Efficient ML-Estimator for Blind Reverberation Time Estimation

Heinrich W. Löllmann, Andreas Brendel, and Walter Kellermann

Chair of Multimedia Communications and Signal Processing
Friedrich-Alexander University Erlangen-Nürnberg
91058 Erlangen, Germany

{heinrich.loellmann,andreas.brendel,walter.kellermann}@fau.de

***Abstract*—A new maximum likelihood (ML) estimator for the blind estimation of the reverberation time (RT) is derived. In contrast to previously proposed ML-based reverberation time estimators, the RT estimate is obtained by a simple closed-form expression, which leads to significant computational savings. Moreover, it is shown that the new estimator is unbiased and reaches the Crámer-Rao lower bound. The proposed RT estimator achieves a similar estimation accuracy but involves a significantly lower computational complexity compared to an ML-based RT estimator that scored among the best at the ACE Challenge.**

I. Introduction

An important quantity for the characterization of acoustic environments is given by the reverberation time (RT) T_{60} [1]. Knowledge about the RT can be exploited, e.g., for enhanced automatic speech recognition (ASR), e.g., [2]–[5], or speech dereverberation, e.g., [6]–[8]. For such applications, the (often time-varying) RT can usually not be determined from a measured room impulse response (RIR) by means of the Schroeder method [9] or its variants, but has to be estimated *blindly* from a reverberant speech signal, which is frequently also distorted by noise. The above-mentioned applications and others have fueled the research interest in blind reverberation time estimation (RTE) and numerous methods were proposed in recent years, e.g., [10]–[23]. The variety of concepts for blind RTE has motivated the Acoustic Characterisation of Environments (ACE) Challenge, which aimed for an objective benchmarking of RT and direct-to-reverberant energy ratio (DRR) estimators [24], [25]. It turned out that the approaches of [26], [27], which are based on a sound decay detection, achieved the highest estimation accuracy for single-channel RT estimation [25]. Many algorithms for RT estimation rely on the detection of sound energy decays of the reverberant speech signal from which the decay rate (and thus the RT) is estimated either by means of a maximum-likelihood (ML) estimator [10], [11], [15], [16], [27] or linear regression [6], [12], [13], [18], [21]. Some require prior training for parameter calibration [13], [18], [21] or perform the sound decay detection in the subband-domain, e.g., [17], [26], [27]. A comprehensive statistical analysis of such blind RT estimators w.r.t. estimation bias and variance is provided by [28].

A drawback of current ML estimators for RTE [10], [11], [15], [16], [27] is that the maximum of the likelihood function can not be determined by a closed-form expression, but has to be found by iterative schemes or a brute force search, which typically causes a high computational load. In this contribution, a novel ML estimator is derived which is given by a simple closed-form expression and is shown to reach the Crámer-Rao lower bound (CRLB).

The paper is organized as follows: In Sec. II, the new ML estimator is derived and its statistical properties are analyzed in Sec. III. The overall system for blind RTE is described in Sec. IV and simulation results are provided by Sec. V. The paper concludes with a summary in Sec. VI.

II. New ML Estimator

A reverberant speech signal $z(k)$ is modeled as a convolution of a speech signal $s(k)$ with a time-varying RIR $h(\eta, k)$ of length L_h

$$z(k) = \sum_{\eta=0}^{L_h-1} s(k-\eta) \cdot h(\eta, k) \tag{1}$$

with discrete time index k and sample index η. If a speech pause begins at an instant $k - L_\mathrm{o}$

$$s(k-\eta) \begin{cases} \approx 0 & \text{for } \eta = 0, 1, \ldots, L_\mathrm{o} - 1 \\ \neq 0 & \text{for } \eta = L_\mathrm{o}, \ldots, L_h - 1, \end{cases} \tag{2}$$

