DATA-DRIVEN STATISTICAL MODELLING OF ROOM IMPULSE RESPONSES IN THE POWER DOMAIN

Clement S. J. Doire[*], *Mike Brookes*[*], *Patrick A. Naylor*[*], *Søren Holdt Jensen*[†]

[*] Electrical and Electronic Engineering, Imperial College London, UK
[†] Department of Electronic Systems, Aalborg University, Denmark

ABSTRACT

Having an accurate statistical model of room impulse responses with a minimum number of parameters is of crucial importance in applications such as dereverberation. In this paper, by taking into account the behaviour of the early reflections, we extend the widely-used statistical model proposed by Polack. The squared room impulse response is modelled in each frequency band as the realisation of a stochastic process weighted by the sum of two exponential decays. Room-independent values for the new parameters involved are obtained through analysis of several room impulse response databases, and validation of the model in the likelihood sense is performed.

Index Terms— Statistical model, Impulse Response, Early Decay

1. INTRODUCTION

The issue of reverberation is important as it degrades the intelligibility of speech in everyday communication scenarios [1], creating the need for effective means of controlling and removing it [2]. Statistical modelling of the measured Room Impulse Response (RIR) is used in many research topics such as dereverberation [3, 4] or reverberation parameter estimation [5,6]. The statistical model used in such algorithms needs to be accurate but should also involve as few parameters as possible.

Following [7], a time-domain reverberation model was introduced in [8], in which the room impulse response for a given source and receiver position is given by a single realisation of a non-stationary stochastic process. The model gives a two-parameter representation of the impulse response, and is valid for t greater than the mixing time t_m and frequencies greater than the Schroeder frequency [9] under the assumption that the room is ergodic. We note that the parameters of the model are normally frequency-band dependent [10]. The restriction $t > t_m$ is relaxed in [11] by dividing the RIR into two segments with different parameter values.

In this paper, we propose an extended model of the RIR that represents both early and late reverberation components in multiple frequency bands. Using a representative set of measured impulse responses, fixed values are derived for the extra parameters introduced in the extended model.

The paper is organised as follows: in Section 2 the problem of statistical modelling is defined and details of estimating its parameters are explained. The relation between the early and late decay is explored in Section 3, and the final model is detailed in Section 4. Section 5 presents the validation of the model's performance in the likelihood sense, while Section 6 concludes this paper.

2. PROBLEM STATEMENT

Polack proposed a time-domain statistical model [8] of the RIR, $h(t)$, excluding the contribution of the direct path:

$$h(t) = \begin{cases} \sigma\, b(t) e^{-\frac{3\log(10)}{T_{60}}t} & t \geq 0 \\ 0 & t < 0 \end{cases}, \tag{1}$$

in which $b(t)$ is zero-mean, unit variance, stationary Gaussian noise and T_{60} is the reverberation time. The squared version of this model for $t \geq 0$ is given by

$$h^2(t) = \sigma^2 b^2(t) e^{-\frac{6\log(10)}{T_{60}}t} \tag{2}$$

with $b^2(t)$ following a χ^2 distribution with 1 degree of freedom or, equivalently, a Gamma distribution $b^2(t) \sim \Gamma(\frac{1}{2}, 2)$.

Converting the model of (2) into discrete time with sample rate f_s, we derive, in each of K third-octave sub-bands, the following model for the squared impulse response:

$$h_k^2(n) = \left(u(n-1)\,\gamma_k e^{\log(\alpha_k)(n-1)} + N_k\right) b^2(n) \tag{3}$$

where N_k is the noise floor of the measured RIR and $b^2(n)$ is the discrete equivalent of the Gamma distributed noise term discussed earlier. $u(n)$ is the unit step function, α_k is the decay constant in sub-band k, related to $T_{60,k}$ through

$$\log(\alpha_k) = \frac{-6\log(10)}{f_s\, T_{60,k}}, \tag{4}$$

The research leading to these results has received funding from the EU 7th Framework Programme (FP7/2007-2013) under grant agreement ITN-GA-2012-316969.

and γ_k is the drop in energy after the direct path, related to the frequency-dependent Direct-to-Reverberant Ratio (DRR) by the equation

$$\mathrm{DRR}_k = \frac{1-\alpha_k}{\gamma_k}. \tag{5}$$

The three parameters of this model, the decay constant α_k, the drop in energy γ_k and the noise floor N_k, can all be estimated from a measured room impulse response. To do so, the RIR is filtered in third-octave sub-bands, then the Hilbert envelope of each sub-band is computed and squared. The method presented in [12] is used to fit a decay+noise model similar to (3) to this squared envelope in the log domain because this algorithm is robust to noise and gives an unbiased estimate of T_{60}. The three parameters are directly estimated by the method using nonlinear optimisation.

