ESTIMATION OF THE REVERBERATION TIME IN NOISY ENVIRONMENTS

Heinrich W. Löllmann and Peter Vary

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

ABSTRACT

This contribution presents a novel approach to determine the reverberation time (RT) from noisy observations by using a maximum likelihood (ML) estimator. It is based on a statistical model for the sound decay in reverberant enclosures which takes additive noise into account. Two estimation methods are presented. The first one considers dedicated excitation signals such as switched-off noise sources or impulsive tone-bursts. The derived ML estimator allows to determine the RT from the received signal even in the presence of background or measurement noise and achieves a significantly higher estimation accuracy than comparable approaches.

The second approach allows to estimate the RT from reverberant and noisy speech signals without a priori knowledge. This blind estimation achieves an estimation accuracy which enable its use for speech enhancement algorithms which suppress the effects of background noise and late reverberation.

Index Terms— reverberation time, maximum likelihood estimation, blind estimation, room acoustics

1. INTRODUCTION

The reverberation time (RT) is an important and well-known measure for the characterization of reverberant enclosures. It quantifies the persistence of sound within a room that is caused by the multiple reflections of sound waves from different surfaces. The RT is defined as the time span in which the energy of a steady-state sound field decays 60 dB below its initial level after switching-off the excitation source, e.g., [1]. Knowledge about the RT is of interest, among others, for the characterization of acoustic environments, predicting the subjective preference of reverberant speech, or for the enhancement of distorted speech signals, cf. [1, 2]. Accordingly, methods for *reverberation time estimation* (RTE) are a subject of interest for acousticians and engineers alike.

The RT can be determined by measuring the sound decay after turning-off the excitation source, e.g., by means of the interrupted noise method [3]. Schroeder has developed a method to calculate the ensemble average of different decay curves from the measured room impulse response (RIR) [4]. The method of Xiang [5] estimates the RT from the sound decay under *noisy* conditions by means of nonlinear regression. The RT is calculated by an iterative procedure which relies very much on a good initial guess for the first iteration and does not necessarily converge.

For speech enhancement systems, the RT must be estimated *blindly* from a reverberant and noisy speech signal. Methods for a semi-blind RT estimation have been proposed, e.g., in [6, 7, 8]. In [6], room characteristics are 'learned' by using a neural network approach. Other semi-blind methods try to detect gaps in the speech signal to measure the sound decay using either one or two microphones [7, 8].

Algorithms for an entirely blind estimation of the RT are presented in [9, 10, 11]. However, all these proposals for a (partly) blind estimation of the RT deal not (explicitly) with the impairments due to additive noise.

In this paper, a new approach to estimate the RT from noisy measurements is developed. The devised *generalized maximum likelihood* (GML) approach allows to estimate the RT from a measured sound decay or RIR degraded by additive noise. It is also shown how the ML approach can be applied for a blind RTE where the reverberant signal is also disturbed by noise.

2. MODEL FOR THE SOUND DECAY

It is assumed that the observed sequence $y(k)$ contains the sound decay due to reverberation $h_{\mathrm{M}}(k)$ and additive noise $n(k)$:

$$y(k) = h_{\mathrm{M}}(k) + n(k) \ . \tag{1}$$

The noise sequence $n(k)$ is assumed to be uncorrelated with $h_{\mathrm{M}}(k)$ and represents i.i.d. random variables with zero mean and normal distribution $\mathcal{N}(0, \sigma_n^2)$. The sound decay is modeled by a discrete random process

$$h_{\mathrm{M}}(k) = A_{\mathrm{r}}\, v(k)\, e^{-\rho\, k\, T_{\mathrm{s}}}\, \epsilon(k) \tag{2}$$

with real amplitude $A_{\mathrm{r}} > 0$. The variable k marks the discrete sample index and $\epsilon(k)$ the unit step sequence. The parameter $T_{\mathrm{s}} = 1/f_{\mathrm{s}}$ represents the sampling period and $v(k)$ is a sequence of i.i.d. random variables with zero mean and normal distribution $\mathcal{N}(0, 1)$. Eq. (2) can also be seen as a simple statistical model for the RIR, which considers only the effects of late reflections and models them as diffuse noise. The energy decay curve for the corresponding time-continuous sound decay model reads

