A Novel Approach for Blind Estimation of Reverberation Time using Rayleigh Distribution Model

AMAD HAMZA*, TARIQULLAH JAN*, AND AMJAD ALI**

RECEIVED ON 09.04.2015 ACCEPTED ON 14.12.2015

ABSTRACT

In this paper a blind estimation approach is proposed which directly utilizes the reverberant signal for estimating the RT (Reverberation Time).For estimation a very well-known method is used; MLE (Maximum Likelihood Estimation). Distribution of the decay rate is the core of the proposed method and can be achieved from the analysis of decay curve of the energy of the sound or from enclosure impulse response. In a pre-existing state of the art method Laplace distribution is used to model reverberation decay. The method proposed in this paper make use of the Rayleigh distribution and a spotting approach for modelling decay rate and identifying region of free decay in reverberant signal respectively. Motivation for the paper was deduced from the fact, when the reverberant speech RT falls in specific range then the signals decay rate impersonate Rayleigh distribution. On the basis of results of the experiments carried out for numerous reverberant signal it is clear that the performance and accuracy of the proposed method is better than other pre-existing methods.

Key Words: Gamma Distribution, Maximum Likelihood Estimation, Reverberation Time, Reverberant Signal Analysis.

1. INTRODUCTION

In the field of signal processing, one of the most crucial research problem is to tackle the room reverberations. The solution to this problem will yield results in ASR (Automatic Speech Recognition). RT estimated value represents the understandably and quality of speech produced in that enclosure. When a sound generating source present inside a room is turned off then a 60dB decrement in sound level is obtained in some time that time taken is called the reverberation time [1-2]. Reflected signals R(n) produced in sound propagation are the main cause of reverberations. At the microphone end the received signal x(n) is represents as

$$x(n) = h(n)\, r(n) + e(n) \tag{1}$$

In Equation (1) e(n) represents the noise inside the room and the room impulse response is represented by h(n). The issues normally associated with the unclean speech which elevates more when we use software like Siri and Cortana for speech recognitions and for hands-free systems.

The RT prediction for an enclosed environment or a room using an empirical formula was first proposed in [3] in early 20th century. The enclosure absorption capabilities

* Department of Electrical Engineering, University of Engineering & Technology, Peshawar.
** Department of Electrical Engineering, Sarhad University of Science & Information Technology, Peshawar.

and its geometry (volume and surface area) are the major ingredients of the proposed formula.

However, to use these empirical formulae the information of the absorptive characteristic and the geometry of the target enclosure were essential requirements. The extraction of these information for an enclosure is not easy task, hence a new method were required to estimate RT using on the impulse response of an enclosure.

In [4] similar approach is used in which the energy decay curves are used in measuring approaches of RT. The desire decay curve can be obtained by turning off a noise producing source after achieving its steady state. Slope of the decay curve is used in the estimation of the RT. The noise source fluctuation leads to multiple decay curves. The reliable value of the RT is achieved by averaging technique. In some other approaches the reverberant signal partitioning approach is used for its gap detection within that signal [5-6] resulting decay curves tracking. This issue is rectified in 1965 by introducing an impulse response based approach [7].This approach prevailed the field of signal processing (speech) for several decades however, on the other hand a technique was required having the capability of estimating the value of the RT of a reverberant signal directly. In other words a blind estimation approach were required which should takes only the reverberant signal as an input. Blind estimation of RT for a reverberant signal can be obtained by numerous approaches [2,8-11]. In these approaches statistical models and Maximum likelihood estimation are the base element for the optimized RT value determination. Some other methods are also developed which make use of the neural networks in order to obtain the enclosure characteristics and are called semi-blind approaches [12-14].

Ratnam, et. al. [2] modelled the characteristics of reverberation of a room by utilizing an approach which is based on the decay curve of the noise, in order to develop an algorithm for estimating the value of RT blindly. In [2] the detection of the sound decay is achieved by utilizing a repetitive process and this repetition process made the algorithm expensive in terms of computation. Ratnam, et. al. [10] increased the computational efficiency of the algorithm by making some changes in it and presented the new version.

Recently Lollmann, et. al. [8] presented a blind approach RT estimation algorithm based on the sounds statistical decay model. A method of pre selection is utilized in order to detect the possible decay in sound signal. This pre selection process made the algorithm more robust in terms of estimation and more efficient computationally. In this paper a Lollmann, et. al. [8] method based approach is proposed for estimating reverberation time blindly along with Rayleigh distribution model for decay rate modelling.