We want to incorporate in (3) the behaviour of the early reflections, modelled in a statistical way. The idea of having two different frequency-band dependent decay rates for the early and late part of the impulse response has been explored in [13, 14] in the context of artificial reverberators. However, adding extra unknown parameters to estimate in the overall model would not be practical in many applications linked with controlling and removing the reverberation. For this particular reason, the early part of the impulse responses will be described as a decaying stochastic process as well, and a relationship will be derived between early and late reverberation decays. In a similar fashion to α_k being linked to T_{60}, the early decay constant $\alpha_{E,k}$ is related to the Early Decay Time (EDT) [15].

3. RELATION BETWEEN EARLY AND LATE DECAY CONSTANTS

In a similar fashion to the statistical model of (3), the behaviour of the early reflections is modelled as a realisation of an exponentially decaying stochastic process. In the remainder of Section 3, the following parameterisation of the early and late decay constants will be used:

$$\rho(\alpha_k) = \log\left(\frac{\alpha_k}{1-\alpha_k}\right). \tag{6}$$

This maps the range $0 < \alpha_k < 1$ to $-\infty < \rho(\alpha_k) < \infty$. We will be looking at the relationship between $\rho(\alpha_k)$ and $\rho(\alpha_{E,k})$ rather than between α_k and $\alpha_{E,k}$ directly.

In the remainder of this paper, the set of impulse responses listed in Table 1 will be used to train and evaluate our new statistical model. In the table are listed the different rooms, along with their measured broadband reverberation time and reference to the database they belong to. *N^o Conf.* refers to the number of different source-receiver configurations, in contrast to *N^o RIRs* which refers to the total number of room impulse responses included. t_m refers to the mixing time and was computed using the physical predictor described in the next paragraph. The training set consists of binaural recordings of RIRs included in the Aachen database [16], the reflective configuration of the MARDY database [17], as well as a random subset of the SMARD database [18]. The RIR database presented in [19] was then used as a validation set. This database was recorded in a room offering variable acoustic properties with an omnidirectional loudspeaker. The two most reverberant conditions and the room impulse responses recorded at the 4th microphone of the linear array were used as the validation dataset. All RIRs were recorded at a sampling frequency $f_s = 48\,\mathrm{kHz}$.

Room	T_{60} (s)	Database	N^o Conf.	N^o RIRs	t_m (ms)
A: Lecture	0.780	[16]	6	24	46.8
B: Meeting	0.230	[16]	5	20	36.3
C: Office	0.430	[16]	3	12	33.5
D: Cathedral	3.43	[16]	11	22	-
E: Stairway	0.890	[16]	13	78	-
F: MARDY	0.447	[17]	9	72	35.6
G: SMARD	0.150	[18]	8	24	38.6
V1: MIRD	0.360	[19]	26	26	31.3
V2: MIRD	0.610	[19]	26	26	31.3

Table 1: Table detailing information about RIRs included in the training (A to G) and validation (V1 -V2) datasets

In order to compute the decay constant of the early part of a RIR, a boundary must be chosen as to when the density of reflections becomes sufficient to consider the diffuse reverberation tail has been reached. This boundary, called the mixing time t_m, is investigated in [20] where different physical predictors of the mixing time are reviewed, compared to more empirical methods, and perceptual studies are conducted to verify their validity. Using, when available, the dimensions of the rooms in the training dataset listed in Table 1, we were able to predict the values of the mixing time t_m.

