 14, 17], the proposed FDR detection enables estimation from highly reverberant (T_{60} up to 6 s), short ($\leq 20\,s$) speech signals.

Key feature of the proposed method is the effective assessment of data regarding its suitability for T_{60} estimation. We introduce *FDR profiles* to describe minimal requirements towards a potentially useful sound decay, for different reverberant environments. Only if a sound decay fulfills all criteria of one *FDR profile*, it is accepted for T_{60} estimation. *FDR profiles* consist of 3 criteria:

1. Minimum amplitude difference
2. Minimum decay length
3. Minimum model fit quality

Figure 2 illustrates all 3 criteria for an exemplary sound decay. The third property defines, how closely a sound decay must resemble the fitted model in order to be regarded suitable. This enables us to detect and sort out data containing large, unexpected disturbances (e.g. due to speaker interference or background noise), that hinder accurate T_{60} estimation.

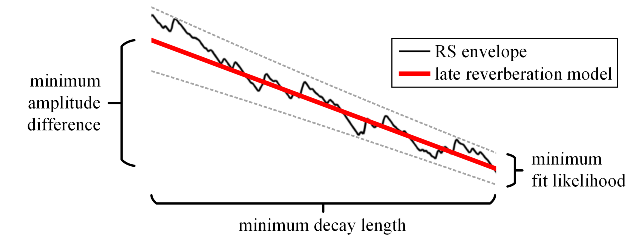


Fig. 2: To assess the suitability of a sound decay for subband T_{60} estimation, we employ *FDR profiles*, each consisting of 3 criteria: 1. Minimum amplitude difference, 2. Minimum decay length, 3. Minimum model fit quality (*fit likelihood*).

3.1. Fit likelihood measure

We measure the quality of a model fit using local likelihood functions [24]. Given the RS envelope of a sound decay $\mathbf{s}[k]$ with $k = 1...K$, and $\mathbf{E}_d[k]$, the WLS-fitted *late reverberation* model from Eq. (4), we calculate the *fit likelihood*

$$p(\mathbf{s}|\mathbf{E}_d) = \left(\prod_{k=1}^{K} p_w(\mathbf{s}[k]|\mathbf{E}_d[k])^{\gamma^{K-k}} \right)^{\frac{1}{\sum_{k=1}^{K} \gamma^{K-k}}} \quad (7)$$

where $p_w(\cdot)$ is a Gaussian probability density function. Intuitively, Eq. (7) describes a weighted mean likelihood to observe a sound decay $\mathbf{s}[\cdot]$, given the fitted model $\mathbf{E}_d[\cdot]$. According to the WLS model fitting in Eq. (5), which uses a forgetting factor γ to emphasize the late samples of a sound decay, Eq. (7) also weights the fit quality of late samples higher. Note that *fit likelihoods* are independent of the sound decay length and thus only assess data structure. High *fit likelihoods* are produced if a sound decay is very similar to the fitted model, while disturbed sound decays produce low *fit likelihoods* due to unexpected irregularities in the signal.

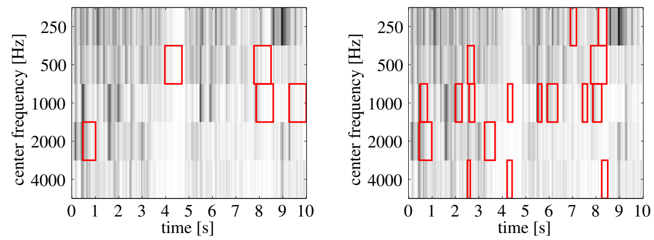
3.2. Defining FDR profiles

FDR profiles enable us to identify free decay regions within subband speech signals in an open loop manner, i.e. without needing to adapt to various reverberant settings adaptively.

We define a *FDR profile* for little reverberation by: A large amplitude difference (property 1), short decays (property 2), high *fit likelihoods* (property 3). In contrast, a *FDR profile* for highly reverberant environments requires long sound decays (property 2), while the amplitude difference is not too relevant (property 1). Signal fluctuations are common in settings with long T_{60} (e.g. due to room modes), therefore a moderate *fit likelihood* threshold (property 3) is used to sort out only data that exhibits major disturbances. *FDR profiles* in between little and high reverberation are defined accordingly.