2. PROPOSED MODEL AND MLE

In this paper Rayleigh distribution based statistical model is used for modelling the energy decay of the target reverberant signal. For a specific RT range the speech signal amplitude distribution is then better modelled by Rayleigh distribution [15]. The method presented in [15] is the motivation for the proposed work. Fig. 1 derived from [15] also shows the motivation. Small enclosure such as ATM booth, Telephone booth etc. are mainly focused in this work in order to estimate the RT value for improving the quality of the reverberant speech signal. In such enclosure RT falls in the range 0-150msand therefore Rayleigh distribution would be utilized for decay curve modelling of the target speech.

Hence, Rayleigh distribution G(x,θ) accompanied by a random variable sequence is used to model the tail of the decaying reverberant sound. In G(x,θ) random variable is represented by x and variance by θ. Mathematically the Rayleigh distribution is expressed as [16]:

$$f(x;\sigma)=\frac{x * e^{-x/2\sigma^2}}{\sigma^2} \tag{2}$$

The model evolution is based on an assumption. In this assumption the tail of the decaying reverberant signal denoted by y is expressed by the expression below:

$$y(n)=x(n) * b(n) \tag{3}$$

Here b and x represents deterministic envelop and the fine structure of the random process respectively.

The y(n) is not similarly distributed and have pdf G(θ,b(n)) due to time variant x(n) although having an independent y(n).

For estimating target (reverberant) signal decay rate a limited sequence, n = 0,1,2,..., M”1, is selected. For

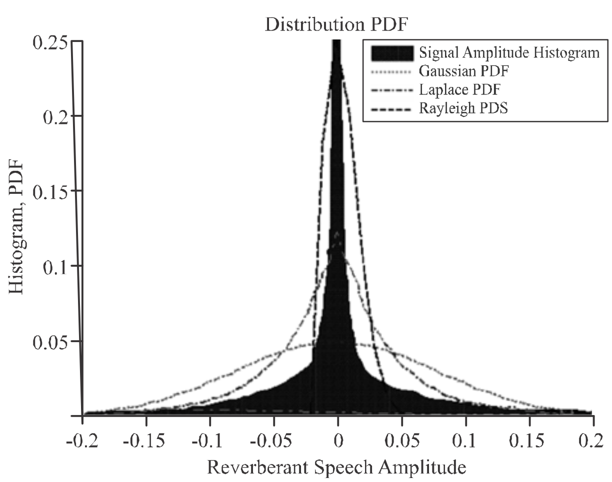

FIG. 1. NUMEROUS DISTRIBUTIONTHEORETICAL PDF (LAPLACE DISTRIBUTION BLUE LINE, GAMMA DISTRIBUTION BLACK LINE, GAUSSIAN DISTRIBUTION PINK LINE) WITH ZERO MEAN ALONG WITH DISTRIBUTION OF ECHOIC SIGNAL

convenience in terms of notation and simplicity **b** and **y** are introduced to represent the b and y vector (M-dimensional). The joint probability density or the function of likelihood for y having the parameters σ and **b** is given as,

$$L(y;b,\theta)=\frac{x(0) * x(1)....x(M-1)}{b(0) * b(1)....b(M-1)}\left(\frac{1}{\sigma^2}\right)^M \times \exp\left(-\frac{\sum_{n=0}^{M-1} y(n)/b(n)}{2\sigma^2}\right) \tag{4}$$

Where the observation of y is used to estimate the unknown values of the parameters (**b**,θ). For simplifying the derived likelihood function in Equation (4) a supposition is done. In this supposition τ (unity decay rate) is used for the damping of the target reverberant sound signal energy envelop, resulting in an expression for b(n) as given in [2].

$$b(n)=\exp\left(-\frac{n}{\tau}\right) \tag{5}$$

A single scalar argument band σ can be utilized to express the multi-dimensional argument b(n). The parameter b is given as:

$$b=\exp\left(-\frac{1}{\tau}\right) \tag{6}$$

By making use of Equations (5-6) we get:

$$b(n)=b^n \tag{7}$$

After integrating Equation (7) with Equation (4) yields:

$$L(y;b,\sigma)=\frac{x(0) * x(1)....x(M-1)}{b^0 * b^1....b^{M-1}}\left(\frac{1}{\sigma^2}\right)^M \times \exp\left(-\frac{\sum_{n=0}^{M-1} y(n)/b(n)}{2\sigma^2}\right) \tag{8}$$