In [21], diffuseness is assumed when each sound particle has been through at least four reflections within the room. This leads to to the physical predictor $t_m = 47\frac{V}{S}$ ms with V the room volume in m^3 and S the total surface area of the room in m^2. Using this formula, the mixing times given in Table 1 are obtained. To confirm these results, we then used the perceptually motivated formulae presented in [20], giving the following range for the mixing time values: $t_m \in [25, 32]$ ms. As we want to make sure we do not include any portion of the late decay when fitting $\alpha_{E,k}$, a lower limit on the possible values of t_m should be chosen. Therefore, we approximate the boundary between early and late reverberation by choosing it to be 25 ms in each sub-band and for all rooms in the training set. The early decay constant $\alpha_{E,k}$ was thus computed using the first 25 ms after the direct path of each RIR.

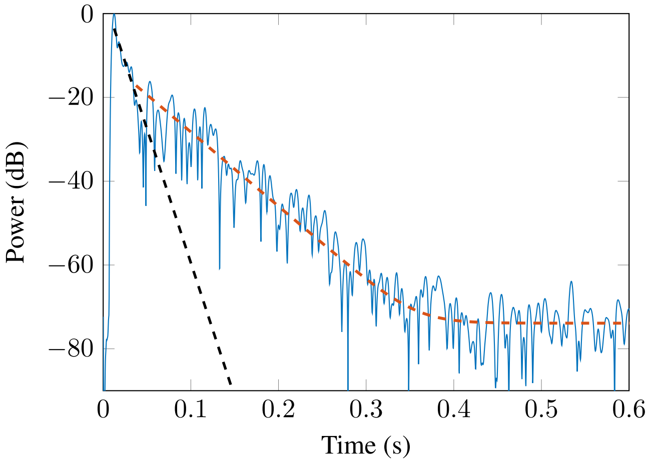

Fig. 1: Hilbert envelope of an example 1 kHz sub-band RIR with fitted early decay $\alpha_{E,k}$ and late decay α_k (dashed lines).

The values of α_k, and therefore $\rho(\alpha_k)$, were obtained from [12] as described in Section 2. To compute $\alpha_{E,k}$, a Least-Squares linear fitting was used on a 25 ms window of the squared Hilbert envelope of each sub-band RIR in the log domain, starting at the next sample after the direct path.

We now consider the ratio $\eta_k = \frac{\rho(\alpha_{E,k})}{\rho(\alpha_k)}$. After computation of this ratio in each of the K third-octave sub-bands, for all RIRs in all rooms, a distribution of its values was obtained for each sub-band. Figure 2 shows the median as well as the 5%, 25%, 75% and 95% quantiles of these distributions plotted against sub-band centre frequency. A value of η_k smaller than 1 indicates that the early decay is faster decaying than the late decay as illustrated in the example of Fig. 1.

Our findings indicate the distribution of η_k is quite narrow for all frequencies. Accordingly, we use a room-independent value for η_k equal to the median in the corresponding sub-band and plotted as the solid curve in Fig. 2. From (6) we therefore obtain

$$\alpha_{E,k} = \frac{\alpha_k^{\eta_k}}{\alpha_k^{\eta_k} + (1-\alpha_k)^{\eta_k}} \tag{7}$$

which will be used to compute $\alpha_{E,k}$ in the remainder of the paper.

4. SUM OF TWO DECAYS

When observing the behaviour of the energy envelope of a room impulse response, one can notice there is not a clear separation between early and late decays. To model this smooth transition, the model of equation (3) is extended to include a sum of two decays:

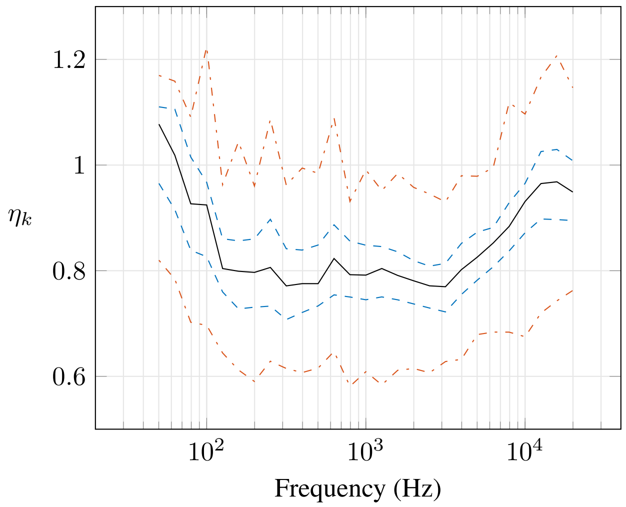

Fig. 2: Distribution of the ratio $\eta_k = \frac{\rho(\alpha_{E,k})}{\rho(\alpha_k)}$ across frequencies. The solid line shows the median value and the dashed lines show the 5%, 25%, 75% and 95% quantiles.

