

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

Amad Hamza[†], Tariqullah Jan*, Asiya Jehangir*, Waqar Shah*,
Haseeb Zafar* and M. Asif**

Abstract – In this paper we proposed an unsupervised algorithm to estimate the reverberation time (RT) directly from the reverberant speech signal. For estimation process we use maximum likelihood estimation (MLE) which is a very well-known and state of the art method for estimation in the field of signal processing. All existing RT estimation methods are based on the decay rate distribution. The decay rate can be obtained either from the energy envelop decay curve analysis of noise source when it is switch off or from decay curve of impulse response of an enclosure. The analysis of a pre-existing method of reverberation time estimation is the foundation of the proposed method. In one of the state of the art method, the reverberation decay is modeled as a Laplacian distribution. In this paper, the proposed method models the reverberation decay as a Gamma distribution along with the unification of an effective technique for spotting free decay in reverberant speech. Maximum likelihood estimation technique is then used to estimate the RT from the free decays. The method was motivated by our observation that the RT of a reverberant signal when falls in specific range, then the decay rate of the signal follows Gamma distribution. Experiments are carried out on different reverberant speech signal to measure the accuracy of the suggested method. The experimental results reveal that the proposed method performs better and the accuracy is high in comparison to the state of the art method.

Keywords: Gamma distribution, Maximum likelihood estimation, Reverberant signal analysis, Reverberation time

1. Introduction

Reverberation is an important problem for all those engineers and scientist whom deals in the processing of audio signals and room acoustics. Reverberation time (RT) is an important parameter in the quality based categorization of an auditory enclosure. The speech quality and intelligibility observed in an enclosure is indicated by the estimated RT of that speech. After switching off a sound source present in an enclosed environment, linger time of sound is the RT of that enclosure. The time taken by the sound level to have a decrement of 60dB than its cessation value is also referred as RT [1, 2]. The main cause of reverberation is the multi-path propagation of reflected signal $b(n)$ from a source to microphone. The signal received at the microphone can be expressed as

$$x(n) = h(n)b(n) + c(n) \quad (1)$$

Where $h(n)$ is the impulse response of the enclosure and $c(n)$ is the noise present in the enclosure. Some of the major issues associated with reverberant speech is sounding distinct, echo, lower quality, intelligibility, distortion in the energy envelop and fine structure of the speech signal. This issue becomes more problematic and heightened in the unwired or hands-free systems and speech recognition software's.

Various algorithms have been developed to estimate the RT. In the beginning of the 20th century Sabine [3] proposed the first empirical formula for the prediction of RT in an enclosed environment e.g. room. The base of the formula is purely laid on the geometry (volume and surface area) of the enclosed environment and the absorption capabilities of the enclosure surfaces. To increase the accuracy of the Sabine RT formula, extensive modification has been done (for modifications details refer [1]). These RT calculating formulae is used in the designing of the places where the sound quality has as great importance like classrooms, anechoic chamber, concert halls. However, to implement these formulae the pre determination of the geometry and the absorptive characteristic of the enclosure is required. As these determinations are not easily possible so new methods is required, based only on the enclosed environment radiated test sound signal.

Such RT measuring methods uses the decay curve, like interrupted noise method [4] which used a narrow or wide

[†] Corresponding Author: Dept. of Electrical Engineering, University of Engineering & Technology, Peshawar, Pakistan. (amadhamza@gmail.com)

* Dept. of Electrical Engineering, University of Engineering & Technology, Peshawar, Pakistan. ({tariqullahjan, waqar.shsh, haseeb}@uetpeshawar.edu.pk, a_jehangir49@yahoo.com)

** Dept. of Electronics, University of Peshawar, Pakistan. (m.asif@upesh.edu.pk)

Received: April 9, 2014; Accepted: November 3, 2015

band noise burst to radiate in the enclosed environment. The decay curve is obtained by switching off the noise source after reaching its steady state. In RT estimation the decay curve slope is used. The fluctuation presence in the noise source will result different decay curves from trial to trial. The averaging technique is applied on a large number of obtained decay curve to get a reliable RT value. In another method the segmentation process is used for the detection of gaps in the reverberant speech [5, 6] which leads to the tracking of sound decay curve. In 1965 Schroeder addressed this problem by developing a method based on the integrated impulse response [7]. In this method the narrow or broad band excitation signal is replaced by a brief pulse. Schroeder reported an integral relationship between the overall average of the interrupted noise method, decay curves and the impulse response of the room, and thus the recurred trials were unnecessary.

