NOISE-ROBUST REVERBERATION TIME ESTIMATION USING SPECTRAL DECAY DISTRIBUTIONS WITH REDUCED COMPUTATIONAL COST

James Eaton*, *Nikolay D. Gaubitch*†, *Patrick A. Naylor**

*Department of Electrical and Electronic Engineering, Imperial College London, UK
†Signal and Information Processing Lab, Delft University of Technology, Netherlands
{j.eaton11, p.naylor}@imperial.ac.uk, N.D.Gaubitch@tudelft.nl

ABSTRACT

Reverberation Time (T_{60}) is an important measure of the acoustic properties of a room. It can provide information about the acoustic environment, the intelligibility, and quality of speech recorded in the room, and help improve the performance of speech processing algorithms with reverberant speech. Where the acoustic impulse response of the room is not available, the T_{60} must be estimated non-intrusively from reverberant speech. State-of-the-art non-intrusive T_{60} estimators have been shown to be strongly biased in the presence of noise. We describe a novel T_{60} estimation algorithm based on spectral decay distributions that provides robustness to additive noise for a range of realistic noise types for signal-to-noise ratios in the range 0 to 35 dB and T_{60}s between 200 and 950 ms. The proposed method also has much reduced computational cost.

Index Terms— speech enhancement, SNR, reverberation time

1. INTRODUCTION

A speech signal, $x(n)$ produced at a given position in a room will follow multiple paths to any observation point comprising the direct path as well as reflections from walls and other surfaces in the room. The reverberant signal, $y(n)$, captured by a microphone in the room is characterised by the Acoustic Impulse Response (AIR), $h(n)$, of the acoustic channel between the source and microphone, such that

$$y(n) = x(n) * h(n) + v(n) \tag{1}$$

where $v(n)$ is additive noise at the microphone. The AIR is a function of the room geometry, the reflectivity of the walls and other surfaces, the location of the microphone, and the distance from the microphone to the source.

Reverberation Time (T_{60}) is defined as the time taken for a sound to decay by 60 dB after the source has abruptly ceased. It can provide important information about the acoustic environment, the intelligibility and quality of speech recorded in the room, and can be used to improve the performance of speech processing algorithms with reverberant speech such as speech recognition [1] and de-reverberation [2, 3, 4, 5, 6]. T_{60} can be characterized by the Sabine or Eyring equations [5, 7], and in contrast to the AIR, T_{60} measured in the diffuse sound field is independent of the source to microphone configuration. Standardized methods exist for estimating T_{60} from a measured AIR [8] such as [9]. In many practical situations, the AIR is not available and so T_{60} must be estimated non-intrusively from reverberant speech. Existing algorithms for estimation of T_{60} non-intrusively include [4, 10, 11, 12, 13, 14]. However, all of these methods except [13] have been shown to give strongly biased estimates of T_{60} in the presence of high levels of additive noise [15], whilst [13] was shown to be robust to noise but has a large variance with respect to different speakers.

The key contributions of this paper are: to propose a T_{60} estimation method employing Spectral Decay Distributions (SDD) which is substantially robust to additive noise thus avoiding problems of bias in existing estimators; to show how the cost of computing the SDD can be reduced; and to show a comparison of the improved method's results with previous research. The baseline method selected for comparison is the algorithm by Wen *et al.* [11] because it performed well in noise-free conditions, and showed little variance with different speakers in the detailed benchmarking [15]. The remainder of this paper is organized as follows: In Section 2 we review the baseline T_{60} estimation method. In Section 3 we discuss the proposed approach to reducing the computational cost of the SDD calculation providing robustness to additive noise. In Section 4 we discuss the experimental approach to testing the proposed algorithm and the results, and presented in Section 5 are the conclusions.

The **relationship to prior work** is presented throughout the paper: in Section 1 we discuss the baseline SDD algorithm for comparison. In Section 2 we review the baseline algorithm. In Section 4 we compare the proposed algorithm with the baseline algorithm , and in the conclusion we summarize the main improvements.