the delayed reverberant part of the signal $d(k)$ can be observed, since

$$z(k) = \underbrace{\sum_{\eta=0}^{L_\mathrm{o}-1} s(k-\eta) \cdot h(\eta, k)}_{\approx 0} + \underbrace{\sum_{\eta=L_\mathrm{o}}^{L_h-1} s(k-\eta) \cdot h(\eta, k)}_{= d(k)} \tag{3}$$

assuming that $h(\eta, k) \neq 0$ for at least one value $L_\mathrm{o} \leq \eta < L_h$. The envelope of the reverberant component $d(k)$ is typically exponentially decaying and, thus, modeled as follows

$$d_\mathrm{m}^{(1)}(k) = A\, v(k)\, e^{-\rho\, k\, T_\mathrm{s}}\, \epsilon(k) \tag{4}$$

This work was partially supported by the DFG under contract no <Ke890/10-1> within the Research Unit FOR2457 "Acoustic Sensor Networks".

ISBN 978-90-827970-1-5 © EURASIP 2018

with amplitude $A > 0$, decay rate ρ, unit step sequence $\epsilon(k)$, and sampling period $T_\mathrm{s} = 1/f_\mathrm{s}$. $v(k)$ represents a sequence of i.i.d. standard-normally distributed random variables with $\mathcal{N}(0,1)$. Eq. (4) is a coarse approximation of the reverberant sound decay and, at the same time, represents a simple statistical model for the RIR which considers only the effects of late reflections and models them as diffuse noise, e.g., [6]. The relation to the RT is given by, e.g., [10]

$$T_{60} = \frac{3}{\rho \log_{10}(e)} \approx \frac{6.908}{\rho} \,. \tag{5}$$

The model of Eq. (4) serves as basis for various ML-based RT estimators [10], [11], [15], [16], [27], see also [28].

In the following, an ML estimator for RT estimation is derived from a novel modified probabilistic model of the sound decay which directly leads to a closed-form solution. The magnitude of the sound decay $|d_\mathrm{m}^{(2)}(k)|$ is modeled as

$$|d_\mathrm{m}^{(2)}(k)| = A\, w(k)\, e^{-k\rho T_\mathrm{s}} \epsilon(k) \tag{6}$$

with $k \in \{0, 1, \ldots, N-1\}$, real amplitude factor $A = \exp\{\alpha\}$ and $w(k)$ denoting a sequence of N i.i.d. log-normally distributed random variables. The model of Eq. (6) can be rewritten as

$$\begin{aligned} |d_\mathrm{m}^{(2)}(k)| &= e^{\alpha}\, e^{-\rho\, k\, T_\mathrm{s}}\, e^{n(k)}\, \epsilon(k) \\ &= \exp\{\underbrace{\alpha + \beta\, k + n(k)}_{= y(k)}\}\, \epsilon(k) \end{aligned} \tag{7}$$

with $\beta = -\rho\, T_\mathrm{s}$. The noise term $n(k)$ denotes a sequence of N i.i.d. random variables following a normal distribution $\mathcal{N}(0, \sigma^2)$. The log-likelihood function for an observation $\boldsymbol{y} = [y(0), y(1), \ldots, y(N-1)]^T$ is given by

$$\ln p(\boldsymbol{y}; \alpha, \beta) = -\frac{N}{2} \ln \left(2\,\pi\,\sigma^2\right) - \sum_{k=0}^{N-1} \frac{(y(k) - \alpha - \beta\, k)^2}{2\,\sigma^2} \,. \tag{8}$$

The ML estimate $\hat{\alpha}$ for the parameter α for a given observation $\boldsymbol{y}$ is obtained by

$$\frac{\partial \ln p(\boldsymbol{y}; \hat{\alpha}, \hat{\beta})}{\partial \hat{\alpha}} = \frac{2}{\sigma^2} \left(\sum_{k=0}^{N-1} y(k) - \hat{\alpha} - \hat{\beta}\, k \right) \overset{!}{=} 0 \tag{9}$$

$$\Rightarrow \quad \hat{\alpha} = \frac{1}{N} \left(\sum_{k=0}^{N-1} y(k) - \hat{\beta}\, k \right) . \tag{10}$$