Mathematical rearrangements and use of mathematical induction made Equation (8) simplify and becomes:

$$L(y;b,\sigma)=\frac{\prod_{n=0}^{M-1} y(n)/b^{n}}{\left(\sigma^{2}*b^{\frac{M-1}{2}}\right)^{M}}\times\exp\left(-\frac{\sum_{n=0}^{M-1}\frac{y(n)}{b^{n}}}{2\sigma^{2}}\right) \tag{9}$$

MLE is used for the b and σ estimation. In this estimation process function of log-likelihood is required which is obtained by taking log of Equation (9):

$$\ln L(y;b,\sigma)=\ln\left(\frac{\prod_{n=0}^{M-1}\frac{y(n)}{b^{n}}}{\left(\sigma^{2}*b^{\frac{M-1}{2}}\right)^{M}}\right)\times\ln\exp\left(-\frac{\sum_{n=0}^{M-1}\frac{y(n)}{b^{n}}}{2\sigma^{2}}\right) \tag{10}$$

Terms rearrangements, mathematical induction and logarithmic identities are used to simplify Equation (10)

$$\ln L(y;b,\sigma)=\sum_{n=0}^{M-1}\ln\left(y^{n}\right)-2\sum_{n=0}^{M-1}\ln\left(b^{n}\right)-2M\ln(\sigma)-\frac{1}{2\sigma^{2}}\sum_{n=0}^{M-1}y(n)/b^{n} \tag{11}$$

For score function [17] represented by Sf_{a} or the maximum of Equation (11) differentiation (with respect to a) is used

$$Sf_{a}L(b;y,\sigma)=\frac{\partial\ln L(y;b,\sigma)}{\partial}=-\frac{2}{b}\sum_{n=0}^{M-1}n+\frac{1}{2\sigma^{2}}\sum_{n=0}^{N-1}n*y(n)*a^{-n-1} \tag{12}$$

The extremum of Equation (12) is obtained below:

$$-\frac{2}{b}\sum_{n=0}^{N-1}n+\frac{1}{2\sigma^{2}}\sum_{n=0}^{N-1}n*y(n)*b^{-n-1}=0 \tag{13}$$

Similarly, for score function of θ same steps are followed and yield following results:

$$Sf_{\theta}L(\sigma;y,b)=\frac{\partial\ln L(y;b,\sigma)}{\partial\sigma}=-\frac{2M}{\sigma}+\frac{1}{\sigma^{3}}\sum_{n=0}^{M-1}y(n)*b^{-n} \tag{14}$$

$$-\frac{2M}{\sigma}+\frac{1}{\sigma^{3}}\sum_{n=0}^{M-1}y(n)*b^{-n}=0 \tag{15}$$

$$\Rightarrow \sigma=\sqrt{\frac{1}{2M}\sum_{n=0}^{M-1}y(n)*b^{-n}} \tag{16}$$

As Equation (16) is an explicit expression and can be solved if b is known. The value of b can be calculated from Equation (6) after estimating σ (time constant)

The observation of mapping b over σ showed that b∈[0,1] maps one-to-one onto τ[0,∞]. For estimating a, quantization is the base and also used in [10] and [2]. Quantization is applied on the given range of b in order to create bins and form its histogram. For assigning a value to a bin in histogram the maximum likelihood value is selected.

Let b∈[0,1] is quantized and K values are produced, represented by b_w, where w=1, 2, 3,...,K. Now Equation (11) is used to calculate the log likelihood for each value of the b_w

$$\ln L\left(b_{w};y\right)=\sum_{n=0}^{M-1}\ln\left(y^{n}\right)-2\sum_{n=0}^{M-1}\ln\left(b_{w}^{n}\right)-2M\ln(\sigma)-\frac{1}{2\sigma^{2}}\sum_{n=0}^{M-1}\frac{y(n)}{b_{w}^{n}} \tag{17}$$

The expression for the $b_{\overline{ML}}$ can be expressed as:

$$b_{\overline{ML}}=\max_{b}\left\{\ln L\left(b_{w}^{b}.\sigma\right)\right\} \tag{18}$$

The value is the used to find by utilizing Equation (9). Finally from [11] the formula for estimating RT (is used to calculate the reverberation time value.

$$\hat{T}_{60}^{ML}=6.908\times\tau_{\overline{ML}} \tag{19}$$

2. EFFECTIVE ESTIMATION OF RT

For enhancing the computational performance of the system those region are identified where the decaying of the sound signal is under no influence. The ML approach is then applied on the identified regions for decay rate estimation. For the sack of achieving this goal the method in [8], proposed by Lollmann, et.al. is used. We also used this approach for ML estimation improvement for the Rayleigh parameters.