2. REVIEW OF SDD T_{60} ESTIMATION

2.1. Room decay model

Room reverberation consists of direct sound, early reflections and late reverberation. The fine structure of late reverberation is typically modelled statistically whilst the decaying envelope of the AIR can be modelled as a deterministic signal parameterized by some damping constant δ [16, 17]. Polack developed a time-domain model that describes the AIR as one realisation of a non-stationary stochastic process [17] described by

$$h(t) = b(t)e^{-\delta t} \quad \text{for} \quad t \geq 0, \tag{2}$$

where $b(t)$ is zero-mean stationary Gaussian noise, and damping constant, δ is related to the reverberation time, T_{60} by

$$\delta = 3\ \log 10/T_{60} \quad \text{or} \quad T_{60} = 3\ \log 10/\delta. \tag{3}$$

The relation between the damping constant δ and the T_{60} is only valid when the sound field in the enclosure is diffuse and the source-microphone distance is greater than the critical distance [16]. The room decay model can be defined using (2) as

$$E\{h^2(t)\} = \sigma_b^2 e^{-2\delta t} = \sigma_b^2 e^{\lambda_h t}, \tag{4}$$

where σ_b^2 denotes the variance of $b(t)$, and the decay rate, $\lambda_h = -2\delta$. The room decay can be extended for frequency dependent decay rates by rewriting (2) as the frequency dependent room decay model:

$$\bar{H}(t,f) = P(f)e^{\lambda_h(f)t} \quad \text{for} \quad t \geq 0 \tag{5}$$

where $\bar{H}(t,f)$ is the energy envelope of AIR at time t and frequency f, $\lambda_h(f)$ is the decay rate at frequency f, and $P(f)$ is the initial Power Spectral Density (PSD) of the noisy reverberant speech signal. Equation (5) can be linearized by taking the natural logarithm:

$$\log \bar{H}(t,f) = \log P(f) + \lambda_h(f)t \quad \text{for} \quad t \geq 0. \tag{6}$$

The decay rate $\lambda_h(f)$ can therefore be estimated by applying a linear fit to the natural logarithm of the time-frequency energy envelope of the reverberant speech signal.

2.2. SDD Method

SDD can be used as in [11] to provide a method of estimating T_{60} by observing the energy envelope of a reverberant speech signal. Frequency dependent decay rates are estimated for each analysis time frame by applying a least-squares linear fit to the log-energy envelope of the signal in each frequency band in the Short Time Fourier Transform (STFT) domain of reverberant speech. As shown in [11] the Negative-Side Variance (NSV), defined as the variance of the negative gradients in the distribution of the decay rates, correlates well with the room decay rate and by using a polynomial mapping function can be used as an estimator for T_{60}. The mapping function must be trained on a suitable clean speech database that has been convolved with AIRs of known T_{60}. The NSV denoted by σ_{x-}^2 is defined as the variance of a symmetrical distribution ($f_x^-(\lambda)$) with the same negative-side distribution of the original distribution ($f_x(\lambda)$) as in

$$f_x^-(\lambda) = \begin{cases} f_x(\lambda) & \text{for} \quad \lambda \leq 0 \\ f_x(-\lambda) & \text{for} \quad \lambda > 0. \end{cases} \tag{7}$$

Let $\lambda(k,l)$ equal the estimated decay rate for frequency band k and time frame l in a signal containing K frequency bands and L frames, and

$$\lambda'(k,l) = \quad \lambda(k,l) \quad \text{for} \quad \lambda(k,l) < 0 \tag{8}$$

where only negative λ are relevant to decays, then in the baseline approach of [11] the NSV is calculated as

$$\sigma_{x-}^2 = \frac{1}{KL}\sum_{k=1}^{K}\sum_{l=1}^{L}(\lambda'(k,l))^2. \tag{9}$$