The ML estimate $\hat{\beta}$ for the parameter β is derived accordingly

$$\frac{\partial \ln p(\boldsymbol{y}; \hat{\alpha}, \hat{\beta})}{\partial \hat{\beta}} = \frac{1}{\sigma^2} \left(\sum_{k=0}^{N-1} (y(k) - \hat{\alpha} - \hat{\beta}\, k)\, k \right) \overset{!}{=} 0 \tag{11}$$

$$\Rightarrow \quad \hat{\beta} = \frac{6}{N\,(N-1)\,(2\,N-1)} \sum_{k=0}^{N-1} k\,(y(k) - \hat{\alpha}) \,. \tag{12}$$

Table I
Number of required arithmetic operations of the proposed ML estimator in comparison to the iterative schemes of [11] and [16].

	Multiplications	Summations	Divisions
New ML estimator	$7\,N + 6$	$7\,N + 5$	3
ML estimator of [11]	$N\,(3\,I + 1) + 2\,I$	$(2\,N - 1)\,I$	-
ML estimator of [16]	$(N + 8)\,J + N$	$(N + 6)\,J$	J

Combining Eq. (10) and Eq. (12) leads after some manipulations to

$$\hat{\alpha} = \frac{2\,(2\,N-1)}{N\,(N+1)} \sum_{k=0}^{N-1} y(k) - \frac{6}{N\,(N+1)} \sum_{k=0}^{N-1} k\, y(k) \tag{13}$$

$$\hat{\beta} = \frac{-6}{N\,(N+1)} \sum_{k=0}^{N-1} y(k) + \frac{12}{N\,(N-1)(N+1)} \sum_{k=0}^{N-1} k\, y(k). \tag{14}$$

The estimated RT is obtained by means of Eq. (5):

$$\widehat{T}_{60} = \frac{-3\, T_\mathrm{s}}{\hat{\beta}\, \log_{10}(e)} \approx -6.908\, \frac{T_\mathrm{s}}{\hat{\beta}} \,. \tag{15}$$

Table I compares the number of required arithmetic operations of the proposed ML estimator with the so-called fast block algorithm presented in [11] and the approach of [16] where the likelihood function is evaluated for J discrete values of the decay rate to find approximately its maximum. The proposed ML estimator needs the logarithm of the magnitude of a detected sound decay $|d(k)|$, which can be obtained, e.g., via look-up tables or a Taylor series approximation. The use of a third-order Taylor series approximation turned out to be sufficient for this purpose and is hence considered in Table I for the new estimator as well as the ML estimator of [16].

In [11], it is suggested that about $I = 5$ iterations are needed for a properly tuned step size parameter (which still needs to be confirmed in practice). The approach of [16] to determine the ML-estimate is also used in [27], where it has been shown that good results can be achieved with a search over $J = 10$ values for the decay rate. However, the evaluation for discrete values does not guarantee that a (global) maximum is found.

The proposed closed-form solution features the lowest complexity of all listed ML estimators where the global maximum of the likelihood function is found in one step without the need for parameter tuning as in [11].

It should be noted that Table I only lists the operations to determine a single ML-estimate but not the overall complexity of the respective algorithms for RTE. The overall computational complexity of the approach of [16] is significantly lower and the estimation accuracy higher than for the method presented in [11] due to an efficient pre-selection scheme for the sound decays.

