Geometry-Informed Estimation of Surface Absorption Profiles from Room Impulse Responses

Stéphane Dilungana, Antoine Deleforge, Cédric Foy, Sylvain Faisan

Geometry-Informed Estimation of Surface Absorption Profiles from Room Impulse Responses

Stéphane Dilungana, Antoine Deleforge
Université de Lorraine, CNRS, Inria, LORIA
Nancy, France
firstname.lastname@inria.fr

Cédric Foy
UMRAE, Cerema, Univ. Gustave Eiffel, Ifsttar. Strasbourg, France
cedric.foy@cerema.fr

Sylvain Faisan
ICube, Université de Strasbourg, CNRS, Strasbourg, France
faisan@unistra.fr

***Abstract*—This paper presents a method to jointly estimate the frequency-dependent absorption coefficients of the walls, ceiling and floor in a room from several impulse response measurements. The principle of the approach is to search among the observations for temporal windows of fixed size in which there is only one manifestation of acoustic reflection, based on the geometry of the setup which is assumed known up to some error. A probablistic procedure inspired by RANSAC that rejects putative outliers is devised for this purpose. Once the windows have been selected, the parameters of interest are estimated from the magnitude spectrograms of room impulse responses by minimizing a constrained cost function. Extensive simulation results on random shoebox rooms reveal that absorption coefficients can be efficiently recovered with the procedure, and that increasing the number of measurements improve the results while enhancing the robustness to noise and to geometrical uncertainty.**

***Index Terms*—Room Impulse Response, Absorption Coefficients, Impedance, Room Acoustics, RANSAC**

I. Introduction

When sound propagates inside a room, its interaction with reflective surfaces such as the walls, the floor and the ceiling leads to the phenomenon of reverberation, which can be an important source of nuisance for the users. The main parameters an acoustician can act on to tackle this issue are the room's *absorption coefficients* $\alpha_k[f] \in [0, 1]$, namely, the ratio of sound energy that is not reflected by surface k, within a given frequency band f. The absorption profiles of building materials are generally provided by the manufacturers after being measured in isolation under laboratory conditions. However, estimating these profiles *in situ* for acoustic diagnosis purposes remains a challenging problem to date.

A coarse but simple approach consists in using the well-known Sabine or Eyring formulas, that relates the surface-weighted mean absorption coefficient $\bar{\alpha}[f]$ to the reverberation time in that frequency band, the volume, and the total boundary surface of the room. If the unknown surfaces in the room are assumed to have similar absorption profiles, they can be roughly deduced in this way. The attractivity of this approach is that the reverberation time is a global acoustic quantity that can be easily obtained using a single measurement from a simple apparatus. However, these formulas critically rely on the assumption of a homogeneous, diffuse sound field, which is violated for most commonly encountered rooms, making the approach unsuitable in practice. An attempt to alleviate this limitation using virtually-supervised deep learning and a single room impulse response (RIR) measurement was recently proposed in [1] and validated on real data, but remains limited to the estimation of $\bar{\alpha}[f]$ rather than individual profiles.

Instead, standard techniques for *in situ* estimation use dedicated apparatus that must be carefully placed at a precisely measured position with-respect to one surface of interest in order to minimize interference with reflections from other surfaces. An extensive review of such techniques is provided in [2] and a more recent approach is presented in [3]. Such measurements need to be repeated for every surface in the room which is time consuming and not always feasible.

Given this state of affair, an attractive research direction to simplify acoustic diagnosis is to devise techniques that can estimate all the absorption profiles in a room jointly from a small set of acoustic measurements at arbitrary locations. This difficult inverse problem has received relatively little attention in the literature. To the best of the authors' knowledge, all previously proposed techniques rely on a discretization of the wave equation using, *e.g.*, boundary element methods or finite-difference time domain methods [4]–[7]. A well-known limitations of these techniques is that their computational and memory costs grow with Δ_s^D, where Δ_s is the space discretization step and D the number of space dimensions, which can quickly become intractable. Because of this, [6], [7] are limited to 2D rooms and [4], [5] are limited to frequencies below 125 Hz and 450 Hz, respectively. Another limitation of these wave-based techniques is their sensitivity to the knowledge of the *exact geometry* of the setup, namely, the positions of the sources, microphones and surfaces, which is always only approximately available in practice. Interestingly, [6] proposes a sparsity-based method to jointly estimate the source position and the walls' impedance, in the 2D setting. Alternatively, some recent methods attempted to directly map a single RIR measurement and the known geometry to the set of absorption coefficients of surfaces using deep learning on a training dataset generated by a room acoustic simulator [8], [9]. However, such method are expected to be very sensitive to the realism and the range of the simulated training set, and their generalisation capabilities have not been tested.

