PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

# Comparative study of single-channel algorithms for blind reverberation time estimation

Heinrich W. Löllmann, Andreas Brendel, and Walter Kellermann

Friedrich-Alexander University Erlangen-Nürnberg, Germany
{Heinrich.Loellmann, Andreas.Brendel, Walter.Kellermann}@FAU.de

**Abstract**

Knowledge about the reverberation time (T60) is exploited by numerous algorithms, e.g., for automatic speech recognition, speech dereverberation, or as a feature for acoustic scene classification. This contribution provides a comprehensive survey about various state-of-the-art methods to estimate the reverberation time blindly from a noisy and reverberant speech signal and compares their estimation performance for different acoustic scenarios. The evaluation considers different acoustic conditions regarding the signal-to-noise ratio (SNR) and direct-to-reverberation ratio (DRR), and uses databases with room impulse responses measured in different acoustic enclosures. The findings of the Acoustic Characterisation of Environments (ACE) challenge are extended by taking into account algorithms and acoustic scenarios which are not considered in this benchmarking campaign, and by using a much larger set of reverberant speech signals.

Keywords: Reverberation time, blind estimation, single-channel, survey

## 1 Introduction

The reverberation time (RT) $T_{60}$ is an important and widely used quantity for the characterization of acoustic enclosures [1]. Knowledge about the RT is exploited for various applications such as automatic speech recognition (ASR), e.g., [2–4] or speech dereverberation, e.g., [5–7]. For such applications, the (possibly time-varying) RT can typically not be obtained by intrusive methods, e.g., by estimating it from a measured room impulse response (RIR) by the Schroeder method [8] or its variants [9], but needs to be estimated *blindly* from a reverberant speech signal, usually in the presence of background noise. Various *reverberation time estimators (RTEs)* to address this challenging task were developed in the past two decades, e.g., [10–34].

The growing research interest in blind RT estimation also fuels the interest in comparative studies of different methods. In [35], three state-of-the-art methods [16, 17, 19] are compared using simulated and measured RIRs. In [30], an analytical and experimental analysis of three common decay rate estimators [10, 12, 18] is presented which are the main building blocks of many modeled-based RTEs. The variety of concepts for blind RT estimation has motivated the Acoustic Characterisation of Environments (ACE) challenge, which provided an objective benchmarking of RT and direct-to-reverberant energy ratio (DRR) estimators [36, 37].

This contribution presents a survey of various methods for blind RT estimation and evaluates the performance of selected methods corresponding to different classes of algorithms for $T_{60}$ estimation and, thus, aims at complementing previous benchmarking efforts. The investigation is based on the evaluation dataset and performance measures used for the ACE challenge to complement its findings by also investigating algorithms which were not evaluated in that challenge. In addition, a second database is used generated by more than 700 measured RIRs to cover a larger range of RTs and DRRs than for the ACE challenge to investigate how well different methods perform for various acoustical scenarios.

The paper is organized as follows: In Sec. 2, a survey of different approaches to blind RT estimation is provided. The datasets used for the evaluation are described in Sec. 3 and Sec. 4 presents evaluation results for various RTEs. The paper concludes with a summary of the main results in Sec. 5.

## 2 Methods for Blind RT Estimation

This section provides a survey over various methods for blind RT estimation from single-channel speech recordings.

### 2.1 Model-based methods without training

Many approaches to model-based blind RT estimation were inspired by the work of Lebart *et al.* [10]. They have in common that speech pauses in which the signal energy decays gradually due to the room reverberation need to be detected followed by an estimation of the RT from the decay behavior. It is furthermore assumed that the signal decay can be described by a statistical model

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

with sample index $k$, amplitude $A > 0$, decay rate $\rho$, unit step sequence $\varepsilon(k)$, sampling period $T_{\mathrm{s}}$ and $v(k)$ representing a sequence of i.i.d. standard-normally distributed random variables. The decay rate $\rho$ is then estimated by linear regression from the logarithm of the smoothed energy envelope of the reverberant speech signal within the detected decay interval.[1] The RT is obtained from the decay rate by the relation $T_{60} = 3/(\rho\, \log_{10}(e))$ such that both terms can be used interchangeably.