3. T_{60} ESTIMATION ROBUST TO NOISE USING SDD WITH REDUCED COMPUTATIONAL COST

3.1. SDD with reduced computational cost

It was shown in [15] that the method of [11] has a high computational cost relative to other estimators such as [4, 13] because it operates in the STFT domain with many frequency bands. Our proposed method follows a perceptually motivated frequency analysis and therefore employs a filter bank with uniformly spaced filters on the Mel frequency scale [18]. The number of Mel-spaced bands is much less than the number of STFT bands so that the computational complexity of the least squares fitting procedure in our algorithm is correspondingly reduced compared to the baseline SDD. Additionally, since the Mel-spaced bands are formed by weighted averaging of STFT bins, this also gives some reduced sensitivity to noise. It can be seen from the Bienaymé formula [19] for uncorrelated random variables

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i), \tag{10}$$

for a population with a small variance (i.e. the noise) and a population with a large variance (i.e. the speech) that in the limit where the ratio of the variances approaches infinity, the variance can be assumed to be the variance of the speech. We will refer to this proposed algorithm as SDD with Mel-spaced frequency bands (SDDMSB).

3.2. SDD T_{60} estimation robust to noise

To reduce further the effect of noise on the T_{60} estimation variance, we will invoke two additional concepts in our algorithm for computing a representative decay rate for the signal given $\lambda(k,l)$, the raw decay rates estimated at each frequency band in each time frame: i) The first concept is to select time-frequency bins where there is more likelihood of speech being present based on their decay rates. ii) The second concept is switching of the selection method depending on the *a posteriori* Signal-to-Noise Ratio (SNR) estimate. To illustrate this approach, consider time-frequency bins containing frames of the log magnitude speech spectrum averaged over q Mel-spaced frequency bands near the fundamental frequency of the speech, typically between 85 and 255 Hz [20, 21]. The most negative gradients of these frames determined using least squares fitting for (6) will follow the decay of the speech with the reverberation tail. Now consider similar time-frequency bins containing noise that is uncorrelated with the speech and independent. The gradients of the frames containing noise will tend to zero. In the case of noisy reverberant speech, T_{60} estimation can be made more robust to the noise by basing it on the first of the above two cases, i.e. by selecting the highest negative decay rate gradients from time-frequency bins from (10).

We next employ mode-switching of the T_{60} estimation algorithm for operation in noise with switching controlled according to the input SNR. The choice of SNR threshold values in the following definitions of the modes will be explained below in Section 4.1. Mode A: for input SNRs better than 30 dB where the energy decay values across all frequency bands are used to determine the NSV and hence the T_{60} as in SDD. Mode B: for input SNRs between 15 and 20 dB a selection is made of energy decay values representative of clean speech is obtained by selecting the most negative gradient within each time frame across all frequency bands to estimate the NSV. Mode C: for input SNRs below 10 dB where noise power may exceed the speech power in a given time-frequency bin and produce large erroneous decays, the frames most likely to contain speech are identified under these conditions and we assume that the fundamental frequency of the speaker will tend to occupy one frequency band for most of the speech signal. To obtain the NSV in mode C we therefore compute the variance of each entire frequency band for all frames, and select the largest of these variances. The NSVs for modes A, B and C respectively are defined as

$$\begin{aligned} \sigma_A^2(\lambda) &= \sigma_{x-}^2 \\ \sigma_B^2(\lambda) &= \tfrac{1}{L}\sum_{l=1}^{L}\left(\min_q(\lambda(q,l))\right)^2 \\ \sigma_C^2(\lambda) &= \max_q\left(\tfrac{1}{L}\sum_{l=1}^{L}(\lambda'(q,l))^2\right). \end{aligned} \tag{11}$$