$$E_{\tilde{h}}(t) \doteq E\left\{\tilde{h}_{\mathrm{M}}^2(t)\right\} = A_{\mathrm{r}}^2\, e^{-2\,\rho\, t}\, \tilde{\epsilon}(t) \tag{3}$$

where the tilde indicates the time-continuous counterparts to the discrete quantities of Eq. (2). A relation between the *decay rate* ρ and the *reverberation time* T_{60} can be established by the requirement

$$10\, \log_{10}\left(\frac{E_{\tilde{h}}(0)}{E_{\tilde{h}}(T_{60})}\right) \stackrel{!}{=} 60 \tag{4}$$

which leads to the equation

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

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

Due to this relation, the terms decay rate and RT will be used interchangeably in the following.

According to our model, $y(k)$ is a random variable with the Gaussian probability density function (PDF)

$$p_{y(k)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}\,\xi(k)} \exp\left\{-\frac{x^2}{2\sigma^2\,\xi^2(k)}\right\} \tag{6}$$

$$\text{with} \quad \xi(k) = \sqrt{A_\mathrm{r}^2 \cdot a^{2\,k} \cdot \epsilon(k) + \sigma_n^2} \quad \text{and} \quad a = e^{-T_\mathrm{s}\,\rho} \ . \tag{7}$$

Hence, the sequence $y(k)$ for $k \in \{0, \ldots, N-1\}$ consists of N independent random variables with zero mean and *non*-identical PDFs having normal distributions $\mathcal{N}(0, \xi^2(k) \cdot \sigma^2)$.

3. MAXIMUM LIKELIHOOD ESTIMATION

The model introduced Sec. 2 enables the use of a maximum likelihood (ML) estimator for the RT, cf. [12]. The likelihood function (joint PDF) for an observed sequence of N (noisy) samples $y(k)$ with $k \in \{0, \ldots, N-1\}$ is derived from Eq. (6):

$$L_f\left(y, \xi, \sigma\right) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}} \prod\limits_{i=0}^{N-1} \xi(i)} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=0}^{N-1} \frac{y^2(i)}{\xi^2(i)}\right\} \ . \tag{8}$$

The *log-likelihood function* (LLF) can now be stated as

$$\begin{aligned} \mathcal{L}\left(y, \xi, \sigma\right) &= \ln\left(L_f\left(y, \xi, \sigma\right)\right) \\ &= -\frac{N}{2} \ln(2\pi\sigma^2) - \sum_{i=0}^{N-1} \ln\left(\xi(i)\right) - \frac{1}{2\sigma^2} \sum_{i=0}^{N-1} \frac{y^2(i)}{\xi^2(i)} \end{aligned} \tag{9}$$

with $\ln(\cdot)$ representing the natural logarithm. The unknown *damping factor* a (and thus T_{60}) can be estimated by the maximum of the LLF

$$\hat{a} = \arg_a \left\{ \underset{A_\mathrm{r}, a, \sigma, \sigma_n}{\text{maximum}} \left\{\mathcal{L}\left(y, \xi, \sigma\right)\right\} \right\} \ . \tag{10}$$

In the following, the dependence of the LLF from the variables (y, ξ, σ) is omitted to simplify the notation.

A solution for Eq. (10) is obtained by setting the partial derivatives of the LLF towards the unknown variables equal to zero

$$\frac{\partial \mathcal{L}}{\partial a}\Big|_{a\,=\,\hat{a}} = -\sum_{i=0}^{N-1} \frac{A_\mathrm{r}^2\, i\, \hat{a}^{2\,i-1}}{\hat{\xi}^2(i)} \left(\frac{y^2(i)}{\sigma^2\,\hat{\xi}^2(i)} - 1\right) \overset{!}{=} 0 \tag{11a}$$

$$\frac{\partial \mathcal{L}}{\partial A_\mathrm{r}}\Big|_{A_\mathrm{r}\,=\,\hat{A}_\mathrm{r}} = -\sum_{i=0}^{N-1} \frac{\hat{A}_\mathrm{r}^2\, a^{2\,i}}{\hat{\xi}^2(i)} \left(\frac{y^2(i)}{\sigma^2\,\hat{\xi}^2(i)} - 1\right) \overset{!}{=} 0 \tag{11b}$$

$$\frac{\partial \mathcal{L}}{\partial \sigma}\Big|_{\sigma\,=\,\hat{\sigma}} = -\frac{N}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^3} \sum_{i=0}^{N-1} \frac{y^2(i)}{\xi^2(i)} \overset{!}{=} 0 \ . \tag{11c}$$