In [12], the model of Eq. (1) is used to derive a *maximum-likelihood (ML) estimator* for the decay rate. Decay rates are estimated continuously from speech frames (of 200 ms duration) shifted by one sample instant. The final decay rate estimate is determined by applying an order statistics filter to a histogram of all decay rate estimates, i.e., the maximum of the histogram yields the final decay rate estimate as detailed in [12]. A benefit of this approach is that no speech decay detection is needed. However, the ML estimate of [12] is not given by a closed-form solution, but has to be calculated by numerical optimization. In [13], two efficient solutions are presented for this: A fast online algorithm is proposed which compares the values of the log-likelihood function for a finite set of candidate decay rates to determine its approximate maximum. As an alternative, a fast block algorithm is proposed (but not evaluated) which finds the maximum of the log-likelihood function iteratively by a gradient-based method. However, a suitable step size has to be found to ensure convergence of the algorithm.

The use of an ML estimator to determine the RT from a *noisy* reverberant speech signal by means of a prior denoising step is discussed in [18]. Based on this work, an efficient scheme to estimate a time-varying RT by ML estimation is presented in [19]. A pre-selection of promising speech decays is performed to reduce the computational complexity and the number of outliers for the $T_{60}$ estimation. In [20], it is proposed to determine the fullband RT by an energy-weighted average of ML-estimates determined from the subband signals of an octave filter bank. As shown in [20, 36], this method provides a higher estimation accuracy at the cost of increased computational complexity relative to the scheme of [19].

Various authors have derived ML estimators for the RT which are based on variants of the statistical decay model of Eq. (1). In [18], an ML estimator is derived which allows to estimate the decay rate from a measured RIR distorted by additive noise. An improvement of this estimator is presented in [30] which achieves a similar estimation accuracy, but involves a significantly lower computational complexity. However, such estimators rely on an accurate estimation of the noise variance. In [21], an ML estimator is derived from a statistical model for the magnitude of the reverberant speech decay which leads to a closed-form solution for the $T_{60}$ estimate. The obtained RTE achieves a similar estimation accuracy as the method of [20], but with a considerably lower computational complexity. In [14], an ML estimator is presented which extends the model of Eq. (1) to account for different decay rates for the early and late reflections. In [24] and [33], ML estimators for the RT are derived which assume a sequence of Laplace- and Gamma-distributed random variables, respectively, for $v(k)$ in Eq. (1).

In [25], a speech model-based algorithm is presented where the decay rate is estimated from the residual signal obtained by a linear prediction (LP) filter. This approach is based on the assumption that the signal after LP

[1] In [10], there is no scheme described for detecting the speech pauses.

filtering contains a residual speech signal convolved with the RIR, but not with the vocal tract filter any more. In this case, the auto-correlation of the signal after LP filtering contains decays, which can be modeled by Eq. (1) and estimated by an ML estimator as presented in [12]. An LP filter is also used in [32] to obtain an estimate of the RIR from which the RT is calculated by the Schroeder method.

Besides ML estimation, the use of *linear regression* is another common method to obtain the RT from detected speech decays [10]. In [22], RT estimation is performed in the discrete magnitude short-time Fourier transform (STFT) domain by estimating the decay rate by linear regression from detected decay regions. The median of the subband estimates is mapped to the fullband estimate by a linear function. A modification of this scheme is presented in [23] to account for additive noise. It should be noted that model-based RTEs which obtained the fullband estimate by a mapping of subband estimates [20,23] scored among the best at the ACE challenge [37].

The RTE presented in [29] operates in the modulation domain. The modulation coefficients are obtained by a discrete Fourier transform (DFT) of the magnitude of the discrete STFT coefficients of the reverberant speech signal calculated over the STFT frequency bands. The decay rate is determined by minimizing the mean-square error (MSE) between the modulation coefficients of the reverberant speech and the representation of the statistical RIR of Eq. (1) in the modulation domain for a detected sound decay.

### 2.2 Methods based on single feature mapping with training

The previously discussed model-based approaches have in common that the decay rate is estimated directly via linear regression or ML estimation from a speech decay interval, an auto-correlation sequence or an estimate of the RIR. In contrast, methods based on feature mappings derive the RT from a distinct parameter calculated from the reverberant speech which changes with the $T_{60}$. The RT is then obtained by a learned mapping function between the calculated feature and the $T_{60}$, e.g., [15–17,26–28].