Schroeder's method ruled the area of voice signal processing for the past few decades however there was a need of a blind RT estimation technique, which has the ability to estimate RT from the available reverberant signals, without using the information of the enclosure or in the absence of the test sound signal. In hands free telephony devices or hearing aids incorporating, the method based on the direct analysis of the sound signal, will be very helpful. Recently many method have been deployed for the blind RT estimation [2, 8, 9, 10-12], uses the reverberant sound (recorded sound) directly for estimation. In all these methods, statistical modelling is the core followed by the MLE to determine the optimized RT. Some semi-blind methods have also been developed [13-16] in which the enclosure characteristics are learned using neural network approaches. In neural networks first the network is trained by a set of data and then tested after that it is used for estimation. The estimation of reverberation characteristics using only the microphone (reverberant) signal is a good tool for the study of reverberations [2].

Ratnam *et al.* [2] used noise decay curve approach to model the room reverberation characteristics in order to develop a blind RT estimation algorithm i.e. completely based on the reverberant sound. The test room sound signal is incessantly processed to achieve a running RT estimate by utilising the ML parameter estimation technique. An order statistics filter is used in decision making step to get the most probable RT from the pool of estimated RT (over a period of time). In [2] the objective of detecting sound decay is obtained by using the iterative technique. This iterative approach for the detection of the sound decay makes the algorithm computationally expensive. Ratnam *et al.* made some changes in his algorithm in [2] in order to improve the computational efficiency of it and presented a new algorithm in [10]. Lollmann *et al.* [8] recently developed an algorithm based on the sound decay statistical model [2] for the blind RT estimation directly from reverberant speech. The main advantage of the

Lollmann *et al.* method is the detection of the possible sound decay via a pre selection method which leads towards an improvement in robustness of estimation and computational efficiency. In this paper we proposed a method for the blind RT estimation based on the Lollmann *et al.* method and a pre selection mechanism for the possible decay rate detection and used Gamma distribution for modelling the decay rate.

The rest of the paper is divided in four sections and is organized as follows. ML estimation procedure and the sound decay model used in the proposed method is discussed in Section 2. In Section 3 the estimation procedure of RT and its efficiency is discussed. Experimental results and proposed algorithm performance is evaluated in Section 4, and followed by Section 5 containing the conclusion.

2. Proposed Model and MLE

In this work, to model the reverberant signal energy decay, the statistical model used is based on the Gamma distribution. When the RT of a reverberant signal falls in a certain range then Gamma distribution is used for the better modelling of the amplitude distribution of the signal [17]. The proposed work is motivated from the method in [17]. A figure has been derived here based on the method in [17] and it shows the main motivation for this work as shown in Fig. 1. As the reverberation time lies in the lower range (0-150ms) then the distribution pattern followed by signal is Gamma. In current work our main focus is on the small chamber such as telephone booth, aeroplane chamber (cockpit), ATM booth, where we have to develop a method to estimate the reverberation time properly such that enhanced speech signal can be achieved at the end. In such small chambers, RT generally lies in the lower range. Therefore reverberant speech decay

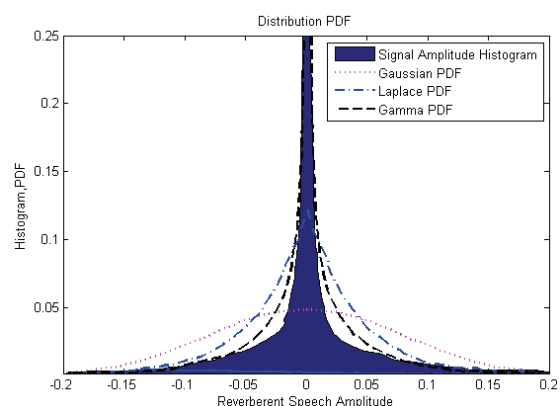


Fig. 1. Reverberant speech distribution and theoretical PDF of different distribution (Laplace distribution blue line, Gamma distribution black line, Gaussian distribution pink line) with zero mean.

curve would be modelled in the work based on Gamma distribution.

