Mean Absorption Coefficient Estimation From Impulse Responses: Deep Learning vs. Sabine

Corto Bastien, Antoine Deleforge, Cédric Foy

MEAN ABSORPTION COEFFICIENT ESTIMATION FROM IMPULSE RESPONSES: DEEP LEARNING VS. SABINE

Corto Bastien[1] **Antoine Deleforge**[1] **Cédric Foy**[2]

[1] Universite de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

[2] Cerema, Ifsttar, UMRAE, 67200 Strasbourg, France

antoine.deleforge@inria.fr, cedric.foy@cerema.fr

ABSTRACT

We consider the problem of estimating the mean absorption coefficients of a room from an impulse response using supervised learning on simulated training sets. Two neural network architectures and two training dataset designs are considered. The proposed approach is shown to yield smaller estimation errors than the classical Sabine and Eyring formulas, despite not relying on any geometrical information on the room. Simulated results demonstrate the robustness of the approach under different challenging acoustic conditions.

1. INTRODUCTION

To improve the listening quality of a room, the main parameters an acoustician can act on are the *absorption coefficients* of its surfaces. These are generally frequency-dependent and are typically expressed within a given octave band, here, $b \in \{.125, .25, .5, 1, 2, 4\}$ kHz. To obtain the acoustic diagnosis of a room and deduce a renovation plan, one would ideally need to know the absorption profile $\alpha_i(b)$ of each surface i in the room. No method is currently known to reliably estimate these quantities *in situ*, so in practice, the *mean absorption coefficients* $\bar{\alpha}(b)$ weighted by the area S_i of each surface are estimated instead. The conventional method to estimate them is to calculate the room's reverberation times $RT_{60}(b)$ by Schroeder integration on a measured room impulse response (RIR) [1] and to use these in the celebrated Sabine's law or its slightly more precise variant from Eyring:

$$\bar{\alpha}_{\text{Eyring}}(b) = -\ln\left(1 - 0.163 \cdot V/(S \cdot RT_{60}(b))\right) \quad (1)$$

where V denotes the room's volume and $S = \sum_i S_i$ its total surface. These simple formulas stemming from reverberation theory [2] hold well under diffuse homogeneous sound fields such as those encountered in approximately cubic rooms with uniform absorption profiles. However, they are no longer accurate in environments such as corridors or acoustically treated rooms. In this study, we investigate the possibility of overcoming these limitations using supervised learning on large simulated training datasets. We show that neural networks trained to map RIRs to their corresponding $\bar{\boldsymbol{\alpha}}$ yield more accurate estimates than Sabine and Eyring, especially in non-homogeneous conditions, despite not needing to know V or S.

2. TRAINING SETS AND NETWORKS

As a good compromise between realism and computational time, we simulated RIRs using the hybrid shoe-box room acoustic simulator Roomsim [3]. One of its attractive features is to combine the image-source method [4] to accurately model early specular reflections and an efficient ray-tracing method called *diffuse rain* to efficiently capture late scattering effects. The training and development datasets are respectively composed of 15k and 5k RIRs corresponding to shoe-box rooms with lengths and widths uniformly drawn at random in $[1.5, 10]$ and heights in $[2.5, 4]$. These ranges incorporate homogeneous, elongated and flat geometries that are typical of housing and office buildings. A (source, receiver) pair spaced by at least 1 m is placed uniformly at random in each simulated room with a minimum distance of 0.5 m to each surface.