In [15], a pitch-based method is proposed for the estimation of short RTs (up to 0.6 s). It is based on the observation that reverberation changes the harmonic structure of voiced speech and the pitch strength measure is introduced to quantify this change. The RT is finally obtained by a learned mapping function which relates the pitch strength to the $T_{60}$.

In [16], the variance corresponding to the negative side of the distribution of measured decay rates is mapped to the RT by a second-order polynomial function whose coefficients are found by training with a few reverberant speech files. The distribution of the decay rates is obtained by estimating decay rates in the discrete magnitude STFT domain via linear regression for each frame without sound decay detection. Various improvements of this algorithm have been proposed: In [26], the method of [16] is extended by including an LP-based prewhitening step to obtain a better fit for the model assumption of Eq. (1). In [28], the estimation accuracy and robustness towards noise is improved by learning a data-driven representation of the decay rates based on training for several room configurations. In [27], the complexity of the RTE of [16] is reduced by calculating the decay rates in Mel frequency bands instead of STFT frequency bands. The robustness towards noise is improved by introducing a dependency on the signal-to-noise ratio (SNR) for the $T_{60}$ calculation.

### 2.3 Methods based on multiple features with training

This class of RTEs maps multiple features to the RT, typically by using an artificial neural network (ANN). In contrast to approaches based on single feature mapping, these methods usually require a much larger training and do typically not rely on the statistical RIR model of Eq. (1).

An early attempt to estimate the RT with ANNs is presented in [11] which was trained on noiseless reverberant speech. In [17], two approaches to estimate the RT are presented where various features are mapped to the RT by support vector regression (SVR). In the first method, the cepstral coefficients of the reverberant speech signal are used for the feature calculation. As an algorithmic variant, the reverberation-to-speech modulation energy ratio (RSMR) is introduced, which is calculated in the modulation domain and used as feature for RT estimation.

More recently, the joint room parameter estimator (jROPE) has been presented in [34] which jointly estimates the RT and early-to-late reverberation ratio (ELR) by an ANN. A multi-layer perceptron maps auditory-inspired temporal modulation features to binned classes for the RT and ELR.

## 3 Datasets

The data corpus of the ACE challenge [36, 37] has become a common dataset for the evaluation of RT and DRR estimators. The evaluation dataset contains 4500 noisy reverberant speech files with a sampling frequency of 16 kHz and SNRs of $-1$ dB, 12 dB and 18 dB. The files were generated by convolving RIRs measured in five different rooms with anechoic speech signals and adding noise recorded in the same rooms. The development dataset contains 288 sound files generated by measured RIRs and recorded noise obtained for two rooms different from those considered for the evaluation dataset.

The evaluation dataset of the ACE challenge provides very realistic recordings, but comprises only a rather small set of acoustic environments due to the cumbersome recording procedure. Therefore, a new, larger dataset with noisy reverberant speech signals was created for this study, termed as large evaluation dataset (LED). It is used in addition to the ACE evaluation dataset to account for a larger range of RTs and DRRs. As for the ACE challenge dataset, only measured RIRs and no simulated RIRs were used to generate the reverberant speech. The measured RIRs were mainly obtained from publicly available databases which do not provide separate noise recordings for the rooms in which the RIRs were measured. Therefore, the ambient noise was generated by convolving the late reverberant part of the RIRs with either white Gaussian noise or babble noise, cf., [34]. Thus, the noisy reverberant speech signals were obtained as follows

$$y_{p,i,j,\alpha}(k) = \sum_{l=0}^{L-1} s_p(k-l)\, h_i(l) + \alpha \sum_{m=0}^{L-1} h_i^{(\mathrm{late})}(m)\, n_j(k-m) \tag{2}$$

with indices $p \in \{1,\dots,4\}$, $i \in \{1,\dots,782\}$ and $j \in \{1,2\}$ denoting the used anechoic speech signals $s_p(k)$, measured RIRs $h_i(k)$ and noise signals $n_j(k)$, respectively. Two male and two female speech files each with a length of $14-16$ s were used obtained by concatenating speech files of the CSTR VCTK1 database [38] downsampled to 16 kHz. Babble noise for an anechoic environment can be created by adding up non-reverberant speech signals, e.g., [34]. Here, 7 male and 7 female speech sequences using concatenated files of the CSTR VCTK1 database were added for this. The temporal segment of the RIR beginning 50 ms after the first dominant peak was taken as late reverberant part $h_i^{(\mathrm{late})}(m)$. The scaling factor $\alpha \geq 0$ was adjusted to obtain a predefined segmental SNR between reverberant speech and additive noise. The active speech level for the segmental SNR calculation was determined according to ITU-T P.56 [39] as provided by the MATLAB toolbox Voicebox [40]. Segmental SNR values of $-5$ dB, 5 dB, 15 dB and the case of no added noise were considered for the generation of the dataset.