Hence, the decaying sound reverberant tail is modeled by employing Gamma distribution $G(x, k, \theta)$ along with a sequence of random variables. In $G(x, k, \theta)$ k is the shape factor and is equal to one for zero mean, θ is the variance of the Gamma distribution and x is the random variable. The general mathematical expression of Gamma distribution probability density is given below [18].

$$f(x; k, \theta) = \frac{x^{k-1} * e^{-x/\theta}}{\theta^k * \Gamma(k)} \tag{2}$$

where $\Gamma(k)$ is the gamma function of k . For zero mean we put $k=1$ in (2) and becomes [18];

$$f(x; 1, \theta) = \frac{x^{1-1} * e^{-x/\theta}}{\theta^1 * \Gamma(1)} \tag{3}$$

after simplification and calculating the value of [(1), (3) reduced to [18],

$$f(x; 1, \theta) = \frac{e^{-x/\theta}}{\theta} \tag{4}$$

In the evolution of model an assumption is made. In this assumption the reverberant decaying sound tail is represented by y which is the product of x and a . Where x is a random process and representing the fine structure, and a represents the deterministic envelope. Furthermore, $x(n)$ is supposed as an independent and identically distribution (i.i.d.) random number sequence valid for $n \geq 0$, having Gamma distribution with zero mean($k=1$) and variance θ , $G(1, \theta)$. Likewise for each and every value of n a deterministic sequence $a(n) > 0$ is defined in order to evolve the room decay (y) model [2]. The mathematical expression of the decay model is given as.

$$y(n) = x(n) \times a(n) \tag{5}$$

After rearranging the terms, (5) can be written as

$$x(n) = y(n)/a(n) \tag{6}$$

As $y(n)$ are independent but due to variation in $x(n)$ with time the $y(n)$ are not identically distributed and have a probability density function $G(1, \theta a(n))$.

Select a finite observation sequence, $n = 0, 1, 2, \dots, N-1$, in order to estimate the decay rate of the reverberant signal. For the sake of simplicity and notational convenience, the N -dimensional vectors of y and a are announced as \mathbf{y} and \mathbf{a} respectively. Thus the likelihood function of \mathbf{y} (the joint probability density), parameterized by \mathbf{a} and θ is,

$$L(\mathbf{y}; \mathbf{a}, \theta) = \frac{1}{a(0) * a(1) \dots a(N-1)} \left(\frac{1}{\theta}\right)^N \times \exp\left(-\frac{\sum_{n=0}^{N-1} y(n)/a(n)}{\theta}\right) \tag{7}$$

where \mathbf{a} and θ are the $(N+1)$ unknown parameters that are required to be estimated from the observation \mathbf{y} . Here the main objective is the modelling of sound decay in an enclosure and to make some simplification to the acquired likelihood function obtained in (7). A supposition is made that the reverberant sound signal energy envelop damping is defined by a single decay rate σ in the intermediate free decay regions (i.e., the regions followed by steep speech-sound offset) instead of the sound ongoing, onset, or gradually declining speech offsets regions. This leads to an expression for $a(n)$ and is determined by [2],

$$a(n) = \exp\left(-\frac{n}{\sigma}\right) \tag{8}$$

As the N -dimensional parameter $a(n)$ can be expressed in terms of σ and a single parameter a , i.e., $a = e^{-1/\sigma}$ by using a scalar parameter a instead of $a(n)$ and expressed as

$$a = \exp(-1/\sigma) \tag{9}$$

By putting the value of Eq. (9) in Eq. (8) we get

$$a(n) = a^n \tag{10}$$

After incorporating Eq. (10), Eq. (7) becomes

$$L(\mathbf{y}; \mathbf{a}, \theta) = \frac{1}{a^0 * a^1 \dots a^{N-1}} \left(\frac{1}{\theta}\right)^N \times \exp\left(\frac{\sum_{n=0}^{N-1} y(n)/a^n}{\theta}\right) \tag{12}$$

We use mathematical induction to simplify Eq. (12) by replacing $0 \rightarrow N-1$ positive consecutive integer by its logical equivalent, so that

$$0 + 1 + 2 + 3 + \dots + (N-1) = \frac{N(N-1)}{2} \tag{13}$$

After making use of Eq. (13) and terms rearrangements, Eq. (12) becomes

$$L(\mathbf{y}; \mathbf{a}, \theta) = \left(\frac{1}{\theta * a^{(N-1)/2}}\right)^N \times \exp\left(-\frac{\sum_{n=0}^{N-1} y(n)/a^n}{\theta}\right) \tag{14}$$

For the estimation of the unknown parameters (a, θ) MLE is used. First step is to obtain the log-likelihood function, for this logarithm is applied on the both side of the Eq. (14)

$$\ln L(\mathbf{y}; \mathbf{a}, \theta) = \ln\left(\frac{1}{\theta * a^{(N-1)/2}}\right)^N \times \ln \exp\left(-\frac{\sum_{n=0}^{N-1} y(n)/a^n}{\theta}\right) \tag{15}$$