The variance σ_n^2 is assumed to be known as it can usually be estimated by the noise floor following the sound decay. Eq. (11c) can be solved for the variance

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=0}^{N-1} \frac{y^2(i)}{\xi^2(i)} \ . \tag{12}$$

Thus, only the two unknown variables $\hat{a}$ and $\hat{A}_\mathrm{r}$ remain in Eq. (11). The sufficient condition for a maximum that the second derivatives of the LLF w.r.t. $\hat{a}$ and $\hat{A}_\mathrm{r}$ are less than zero can not be proven analytically for all values and must hence be verified numerically.

Inserting Eq. (12) into Eq. (9) yields the new LLF

$$\mathcal{L} = -\frac{N}{2} \left(\ln\left(\frac{2\pi}{N} \sum_{i=0}^{N-1} \frac{y^2(i)}{\xi^2(i)}\right) + 1\right) - \sum_{i=0}^{N-1} \ln\left(\xi(i)\right) \ . \tag{13}$$

Thus, the estimation of the RT T_{60} or the damping factor a, respectively, can be done by either calculating the zeros of the score functions given by Eq. (11) or finding the maximum of Eq. (13).

The exact determination of the parameters $\hat{a}$ and $\hat{A}_\mathrm{r}$ requires a high algorithmic complexity since there exists no simple closed-form solution. We solve this problem by an iterative procedure based on an *expectation-maximization* (EM) approach: In an initial step ($j = 0$), a guess for the amplitude $\hat{A}_\mathrm{r}^{(0)}$ is made. In iteration step j, Eq. (11a) is solved for $\hat{a}^{(j)}$ with a fixed value $\hat{A}_\mathrm{r}^{(j-1)}$. Afterwards, Eq. (11b) is solved with the obtained value $\hat{a}^{(j)}$ to gain the new estimate $\hat{A}_\mathrm{r}^{(j)}$. This iterative procedure is aborted when no further improvements are achieved or a maximum number of iterations is reached. Instead of Eq. (11), it is also possible to use the LLF of Eq. (13) for this iterative procedure. The devised EM procedure is suboptimal in comparison to an exact solution of Eq. (11), but it provides good estimation results in practice as shown later.

If the interfering noise is not too strong, the value for A_r can also be estimated by taking the mean

$$\hat{A}_\mathrm{r} = \sqrt{\frac{1}{L} \sum_{i=0}^{L} y^2(i)} \ . \tag{14}$$

The value for L should cover a period of about 20 ms or less so that the sound decay has no significant influence. By this, the RT can be calculated directly by Eq. (11a), termed as *non-iterative* GML RTE.

An important special case is given, if no additive noise is assumed to be present, i.e., $\sigma_n = 0$. Using the identity $\sum_{i=0}^{N-1} i = N\,(N-1)/2$, it is straightforward to show that the LLF of Eq. (13) simplifies to the expression

$$\mathcal{L} = -\frac{N}{2} \left((N-1)\,\ln(a) + \ln\left(\frac{2\pi}{N} \sum_{i=0}^{N-1} a^{-2\,i}\, y^2(i)\right) + 1\right) \ . \tag{15}$$

Accordingly, the problem of Eq. (11) reduces to the zero search

$$-\frac{N\,(N-1)}{2\,\hat{a}} + \frac{1}{\hat{a}\,\hat{\sigma}^2} \sum_{i=0}^{N-1} y^2(i)\, i\, \hat{a}^{-2\,i} \overset{!}{=} 0 \tag{16a}$$

$$\text{with} \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{i=0}^{N-1} y^2(i)\, \hat{a}^{-2\,i} \tag{16b}$$

where $\partial^2\mathcal{L}/\partial a^2 < 0\,\forall\, a$. Thus, the parameter $\hat{A}_\mathrm{r}$ drops out and only the parameter $\hat{a}$ needs to be determined. In this case, the new generalized ML estimator simplifies to the ML estimator of [9].

It is important to notice that the devised algorithm can also be used to estimate a frequency dependent RT in the subband domain. In this case, the GML estimator is applied individually to the non-subsampled output signals of an analysis filter-bank.