III. Crámer-Rao Lower Bound (CRLB)

The derived estimators of Eq. (13) and Eq. (14) are unbiased since it is easily shown that $E\{\hat{\alpha}\} = \alpha$ and $E\{\hat{\beta}\} = \beta$. A lower bound for the variance can be obtained in this case by

ISBN 978-90-827970-1-5 © EURASIP 2018

the CRLB. Calculating the inverse of the Fisher information matrix for the derived ML estimator leads to [29]

$$\boldsymbol{\mathcal{J}}_{\alpha,\beta}^{-1} = \begin{bmatrix} \dfrac{\partial^2 \ln p(\boldsymbol{y};\alpha,\beta)}{\partial\alpha^2} & \dfrac{\partial^2 \ln p(\boldsymbol{y};\alpha,\beta)}{\partial\alpha\partial\beta} \\ \dfrac{\partial^2 \ln p(\boldsymbol{y};\alpha,\beta)}{\partial\beta\partial\alpha} & \dfrac{\partial^2 \ln p(\boldsymbol{y};\alpha,\beta)}{\partial\beta^2} \end{bmatrix}^{-1}$$

$$= \frac{1}{\sigma^2}\begin{bmatrix} N & \dfrac{N\,(N-1)}{2} \\ \dfrac{N\,(N-1)}{2} & \dfrac{N\,(N-1)\,(2\,N-1)}{6} \end{bmatrix}^{-1} \tag{16}$$

$$= \sigma^2 \begin{bmatrix} \dfrac{2\,(2\,N-1)}{N\,(N+1)} & -\dfrac{6}{N\,(N+1)} \\ \dfrac{-6}{N\,(N+1)} & \dfrac{12}{N\,(N^2-1)} \end{bmatrix}. \tag{17}$$

It follows from Eq. (17) that the CRLB for the variance of the estimator for β is given by

$$\mathrm{var}\{\hat{\beta}\} \geq \frac{12\,\sigma^2}{N\,(N^2-1)}. \tag{18}$$

After some calculations, it can be shown that

$$\begin{bmatrix} \dfrac{\partial \ln p(\boldsymbol{y};\alpha,\beta)}{\partial\alpha} \\ \dfrac{\partial \ln p(\boldsymbol{y};\alpha,\beta)}{\partial\beta} \end{bmatrix} = \boldsymbol{\mathcal{J}}_{\alpha,\beta} \cdot \begin{bmatrix} \hat{\alpha}-\alpha \\ \hat{\beta}-\beta \end{bmatrix} \tag{19}$$

with $\hat{\alpha}$, $\hat{\beta}$, and $\boldsymbol{\mathcal{J}}_{\alpha,\beta}$ given by Eq. (13), Eq. (14), and Eq. (16), respectively. Thus, the derived ML estimator for the parameter β from a model decay $y(k)$ attains the CRLB and is hence (statistically) *efficient*, cf., [29].

As shown in [10], [28], the ML estimators based on the model of Eq. (4) [10], [11], [16] also reach the CRLB but they are only asymptotically unbiased, whereas the proposed approach is a minimum variance unbiased (MVU) estimator for a *finite* observation length N.

IV. Blind RT Estimation

The overall algorithm to estimate the RT is based on [27]. The noisy and reverberant speech signal is first denoised using the spectral minimum mean-square error (MMSE) estimator of [30]. For this, a discrete Fourier transform (DFT) of transform length 512 and half-overlapping frames are considered. The needed noise power spectral density (PSD) is obtained by the MMSE-based estimator proposed in [31].

The denoised signal is split into $\mu = n_{\mathrm{oct}}, \ldots, N_{\mathrm{oct}}$ subbands $Y_\mu(k)$ by means of a 1/3-octave filter-bank using the design described in [24]. The first n_{oct} subbands are not considered since they have a very narrow bandwidth and, hence, provide low signal energy only. Each subband signal is processed in signal frames of M samples shifted by M_Δ sample instants

$$x_\mu(\lambda, m) = Y_\mu(\lambda\, M_\Delta + m) \quad \text{with} \quad m = 0, 1, \ldots, M-1 \tag{20}$$

and frame index $\lambda \in \mathbb{N}$. In a first step, a pre-selection is conducted for each subband signal μ to detect potential frames with sound decays, cf., Eq. (3). For this, the current frame $x_\mu(\lambda, m)$ is divided into $L_{\mathrm{f}} = M/P \in \mathbb{N}$ sub-frames