$$h_k^2(n) = \Big[u(n-1)\,\frac{\gamma_k}{\beta}\Big(\beta e^{\log(\alpha_k)(n-1)} + (1-\beta)e^{\log(\alpha_{E,k})(n-1)}\Big) + N_k\Big]\,b^2(n) \tag{8}$$

with β the weight of the late decay term. In the following, we will differentiate the original model of (3) and the extended model of (8) using the superscripts $^{(1)}$ and $^{(2)}$ respectively. Let $c_k^{(i)}(n)$ be the deterministic term in each model so that

$$c_k^{(1)}(n) = u(n-1)\,\gamma_k e^{\log(\alpha_k)(n-1)} + N_k, \tag{9}$$

$$c_k^{(2)}(n) = u(n-1)\,\frac{\gamma_k}{\beta}\Big(\beta e^{\log(\alpha_k)(n-1)} + (1-\beta)e^{\log(\alpha_{E,k})(n-1)}\Big) + N_k. \tag{10}$$

To construct the likelihood of each model, we use

$$b^2(n) \sim \Gamma(\frac{1}{2},2) \Rightarrow c_k^{(i)}(n)\,b^2(n) \sim \Gamma(\frac{1}{2},2c_k^{(i)}(n)) \tag{11}$$

and assume the realisations of the squared impulse response to be independent of each other. $\widetilde{h_k^2}(n)$ is used to denote the actual realisations of the RIR in the sub-band k, with $n=1$ corresponding to the sample after the direct path and N the total length of the impulse response. Using

$$f_\Gamma(x;\kappa,\theta) = \frac{x^{\kappa-1}e^{-\frac{x}{\theta}}}{\theta^\kappa\Gamma(\kappa)} \tag{12}$$

we have, in each sub-band:

$$\mathcal{L}_k^{(i)} = \prod_{n=1}^{N} f_\Gamma\left(\widetilde{h_k^2}(n);\,\frac{1}{2},2c_k^{(i)}(n)\right). \tag{13}$$

Assuming the likelihoods of each model are independent between sub-bands, we can compute the joint-likelihood

$$\mathcal{L}^{(i)} = \prod_{k=1}^{K} \prod_{n=1}^{N} f_\Gamma \left(\widetilde{h_k^2}(n); \frac{1}{2}, 2c_k^{(i)}(n) \right). \tag{14}$$

To determine the optimal value for β, we need to maximise (14) or, equivalently, minimise its negative logarithm. This is a one-dimensional constrained optimisation problem that can be solved using an interior-point algorithm [22]. The constraints are $0 \leq \beta \leq 1$ and the starting point of the optimisation method can be chosen to be $\beta = 1$ so that the likelihood is initialised to $\mathcal{L}^{(1)}$. The output of the optimisation procedure necessarily results in an improvement over the original model.

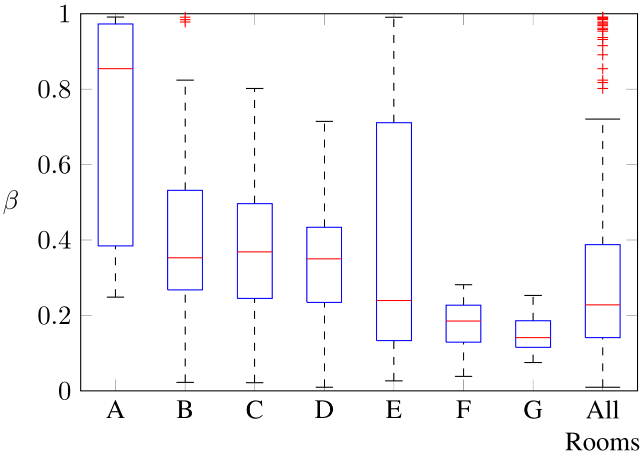

Fig. 3: Distribution of the β parameter from (8) for each room found through 1-dimensional maximum likelihood optimisation. The box indicates the median and inter-quartile range, the whiskers show the full inlier range and outliers are plotted individually.