The algorithm takes the speech signal (f(n)) as an input and process it frame-wise. Each frame consists of S samples and a movement factor ΔS [8]. The free decaying sound regions are obtained by this process and is given as:

$$F(\Upsilon,s) = f(\Upsilon*\Delta S+s) \text{ with } s= 0,1, ...,S"1 \tag{20}$$

where $\Upsilon \in N$. The $F(\Upsilon,s)$ is divided into $V=S/R \in N$ sub-frames for spotting the possible sound decay and is given by:

$$E(\gamma, g_{sub}, \ell_{sub}) = E(\gamma, g_{sub} R + \ell_{sub}) \tag{21}$$

where $l_{sub} = 0,1,2, ...,R-1$ and $g_{sub} = 0,1,2, ...,V-1$ symbolize the indices of thesub-frame. The energy boundary limits of the sub-frames are examined in the next step and also compared it to the next sub-frame value for divergence as done in [9].

$$\sum_{\ell_{sub}}^{R-1} = E^2(\gamma, g_{sub}, \ell_{sub}) > \partial_{g_{sub}} . \sum_{\ell_{sub}=0}^{Q-1} = E^2(\gamma, \ell_{sub}+1, \ell_{sub}) \tag{22}$$

$$\max_{\ell_{sub}} \{E(\gamma, g_{sub}, \ell_{sub})\} > \partial_{g_{sub}} .\max_{\ell_{sub}} \{E(\gamma, g_{sub}+1, \ell_{sub})\} \tag{23}$$

$$\min_{\ell_{sub}} \{E(\gamma, g_{sub}, \ell_{sub})\} < \partial_{g_{sub}} .\min_{\ell_{sub}} \{E(\gamma, g_{sub}+1, \ell_{sub})\} \tag{24}$$

Where $0 < \delta_{isub} < 1$ and represents weights. In some cases when the minimum value ($1 < g_{submin} < V-2$) of the counter (g_{sub}) is reached expression Equations (22-24) might fail in other than that case the non-equivalence check is halted and computation of next signal frame $G(\Upsilon+1,b)$ starts. On the other hand the satisfying of the above expressions by the sequence of the sub-frame might represent decaying sound region. After spotting the desired region Equations (6,17,9) are utilized to compute RT. The process of recursive smoothing is also utilized for suppressing the effect of pre-selection process on the variance of computed value of RT. This leads to a modified expression, given below:

$$\widetilde{T}_{60}^{ML}(\gamma) = \lambda.\widetilde{T}_{60}^{ML}(\gamma-1) + (1-\lambda).\widetilde{T}_{60}^{ML}(\gamma) \tag{25}$$

where 0.9 << 1 . The final resultant value for RT is obtained by:

$$\widetilde{T}_{60}^{ML} = \text{mean}\left(\widetilde{T}_{60}^{ML}(\gamma)\right) \tag{26}$$

3. RESULTS AND DISCUSSION

The proposed method is evaluated by carrying out numerous Matlab simulations. Ten dissimilar test signals (echoic signals) were generated by convolving RIR (Room Impulse Responses) from AIR database [18] and clean speech signals (5 male, 5 female) picked from TIMIT database. Dissimilar enclosure i.e. office, booth, meeting room, lecture room, IIR were selected. At two different distance $\{S_{d1},S_{d2}\}$ a source microphone were placed in eachenvironment respectively i.e. {1,3m}, {0.5, 1.5m}, {1.45, 2.8m} and {2.25,7.1m}. The technical specification of the system on which the experiments are carried out are, CPU: Intel Core i5-3317U 1.6 GHz (Laptop), RAM: 4 GB and Operating System: Windows 7 (64-bit). The other arguments values are K= 10, V= 7, $g_{submin} = 3$, $\lambda = 0.995$, S= 1631 (approx. time span of 0.10 s), R= 233, Δd= 67 (approximately frame shift of 0.0042s), and δ_{gsub} .