In this paper, a new optimization-based approach is proposed that relies on a set of RIRs obtained from source-microphone pairs at arbitrary locations in a room. The geometry of the setup is assumed known but only up to some errors of, *e.g.*, a few centimeters. This can be used to approximately estimate the times of arrival (TOAs) of the direct path and of the manifestations of acoustic reflections inside the RIRs, refered to as *echoes*, using the image source method [10]. To alleviate the difficulty of modeling the spectral phases of acoustic echoes, both due to TOA errors and to the unknown phase responses of surfaces, the RIRs are represented in the magnitude time-frequency domain. The task is then formulated as an optimization problem over a selected subset of RIR windows that are expected to contain an isolated echo, under the multiplicative constraints arising from the image source model. To improve the selection of RIR windows, a robust, probabilistic approach inspired by the random sample consensus (RANSAC) method [11] is devised. Experiments on 500 simulated random shoebox rooms reveal that the approach can jointly recover 80% of the absorption coefficients of the 6 surfaces in 16 linearly spaced frequency bands with an absolute error under 0.1, using as little as 4 RIRs. Further, increasing the number of RIRs is shown to improve the results while enhancing the robustness to noise and to geometrical uncertainty.

II. Acoustic and Signal Models

Let us consider S omnidirectional sound sources placed at known locations $\{\boldsymbol{r}_s\}_{s=1}^{S} \subset \mathbb{R}^3$ and M omnidirectional microphones placed at known locations $\{\boldsymbol{r}'_m\}_{m=1}^{M} \subset \mathbb{R}^3$ inside a room bounded by K surfaces with known geometry. It is assumed that the setup is *calibrated*, so that the discrete-time-domain effect of every source-microphone pair on a source signal can be modeled by the same, unknown, finite impulse response (FIR) $\widetilde{d}[t]$. Likewise, the effect of each surface k on an impinging signal is assumed independent of the angle of incidence and modeled by an unknown FIR $\widetilde{w}_k[t]$. Here, t is a discrete time index and the microphones' frequency of sampling in Hz is denoted by $\bar{f}$. Sound propagation from source s to microphone m is modeled using the *image source method* (ISM) [10]. For convenience, the images of a given source s are indexed by a multidimensional tuple $\boldsymbol{k} \in \mathcal{K}$ defined as follows[1]:

- $\boldsymbol{k} = ()$ (the empty tuple) denotes the true source s itself, *i.e.*, the direct propagation path,
- $\boldsymbol{k} = (k)$ denotes the image source associated to a first-order reflection on surface k,
- $\boldsymbol{k} = (k_1, k_2)$ denotes the image source associated to a second-order reflection on surface k_1 and then k_2,
- etc., up to reflection order $Q_{\max}$.

If $\boldsymbol{r}_{s,\boldsymbol{k}} \in \mathbb{R}^3$ denotes the position of the $\boldsymbol{k}$-th image of source s according to the ISM, the time of arrival (TOA) of this image source at microphone m expressed in samples is given by:

$$\tau_{s,m,\boldsymbol{k}} = \bar{f}\|\boldsymbol{r}_{s,\boldsymbol{k}} - \boldsymbol{r}'_m\|/c. \tag{1}$$

[1]Here, $\mathcal{K}$ only contains image sources that are *not occluded*. For the sake of simplicity occlusions are assumed independent of the microphone positions, although this may not be the case in non-convex rooms.