Averaging is employed in order to provide a smooth transition between modes. The estimated T_{60} is therefore given by

$$\widehat{T_{60}} = -3\log 10 \begin{cases} \frac{1}{m(\sigma_A^2(\lambda))} & \text{for} \quad \xi > 30 \\ \frac{2}{m(\sigma_A^2(\lambda))+m(\sigma_B^2(\lambda))} & \text{for} \quad 20 \le \xi < 30 \\ \frac{1}{m(\sigma_B^2(\lambda))} & \text{for} \quad 15 \le \xi < 20 \\ \frac{2}{m(\sigma_B^2(\lambda))+m(\sigma_C^2(\lambda))} & \text{for} \quad 10 \le \xi < 15 \\ \frac{1}{m(\sigma_C^2(\lambda))} & \text{for} \quad \xi < 10 \end{cases} \tag{12}$$

where $m(\cdot)$ is a mapping function between the NSV and δ (as defined in (3)), and ξ is the SNR of the reverberant speech in decibels obtained from a noise estimator. In our tests, mode C was found experimentally to be most effective with more than 30 frequency bands in the Mel frequency filter bank. Estimation accuracy of the algorithm was found to converge within three TIMIT utterances (approximately 8 s of speech) when trained on a single utterance. Therefore to be able to process realistic length audio streams, signals were split into 8 s blocks with T_{60} estimates found from the average across all blocks. Note that the consideration of bias within this estimator is beyond the scope of this paper. We will refer to this algorithm as SDD with Mel-spaced frequency bands and selective averaging (SDDSA).

4. PERFORMANCE EVALUATION

4.1. Test and training

Speech signals $x(n)$ were randomly selected from the training and test partitions of TIMIT [22] to produce exclusive training and test datasets. These were convolved with AIRs $h(n)$ generated separately for training and testing using the source-image method [23, 24] for a room with dimensions $5 \times 4 \times 6$ m, a source-microphone distance of 2 m, and T_{60} values from 200 to 950 ms in 150 ms intervals. Training signals comprised single utterances from each of four different male and four different female speakers, whilst test signals comprised six utterances concatenated from each of four different male and four different female speakers to provide realistic tests on long sentences, and to avoid any per-speaker bias. The source-microphone distance was always greater than the critical distance as shown in Table 1 for estimated Direct-to-Reverberant Ratios (DRRs) of the impulse responses $h(n)$ computed by comparing the energy before and after 2.5 ms beyond the approximate arrival time of the direct path component [5]. Babble, White, Factory1 and Volvo noise signals $v(n)$ from NOISEX-92 [25] were added to the reverberant speech test signals to simulate realistic noisy conditions. To obtain the desired test SNR ξ, the noisy reverberant speech test signals $y(n)$ were constructed using ITU-T P.56 Method B [26]. A one-time offline training procedure was used to determine the mapping function $m(\cdot)$ for the relationship between the NSV and δ (as defined in (3)) derived using a fourth order polynomial fit using the training dataset and the oracle T_{60} without noise. SDDSA was tested with the oracle SNR as well as the SNR from two state-of-the-art noise estimators: an implementation [27] of Gerkmann [28]; and Hendriks [29], henceforth referred to as SDDSA-G and SDDSA-H respectively. In addition, for each estimator the mode switching thresholds in (12) were optimized to minimize overall T_{60} Root Mean Square (RMS) error in white noise. The risk of training the thresholds too specifically to the training data used was mitigated by reviewing the results across all noise types. To evaluate SDDSA we compare with the SDD, SDDMSB, SDDSA-G, and SDDSA-H methods in terms of RMS T_{60} estimation error as a function of either the oracle T_{60} or SNR. In addition, the estimated Real-Time Factor (RTF) for each algorithm was determined by measuring the elapsed processing time using the Matlab *cputime* function for each call and calculating the mean time per algorithm divided by the mean speech file duration. Tests were performed in Matlab on a 2.3 GHz Intel i5 Core processor with 4 GB 1.333 GHz DDR3 memory.