Two different approaches are considered to sample the acoustic properties of rooms. In the *Unif.* dataset, the 36 absorption coefficients (6 surfaces, 6 octave bands) are drawn uniformly and independently at random between 0 and 1. While this approach seems natural, we observed that over 90% of resulting $RT_{60}(b)$ values were lower than 0.4 s, which is not representative of commonly encountered room acoustics. To correct this, we designed a *reflectivity-biased* (RB) dataset as follows: (i) for each surface type (walls, floor, ceiling), toss a coin, (ii) on heads, draw a *reflective*, frequency-independent absorption profile uniformly at random in $[0.1, 12]$, (iii) on tails, draw a *non-reflective*, frequency-dependent absorption profile, uniformly at random within ranges depending on the surface type, as defined in Table 1. These ranges were chosen to incorporate about 100 commonly encountered absorption profiles manually selected from available lab measurements [1] [2] [3]. The resulting dataset features better spread $RT_{60}s$ from 0 to 2 s. Finally, since scattering coefficients are harder quantities to define and measure in practice, a

[1] https://www.acoustic-supplies.com/absorption-coefficient-chart/

[2] https://www.acoustic.ua/st/web_absorption_data_eng.pdf

[3] http://www.acouphile.fr/materiaux.html

freq. (Hz)		125	250	500	1k	2k	4k
Walls	$\alpha_i^{\min}$	0.01	0.01	0.01	0.01	0.01	0.01
	$\alpha_i^{\max}$	0.50	0.50	0.30	0.12	0.12	0.12
Floor	$\alpha_i^{\min}$	0.01	0.01	0.05	0.15	0.25	0.30
	$\alpha_i^{\max}$	0.20	0.30	0.50	0.60	0.75	0.80
Ceiling	$\alpha_i^{\min}$	0.01	0.15	0.40	0.40	0.40	0.30
	$\alpha_i^{\max}$	0.70	1.00	1.00	1.00	1.00	1.00

Table 1. Absorption coefficients ranges in the RB dataset.

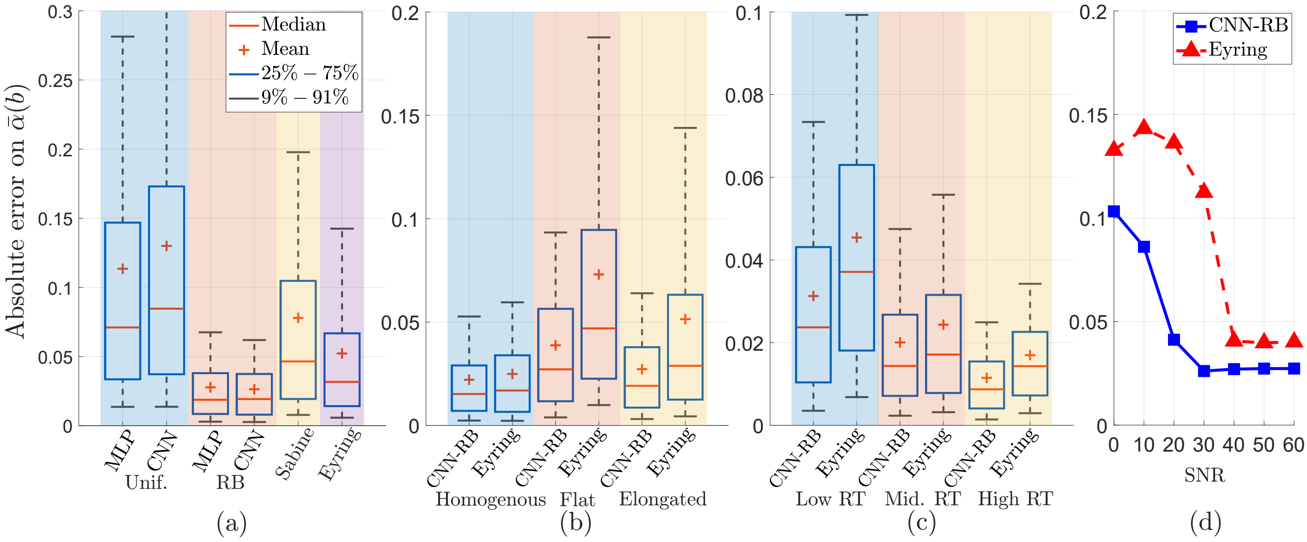

Figure 1. Errors on mean absorption coefficients using the proposed neural networks (MLP, CNN) trained on two datasets (Unif., RB) vs. reverberation theory (Sabine, Eyring). (a) Realistic absorption profiles, (b) Homogenous vs. flat vs. elongated room geometries, (c) $RT_{60}(b) < 0.5$ s vs. $RT_{60}(b) \in [0.5, 1.5]$ s vs. $RT_{60}(b) > 1.5$ s, (d) Varying SNR.