Eq. (15) is then simplified by using logarithm identities and terms rearrangements

$$\ln L(\mathbf{y}; \mathbf{a}, \theta) = -\frac{N(N-1)}{2} \ln(a) - N \ln(\theta) - \frac{1}{\theta} \sum_{n=0}^{N-1} \frac{y(n)}{a^n} \tag{16}$$

By making use of mathematical induction, Eq. (16) is more simplified and becomes

$$\ln L(\mathbf{y}; a, \theta) = -\sum_{n=0}^{N-1} \ln(a^n * \theta) - \frac{1}{\theta} \sum_{n=0}^{N-1} \frac{y(n)}{a^n} \quad (17)$$

For the maximum of $\ln(L(\mathbf{y}; a, \theta))$, we differentiate Eq. (17) with respect to a to get the score function Sf_a [19].

$$\begin{aligned} Sf_a L(a; \mathbf{y}, \theta) &= \frac{\partial \ln L(\mathbf{y}; a, \theta)}{\partial a} \\ &= -\frac{1}{a} \sum_{n=0}^{N-1} n + \frac{1}{\theta} \sum_{n=0}^{N-1} n * y(n) * a^{-n-1} \end{aligned} \quad (18)$$

To get the extremum achieved by the log-likelihood function, we put the differentiation resultant (Eq. (18)) equals to zero, given as

$$-\frac{1}{a} \sum_{n=0}^{N-1} n + \frac{1}{\theta} \sum_{n=0}^{N-1} n * y(n) * a^{-n-1} = 0 \quad (19)$$

The estimated value of a , or in other words the zero of the scoring function is denoted by a_{sf_0} . The value of a_{sf_0} should satisfy Eq. (19). For verification that the estimated value of a (a_{sf_0}) maximizes the log-likelihood function, the second derivate test $\left. \frac{\partial^2 \ln L(\mathbf{y}; a, \theta)}{\partial a^2} \right|_{a=a_{sf_0}} < 0$, can be performed.

Similarly, the score function is obtained for θ by following the same procedure as followed for a . The log likelihood function in Eq. (17) is partially differentiated with respect to θ .

$$Sf_\theta L(\theta; \mathbf{y}, a) = \frac{\partial \ln L(\mathbf{y}; a, \theta)}{\partial \theta} = -\frac{N}{\theta} + \frac{1}{\theta^2} \sum_{n=0}^{N-1} y(n) * a^{-n} \quad (20)$$

The score function ($Sf_\theta L(\theta; \mathbf{y}, a)$) is then put equal to zero to obtain an expression for θ .

$$-\frac{N}{\theta} + \frac{1}{\theta^2} \sum_{n=0}^{N-1} y(n) * a^{-n} = 0 \quad (21a)$$

$$\Rightarrow \theta = \frac{1}{N} \sum_{n=0}^{N-1} y(n) * a^{-n} \quad (21b)$$

The estimated value of θ is denoted by θ_{sf_0} and should satisfy Eq. (21a). For the value of θ given in Eq. (21b) the log-likelihood function will achieve an extremum, for verification that the estimated value of θ maximizes the log-likelihood function, second derivative test $\left. \frac{\partial^2 \ln L(\mathbf{y}; a, \theta)}{\partial \theta^2} \right|_{\theta=\theta_{sf_0}} < 0$, can be performed

By observing estimation expression of a and θ represented by Eq. (19) and Eq. (21b) respectively, it is clear that Eq. (19) belongs to the category of the implicit expressions and its explicit solution will not exist, on the other hand Eq. (21b) fall in the category of explicit expressions and its explicit solution exists ,if a is known. It is already defined in Eq. (9), σ is the

required time constant and to be estimated.

The mapping of a over the σ is observed and it revealed that $a \in [0,1)$ maps one-to-one onto $\sigma \in [0, \infty)$. The method that we used here and the method used in [2] and [10] are similar because in both method quantization is the base for the estimation of a . To form the histogram of a , first the bins are created by quantizing the given range of a . For assigning values to these bin likelihood is used. When the likelihood values are calculated the maximum likelihood value is then assigned to that bin in the histogram.