Table 1. *Approximate estimated DRR*

T_{60} (ms)	200	350	500	650	800	950
DRR (dB)	−0.39	−7.0	−10	−12	−14	−15

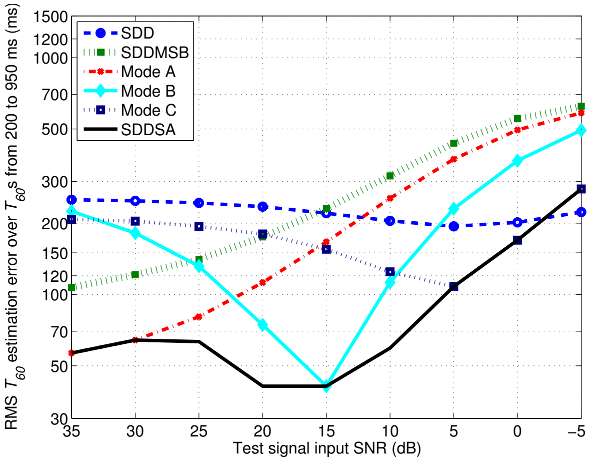

Fig. 1. *T_{60} RMS estimation error by SNR and computation method in Babble noise using TIMIT speech in T_{60}s from* 200 *to* 950 ms

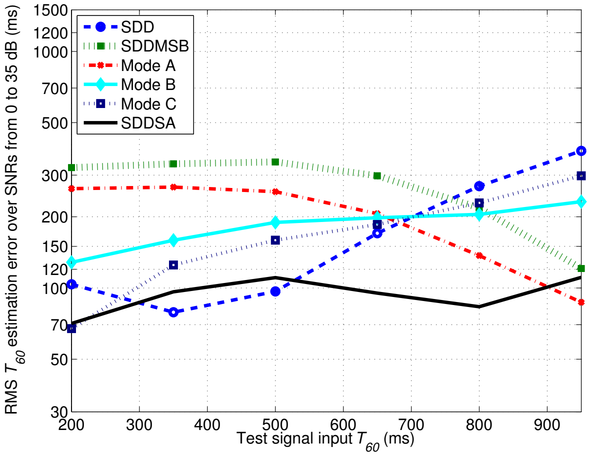

Fig. 2. *T_{60} RMS estimation error by T_{60} and computation method in babble noise using TIMIT speech in SNRs from* 0 *to* 35 dB

4.2. Results

We begin by comparing SDDSA and SDDMSB with SDD. Figs. 1 and 2 show the RMS T_{60} estimation errors for the each method in Babble noise by SNR and T_{60} respectively. RMS T_{60} estimation

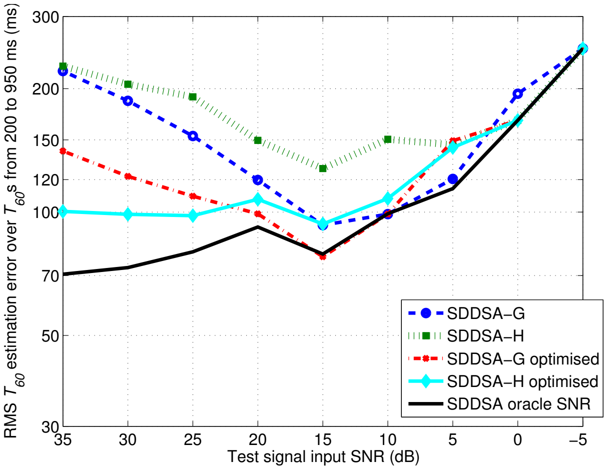

Fig. 3. *T_{60} RMS estimation error by SNR and computation method in Babble noise using TIMIT speech in T_{60}s from* 200 *to* 950 ms

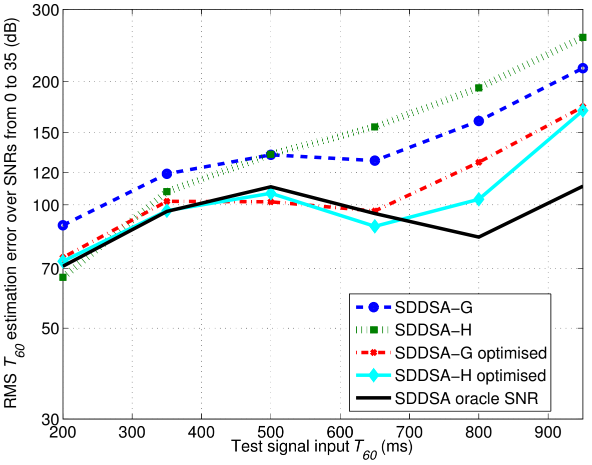

Fig. 4. *T_{60} RMS estimation error by SNR and computation method in Babble noise using TIMIT speech in SNRs from* 0 *to* 35 dB