This operation is performed for all the impulse responses from each room. Box and whisker plots of the resulting values of β are shown in Fig. 3. The distribution of β across all rooms in the training set is relatively narrow, with a median value of 0.23 and a mean value of 0.32. In the remainder of the paper, β will therefore be set to the mean value found through optimisation $\beta = 0.32$.

5. VALIDATION

In order to validate the improved accuracy of the extended model given in (8) compared to the original model described in (3), their respective likelihood functions, $\mathcal{L}^{(2)}$ and $\mathcal{L}^{(1)}$, were computed for each room impulse response in the training database. The logarithm of the likelihood ratio was then computed:

$$\log \left(\frac{\mathcal{L}^{(2)}}{\mathcal{L}^{(1)}} \right) = \log(\Lambda). \tag{15}$$

A positive value means that the extended model in (8), which incorporates early decay modelling, leads to an improvement in fitting the data in the likelihood sense. In order to be able to compare the results between the different rooms in the dataset, the same length N was used when computing the joint-likelihoods. As the difference between the original and extended model resides in the earlier part of the RIRs, they were all truncated to a duration of 520 ms, corresponding to the length of the shortest recorded room impulse response. Box and whisker plots of the obtained log likelihood ratios are plotted Fig. 4 for each room in both training and validation dataset.

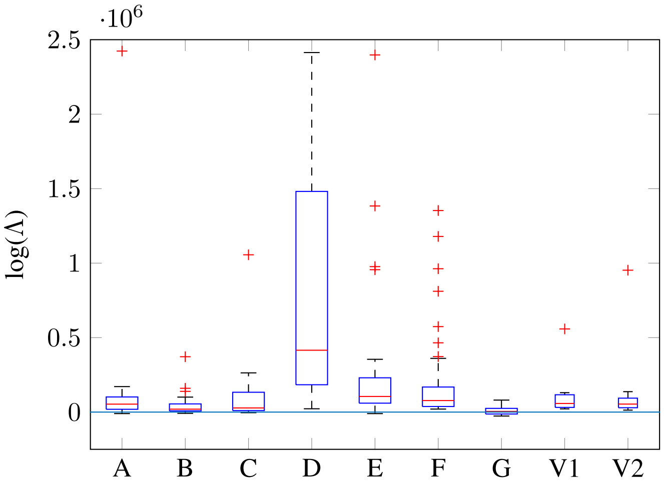

Fig. 4: Distribution of the logarithm of the likelihood ratio Λ for all RIRs in each room. A positive value means an improvement compared to the original model.

The lower quartile is positive for all rooms except the SMARD room, G, which presents a median slightly above zero. Even though there are, for each room in the training set, a few cases where the original model fits the data slightly better, it is seen that the extended model almost always increases the joint-likelihood. Moreover, for all room impulse responses and both room settings of the validation dataset, the joint-likelihood of the data is higher using the extended model described by (8).

6. CONCLUSION

In this paper we have presented a new statistical model of squared room impulse responses. It extends the well-known Polack model by incorporating an approximation of the early reflections' behaviour. Using an extensive training set of impulse responses, room-independent values were determined for the additional parameters included in the new model. A likelihood comparison over the training set and a validation set showed a substantial improvement in fitting the measured data.

REFERENCES

[1] Erwin L. J. George, Joost M. Festen, and Tammo Houtgast, "The combined effects of reverberation and nonstationary noise on sentence intelligibility," *Journal of the Acoustical Society of America*, vol. 124, no. 2, pp. 1269–1277, 2008.

[2] Keisuke Kinoshita, Marc Delcroix, Takuya Yoshioka, Tomohiro Nakatani, Emanuel Habets, Reinhold Haeb-Umbach, Volker Leutnant, Armin Sehr, Walter Kellermann, Roland Maas, Sharon Gannot, and Bhiksha Raj, "The REVERB challenge," http://reverb2014.dereverberation.com/, 2014.