$$f_\mu(\lambda, l, \kappa) = x_\mu(\lambda, l\,P + \kappa) \tag{21}$$

of length P with $\kappa \in \{0, 1, \ldots, P-1\}$ and $l \in \{0, 1, \ldots, L_{\mathrm{f}}-1\}$. It is then tested for frame λ whether the energy, maximum and minimum value of a sub-frame $f_\mu(\lambda, l, \kappa)$ deviate relative to the successive sub-frame $f_\mu(\lambda, l+1, \kappa)$ according to

$$w \cdot \sum_{\kappa=0}^{P-1} f_\mu^2(\lambda, l, \kappa) > \sum_{\kappa=0}^{P-1} f_\mu^2(\lambda, l+1, \kappa) \tag{22a}$$

$$w \cdot \max\{f_\mu(\lambda, l, \kappa)\} > \max\{f_\mu(\lambda, l+1, \kappa)\} \tag{22b}$$

$$w \cdot \min\{f_\mu(\lambda, l, \kappa)\} < \min\{f_\mu(\lambda, l+1, \kappa)\} \tag{22c}$$

for $l = 0, 1, \ldots, L_{\mathrm{f}}-2$ and weighting factor $0 \ll w \leq 1$ to control the required decay. If the conditions of Eq. (22) are fulfilled for $\tilde{l} \geq l_{\mathrm{min}}$ consecutive sub-frames, a possible sound decay is detected within a frame. In this case, the RT is calculated from the segment composed of these $\tilde{l}$ consecutive sub-frames

$$y_\mu(i) = \ln(|x_\mu(\tilde{\lambda}, i)| + \delta) \quad \text{with} \quad i = 0, 1, \ldots, \tilde{l}\,P - 1 \tag{23}$$

with $0 \leq \delta \ll 1$ by means of Eq. (14) and Eq. (15), where $\tilde{\lambda}$ marks a frame in which a sound decay is detected. If no sound decay is detected, the next signal frame $x_\mu(\lambda, m)$ is tested for a decay. A new ML estimate is used to update a histogram calculated for the latest K_{f} ML estimates of the RT. The RT value associated with the maximum of the histogram is taken as instantaneous estimate $\widehat{\mathcal{T}}_{60}^{(\mathrm{inst})}(\mu, \tilde{\lambda})$. The variance for the instantaneous estimate $\widehat{\mathcal{T}}_{60}(\mu, \tilde{\lambda})$ is reduced by recursive smoothing

$$\widehat{\mathcal{T}}_{60}(\mu, \tilde{\lambda}) = \gamma \cdot \widehat{\mathcal{T}}_{60}(\mu, \tilde{\lambda}-1) + (1-\gamma) \cdot \widehat{\mathcal{T}}_{60}^{(\mathrm{inst})}(\mu, \tilde{\lambda}) \tag{24}$$

with forgetting factor $0 \leq \gamma < 1$. If no decay was detected, $\widehat{\mathcal{T}}_{60}(\mu, \lambda) = \widehat{\mathcal{T}}_{60}(\mu, \lambda-1)$ and $\widehat{\mathcal{T}}_{60}(\mu, \lambda) = \widehat{\mathcal{T}}_{60}(\mu, \tilde{\lambda})$ otherwise.

The fullband estimate for the RT is obtained by a weighted averaging of all subband estimates

$$\widehat{\mathcal{T}}_{60}(\lambda) = \frac{1}{N_{\mathrm{oct}} - n_{\mathrm{oct}} + 1} \sum_{\mu=n_{\mathrm{oct}}}^{N_{\mathrm{oct}}} w_\mu(\lambda)\, \widehat{\mathcal{T}}_{60}(\mu, \lambda), \tag{25}$$

which will then obviously mask a possibly given frequency dependency of the RT. The weighting factors $w_\mu(\lambda)$ are given by the ratio of the subband signal energy to the sum of all considered subband energies calculated for signal segments of length L_{w}