4. BLIND ESTIMATION

The ML estimation can also be used for a blind RTE from a noisy and reverberant speech signals. It turned out that the direct estimation of the RT from a noisy and reverberant signal is difficult to perform. Instead, it is feasible to denoise the degraded speech signal first. This can be achieved by common speech enhancement techniques such as spectral subtraction or Wiener filtering, cf. [13].

Afterwards, a blind RTE is performed by ML estimation and order-statistics filtering similar to the approach of [9]. The ML estimation of Eq. (15) or Eq. (16) is performed at intervals of R sample instances to a frame $y(\lambda R - N + 1 + i)$ with $\lambda = \lfloor k/R \rfloor$ and $i = 0, 1, \ldots, N-1$. A correct RT estimate can be obtained, if the current segment captures a free decay period following the (sharp) offset of a speech sound. Otherwise, an incorrect RT is obtained, e.g., for segments with ongoing speech, speech onsets or gradually declining speech offsets. Such estimates can be expected to overestimate the RT since the damping of sound cannot occur at a rate faster than the free decay. However, taking the minimum of the last K_1 ML estimates is likely to underestimate the RT since the estimation procedure is a stochastic process. A more robust strategy is to built the histogram of the last K_1 ML estimates and to take the first local maximum as RT $\widehat{T}_{60}^{(\text{peak})}(\lambda)$, termed as order-statistics filtering. The effects of outliers are efficiently reduced by recursive smoothing

$$\widehat{T}_{60}(\lambda) = \beta \widehat{T}_{60}(\lambda - 1) + (1-\beta)\widehat{T}_{60}^{(\text{peak})}(\lambda);\ 0.9 < \beta < 1. \quad (17)$$

This blind ML RTE differs significantly from the recent proposal of [11], which establishes a relation between the negative-side variance of the reverberant speech decay rate distribution and the true decay rate. This relation is established by a second order mapping function whose parameters are determined by a calibration procedure (which is not described in detail).

The blind ML algorithm exploits the fact that the observed signal contains occasionally small pauses of some hundred milliseconds, which is always fulfilled for speech signals. In contrast to the algorithm of [10], it is also possible to estimate larger RTs ($T_{60} > 0.6$ s).

5. SIMULATION RESULTS

5.1. Estimation with Excitation Signals

In a first experiment, a sound decay is generated by convolving a limited white Gaussian noise sequence ($\sigma_w = 0.45$) followed by zeros with a RIR for $f_s = 16$ kHz. The RIR has been measured (with $f_s = 44.1$ kHz) in a public building by a system described in [14]. The emulated sound decay is disturbed by additive, white Gaussian noise ($\sigma_n = 0.02$). The obtained signals are plotted in Fig. 1, and the used RIR is shown in Fig. 2-a.[1]

For reference, the RT has been determined by the (modified) Schroeder method [4]. The logarithm of the Schroeder integral is approximated by a linear function $f_1(t)$ according to

$$\tilde{I}_S(t) = 10 \log_{10}\left(\int_t^{\infty} \tilde{y}^2(\tau)\,\mathrm{d}\tau\right) \approx f_1(t) = b \cdot t + c \qquad (18)$$

so that the RT is given by $\widehat{T}_{60} = 60/b[\text{s}]$. The parameters b and c are determined by a least-squares fit for a chosen interval $t_0 \le t \le t_1$ using the MATLAB function `polyfit`. Fig. 1 and Fig. 2 show the normalized Schroeder integrals (summations) $\bar{I}_S(t) = I_S(t) - I_S(0)$.

[1] All discrete sequences are plotted over time for the sake of clearness.

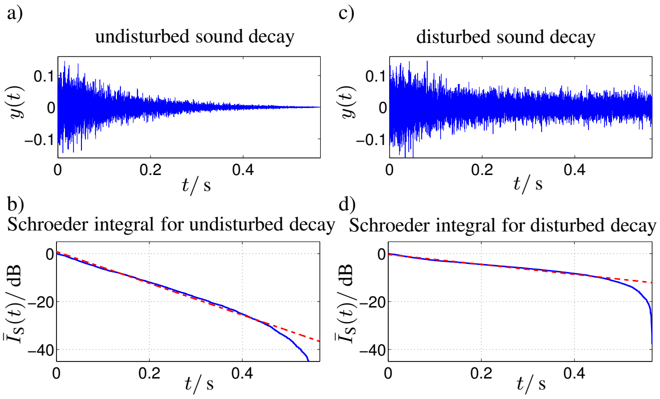

Fig. 1. Sound decay obtained by convolving a limited noise sequence with a RIR, distorted by additive white Gaussian noise ($\sigma_n = 0.02$, $f_s = 16$ kHz). The lower plots show the normalized Schroeder integrals with linear regression curves (dashed red lines) determined for $t_0 = 0$ s and $t_1 = 0.5$ s.