[3] Emanuel A. P. Habets, *Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement*, Ph.D. thesis, Technische Universiteit Eindhoven, 2007.

[4] Patrick A. Naylor and Nikolay D. Gaubitch, *Speech Dereverberation*, Signals and Communication Technology. Springer, 2010.

[5] James Eaton, Nikolay D. Gaubitch, and Patrick A. Naylor, "Noise-robust reverberation time estimation using spectral decay distributions with reduced computational cost," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vancouver*, pp. 161–165, 2013.

[6] Clement S. J. Doire, Mike Brookes, Patrick A. Naylor, Dave Betts, Christopher M. Hicks, Mohammad A. Dmour, and Søren H. Jensen, "Single-channel blind estimation of reverberation parameters," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Brisbane*, 2015.

[7] Manfred R. Schroeder, "Statistical parameters of the frequency response curves of large rooms," *J. Audio Eng. Soc.*, vol. 35, no. 5, pp. 299–306, 1987.

[8] Jean-Dominique Polack, *La transmission de l'énergie dans les salles*, Ph.D. thesis, Université du Maine, Le Mans, France, 1988.

[9] Jean-Marc Jot, Laurent Cerveau, and Olivier Warusfel, "Analysis and synthesis of room reverberation based on a statistical time-frequency model," *Proceedings of the 103rd AES convention, New York*, 1997.

[10] Heinrich Kuttruff, *Room Acoustics*, Spon Press, fifth edition, 2009.

[11] Emanuel A. P. Habets, "Single-channel speech dereverberation based on spectral subtraction," *Workshop on Circuits, Systems and Signal Processing*, pp. 250–254, 2004.

[12] Matti Karjalainen, Poju Antsalo, Aki Mäkivirta, Timo Peltonen, and Vesa Välimäki, "Estimation of modal decay parameters from noisy response measurements," *Proceedings of the 110th AES convention, Amsterdam*, pp. 12–15, May 2001.

[13] Esa Piirilä, Tapio Lokki, and Vesa Välimäki, "Digital signal processing techniques for non-exponentially decaying reverberation," *Proceedings of the 1998 Digital Audio Effects Workshop (DAFX-98), Barcelona*, pp. 21 – 24, 1998.

[14] Jonathan S. Abe and Keun-Sup Lee, "A reverberator with two-stage decay and onset time controls," *Proceedings of the 129th AES Convention, San Francisco*, 2010.

[15] Vilhelm Lassen Jordan, "Room acoustics and architectural acoustics development in recent years," *Applied Acoustics*, vol. 2, no. 1, pp. 59–81, 1969.

[16] Marco Jeub, Magnus Schäfer, and Peter Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," *Proceedings of International Conference on Digital Signal Processing (DSP)*, pp. 1–4, July 2009.

[17] Jimi Y. C. Wen, Nikolay D. Gaubitch, Emanuel A. P. Habets, Tony Myatt, and Patrick A. Naylor, "Evaluation of speech dereverberation algorithms using the mardy evaluation of speech dereverberation algorithms using the MARDY database," *Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC), Paris*, 2006.

[18] Jesper Kjaer Nielsen, Jesper Rindom Jensen, Søren Holdt Jensen, and Mads Graesbøll Christensen, "The single- and multichannel audio recordings database (SMARD)," *Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC), Antibes - Juan les Pins*, pp. 40–44, 2014.

[19] Elior Hada, Florian Heese, Sharon Gannot, and Peter Vary, "Multichannel audio database in various acoustic environments," *Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC), Antibes - Juan les Pins*, September 2014.

[20] Alexander Lindau, Linda Kosanke, and Stefan Weinzierl, "Perceptual evaluation of model- and signal-based predictors of the mixing time in binaural room responses," *Journal of the Audio Engineering Society*, vol. 60, no. 11, pp. 887–898, November 2012.

[21] P. Rubak and L. G. Johansen, "Artificial reverberation based on a pseudo-random impulse response II," *Proceedings of the 106th AES Convention, Munich*, 1999.

[22] Sanjay Mehrotra, "On the implementation of a primal-dual interior point method," *SIAM Journal on Optimization*, vol. 2, no. 4, pp. 575–601, 1992.