Let $\widetilde{x}_{s,m}[t]$ denote a measured room impulse response (RIR) from source s to microphone m. Given the above assumptions, it can be expressed as :

$$\widetilde{x}_{s,m}[t] = \sum_{\boldsymbol{k}\in\mathcal{K}} \widetilde{h}_{s,m,\boldsymbol{k}} \cdot (\widetilde{\phi}_{s,m,\boldsymbol{k}} * \widetilde{v}_{\boldsymbol{k}})[t] + \widetilde{e}_{s,m}[t] \tag{2}$$

where $*$ denotes discrete-time convolution and:

- $\widetilde{h}_{s,m,\boldsymbol{k}}$ is a scalar gain capturing the attenuation due to sound propagation in free field. In this paper we use $\widetilde{h}_{s,m,\boldsymbol{k}} = \bar{f}/c\tau_{s,m,\boldsymbol{k}}$. This could be modified to account for atmospheric attenuation, which is here neglected.
- $\widetilde{v}_{\boldsymbol{k}}[t]$ is an FIR referred to as an *echo*, capturing the joint effect of the source, the receiver and surfaces inside of $\boldsymbol{k}$ on the received source signal.
- $\widetilde{\phi}_{s,m,\boldsymbol{k}}$ is an all-pass fractional delay filter delaying the echo $\widetilde{v}_{\boldsymbol{k}}$ by $\tau_{s,m,\boldsymbol{k}}$, namely, $\widetilde{\phi}_{s,m,\boldsymbol{k}}[t] = \mathrm{sinc}(t - \tau_{s,m,\boldsymbol{k}})$ where $\mathrm{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$.
- $\widetilde{e}_{s,m}[t]$ is a residual term capturing the effect of image sources of higher order than $Q_{\max}$ and measurement noise.

Each echo can be further decomposed as follows:

$$\widetilde{v}_{()} = \widetilde{d}, \quad \widetilde{v}_{(k)} = \widetilde{d} * \widetilde{w}_k, \quad \widetilde{v}_{(k_1,k_2)} = \widetilde{d} * \widetilde{w}_{k_1} * \widetilde{w}_{k_2}, \ \ldots \tag{3}$$

and is assumed to be of length at most L_v in samples. Let $x_{s,m,t}[f]$ be the squared magnitude of the discrete Fourier transform (DFT) of the windowed RIR $\widetilde{x}_{s,m}[t : t + 2F - 1] \in \mathbb{R}^{2F}$, where f is a discrete frequency index. The window is here assumed to be larger than the length of echoes, *i.e.*, $2F > L_v$. For each image source $\boldsymbol{k}$, the set of *relevant RIR windows* is then defined as follows:

$$\mathcal{J}^*_{\boldsymbol{k}} = \left\{(s,m,t) | \tau_{s,m,\boldsymbol{k}} \in \mathcal{W}^-_t, \tau_{s,m,\boldsymbol{k}'\neq\boldsymbol{k}} \notin \mathcal{W}^+_t\right\} \tag{4}$$

where $\mathcal{W}^-_t = [t : t{+}2F{-}L_v]$ and $\mathcal{W}^+_t = [t{-}L_v : t{+}2F{-}1]$. In words, $\mathcal{J}^*_{\boldsymbol{k}}$ is the set of (s,m,t) triplets for which $\boldsymbol{k}$ is the *only visible image source* in the windowed RIR $\widetilde{x}_{s,m}[t : t{+}2F{-}1]$, and hence the only non-zero term in the sum over $\mathcal{K}$ in (2). By taking the squared magnitude of the DFT of (2) and using the (approximate) discrete convolution theorem, we obtain the simple expression:

$$\forall\,(s,m,t) \in \mathcal{J}^*_{\boldsymbol{k}}, \quad x_{s,m,t}[f] = h_{s,m,\boldsymbol{k}} v_{\boldsymbol{k}}[f] + e_{s,m,t}[f] \tag{5}$$