The RIRs to generate the LED were mainly obtained from publicly available databases, namely, the Aachen impulse response (AIR) database [41], the multichannel acoustic reverberation database at York (MARDY) [42], the multichannel impulse reponse database (MIRD) [43] and the OpenAirLib database [44]. Only RIRs with a RT of less than 1.5 s were considered. Moreover, RIRs measured in the acoustic laboratory, a seminar room and a hallway of the Chair of Multimedia Communication and Signal Processing of the University Erlangen-Nuremberg have been used with RTs ranging from 0.2 to 0.9 s.

A scatter plot showing the RTs and DRRs of the 782 RIRs included in the LED is provided by Fig. 1. The $T_{60}$ was determined by the Schroeder method [8] using least-squares (LS) fitting. The DRR was determined according to [37] (but without equalization prior to the peak search). The RIRs of the ACE challenge were not used for generating the LED, but their RTs and DRRs are shown in Fig. 1 for comparison. The used databases often provide various RIRs measured in the same room at different positions which is reflected by the different DRR values for the same RT.

Four scenarios were considered for generating the LED: A male or female speech signal in the presence of either diffuse noise or babble noise. Hence, the database for each segmental SNR value comprised 3128 audio files and the complete database with its 12512 files has a total duration of 75 hours.

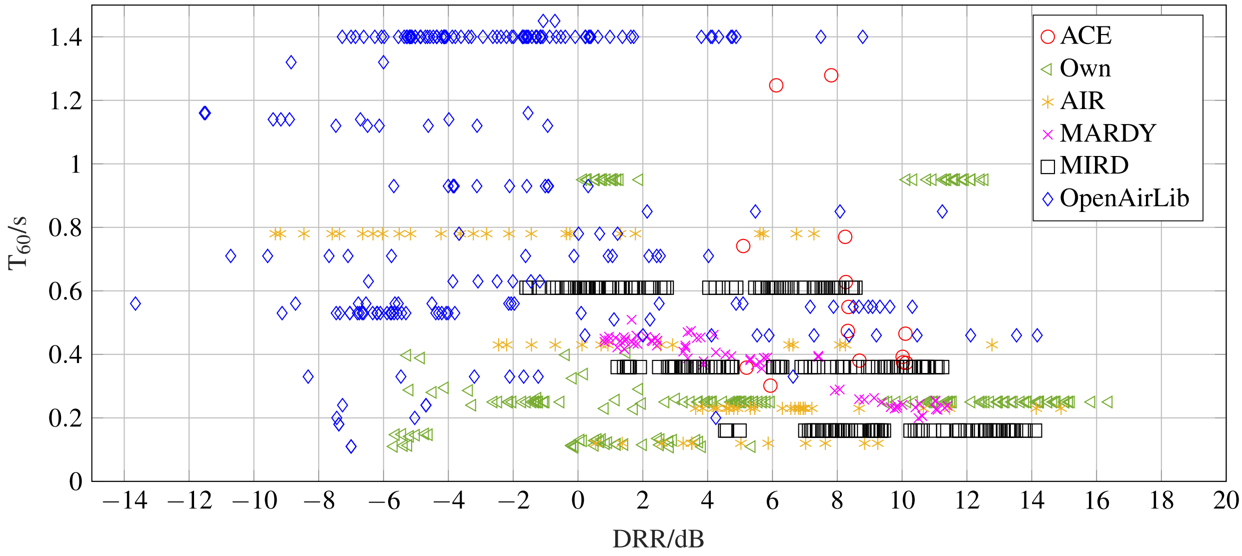


Figure 1. RTs and DRRs of the RIRs used for the large evaluation dataset (LED) and the ACE challenge data corpus (evaluation and development dataset).