Let X values are obtained after the quantization of a having the range $a \in [0,1)$. The quantized values of a is then represented by a_w , where $w=1, 2, 3, \dots, X$ and for each a_w the log-likelihood is calculated using Eq. (17) ad is given as;

$$nL(a_w; \mathbf{y}) = -\sum_{n=0}^{N-1} \ln(a_w^n * \theta) - \frac{1}{\theta} \sum_{n=0}^{N-1} \frac{y(n)}{a_w^n} \quad (22)$$

And a_{ML}^- can be decides as

$$\sigma_{ML}^- = \max_a \{ \ln L(a_w^n, \theta) \} \quad (23)$$

The result obtained from Eq. (23) is then put in Eq. (9) to obtain the estimated decay rate (σ_{ML}^-). At the end the value of RT (T_{60}^{-ML}) is calculated by using the formula from [11].

$$T_{60}^{-ML} = 6.908 \times \sigma_{ML}^- \quad (24)$$

3. Effective Estimation of RT

A looping approach is used in the parent method proposed in [2] to estimate the decay rate of the sound, but the computational resources requirement for that algorithm is significantly high. The computational efficiency of the parent technique is enhanced by the algorithm introduced in [10] in which the recorded signal is divided into frames and then processed instead of processing the whole signal to find the free sound decay regions for maximum likelihood estimation of the reverberant sound decay rate. The computational efficiency can be further enhanced by first capturing the reverberant signal free sound decay regions and then using those detected region for the ML estimation of the decay rate. That is the reason that this approach leads to the increment in computational efficiency by using only small portion of the reverberant signal for processing. This goal can be achieves by an estimation procedure proposed by Lollmann *et al.* [8] and have a residue benefit of diluting the outliers effects on RT estimated value. To enhance the Maximum Likelihood estimation of the Gamma parameters for our proposed method, we have used this procedure.

The reverberant speech signal is used as an input to the algorithm and is represented by $g(n)$. The $g(n)$ is a discrete

Table 1. The proposed RT value estimation method

<p>Objective: To use Gamma distribution based energy decay model for RT value estimation.</p> <p>System Input: Reverberant speech signal, i.e. $g(n)$.</p> <p>Desired Output: Estimated RT, i.e. T_{60}^-.</p> <p>Initialization: 1) In (25), $D=1631$ and $\Delta D=67$ are used. 2) In (26), $Q=233$ is used. 3) In (28), $\beta=0.995$ is used. 4) In (22) and (23), $w = 1, \dots, X$ while $X = 10$ is used.</p> <p>Scenario: The aim is to obtain estimated RT value from reverberant speech signal using following steps. 1) Use (25)-(27) for free decay regions detection, indexed by frame number Υ . 2) For the detected regions, use (22), (23), (9), and (24) to obtain $T_{60}^{-ML}(\cdot)$. 3) Apply recursive smoothing via (28) to the estimated RT values, i.e. $T_{60}^{-ML}(\cdot)$.</p> <p>Output: Use (29) to calculate T_{60}^-</p>
--

time signal and the n represents its index. The $g(n)$ has been processed frame-wise. The samples sequence is divided into frame and each frame has D samples, moved by an instant of ΔD [8]. The purpose of this division is to detect the free sound decay regions and is given as

$$G(\Upsilon, d) = g(\Upsilon * \Delta D + d) \text{ with } d = 0, 1, \dots, D-1 \quad (25)$$

where $\Upsilon \in \mathbb{N}$. To spot the potential sound decays, pre selection is carried out in the first step. This objective is achieved by the division of $G(\Upsilon, d)$ into $T = D/Q \in \mathbb{N}$ sub-frames.

$$U(\Upsilon, i_{sub}, j_{sub}) = G(\Upsilon, i_{sub}Q + j_{sub}) \quad (26)$$

where $j_{sub} = 0, 1, 2, \dots, Q - 1$ and $i_{sub} = 0, 1, 2, \dots, T - 1$ represents subframe indices. In next step we examined the maxima and minima of the sub-frame energy to find whether these values diverts from the next sub-frame values as done in [9].

$$\sum_{j_{sub}=0}^{Q-1} U^2(\Upsilon, i_{sub}, j_{sub}) > \tau_{i_{sub}} \cdot \sum_{j_{sub}=0}^{Q-1} U^2(\Upsilon, i_{sub} + 1, j_{sub}) \quad (27a)$$

$$\max_{j_{sub}} \{U(\Upsilon, i_{sub}, j_{sub})\} > \tau_{i_{sub}} \cdot \max_{j_{sub}} \{U(\Upsilon, i_{sub} + 1, j_{sub})\} \quad (27b)$$

$$\min_{j_{sub}} \{U(\Upsilon, i_{sub}, j_{sub})\} < \tau_{i_{sub}} \cdot \min_{j_{sub}} \{U(\Upsilon, i_{sub} + 1, j_{sub})\} \quad (27c)$$

where $\tau_{i_{sub}}$ is used as a weight and have a range of $0 \leq \tau_{i_{sub}} \leq 1$. For some frames the above equations (27a, 27b, 27c) may not be satisfied due to the counter i_{sub} , when it touches its minimum value $1 < i_{submin} < T - 2$ and if this is not the case, the inequality check is ceased and the incoming signal frame $G(\Upsilon + 1, b)$ is computed. On the other hand the sub-frame sequence satisfying Eq. (27a),