For each distance $\{S_{d1},S_{d2}\}$ ten test signals were generated in each environment. RT is estimated for each test signal and the final value is calculated by taking average of the ten result for each environment. Figs. 2-3 shows these results. In [7] RIR is utilized for RT value calculation and in [19] Laplace distribution model is used. For the evaluation of the algorithm the RT estimated value is compared with the results of the above mention algorithms.

Results for the shorter distance and comparatively larger distance are shown in Figs. 1-2 respectively. By observing the results it is revealed that for small distance the results of the proposed methods are closed to the results of the RIRs (0.49 and 0.429sec). Over all the results produced by the proposed method is in close agreement to the baseline (Schroeder) methods.

4. CONCLUSION

The proposed method is a novel approach for estimating RT blindly. Statistical model based on Raleigh distribution and ML scheme is utilized for representing echoic signal decay and arguments estimation respectively. Experiment exhibited that the results from the proposed arrangements are in close concession with the results acquired by the up-to-the-minute approach. The current statistical model produce good results when the RT is in lower range (0-300ms) in future we would try to modify it so that it could be used for wide range of TR values.

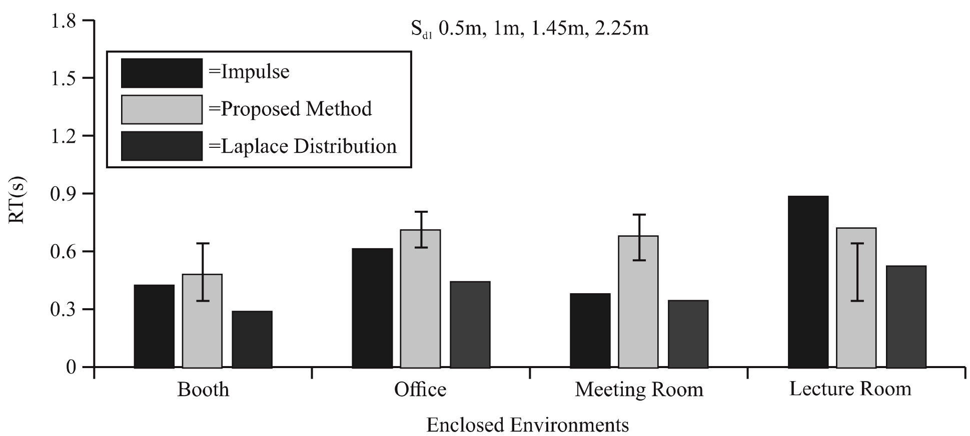

FIG. 2. BLUE BARS = BY SCHROEDER'S METHOD [7] (RIRS RESULTS),GREEN BARS =PROPOSED METHOD, BROWN BARS =LAPLACE DISTRIBUTION [19]. ORANGE LINES= STANDARD DEVIATIONS

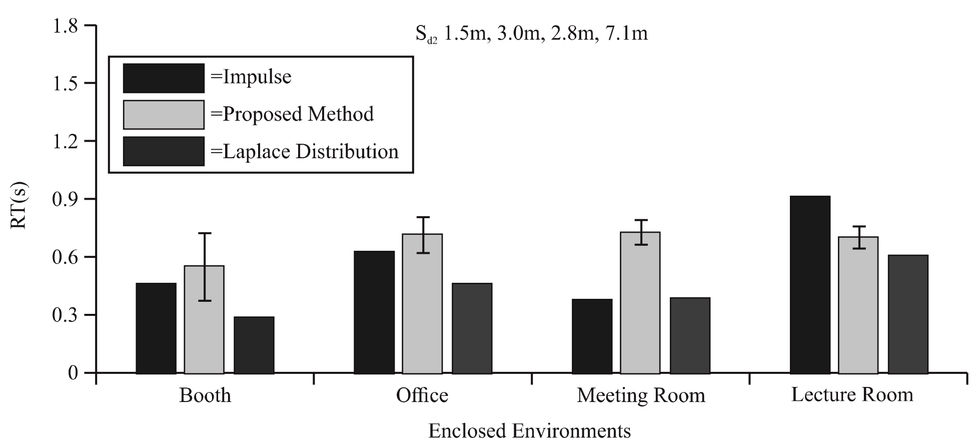

FIG. 3. BLUE BARS = BY SCHROEDER'S METHOD [7] (RIRS RESULTS), GREEN BARS =PROPOSED METHOD, BROWN BARS = LAPLACE DISTRIBUTION [19]. ORANGE LINES= STANDARD DEVIATIONS

ACKNOWLEDGMENTS

Authors would like to thank the unknown referees for their valuable comments. The constructive comments and suggestions of the anonymous referee helped greatly in improving the clarity of the manuscript, and in providing the necessary perspective for evaluating the work with respect to Schroeder's method. We would also thank the CISNR (Centre for Intelligent System and Network Research), University of Engineering & Technology, Peshawar, Pakistan, for providing technical assistance when needed.