$$w_\mu(\lambda) = \frac{\sum\limits_{i=0}^{L_{\mathrm{w}}-1} Y_\mu^2(\lambda\, M_\Delta + i)}{\sum\limits_{\mu=n_{\mathrm{oct}}}^{N_{\mathrm{oct}}} \sum\limits_{i=0}^{L_{\mathrm{w}}-1} Y_\mu^2(\lambda\, M_\Delta + i)} \tag{26}$$

with $\mu = n_{\mathrm{oct}}, \ldots, N_{\mathrm{oct}}$. This weighting is motivated by the rationale that a low signal energy in a subband is assumed to be associated with increased variance of the estimate.

Table II
Algorithmic parameters.

M	M_Δ	P	$l_{\min}$	L_{f}	w	K_{f}	γ	N_{oct}	n_{oct}
4923	137	547	3	9	1	800	0.95	30	20

Table III
Estimation performance of different ML-based RT estimators.

Algorithm	Bias	MSE	ρ_{p}
RT estimator of [16]	-0.05	0.11	0.38
RT estimator of [27]	-0.11	0.09	0.67
New RT estimator	-0.13	0.09	0.69

The devised algorithm allows to estimate a *time-varying* RT. If this is not needed, as for the simulations in Sec. V, the average of all estimates $\widehat{T}_{60}(\lambda)$ can be used to reduce the variance of the estimate and the weights of Eq. (26) are calculated for the entire signal duration.

V. Simulation Results

The proposed algorithm is compared with the closely-related ML-based RT estimators presented in [16] (without fast adaptation) and [27]. The algorithmic parameters of Table II are used for all three approaches to ease the comparison of the proposed ML estimator. A bin size of $0.1\,\mathrm{s}$ and bin centers $0.2\,\mathrm{s}, 0.3\,\mathrm{s}, \ldots, 1.1\,\mathrm{s}$ are used for the histogram of RT estimates. For the evaluation, the single-channel evaluation (Eval) dataset of the ACE Challenge [24], [25] was generated according to [32]. This dataset contains 4500 noisy, reverberant speech signals with a sampling frequency of $f_{\mathrm{s}} = 16\,\mathrm{kHz}$. The recordings are obtained for different acoustic environments with varying DRRs and signal-to-noise ratios (SNRs) of $-1\,\mathrm{dB}$, $12\,\mathrm{dB}$ and $18\,\mathrm{dB}$. For all considered approaches, the mean of the time-varying RT estimates is taken as RT estimate for each recording.

As in [25], the evaluation of the estimation accuracy is done by means of the bias, the mean-square error (MSE) and the Pearson correlation coefficient ρ_{p} between the RT estimates and ground-truth values. The obtained values are listed in Table III. The approach of [16] achieves the lowest bias but features also the lowest value for ρ_{p}, i.e., this approach shows the lowest correlation between the estimates and ground truth for all considered algorithms. The proposed approach features a lower MSE than the approach of [16] and a marginally higher bias and same MSE than the approach of [27].[1] Thereby, the correlation between the estimated RTs and ground-truth values is significantly higher than for the approach of [16] and slightly higher than for the approach of [27].

A comparison of the execution times of the respective MATLAB implementations (on an Intel Core i7 processor) is taken as a rough assessment for the computational complexity of the algorithms.[2] The approach of [27] (which was optimized w.r.t. estimation accuracy but not computational complexity for the ACE Challenge) achieves a higher estimation accuracy in contrast to [16], but its execution time was about 10 times higher, mainly due to the calculation of the RT within subbands. The estimation performance of the new approach is similar than for the approach of [27], but the execution time is only about 3 times higher than for the approach of [16].

The computational complexity of the approach in [16] can be reduced by using the new closed-form ML estimator. An evaluation of this modified approach by means of the Eval dataset led to similar results for the estimation accuracy as listed in Table III for the original approach of [16], but a reduction of the execution time by about 40%. Thus, the new closed-form ML estimator achieves a significant reduction of the computational complexity compared to previous ML-based RT estimators, while achieving the same (or even slightly better) estimation accuracy.