where $e_{s,m,t}[f]$ is an error term, $h_{s,m,\boldsymbol{k}} = |\widetilde{h}_{s,m,\boldsymbol{k}}|^2$ and $v_{\boldsymbol{k}} = |\,\mathrm{DFT}\{\widetilde{v}_{\boldsymbol{k}}\}|^2$. Crucially, since the fractional delay filters $\widetilde{\phi}_{s,m,\boldsymbol{k}}[t]$ are all-pass, this makes their expressions in the short-term magnitude Fourier domain disappear. Let $w_1, \ldots, w_K$ and d be the squared magnitude DFTs of $\widetilde{w}_1, \ldots, \widetilde{w}_K$ and $\widetilde{d}$, respectively. Note that w_k corresponds to the *reflection profile* of the surface k while $\alpha_k \overset{\text{def}}{=} 1 - w_k$ corresponds to its *absorption profile*, both taking values in $[0,1]^{F+1}$ (only no-negative frequencies are considered). In acoustic standards, absorption profiles are usually defined over 6 logarithmically-spaced octave bands centered at $.125, .250, \ldots, 4$ kHz, for perceptual reasons. Here, they are defined in the discrete Fourier domain over linearly-spaced positive frequencies, as this is more natural from a signal-processing and information-theoretic perspective. Nevertheless, one scale can be converted to the other using appropriate interpolation schemes. The following multiplicative relations derived from (3) can now be written:

$$v_{()} = d, \quad v_{(k)} = d \cdot w_k, \quad v_{(k_1,k_2)} = d \cdot w_{k_1} \cdot w_{k_2}, \; \dots \quad (6)$$

For each f, estimating the reflection coefficients $w_1[f], \dots, w_K[f]$ and the source-microphone gain $d[f]$ given $\{x_{s,m,t}[f]\}_{s,m,t}$ and the known TOAs (1) can now be cast as the following minimization problem[2]:

$$\underset{\substack{d[f]\geq 0,\\ w_{1:K}[f]\in[0,1]}}{\operatorname{argmin}} \sum_{\substack{\boldsymbol{k}\in\mathcal{K},\\ q(\boldsymbol{k})\leq Q}} \sum_{(s,m,t)\in\mathcal{J}^*_{\boldsymbol{k}}} \left| x_{s,m,t}[f] - h_{s,m,\boldsymbol{k}} v_{\boldsymbol{k}}[f] \right|^2 \quad (7)$$

with expressions for $v_{\boldsymbol{k}}[f]$ given by (6) and $q(\boldsymbol{k}) \in [0, Q_{\max}]$ denoting the order of image source $\boldsymbol{k}$. Note that Q, the maximum order used for the estimation, may be lower than $Q_{\max}$, the maximum order used to identify relevant RIR windows in (5). This is a non-convex, non-linear, constrained optimization problem that amounts to constrained least-square problems in individual variables. It can be solved using a properly initialized nonlinear programming solver[3]. In practice, an initial estimate of $d[f]$ can be obtained using $Q = 0$. The solution is then a simple weighted average across all relevant direct-path RIR windows. An initial estimate for each $w_k[f]$ can then be similarly obtained by fixing $d[f]$ and using $Q = 1$.

III. Geometrical Uncertainty and Robustness

In practice, the positions of the sources, microphones and surfaces in the room are never known exactly, but up to some uncertainty. Even errors of a few centimeters will have a non-negligible impact on the TOAs estimated by (1), and hence on the computation of $\mathcal{J}^*_{\boldsymbol{k}}$ which may then include spurious RIR windows that either contain a wrong echo or multiple interfering echoes, or miss some relevant RIR windows. These effects will be strengthen if the echo length L_v is underestimated. This will in turn severely degrade the results obtained by minimizing (7). To account for this, the positions of the sources, microphones and surfaces are assumed to be available up to an error with standard deviation (std) σ_{geo}, in centimeters. A Gaussian probabilistic model on TOAs is then assumed:

$$p(\tau_{s,m,\boldsymbol{k}}) = \mathcal{N}(\tau_{s,m,\boldsymbol{k}}; \mu_{s,m,\boldsymbol{k}}, \sigma^2_{\boldsymbol{k}}) \quad (8)$$

where $\mu_{s,m,\boldsymbol{k}}$ is the theoretical TOA calculated from (1) and $\sigma_{\boldsymbol{k}} = (q(\boldsymbol{k}) + 2)\sigma_{\text{geo}}\bar{f}/c$ is the std of TOA errors (assumed independently distributed), which increases linearly with the reflection order $q(\boldsymbol{k})$. The set of *most likely* relevant RIR windows can now be defined as $\mathcal{J}^{\text{top}}_{\boldsymbol{k}} = \{\text{argmax}_{(s,m,t)} \; \pi_{s,m,t,\boldsymbol{k}}\}$, where

$$\pi_{s,m,t,\boldsymbol{k}} = p(\tau_{s,m,\boldsymbol{k}} \in \mathcal{W}^-_t \cap \tau_{s,m,\boldsymbol{k}'\neq\boldsymbol{k}} \notin \mathcal{W}^+_t), \quad (9)$$

[2]Note that the nonnegativity constraints can in fact be dropped thanks to the nonnegativity of h, v, and the absence of echo mixing.