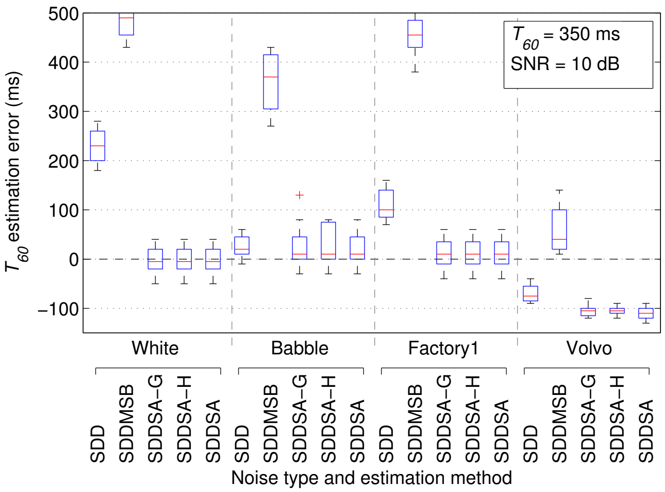

Fig. 5. *T_{60} estimation error by noise type and computation method at a specific operating point (T_{60} =* 350 ms *and SNR =* 10 dB*)*

error for SDD is typically over 200 ms rising to over 250 ms above 25 dB SNR whereas with SDDSA the RMS error is reduced to within 120 ms at 5 dB SNR and above. The justification and advantages of the switching scheme (12) can be seen in Fig. 1 which shows the T_{60} RMS estimation error for each mode as a function of SNR and that the lowest errors occur at different SNRs for each mode. Figs. 3 and 4 show the estimation error of the SDDSA, SDDSA-G and SDDSA-H, the latter two both before and after threshold optimization. The oracle case where the SNR is known *a priori* gives the lowest RMS T_{60} estimation error overall, whilst using optimized thresholds with noise estimators brings the RMS T_{60} estimation errors to within approximately 50 ms of the oracle case, with noise estimation errors having the greatest impact on the T_{60} estimate at SNRs of 20 dB and above. Noise estimation error is typically larger when the SNR is very large or very small hence the wide variation in results at 35 dB SNR shown in Fig. 3. Fig. 5 shows the variation in estimation error at an example operating point with a T_{60} of 350 ms and an SNR of 10 dB in a range of noise types illustrating the effect of different speakers on the algorithm. Errors greater than 500 ms are off-the-scale, too great to be of interest. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

Table 2 shows the estimated RTF for each method. Whilst the RTF is larger than with the SDDMSB method, a fourfold (4×) improvement over the SDD method is obtained. The higher RTF of SDDSA when compared to SDDMSB is due to the larger number of frequency bands needed to be able to differentiate the speech from the noise at low SNRs.

Table 2. *Comparison of estimated RTF*

SDD	SDDMSB	SDDSA	SDDSA-G	SDDSA-H
2.97	0.26	0.67	0.68	1.51

5. CONCLUSION

Non-intrusive estimation of reverberation time T_{60} from reverberant speech has been an important research topic for several years. It was shown in [15] that the method of Wen *et al.* [11] and two other state-of-the-art non-intrusive T_{60} estimation algorithms are strongly biased in the presence of additive noise. We have presented a novel SDD-based T_{60} estimation algorithm[1] that provides increased accuracy, reduced computational cost by a factor of four, and substantial robustness to additive noise with RMS T_{60} estimation errors of 120 ms or better for SNRs 5 dB and above over a wide range of realistic noises degrading to around 250 ms at 0 dB. We have also shown that recent noise estimators an be used instead of the oracle SNR without significantly degrading the T60 estimation accuracy.