It should be noted that the algorithm presented in [27] has scored among the best in the ACE-Challenge for single-channel RT estimation as detailed in [25] where a more comprehensive comparison of various RT estimators can be found.

VI. Conclusions

A novel method for ML-based RT estimation is proposed. In contrast to related ML-based blind RT estimators [10], [11], [15], [16], it is derived from a modified statistical model for the sound decay which leads to a simple closed-form solution for the ML estimates. This results in a significant lower computational complexity compared to previous ML-based RT estimators, which use iterative schemes [11] or a search over a finite set of decay rates [16] to determine the ML estimate. Moreover, it is shown that the underlying ML estimator to determine the decay rate is an MVU estimator for observation intervals of finite length, which is not the case for current ML-based RT estimators. An evaluation of the new approach for the evaluation dataset of the ACE Challenge shows that the proposed RT estimator achieves a similar estimation accuracy but with a significantly lower computational complexity compared to the ML-based RT estimator of [27], which scored among the best at the ACE Challenge for single-channel RT estimation. The estimation and incorporation of a priori knowledge about the RT is a promising approach to further improve the estimation accuracy of the presented estimator.

Acknowledgment

The authors gratefully acknowledge the support of Adhithyan Ramadoss in evaluating the various estimators.

[1]The values for the RT estimator of [27] in Table III are not identical to those listed in [25] since the random noise used to generate the ACE evaluation dataset is not identical to the random noise used to generate the evaluation dataset of the ACE Challenge, see also [32].

[2]The calculation of the logarithm has been performed for all approaches by the MATLAB function `log`.

REFERENCES

[1] H. Kuttruff, *Room Acoustics*, Taylor & Francis, London, UK, 4th edition, 2000.

[2] T. Yoshioka, A. Sehr, M. Delcroix, K. Kinoshita, R. Maas, T. Nakatani, and W. Kellermann, "Making Machines Understand Us in Reverberant Rooms: Robustness Against Reverberation for Automatic Speech Recognition," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 114–126, 2012.

[3] R. Maas, E. A. P. Habets, A. Sehr, and W. Kellermann, "On the Application of Reverberation Suppression to Robust Speech Recognition," in *Proc. of Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 297–300.

[4] P. P. Parada, D. Sharma, P. A. Naylor, and T. van Waterschoot, "Reverberant Speech Recognition: A Phoneme Analysis," in *Proc. of IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Atlanta (Georgia), USA, Dec. 2014, pp. 567–571.

[5] F. Xiong, S. Goetze, and B. T. Meyer, "Estimating Room Acoustic Parameters for Speech Recognizer Adaptation and Combination in Reverberant Environments," in *Proc. of Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 5522–5526.

[6] 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.

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

[8] H. W. Löllmann and P. Vary, "Low Delay Noise Reduction and Dereverberation for Hearing Aids," *EURASIP Journal on Applied Signal Processing*, vol. 2009, pp. 1–9, 2009.

[9] M. R. Schroeder, "New Method of Measuring Reverberation Time," *Journal of the Acoustical Society of America*, vol. 37, pp. 409–412, 1965.

[10] 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, Nov. 2003.

[11] R. Ratnam, D. L. Jones, and W. D. O'Brien, "Fast Algorithms for Blind Estimation of Reverberation Time," *IEEE Signal Processing Letters*, vol. 11, no. 6, pp. 537–540, June 2004.

[12] M. Wu and D. Wang, "A Pitch-Based Method for the Estimation of Short Reverberation Time," *Acta Acustica united with Acustica*, vol. 92, pp. 337–339, 2006.

[13] 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 *Proc. of Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Las Vegas (Nevada), USA, Apr. 2008, pp. 329–332.

[14] T. Falk and W.-Y. Chan, "A Non-Intrusive Quality Measure of Dereverberated Speech," in *Proc. of Intl. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle (Washington), USA, Sept. 2008.