4. FREQUENCY BAND BOOSTING

Subband T_{60} estimation from speech is especially challenging in low frequency bands [12, 17], where, commonly, the longest T_{60} are observed. To this end, we propose a new method prior to subband T_{60} estimation: The idea is to smooth narrow-band signal fluctuations, while preserving each Frequency band's characteristic decay rate. This is achieved by using L



(a) Method in [17, 20] (b) Proposed method

Fig. 3: Detected free decay regions (FDR), i.e. sound decays used for T_{60} estimation, in 5 octave bands for 10 s of speech. The robustness of T_{60} estimates is mainly determined by the number of correctly identified FDRs in each subband.

bandpass filters within a Frequency band to compute narrow-band energy signals $\tilde{\mathbf{x}}_1^2[\cdot], ..., \tilde{\mathbf{x}}_L^2[\cdot]$ and then summing up these narrow-band energy signals to obtain the power spectral density (PSD) of the Frequency band $\tilde{\mathbf{x}}^2[\cdot] = \sum_{l=1}^{L} \tilde{\mathbf{x}}_l^2[\cdot]$. Akin to boosting methods in statistics, by averaging a large number of frequency-domain energy estimators, a more refined estimate for the complete subband should be obtained. Indeed in our experiments, application of this *Frequency band boosting* results in significantly improved accuracy in T_{60} estimation within low octave bands (250, 500, 1000 Hz center frequencies).

Frequency band boosting is exploited efficiently by using a *high-order* Fast Fourier Transform (FFT) to calculate the short-time PSD of a speech signal, e.g. a 1024-point or 2048-point FFT. As high-order FFTs decompose a signal into a great number of narrow-band bins, *Frequency Band Boosting* is performed when summing up narrow-band energies to calculate the short-time PSD.

5. EXPERIMENTAL SETUP

We validate the techniques introduced in Sections 2-4 using the proposed algorithm that estimates subband T60 in an online fashion at low computational cost.

In a first step, the recorded sound signal is decomposed into its frequency components using a 2048-point Short-Time Fourier Transform (STFT) with 75% overlap at 22050 Hz signal sampling rate. The short-time power spectral densities (PSD) of 5 octave bands (with center frequencies 250, 500, 1000, 2000, 4000 Hz) are calculated by summing up the squared signal components of the corresponding STFT bins, which exploits the *Frequency band boosting* effect (Section 4). Subsequently, the RS envelope function is applied to each subband signal (Section 2). RS only preserves decay rates above a minimum T_{60}, depending on the smoothness parameter α. In our experiments, α is chosen so that it allows estimation of $T_{60} \sim 0.3...6.0\,s$. Blind T_{60} estimates are obtained by recursively minimizing the WLS between *late reverberation* model and RS envelope signal (Section 2). We set the forgetting factor to $\gamma = 0.98$, which gives best performance in our experiments. For FDR detection we use 3 *FDR profiles* with manually chosen properties, characterizing scenarios with little, medium and high reverberation. In case a fitted model meets all requirements of a *FDR profile* (Section 3), the subband T_{60} estimate is included into a pool

T_{60}	5 s speech	10 s speech	20 s speech	30 s speech
1.5-6.0	0.70 (±0.30)*	0.48 (±0.21)	0.36 (±0.19)	0.36 (±0.19)
0.75-1.5	0.31 (±0.11)	0.27 (±0.12)	0.25 (±0.11)	0.23 (±0.10)
0.0-0.75	0.12 (±0.03)	0.12 (±0.02)	0.11 (±0.03)	0.12 (±0.03)

Table 1: Mean absolute error of subband T_{60} estimation [s] using the proposed algorithm, for different speech lengths. (*No estimation available for 1 speaker in the data set, as no FDR was detected in 1 subband.)

Oct.band	128-point	1024-point	2048-point
250 Hz†	0.42	0.33	**0.21**
500 Hz†	0.34	0.28	**0.20**
1000 Hz†	0.27	0.26	**0.20**

Table 2: *Frequency band boosting* reduces the mean absolute error [s]. Results after 20 s of speech with 128-, 1024-, 2048-point FFT. (†center frequency)

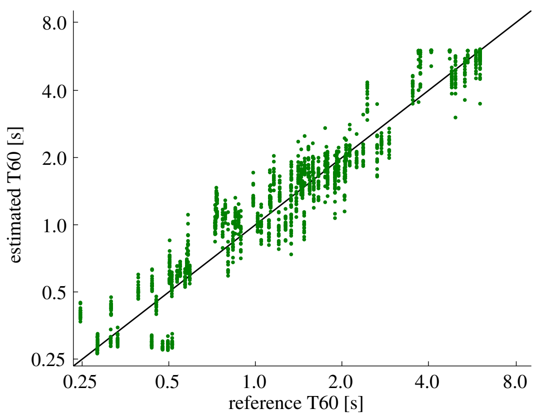


Fig. 4: Subband T_{60} estimation using proposed algorithm in 5 octave bands (center frequencies 250-4000 Hz). 22 real RIRs convolved with 16 speech signals of 20 s length.