Table 1. Overall estimation accuracy of various fullband RTEs.

| | ACE evaluation dataset | | | Large evaluation dataset | | |
|---|---|---|---|---|---|---|
| Algorithm | Bias | MSE | $C_p$ | Bias | MSE | $C_p$ |
| FAO [13] | -0.161 | 0.193 | 0.169 | -0.215 | 0.168 | 0.572 |
| LM RTE [24] | 0.294 | 0.216 | 0.285 | 0.220 | 0.140 | 0.635 |
| ML RTE [19] | -0.046 | 0.111 | 0.384 | 0.010 | 0.065 | 0.770 |
| OSB RTE [20] | -0.106 | 0.072 | 0.751 | 0.046 | 0.066 | 0.758 |
| SFM RTE [27] | -0.051 | 0.095 | 0.502 | -0.010 | 0.062 | 0.778 |
| jROPE II [34] | 0.124 | 0.129 | 0.661 | 0.312 | 0.374 | 0.489 |

## 4 Evaluation

As representatives for the large class of model-based approaches without training, the fast online algorithm (FAO) of [13], the Laplace distribution-based estimator (LM RTE) of [24], the ML estimation-based RTE (ML RTE) of [19] (which served as baseline algorithm in the ACE challenge) and the octave subband-based RTE (OSB RTE) of [20] are considered in this evaluation.[2] The single feature mapping-based RTEs are represented by the algorithm of [27] (SFM RTE) and the jROPE II algorithm [34, 46] represents a method for RT estimation based on multiple feature mapping. The estimation performance is assessed by the bias, MSE and Pearson correlation coefficient $C_p$ as used also for the ACE challenge [37] which also allows to compare the evaluation results with those of [37]. The Pearson correlation coefficient quantifies how closely the estimation results are correlated with the ground truth values, cf., [37].

The evaluation results for the considered estimators and datasets are listed in Table 1. The OSB RTE and SFM RTE show a high estimation performance for both datasets where the OSB RTE achieves the highest performance of all RTEs for the ACE challenge dataset w.r.t. MSE and Pearson correlation coefficient, and

[2] A MATLAB implementation of the ML RTE is available online [45] where the results for the ACE challenge [37] were created with a modified version which used a different parameter setting and included a prior noise suppression.

Table 2. Estimation accuracy for different segmental SNRs and DRRs for the LED.

| | | segmental SNR [dB] | | | | DRR | |
|---|---|---|---|---|---|---|---|
| Algorithm | Measure | -5 | 5 | 15 | $\infty$ | $<5\,$dB | $>5\,$dB |
| OSB RTE [20] | Bias | 0.188 | 0.055 | -0.010 | -0.050 | 0.017 | 0.089 |
| | MSE | 0.131 | 0.054 | 0.039 | 0.040 | 0.075 | 0.052 |
| | $C_\mathrm{p}$ | 0.606 | 0.827 | 0.889 | 0.899 | 0.770 | 0.525 |
| SFM RTE [27] | Bias | 0.096 | -0.018 | -0.028 | -0.092 | -0.047 | 0.043 |
| | MSE | 0.066 | 0.048 | 0.055 | 0.079 | 0.079 | 0.037 |
| | $C_\mathrm{p}$ | 0.799 | 0.851 | 0.831 | 0.754 | 0.772 | 0.586 |
| jROPE II [34] | Bias | 0.591 | 0.452 | 0.179 | 0.025 | 0.285 | 0.351 |
| | MSE | 0.638 | 0.502 | 0.239 | 0.119 | 0.342 | 0.422 |
| | $C_\mathrm{p}$ | 0.372 | 0.450 | 0.631 | 0.689 | 0.533 | 0.312 |

the SFM RTE achieves the highest estimation accuracy of all treated estimators for the LED. In contrast, the estimation performance of the ML RTE is less consistent for both databases as it achieves a marginally better estimation performance than the OSB RTE for the LED, but a significantly lower accuracy for the ACE challenge dataset.

Like the ML RTE, the FAO and the LM RTE are also ML-based methods, but do not account for additive noise which explains their lower estimation performance in comparison to the ML RTE that performs a noise reduction prior to the RT estimation. The jROPE II estimator achieves a rather good estimation performance for the ACE challenge dataset with the second highest Pearson correlation coefficient of all estimators, but the lowest estimation performance of all approaches for the LED. The LED contains a much larger variety of DRRs and RTs than the ACE challenge dataset such that the mismatch between training and evaluation data can become more pronounced.