In addition to the Schroeder method, the RTE is also conducted by the ML estimation of Ratnam et al. [9] and the new GML approaches. The results are listed in the middle columns of Table 1.

approach	sound decay noiseless	sound decay noisy	RIR noiseless	RIR noisy
Schroeder method [4]	0.91 s	2.90 s	0.97 s	3.20 s
ML RTE [9]	0.93 s	2.00 s	1.01 s	2.00 s
EM GML RTE	0.93 s	1.04 s	1.01 s	1.03 s
non-iterative GML RTE	0.93 s	1.03 s	1.01 s	1.07 s

Table 1. Reverberation times determined by different reverberation time estimation (RTE) methods from input signals $y(k)$ shown in Fig. 1 and Fig. 2.

If no noise is present, the results of all ML approaches are identical since the new GML algorithm is equivalent to the ML RTE in this case according to Eq. (15) or (16), respectively. For noisy observations, however, Schroeder method and ML RTE yield insufficient results, where the new GML approaches achieve still an estimation accuracy of about 10%.[2] For this example, the GML with 4 EM iterations and an initial value of $\hat{A}_r^{(0)} = 0.0005$ yields almost the same results as the non-iterative GML approach.

In a second experiment, the RT is estimated from a measured RIR, which can be seen as ideal response to an impulsive tone-burst covering a broad range of frequencies. The used RIRs with and without additive noise are shown in Fig. 2 along with their Schroeder integrals. The estimated RTs are compiled in the right columns of Table 1. Again, the new GML approaches achieve a good estimation accuracy for all cases, whereas Schroeder method and ML RTE show strong deviations form the true value for noisy observations. The non-iterative GML estimator achieves almost the same result as the EM GML approach (with 4 iterations and $\hat{A}_r^{(0)} = 0.0005$).

5.2. Blind Estimation

The blind RTE devised in Sec. 4 has been applied to a distorted speech signal as shown in Fig. 3. The speech signal is first convolved with the RIR plotted in Fig. 2-a and then distorted by adding babble noise taken from the NOISEX-92 database (see Fig. 3-a). The denoising has been performed by the spectral subtraction rule based on

[2] An upper limit of $\widehat{T}_{60} \le 2.0$ s has been taken for all ML approaches.

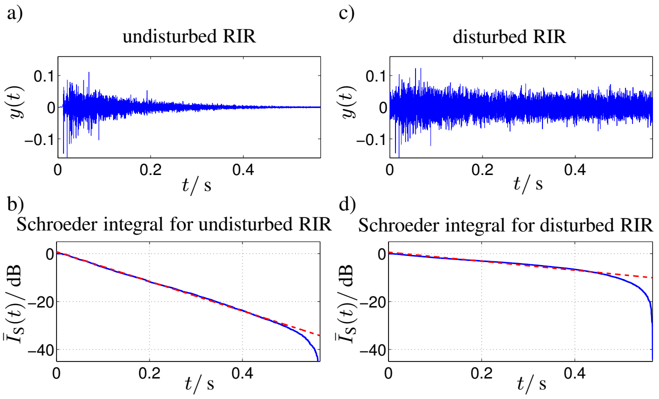

Fig. 2. Measured RIR distorted by white Gaussian noise ($\sigma_n = 0.02$, $f_s = 16$ kHz). The lower plots show the normalized Schroeder integrals with linear regression curves (dashed red lines) for $t_0 = 0.05$ s and $t_1 = 0.5$ s.

a noise power estimation by minimum statistics, cf. [13] (see Fig. 3-b). The histogram for the blind RTE is determined by the 400 most recent ML estimates for a bin size (resolution) of 0.11 s. For the ML estimation, a time span of 0.19 s and a frame shift of 0.025 s are taken. A factor of $\beta = 0.995$ is used for Eq. (17).