[3]For this paper, we used the `fmincon` solver of Matlab.

Algorithm 1 RANSAC-inspired approach

Input: $f, \mathcal{K}, N_{\text{iter}}, \{x_{s,m,t}[f], h_{s,m,\boldsymbol{k}}, \pi_{s,m,t,\boldsymbol{k}}\}_{s,m,t,\boldsymbol{k}}$

Output : $\{\mathcal{J}^{\text{ransac}}_{\boldsymbol{k},f}\}_{\boldsymbol{k}\in\mathcal{K}}$

1: $S := -\infty$; *// Initial score*

2: **for** $n = 1, 2, \dots, N_{\text{iter}}$ **do**

3: Draw (s, m, t) with probability prop. to $\pi_{s,m,t,()}$;

4: $\hat{d}[f] := \frac{x_{s,m,t}[f]}{h_{s,m,()}}$; *// Direct path*

5: **for** $k = 1, 2, \dots, 6$ **do**

6: Draw (s, m, t) with probability prop. to $\pi_{s,m,t,(k)}$;

7: $\hat{w}_k[f] := \frac{x_{s,m,t}[f]}{\hat{d}[f] h_{s,m,(k)}}$; *// Reflection coefficient*

8: **end for**

9: **for** $\boldsymbol{k} \in \mathcal{K}$ **do**

10: $\hat{v}_{\boldsymbol{k}}[f] := \hat{d}[f] \prod_{k\in\boldsymbol{k}} \hat{w}_k[f]$;

11: $\hat{\mathcal{J}}^{\text{ransac}}_{\boldsymbol{k},f} := \left\{ (s, m, t) \;\middle|\; \left| \frac{x_{s,m,t}[f] - \hat{v}_{\boldsymbol{k}}[f] h_{s,m,\boldsymbol{k}}}{\hat{v}_{\boldsymbol{k}}[f] h_{s,m,\boldsymbol{k}}} \right| < 0.1 \right\}$;

12: **end for**

13: $\hat{S} := \sum_{\boldsymbol{k}\in\mathcal{K}} \sum_{(s,m,t)\in\hat{\mathcal{J}}^{\text{ransac}}_{\boldsymbol{k},f}} h_{s,m,\boldsymbol{k}}$; *// Compute score*

14: **if** $\hat{S} > S$ **then**

15: $S := \hat{S}$;

16: $\forall \boldsymbol{k}, \mathcal{J}^{\text{ransac}}_{\boldsymbol{k},f} := \hat{\mathcal{J}}^{\text{ransac}}_{\boldsymbol{k},f}$;

17: **end if**

18: **end for**

and be used instead of $\mathcal{J}^*_{\boldsymbol{k}}$ in (7). Probabilities $\pi_{s,m,t,\boldsymbol{k}}$ in (9) can be computed in closed form using (8) and the erf function. In practice, windows with probabilities lower than 0.1% are discarded.

However, purely relying on $\mathcal{J}^{\text{top}}_{\boldsymbol{k}}$ may be limited as it corresponds to a much smaller set of observations than $\mathcal{J}^*_{\boldsymbol{k}}$. Moreover, regardless of geometrical uncertainty, some of the observations indexed by $\mathcal{J}^*_{\boldsymbol{k}}$ may contain faulty information due to interference with echoes of order higher than $Q_{\max}$ or due to noise. To tackle this issue, an approach combining the probabilistic geometrical model (8) with a robust window selection procedure based on the RANSAC algorithm [11] is proposed. For a fixed f, at each iteration, a set of 7 triple-indices (s, m, t) is drawn from probabilities proportional to $\pi_{s,m,t,\boldsymbol{k}}$ (for $\boldsymbol{k} = (), (1), \dots, (6)$), thus obtaining a minimal set from which it is possible to compute a tentative *room acoustic model* $(\hat{d}[f], \hat{w}_1[f], ..., \hat{w}_6[f])$ according to (7). Then, for each $\boldsymbol{k}$, the subset $\hat{\mathcal{J}}^{\text{ransac}}_{\boldsymbol{k},f}$ is computed so that it contains the triple-indices (s, m, t) for which $x_{s,m,t}[f]$ matches the room acoustic model within a relative error of 10%. $\hat{\mathcal{J}}^{\text{ransac}}_{\boldsymbol{k},f}$ is by definition the the set of *inliers*. Finally, the subset $\mathcal{J}^{\text{ransac}}_{\boldsymbol{k},f}$ containing the highest weighted number of inliers across all iterations is selected. This procedure is summarized in pseudo-code in Algorithm 1.