More detailed results for the OSB RTE, the SMF RTE and the jROPE-II estimator (as representatives for the previously discussed three classes) for the LED are provided by Table 2. The estimation performance of the jROPE II estimator decreases for decreasing SNRs most likely due to a mismatch between the noise used for the LED and the noise used in the training, cf., [34]. A mismatch between training and evaluation dataset could also explain the lower estimation performance for higher DRRs than for lower DRRs.[3] A similar tendency can also be observed for the OSB RTE whose estimation performance decreases for low SNR values of 5 and $-5\,$dB. The prior noise suppression can only partly remove the noise, especially at low SNRs, and the remaining noise contained in the detected speech decays leads to an increased bias and estimation error since the assumed statistical model of Eq. (1) becomes less valid. This statistical model does also not model early reflections which explains the lower Pearson correlation coefficient for higher DRRs. The same can be observed for the SFM RTE which is also derived based on the statistical model of Eq. (1). However, the estimation performance of this RTE does not decrease significantly for low SNRs compared to the OSB RTE which can be attributed to the SNR-dependent RT calculation.

## 5 Conclusions

This contribution has presented a comprehensive survey on single-channel algorithms for blind RT estimation and an evaluation of selected algorithms representing different approaches to RT estimation by using the evaluation dataset of the ACE challenge and a second dataset which covers a larger range of RTs and DRRs. Most model-based approaches for RT estimation are derived by assuming a statistical model for the sound decay and require no training phase. The evaluation of four such approaches [13, 19, 20, 24] has shown that the energy

[3] The non-blind RT estimation from a measured RIR by the Schroeder method [8] is also less reliable for a high DRR as the assumption of a logarithmic energy decay is less valid.

weighted averaging of subband estimates [20] leads to a higher overall estimation performance than the RT estimation in the time-domain as done in [13, 19, 24] (at the cost of a higher computational complexity). RTEs based on single feature mapping require typically a training phase for a small set of room configurations such as [27]. This RTE and the RTE of [20] have achieved a rather high estimation performance for both considered database. However, both methods are derived by a statistical model for the sound decay and show performance degradations if this model becomes less valid due to noise or for high DRRs. The evaluation has shown that adapting the RT calculation based on the SNR in [27] achieves a higher noise robustness than a prior denoising as used in [20].

RTEs based on the mapping of multiple features to the reverberation time, e.g., by an ANN, require much more training data than single feature-based approaches, but do not rely on an explicit signal model. The evaluation of the jROPE II estimator [34] has shown that a mismatch between training and evaluation dataset can lead to a significant decrease of the estimation accuracy.

## Acknowledgment

The authors gratefully acknowledge the contributions of Apurva Kriti in evaluating the various estimators. The authors are also thankful to Feifei Xiong (University of Oldenburg) for providing the code for the jROPE II RTE [34], and James Eaton (Imperial College) for providing a MATLAB implementation for [27] and the ACE challenge data corpus. This work was partly supported by DFG under contract no. <Ke890/10-1> within the Research Unit FOR2457 "Acoustic Sensor Networks".

## 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] 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.

[5] K. Lebart, *Speech Dereverberation applied to Automatic Speech Recognition and Hearing-Aids*, Ph.D. thesis, University of Rennes, France, 1999.

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

[7] 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.

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

[9] ISO 3382:2008, “Acoustics – Measurement of the Reverberation Time of Rooms with Reference to other Acoustical Parameters,” 2008.

[10] 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.

[11] T. J. Cox, F. Li, and P. Dalington, “Extracting Room Reverberation Time From Speech Using Artificial Neural Networks,” *Journal of the Audio Engineering Society*, vol. 49, no. 4, pp. 219–230, 2001.

[12] 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.

[13] 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.

[14] 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*, vol. 93, pp. 760–770, 2007.

[15] 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.

[16] 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.