Fig. 3-c shows that the devised blind RTE achieves an error of less than ± 0.2 s. Such an estimation accuracy is usually sufficient for speech enhancement algorithms which aim at a joint suppression of background noise and late reverberation, cf. [2].

6. CONCLUSIONS

A generalized maximum likelihood (GML) approach to estimate the RT from noisy measurements is proposed. It is derived from a statistical model for the sound decay in reverberant rooms which takes impairments due to additive noise into account. The new approach allows to estimate the RT from a measured sound decay or room impulse response distorted by additive background or measurement noise. The needed noise power estimate can be easily obtained from the observed sequence. The other model parameters (damping factor a and amplitude A_r) can be calculated by an iterative EM approach. If the interfering noise is not too strong, the amplitude A_r can also be estimated directly from the observed sequence so that the RT can be determined without further iterations.

It is also shown how the ML estimation can be used for a blind estimation of the RT from a reverberant and noisy speech signal. After applying a conventional noise reduction system, the RT is estimated by means of a continuous ML estimation followed by order-statistics filtering to select the most likely RT estimate. This new blind RT estimator can achieve an accuracy of less than ± 0.2 s, which makes this approach of interest for speech enhancement applications.

Acknowledgment - Thanks to Ingo Witew of the Institute of Technical Acoustics of RWTH Aachen University for providing the measured RIRs.

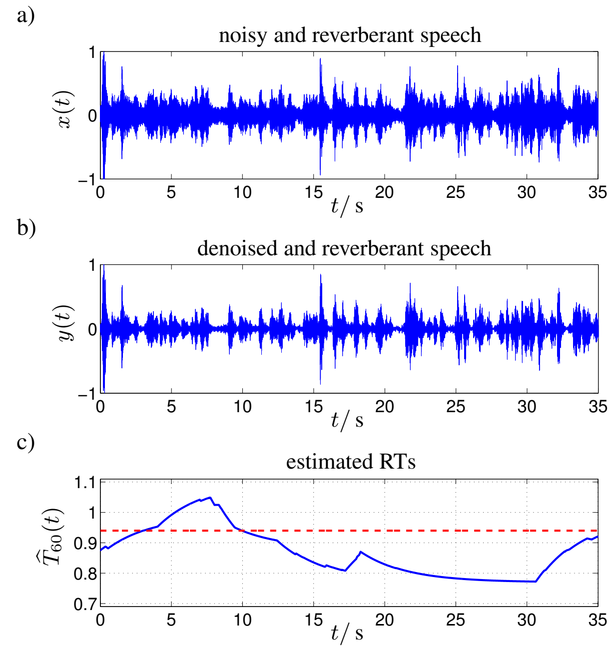

Fig. 3. Adaptive, blind estimation of the RT from a reverberant speech signal ($T_{60} = 0.94$ s) distorted by babble noise (SNR = 5 dB, $f_s = 16$ kHz).

7. REFERENCES

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

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

[3] ISO-3382, "Acoustics-Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters," International Organization for Standardization, Geneva, Switzerland, 1997.

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

[5] N. Xiang, "Evaluation of Reverberation Times Using a Nonlinear Regression Approach," *Journal of the Acoustical Society of America*, vol. 98, no. 4, pp. 2112–2121, Oct. 1995.

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

[7] K. Lebart, J. Boucher, and P. N. Denbigh, "A New Method Based on Spectral Subtraction for Speech Dereverberation," *Acta Acoustica*, vol. 87, no. 3, pp. 359–366, 2001.

[8] S. Vesa and A. Härmä, "Automatic Estimation of Reverberation Time From Binaural Signals," in *Proc. of Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia (Pennsylvania), USA, Mar. 2005, vol. 3, pp. 281–284.

[9] 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.

[10] 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.

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

[12] A. Papoulis and S. U. Pillai, *Probabilities, Random Variables and Stochastic Processes*, McGraw-Hill, New York, 4th edition, 2002.

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

[14] I. Witew and G. K. Behler, "Low-Mid Frequency Measurement of Single Number Parameters in Room Acoustics with Multi-Way Dodecahedron Speakers," in *Proc. of Inter-Noise*, Rio de Janeiro, Brazil, Aug. 2005.