REFERENCES

[1] Kuttruff, H, "Room Acoustics", Elsevier Science Publishers Ltd., Lindin, 3rd Edition, 1991.

[2] Ratnam, R, Jones, D.L., Wheeler, B.C., Obrien, W.D., Lansing, C.R., and Feng, A.S., "Blind Estimation of Reverberation Time", Journal of Acoustic Society of American, Volume 114, pp. 287702892, 2003.

[3] Sabine, W.C., "Collected Papers on Acoustics", Harvard University Press, 1922.

[4] ISO (International Organization for Standardization), "Acoustics-Measurements of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters", Geneva, 1997.

[5] Lebart, K., Boucher, J., and Denbigh, P., "A New Method Based on Spectral Subtraction for Speech Dereverberation", Acta Acustica, Volume 87, pp. 359-366, 2001.

[6] Vesa, S., and Harma, A., "Automatic Estimation of Reverberation Time from Binaural Signals", Proceedings of IEEE International Conference on Acoustic, Speech, Signal Processing, Volume 3, pp. 281-284, 2005.

[7] Schroeder, M.R., "New Method of Measuring Reverberation Time", Journal of Acoustic Society of American, pp. 409-412, 1965.

[8] Lollmann, H.W., Yilmaz, E., Jeub, M., and Vary, P., "An Improved Algorithm for Blind Reverberation Time Estimation", Proceedings of International Workshop Acoustic Echo and Noise Control, Tel Aviv, Israel, August, 2010.

[9] Lollmann, H.W., and Vary, P., "Estimation of the Reverberation Time in Noisy Environments", Proceedings of International Workshop on Acoustic Echo and Noise Control, Washington USA, September, 2008.

[10] Ratnam, R., Jones, D.L., and O'Brien, W.D., "Fast Algorithms for Blind Estimation of Reverberation Time", IEEE Signal Process Letters, Volume 11, pp. 537-540, June, 2004.

[11] Wen, J.Y.C., Habets, E.A.P., and Naylor, P.A., "Blind Estimation of Reverberation Time Based on the Distribution of Signal Decay Rates", Proceedings of IEEE International Conference on Acoustic, Speech, and Signal Processing, pp. 329-332, 2008.

[12] Cox, T.J., Li, F., and Darlington, P., "Extracting Room Everberation Time from Speech Using Artificial Neural Networks", FJ Audio Engineering Society, pp. 219-230, 2001.

[13] Nannariello, J., and Fricke, F., "The Prediction of Reverberation Time Using Neural Network Analysis", Applied Acoustic, Volume 58, pp. 305-325, 1999.

[14] Tahara, Y., and Miyajima, T., "A New Approach to Optimum Reverberation Time Characteristics", Applied Acoustic, Volume 54, pp. 113-129, 1998.

[15] Petsatodis, T., Boukis, C., Talantzis, F., Tan, Z., and Prasad, R., "Convex Combination of Multiple Statistical Models with Application to VAD", IEEE Transactions on Audio, Speech, and Language Processing, pp. 2314-2327, 2011.

[16] Bean, M.A., "Probability: The Science of Uncertainty with Application to Investments, Insurance and Engineering [Online], 2001. (Available: books.google.com.pk/books?isbn=0821847929)

[17] Poor, V., "An Introduction to Signal Detection and Estimation", Springer-Verlag, New York, 1994.

[18] Jeub, M., Schafer, M., and Vary, P., "A Binaural Room Impulse Response Database for the Evaluation of Dereverberation Algorithms", Proceedings of International Conference on Digital Signal Processing, Santorini, Greece, 2009.

[19] Jan, T., and Wang, W., "Blind Reverberation Time Estimation Based on Laplace Distribution", 20th EUSIPCO, 2012.