T_{60}	Löllmann	Prego	Proposed
1.5-6.0 s	2.40 (±1.77)	2.18 (±1.85)	0.36 (±0.19)
0.75-1.5 s	0.39 (±0.19)	0.22 (±0.06)	0.25 (±0.11)
0.0-0.75 s	0.10 (±0.04)	0.16 (±0.07)	0.11 (±0.03)

Table 3: Mean absolute error of subband T_{60} estimation [s] in 5 octave bands (center frequencies 250-4000 Hz), compared with Löllmann [14], Prego [11]. Results after 20 s of speech.

of 25 recent subband T_{60} estimates. We use the median of the pool as the current T_{60} estimate of the octave band (see [11]). Post-processing of the raw T_{60} estimates is performed using penalized least squares regression across subbands [25], for moderate cross-frequency smoothing and outlier detection. This last step improves the robustness of estimation for short speech signals or with band-limited background noise.

5.1. Data

Experiments are conducted using real room impulse responses and natural human speech. To this end a comprehensive set of 22 RIRs is obtained from the Open Acoustic Impulse Response (Open AIR) Library [26]. The featured RIRs comprise a large variety of broadband T_{60} ranging from 0.1 to 15.3 s with source-microphone distances above the critical distance. Many RIRs exhibit non-diffuse structure with an EDT that is not directly correlated to T_{60}. This is common in everyday life, but makes it challenging to estimate T_{60} accurately. Each RIR is convolved with 16 single-channel signals of natural human speech (male and female speakers), which makes a total of 352 reverberant speech files in the data set. Reference values are measured using Schroeder backward integration [27]. All subband $T_{60} > 6$ s (reference values and estimates) are set to the maximum considered value $T_{60,\max} = 6$ s.

6. EXPERIMENTAL RESULTS

Figure 4 shows the algorithm's performance in subband T_{60} estimation. Each point represents a T_{60} estimate in one of 5 octave bands, from reverberant speech after 20 s. Compared to methods by Löllmann et al. [14] and Prego et al. [11], the proposed algorithm produces state-of-the-art performance for T_{60} below 0.75 s and T_{60} between 0.75 and 1.5 s (see Table 3). For T_{60} longer than 1.5 s the proposed algorithm clearly outperforms the other two algorithms, which were designed to estimate T_{60} only from *diffuse* reverberation tails. Table 1 shows the estimation accuracy of the proposed algorithm for speech signals of 5 to 30 s length. As expected, both mean and standard deviation of the estimation error decrease for longer speech signals, due to a greater number of FDRs available.

Frequency band boosting is the key factor responsible for improved estimation accuracy in low frequency bands, which is shown in Table 2. When using a 128-point STFT to calculate the short-time PSD, only one or few bins are available in low octave bands, which compares to a common band-pass filter. The more bins make up an octave band, the better signal fluctuations are smoothed, while the decay rate is preserved. Thus high-order STFTs with 1024-point or 2048-point improve estimation accuracy in low octave bands significantly. STFT orders higher than 2048-point only give minor improvements in our experiments. Estimation accuracy across all octave bands can be balanced by adapting the smoothness parameter α of the RS envelope.

7. CONCLUSIONS

In this work, an online algorithm for blind estimation of T_{60} was proposed. Experiments with a variety of acoustic environments and different speakers show superior overall estimation accuracy compared to two state-of-the-art algorithms. This is enabled by three main contributions: First, a technique for efficient, yet accurate *late reverberation* estimation. Second, an improved free decay region detection method that assesses data according to its suitability for T_{60} estimation, and third, *Frequency band boosting* for improved performance in low frequency bands. Due to its low computational cost and quick availability of new estimates, the proposed algorithm has wide applicability for various real-time and mobile applications and may be used e.g. to support dereverberation algorithms in signal enhancement or distant-talking speech recognition.

8. REFERENCES

[1] A. Warzybok, J. Rennies, S. Doclo, and B. Kollmeier, "Influence of early reflections on speech intelligibility under different noise conditions," in *Forum Acusticum 2011. Proceedings.*, 2011, pp. 1149–1154.

[2] J. S. Bradley, H. Sato, and M. Picard, "On the importance of early reflections for speech in rooms.," *J. Acoust. Soc. Am. 113(6)*, p. 3233–3244, 2003.

[3] I. Tashev and D. Allred, "Reverberation reduction for improved speech recognition.," in *Proc. of HSCMA*, 2005.

[4] M. Jeub, H.W. Löllmann, and P. Vary, "Blind dereverberation for hearing aids with binaural link," in *Sprachkommunikation 2010 (ITG-FB 225)*, 2010.