[1] Code available on http://www.commsp.ee.ic.ac.uk/∼sap/projects/blind-estimation-of-acoustic-parameters-from-speech/blind-t60-estimator/ password:bBs96K

6. REFERENCES

[1] L. Couvreur and C. Couvreur, "On the use of artificial reverberation for ASR in highly reverberant environments," in *Proc. 2nd IEEE Benelux Signal Processing Symposium (SPS-2000)*, Hilvarenbeek, The Netherlands, Mar. 2000, IEEE, pp. S001–S004.

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

[3] E. A. P. Habets, *Single- and multi-microphone speech dereverberation using spectral enhancement*, Ph.D. Thesis, Technische Universiteit Eindhoven, June 2007.

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

[5] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*, Springer, 2010.

[6] Xiang (Shawn) Lin, Andy W. H. Khong, and Patrick A Naylor, "A forced spectral diversity algorithm for speech dereverberation in the presence of near-common zeros," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 3, pp. 888–899, Mar. 2012.

[7] C. F. Eyring, "Reverberation time in "dead" rooms," *J. Acoust. Soc. Am.*, vol. 1, no. 2A, pp. 168, 1930.

[8] ISO, "Acoustics-measurement of the reverberation time of rooms with reference to other acoustical parameters," May 2009.

[9] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am.*, vol. 37, pp. 409–412, 1965.

[10] R. Ratnam, D. L. Jones, B. C. Wheeler, W. D. O'Brien, Jr., C. R. Lansing, and A. S. Feng, "Blind estimation of reverberation time," *J. Acoust. Soc. Am.*, vol. 114, no. 5, pp. 2877–2892, Nov. 2003.

[11] 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. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, USA, Apr. 2008.

[12] T. Falk, C. Zheng, and W.-Y. Chan, "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 7, pp. 1766–1774, Sept. 2010.

[13] T. H. Falk and W.-Y. Chan, "Temporal dynamics for blind measurement of room acoustical parameters," *IEEE Trans. Instrum. Meas.*, in press.

[14] 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. Acoust. Soc. Am.*, vol. 131, no. 4, pp. 2811–2816, Apr. 2012.

[15] 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. Intl. Workshop Acoust. Signal Enhancement (IWAENC)*, Aachen, Germany, Sept. 2012.

[16] H. Kuttruff, *Room Acoustics*, Taylor & Francis, London, fourth edition, 2000.

[17] J. D. Polack, *La transmission de l'énergie sonore dans les salles*, Ph.D. thesis, Université du Maine, Le Mans, France, 1988.

[18] L. R. Rabiner and R. W. Schafer, Eds., *Theory and Applications of Digital Signal Processing*, Pearson, 2010.

[19] I.-J. Bienaymé, "Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés," *Comptes rendus des Séances de l'Académie des Sciences*, vol. 37, pp. 309–324, Aug. 1853.

[20] I. R. Titze, *Principles of Voice Production*, Prentice Hall, 1994.

[21] R. J. Baken, *Clinical Measurement of Speech and Voice*, Taylor and Francis Ltd, London, UK, 1987.

[22] J. S. Garofolo, "Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database," Technical report, National Institute of Standards and Technology (NIST), Gaithersburg, Maryland, Dec. 1988.

[23] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[24] E. A. P. Habets, "Room impulse response generator for MATLAB," `http://home.tiscali.nl/ehabets/rir_generator.html`, 2003.

[25] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition II: NOISEX-92: a database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol. 3, no. 3, pp. 247–251, July 1993.

[26] ITU-T, "Objective measurement of active speech level," Mar. 1993.

[27] D. M. Brookes, "VOICEBOX: A speech processing toolbox for MATLAB," `http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html`, 1997.

[28] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 4, pp. 1383 –1393, May 2012.

[29] R.C. Hendriks, R. Heusdens, and J. Jensen, "MMSE based noise PSD tracking with low complexity," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2010, pp. 4266–4269.