unique coefficient for all surfaces in each octave band was drawn uniformly at random, in $[0, 1]$ for the Unif. dataset, and in $[0, 0.3]$ for octave bands $< 1\ kHz$ and $[0.2, 1]$ otherwise for the RB dataset, following measures in [5].

Simulated RIRs are downsampled to 16 kHz, cropped to 0.5 s, corrupted by additive white Gaussien noise with a 30 dB signal-to-noise ratio (SNR) and normalized to have maximum value of 1, yielding input vectors in $\mathbb{R}^{8000}$. Two neural network architectures are considered. The first one is a multilayer perceptron (MLP) with 3 fully connected hidden layers of successive dimensions 128, 64 and 32, each followed by exponential linear units (ELUs). The second one is a convolutional neural network (CNN) starting with 3 1D-convolutional hidden layers with respective filter sizes 33, 17, 9 and number of filters 64, 32 and 16, each followed by a maxpool-4 layer and ELUs. The resulting output of dimension 2000 is then passed through a fully connected hidden layer of size 32 with ELUs. For each network, a fully-connected output layer followed by sigmoids is used to yield the desired output $\bar{\boldsymbol{\alpha}} \in [0, 1]^6$, with a mean squared error loss. Networks are optimized over training sets using the default Pytorch implementation of ADAM and batches of size 1000. Parameters yielding the lowest loss on development sets over 200 epochs are kept.

3. RESULTS AND CONCLUSION

Absolute errors on absorption coefficients over different test sets containing 500 simulated RIRs each are reported in Fig. 1. In (a), only absorption profiles corresponding to real materials [1 2 3] are used. We see that networks trained on the naive Unif. set do not succeed in outperforming classical approaches. However, estimation errors twice smaller as Eyring and with much less variance are obtained using the RB training set. In (b), we see that as expected, the advantage of supervised learning having access to full RIRs is clearest under room geometries departing from the homogenous assumption of reverberation theory. In (c), we see that performances of CNN-RB degrade under low RT_{60}'s, which is expected as less information is then available in RIRs, and in (d), that it degrades under higher noise. Nevertheless, the supervised learning approach preserves a significant advantage in all explored conditions. Finally, we constructed a test set with similar RT_{60} and geometry as RB, but using the pyroomacoustic simulator [6], which significantly differs from Roomsim by handling only frequency-independent absorption profiles, not incorporating diffusion, and using sinc kernels instead of finite impulse response filters. The mean errors of Eyring and CNN-RB were respectively 0.093 and 0.085 on this test set. This encouragingly suggests that the supervised learning approach did not severely overfit the specific training simulator. Future work will include testing the approach on measured RIRs and extending it to the much more challenging estimation of individual surface absorption profiles, by leveraging multiple RIRs per room.

4. REFERENCES

[1] M. R. Schroeder, "New method of measuring reverberation time," *The Journal of the Acoustical Society of America*, vol. 37, pp. 409–412, 1965.

[2] H. Kuttruff, *Room Acoustics - Fifth edition*. Oxfordshire, England: Spon Press, 2009.

[3] S. M. Schimmel, M. F. Muller, and N. Dillier, "A fast and accurate "shoebox" room acoustics simulator," in *Proc. of IEEE ICASSP*, pp. 241–244, 2009.

[4] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[5] M. Vorländer and E. Mommertz, "Definition and measurement of random-incidence scattering," *Applied Acoustics*, vol. 60, no. 2, pp. 187–199, 2000.

[6] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A python package for audio room simulation and array processing algorithms," in *Proc. of IEEE ICASSP*, pp. 351–355, 2018.