[5] P.A. Naylor and N.D. Gaubitch, "Speech dereverberation," Springer, 2010.

[6] E.A.P. Habets, *Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement*, Ph.D. thesis, Technische Universiteit Eindhoven, 2007.

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

[8] A.W. Mohammed, M. Matassoni, H. Maganti, and M. Omologo, "Acoustic model adaptation using piecewise energy decay curve for reverberant environments," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, 2012, pp. 365 – 369.

[9] H.-G. Hirsch and H. Finster, "A new approach for the adaptation of HMMs to reverberation and background noise," in *Speech Communication, vol. 50, no. 3*, 2008, p. 244–263.

[10] R. Talmon and E.A.P. Habets, "Blind reverberation time estimation by intrinsic modeling of reverberant speech," in *Proc. of the IEEE Int. Conf. on Acoustic., Speech, and Signal Processing (ICASSP)*, 2013.

[11] T. de M. Prego, A.A. de Lima, S.L. Netto, B. Lee, A. Said, R.W. Schafer, and T. Kalker, "A blind algorithm for reverberation-time estimation using subband decomposition of speech signals," *J. Audio Eng. Soc.*, 2012.

[12] A. Keshavarz, S. Mosayyebpour, M. Biguesh, T.A. Gulliver, and M. Esmaeili, "Speech-model based accurate blind reverberation time estimation using an LPC filter," *IEEE Transactions on Audio, Speech, and Language Processing (Volume:20, Issue: 6)*, pp. 1884–1893, 2012.

[13] N. Lopez, Y. Grenier, and I. Bourmeyster, "Low variance blind estimation of the reverberation time," in *Acoustic Signal Enhancement; Proceedings of IWAENC 2012; International Workshop on*, 2012.

[14] H.W. Löllmann, E. Yilmaz, M. Jeub, and P. Vary, "An improved algorithm for blind reverberation time estimation," in *Proc. Int. Workshop Acoust. Echo Noise Control*, Tel Aviv, Israel, 2010, pp. 1–4.

[15] J.Y.C. Wen, E.A.P. Habets, and P.A. Naylor, "Blind estimation of reverberation time based on the distribution of signal decay rates," in *ICASSP*, 2008, pp. 329–332.

[16] R. Ratnam, D.L. Jones, and W.D. Jr. O'Brien, "Fast algorithms for blind estimation of reverberation time," *IEEE Signal Processing Letters vol. 11, no. 6*, p. 537–540, 2004.

[17] P. Kendrick, F.F. Li, and T.J. Cox, "Blind estimation of reverberation parameters for non-diffuse rooms," *Journal of the Acoustical Society of America*, p. 760–770, 2007.

[18] J.Y.C. Wen, A. Sehr, P.A. Naylor, and W. Kellermann, "Blind estimation of a feature-domain reverberation model in non-diffuse environments with variance adjustment," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2009, pp. 175–178.

[19] A. Sehr, J.Y.C. Wen, W. Kellermann, and P.A. Naylor, "A combined approach for estimating a feature-domain reverberation model in non-diffuse environments," in *Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, 2008.

[20] Y. Zhang, J.A. Chambers, F.F. Li, P. Kendrick, and T.J. Cox, "Blind estimation of reverberation time in occupied rooms," in *14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy*, 2006.

[21] M. Karjalainen, P. Antsalo, A. Mäkivirta, T. Peltonen, and V. Välimäki, "Estimation of modal decay parameters from noisy response measurements," *J. Audio Eng. Soc.*, vol. 50, 2002.

[22] A. Lundeby, T.E. Vigran, H. Bietz, and M. Vorländer, "Uncertainties of measurements in room acoustics," *Appl. Acoust., vol. 81*, pp. 344–355, 1995.

[23] H.-A. Loeliger, J. Dauwels, Junli Hu, S. Korl, Li Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," vol. 95, no. 6, pp. 1295–1322, 2007.

[24] L. Bruderer, H.-A. Loeliger, and N. Zalmai, "Local statistical models from deterministic state space models, likelihood filtering, and local typicality," in *ISIT 2014, IEEE International Symposium on Information Theory (Honolulu, HI, USA)*, 2014.

[25] D. Garcia, "Robust smoothing of gridded data in one and higher dimensions with missing values," *Computational Statistics and Data Analysis 54*, pp. 1167–1178, 2010.

[26] S. Shelley and D.T. Murphy, "OpenAIR: An interactive auralization web resource and database," in *Proc. of the 129th AES Convention.*, 2010.

[27] M. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am.*, pp. 407–412, 1965.