(27b), (27c) leads towards potential sound decay. RT (T_{60}^{-ML}) of the spotted frame is calculated for a finite band of RT values by making use of Eq. (22), (23), (9), (24).

The estimated values of RT i.e. T_{60}^{-ML} is the assigned to the bins in order to generate the histogram. The RT (T_{60}^{-ML}) value is update every time when a new value is generated. The bin size of the histogram is kept 10 for the sake of reducing the computational complexity and enhancing estimation accuracy.

Instead of the first peak, the maximum of the histogram is associated to the current estimated RT (T_{60}^{-ML}), because of the reduced number of outliers (result of the frame pre-selection process). The frames pre-selection also effected the estimated RT variance, reduced it by recursive smoothing, and modified the estimation expression. The modified RT expression is given as

$$T_{60}^{-ML}(\Upsilon) = \beta \cdot T_{60}^{-ML}(\Upsilon - 1) + (1 - \beta) \cdot T_{60}^{-ML}(\Upsilon) \quad (28)$$

where $0.9 < \beta < 1$. The RT value is finally estimated by

$$T_{60}^{-ML} = \text{mean}(T_{60}^{-ML}(\Upsilon)) \quad (29)$$

Table 1 summarizes the blind RT estimation technique utilizing the Gamma distribution based statistical model for the decay of reverberant sound signal.

4. Experimental Result and Discussion

Matlab simulations are carried out to evaluate the performance of proposed method for the blind estimation of RT. For simulation our first goal was to generate 10 different reverberant speech signal. For this we convolved 10 anechoic speech signals with RIRs to get the desired signals (reverberant signals). These 10 anechoic speech signal are randomly selected from the TIMIT database. Five of these are uttered by male and five by female, sampled at 16 KHz. The RIRs are obtained from the AIR

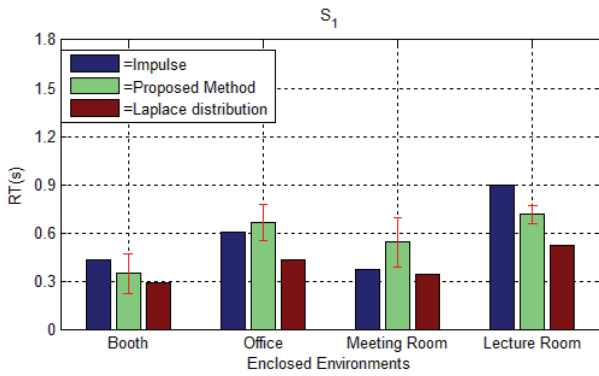


Fig. 2. Different RT estimation methods performance evaluation in terms of accuracy achieved for different room environments from the AIR database. Blue bars represents the RT values estimated from the RIRs by Schroeder's method [7], green bars shows RT values estimated by our proposed method and the brown bars shows RT estimated by using Laplace distribution [21]. The source to microphone distances for all of the four rooms are $S_1 = \{0.5, 1.0, 1.45, 2.25\}$ m respectively. The standard deviations are represented by short orange lines, plotted on top of the green bars.

database [20]. To check the performance of the algorithm for different room environment, the RIRs selected were recorded in different environment (booth, lecture room, office, meeting room). A source microphone distance pair $\{S_1, S_2\}$ m is selected for each environment respectively, i.e. $\{0.5, 1.5\}$, $\{2.25, 7.1\}$, $\{1, 3\}$, and $\{1.45, 2.8\}$. The values of other parameter used in simulation are given as: $X=10$, $T=7$, $i_{submin}=3$, $\beta=0.995$, $D=1631$ (corresponds approximately to a time span of 0.10 s), $Q=233$, $\Delta D=67$ (corresponds approximately to a frame shift of 0.0042 s), and $i_{sub}=1$.