[17] T. H. Falk and W. Chan, "Temporal Dynamics for Blind Measurement of Room Acoustical Parameters," *IEEE Trans. on Instrumentation and Measurement*, vol. 59, no. 4, pp. 978–989, Apr. 2010.
[18] 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, pp. 1–4.
[19] 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, pp. 1–4.
[20] 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, pp. 1–5, available on arXiv: 1511.04063.
[21] H. W. Löllmann, A. Brendel, and W. Kellermann, "Efficient ML-Estimator for Blind Reverberation Time Estimation," in *Proc. of European Signal Processing Conference (EUSIPCO)*, Rome, Italy, Sept. 2018, pp. 2209–2213.
[22] 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.
[23] 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.
[24] T. Jan and W. Wang, "Blind Reverberation Time Estimation Based on Laplace Distribution," in *Proc. of European Signal Processing Conference (EUSIPCO)*, Bucharest, Romania, Aug. 2012, pp. 2050–2054.
[25] 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 Trans. on Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1884–1893, Aug. 2012.
[26] 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, pp. 1–4.
[27] 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.
[28] 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.
[29] F. Lim, M. R. P. Thomas, and I. J. Tashev, "Blur Kernel Estimation Approach to Blind Reverberation Time Estimation," in *Proc. of Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brisbane, Austrailia, Apr. 2015, pp. 41–45.
[30] 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.
[31] C. Schüldt 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.
[32] L. Wu, X. Qiu, I. Burnett, and Y. Guo, "Reverberation Time Estimation from Speech Signals based on Blind Room Impulse Response Identification," *Journal of the Audio Engineering Society*, vol. 138, no. 2, pp. 731–734, Aug. 2015.
[33] A. Hamza, T. Jan, A. Jehangir, W. Shah, H. Zafar, and M. Asif, "A Novel Approach for Blind Estimation of Reverberation Time using Gamma Distribution Model," *Journal of Electrical Engineering and Technology*, vol. 11, no. 2, pp. 529–536, Mar. 2016.
[34] F. Xiong, S. Goetze, B. Kollmeier, and B. T. Meyer, "Joint Estimation of Reverberation Time and Early-To-Late Reverberation Ratio From Single-Channel Speech Signals," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 27, no. 2, pp. 255–267, Feb. 2019.
[35] N. D. Gaubitch, H. W. Löllmann, M. Jeub, T. H. Falk, P. A. Naylor, P. Vary, and M. Brookes, "Performance Comparison of Algorithms for Blind Reverberation Time Estimation from Speech," in *Proc. of Intl. Workshop on Acoustic Signal Enhancement (IWAENC)*, Aachen, Germany, Sept. 2012, pp. 1–4.
[36] 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, pp. 1–5.
[37] 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.
[38] C. Veaux, J. Yamagishi, and K. MacDonald, "English Multi-speaker Corpus for CSTR Voice Cloning Toolkit," [Online, accessed 21.09.2017] http://homepages.inf.ed.ac.uk/jyamagis/page3/page58/page58.html.
[39] ITU-T Rec. P.56, "Objective Measurement of Active Speech Level," Mar. 1993.
[40] M. Brookes, "VOICEBOX: Speech Processing Toolbox for MATLAB," [Online, accessed 04.03.2019] http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html.
[41] M. Jeub, M. Schäfer, and P. Vary, "A Binaural Room Impulse Response Database for the Evaluation of Dereverberation Algorithms," in *Proc of. Intl. Conference on Digital Signal Processing (DSP)*, Santorini, Greece, July 2009, pp. 1–5.
[42] J. Y. C. Wen, N. D. Gaubitch, E. A. P. Habets, T. Myatt, and P. A. Naylor, "Evaluation of Speech Dereverberation Algorithms Using the MARDY Database," in *Proc. of Intl. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Paris, France, Sept. 2006, pp. 1–4.
[43] E. Hadad, F. Heese, P. Vary, and S. Gannot, "Multichannel Audio Database in Various Acoustic Environments," in *Proc. of Intl. Workshop on Acoustic Signal Enhancement (IWAENC)*, Antibes, France, Sept. 2014, pp. 1–5.
[44] "OpenAirLib Database," [Online, accessed 21.09.2017] http://www.openairlib.net.
[45] M. Jeub, "Blind Reverberation Time Estimation," Matlab Central [Online, published 20.03.2012] http://www.mathworks.com/matlabcentral/fileexchange.
[46] "jROPE II Version 0.1 ," [Online, accessed 13.05.2019] https://github.com/ffxiong/jrope_matlab.