[15] 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.

[16] H. W. Löllmann, E. Yilmaz, M. Jeub, and P. Vary, "An Improved Algorithm for Blind Reverberation Time Estimation," in *Proc. of Intl. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, Aug. 2010.

[17] 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," *Journal of the Acoustical Society of America*, vol. 4, no. 131, pp. 2811–2816, Apr. 2012.

[18] J. Eaton, N. D. Gaubitch, and P. A. Naylor, "Noise-Robust Reverberation Time Estimation Using Spectral Decay Distributions with Reduced Computational Cost," in *Proc. of Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013, pp. 161–165.

[19] F. Xiong, S. Goetze, and B. T. Meyer, "Blind Estimation of Reverberation Time Based on Spectro-Temporal Modulation Filtering," in *Proc. of Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013, pp. 443–447.

[20] N. López, Y. Grenier, G. Richard, and I. Bourmeyster, "Low Variance Blind Estimation of the Reverberation Time," in *Proc. of Intl. Workshop on Acoustic Signal Enhancement (IWAENC)*, Aachen, Germany, Sept. 2012.

[21] R. Talmon and E. A. P. Habets, "Blind Reverberation Time Estimation by Intrinsic Modeling of Reverberant Speech," in *Proc. of Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013, pp. 156–160.

[22] S. Diether, L. Bruderer, A. Streich, and H.-A. Loeliger, "Efficient Blind Estimation of Subband Reverberation Time from Speech in Non-Diffuse Environments," in *Proc. of Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015, pp. 743–747.

[23] C. Schlüdt and P. Händel, "Noise Robust Integration for Blind and Non-Blind Reverberation Time Estimation," in *Proc. of Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015, pp. 56–60.

[24] J. Eaton, A. H. Moore, N. D. Gaubitch, and P. A. Naylor, "The ACE Challenge - Corpus Description and Performance Evaluation," in *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz (New York), USA, Oct. 2015.

[25] J. Eaton, N. D. Gaubitch, A. H. Moore, and P. A. Naylor, "Estimation of Room Acoustic Parameters: The ACE Challenge," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 24, no. 10, pp. 1681–1693, Oct. 2016.

[26] T. de M. Prego, A. A. de Lima, R. Zambrano-López, and S. L. Netto, "Blind Estimators for Reverberation Time and Direct-To-Reverberant Energy Ratio Using Subband Speech Decomposition," in *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz (New York), USA, Oct. 2015, pp. 1–5.

[27] H. W. Löllmann, A. Brendel, P. Vary, and W. Kellermann, "Single-Channel Maximum-Likelihood T60 Estimation Exploiting Subband Information," in *Proc. of ACE Challenge Workshop - a satellite event of IEEE-WASPAA 2015*, New Paltz (New York), USA, Oct. 2015, available on arXiv: 1511.04063.

[28] C. Schüldt and P. Händel, "Decay Rate Estimators and Their Performance for Blind Reverberation Time Estimation," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 22, no. 8, pp. 1274–1284, Aug. 2014.

[29] S. M. Kay, *Fundamentals of Statistical Signal Processing - Estimation Theory*, Prentice Hall Signal Processing Series, Upper Saddle River, New Jersey, 1993.

[30] J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, "Minimum Mean-Square Error Estimation of Discrete Fourier Coefficients with Generalized Gamma Priors," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 15, no. 6, pp. 1741–1752, Aug. 2007.

[31] R. C. Hendriks, R. Heusdens, and J. Jensen, "MMSE Based Noise PSD Tracking with Low Complexity," in *Proc. of Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Dallas (Texas), USA, Mar. 2010, pp. 4266–4269.

[32] J. Eaton and P. A. Naylor, *Acoustic Characterization of Environments (ACE) Corpus Software Instructions*, Imperial College London, 2016, http://www.ee.ic.ac.uk/naylor/ACEweb/index.html.

ISBN 978-90-827970-1-5 © EURASIP 2018