The process of reverberant speech generation is repeated 20 time for each enclosed environment. 10 reverberant speech signals were generated with respect to source microphone distance S_1 and 10 reverberant speech signals were generated with respect to source microphone distance S_2 . For four different test environments a total of 80 reverberant speech signals were generated. All the reverberant signals were then tested for the RT estimation and an average over the 10 different result were calculated for each source microphone distance and room type. These average values of the estimated RTs is given in Fig. 2 and 3 respectively. The performance of the proposed method is evaluated by comparing the estimated RT value with the RT values estimated directly from the RIRs as done in [7] and with the RT values estimated using Laplace distribution [21].

Fig. 2 shows the results obtained for the shorter source-microphone distances i.e. S_1 from the above used pairs, while Fig. 3 shows the result obtained from the longer source-microphone distances i.e. S_2 . It can be observed that for different environments the difference between the

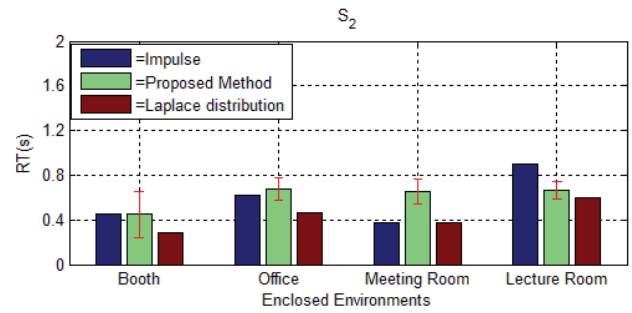


Fig. 3. Different RT estimation methods performance evaluation in terms of accuracy achieved for different room environments from the AIR database. Blue bars represents the RT values estimated from the RIRs by Schroeder's method [7], green bars shows RT values estimated by our proposed method and the brown bars shows RT estimated by using Laplace distribution [21]. The source to microphone distances for all of the four rooms are $S_2 = \{1.5, 3.0, 2.8, 7.1\}$ m respectively. The standard deviations are represented by short orange lines, plotted on top of the green bars.

RT estimated from the RIRs (represented by blue bars) and the RT estimated by our proposed method (represented by green bar) is small.

For example, for the booth at the source to microphone distance S_1 the estimated RT value obtained by our proposed method is 0.347 seconds and the RT value estimated from the RIRs is 0.428 seconds. Similarly for the booth at the source to microphone distance S_2 the estimated RT value obtained by our proposed method is 0.451 seconds and the RT value estimated from the RIRs is 0.456 seconds. Overall, the proposed method produced comparable results to the baseline method by Schroeder.

5. Conclusion

A novel approach for the blind estimation of reverberation time has been presented. The method is established on a Gamma distribution based statistical model for the decay of reverberant sound signal and a maximum likelihood approach for estimation of parameter. Experiments showed that the results obtained by utilizing our proposed method with real data (for booth RT value for S_1 and S_2 is 0.347 second and 0.451 second respectively) using speech signal samples are in close agreement with measured RT values (for booth RT value for S_1 and S_2 is 0.428 second and 0.456 second respectively). The Gamma distribution based statistical model produced acceptable results whenever the RT resided in the lower range (few hundred millisecond's). The results achieved by the proposed algorithm is comparable to the results obtained by the state of the art method and better in few cases. For

estimating the RT values over a wide range we would add some modifications to the algorithm in future, to make it possible.

References

- [1] H. Kuttruff, Room Acoustics, Elsevier Science Publishers Ltd., Lindin, 3rd ed., 1991.
- [2] R. Ratnam, D. L. Jones, B. C. Wheeler, W. D. O'Brien, C. R. Lansing and A. S. Feng, "Blind estimation of reverberation time," J. Acoust. Soc. Am., vol. 114, pp. 2877-2892, Nov. 2003.
- [3] W. C. Sabine, "Collected Papers on Acoustics," 1922.
- [4] International Organization for Standardization (ISO), Geneva, Acoustics- Measurements of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters, 1997
- [5] K. Lebart, J. Boucher, and P. Denbigh, "A new method based on spectral subtraction for speech deriverberation," Acta Acustica, vol. 87, pp. 359-366, 2001.
- [6] S. Vesa and A. Harma, "Automatic estimation of reverberation time from binaural signals," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., vol. 3, 2005, pp. 281-284.
- [7] M. R. Schroeder, "New method of measuring reverberation time," J. Acoust. Soc. Am., pp. 409-412, 1965.
- [8] H. W. Lollmann, E. Yilmaz, M. Jeub and P. Vary, "An improved algorithm for blind reverberation time estimation," Proc. Int. Workshop Acoust. Echo and Noise Control (IWAENC), Aug. 2010, Tel Aviv, Israel
- [9] H. W. Lollmann and P. Vary, "Estimation of the Reverberation Time in Noisy Environments," Proc. Int. Workshop Acoust. Echo and Noise Control (IWAENC), Sep. 2008, Washington USA.
- [10] R. Ratnam, D. L. Jones and W. D. O'Brien, "Fast algorithms for blind estimation of reverberation time," IEEE Signal Process. Letters, vol. 11, pp. 537-540, Jun. 2004
- [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," Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process., pp. 329-332, 2008.
- [12] Scharrer, Roman, and M. Vorländer. "Blind reverberation time estimation." Proceedings of the International Conference on Acoustics, Sydney, Australia. 2010.
- [13] T. J. Cox, F. Li and P. Darlington, "Extracting room reverberation time from speech using artificial neural networks," FJ. Audio Engineering Soc., pp. 219-230, 2001.
- [14] J. Nannariello and F. Fricke, "The prediction of reverberation time using neural network analysis," Applied Acoust., vol. 58, pp. 305-325, 1999.
- [15] Y. Tahara and T. Miyajima, "A new approach to optimum reverberation time characteristics," Applied Acoust., vol. 54, pp. 113-129, 1998.
- [16] Aliabadi, M., Golmohammadi, R., Ohadi, A., Mansoorizadeh, Z., Khotanlou, H., & Sarrafzadeh, M. S. (2014). Development of an Empirical Acoustic Model for Predicting Reverberation Time in Typical Industrial Workrooms Using Artificial Neural Networks. Acta Acustica united with Acustica, 100 vol. 6, pp.1090-1097.
- [17] T. Petsatodis, C. Boukis, F. Talantzis, Z. Tan and R. Prasad, "Convex combination of multiple statistical models with application to VAD," IEEE Trans. Audio, Speech, and Lang. Process., pp. 2314-2327, 2011.
- [18] M. A. Bean (2001). Probability: The science of uncertainty with application to investments, insurance and engineering [Online]. Available: books.google.com.pk/books?isbn=0821847929
- [19] V. Poor, "An Introduction to Signal Detection and Estimation," Springer-Verlag, New York, 1994.
- [20] M. Jeub, M. Schafer and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," Proc. Int. Conf. Digital Signal Process. (DSP), 2009, Santorini, Greece.
- [21] T. Jan, W. Wang, "Blind reverberation time estimation based on laplace distribution," 20th EUSIPCO 2012



Amad Hamza received his BSc degree in Electrical (Electronics) Engineering from Air University Islamabad, Pakistan in 2011 and his MS degree in Communication and Electronics from Univ. of Eng. & Tech. Peshawar, Pakistan in 2015. His research interests are audio/video processing, ANN and CGP.



Tariqullah Jan received his B.Sc. degree in Elect. Eng. from Pakistan in 2002 and PhD in the field of Electronic Engineering from UK in 2012. His Research interest includes Blind signal processing, machine learning, blind reverberation time estimation, multi-modal based approaches for the blind source separation, compressed sensing, and Non-negative matrix/tensor factorization for the blind source separation.

Asiya Jehangir received her B.Sc. degree in Electrical Engineering and M.S degree in Communication & electronics from Univ. of Eng. & Tech. Peshawar, Pakistan in 2007 and 2012 respectively. Her research interest includes Wireless Sensor Networks, Blind Signal Processing, and Applications of Signal processing in the field of Networks.



Syed Waqar Shah received his BSc degree in Elect. Eng. from Pakistan in 1994, MS degree from Pakistan in 2000 and PhD degree from UK in 2005. His research interests include Coordination in Mobile Agent Environment, Channel Estimation / Equalization of Wireless channels, Error Correction etc.



M.Haseeb Zafar received his BSc degree in Elect. Eng. from Pakistan in 1996, MS degree in Telecom. and Computers from USA in 2003 and PhD degree from UK in 2009. He has more than 18 years of teaching, research and industrial experience. His research interests include Wireless Sensor Networks, Smart Grids, Software Defined Networks, Genetic Algorithms, Intelligent Transportation Systems and Signal Processing.



M. Asif received his B.Sc degree in Computer Systems Eng. from Pakistan in 2006 and Ph.D degree in Electronics Engineering from UK, in 2012. He is currently working as an Assistant Professor in Department of Electronics, University of Peshawar, Pakistan. His main research interests include MANETs, Sensor Networks, Signal Processing and Renewable Energy.