The score used in line 13 favors models that provide a large amount of inliers with short TOAs, thanks to the weights $h_{s,m,\boldsymbol{k}} = (\bar{f}/c\tau_{s,m,\boldsymbol{k}})^2$. Encouraging short TOAs in this way showed to improve results, because the echo density in earlier parts of RIRs is known to be lower [12], hence limiting the risk of interference between echoes. It also participates in tackling noise as the signal-to-noise ratio (SNR) of RIRs tends to increase over time. The set $\mathcal{J}_{\boldsymbol{k},f}^{\text{ransac}}$ hence obtained, which may differ for each f, can then be used as a replacement for $\mathcal{J}_{\boldsymbol{k}}^{*}$ in (7).

IV. Experiments and results

The following experiments are carried out on a dataset of synthetic shoebox ($K = 6$) RIRs simulated with the room acoustics simulator ROOMSIM [13]. Simulated RIRs are cropped to $250ms$, sampled at a rate $\bar{f} = 16$ kHz and include specular reflections up to order 20. 500 rooms are simulated with length, width and height sampled uniformly in $[3, 10] \times [3, 10] \times [2, 5]$ in meters. Each room contains S sources and M microphones whose positions are sampled uniformly in the room under the constraints of non-closeness to surfaces (1 meter) and non-mutual-closeness (1 meter). Absorption profiles defined by 6 absorption coefficients in logarithmically-spaced octave bands centered at $.125, .250, \ldots, 4$ kHz are randomly drawn as described in [1] and [9] in order to simulate realistic, diverse and representative room acoustics. The corresponding ground-truth DFT-domain profiles are obtained with an appropriate interpolation scheme. A simple Dirac is used for the source-microphone direct-path FIR of every RIR. It is not assumed known but its estimation by the proposed methods, although important to correctly estimate $w_1, \ldots, w_6$, is not specifically evaluated here as the focus of this study is the estimation of absorption profiles. White Gaussian noise is added to the RIRs to achieve a fixed peak signal-to-noise (PSNR). In order to model geometrical uncertainty, the positions of the sources and of the microphones as well as the dimensions of the room that are used for computing TOAs in (1) are not those used for generating the data. Their differences follow a centered normal distribution of std σ_{geo}. In all experiments, the true value of σ_{geo} was used to compute the sets $\mathcal{J}_{\boldsymbol{k}}^{\text{top}}$ and $\mathcal{J}_{\boldsymbol{k}}^{\text{ransac}}$, implying that the precision of the geometrical measurement device is known.

Spectrograms are computed using DFT windows of size $2F = 32$ samples (2 ms) and a hop size of 1 sample. In this study, echoes are assumed to be of length $L_v = 8$ samples (0.5 ms). Image sources in $\mathcal{K}$ are computed up to order $Q_{max} = 2$ only, as explicitly modeling higher order echoes tended to degrade results. The ground truth absorption profiles $\alpha[f] = 1 - w[f]$ for $f = 0, \ldots, F$ (DFT resolution) are obtained by linear interpolation of the octave-band absorption profiles in the same way as ROOMSIM. The corresponding linearly-spaced frequency bands are approximately centered at 0 Hz, 470 Hz, …, 8 kHz. Preliminary experiments revealed that none of the proposed approaches were able to correctly estimate absorption values at the lowest (DC) frequency band $f = 0$, yielding values close to random. Hence, this frequency band is omitted in the following results. This limitation can be explained by the relatively short DFT window size employed in this study. On the other hand, increasing window sizes showed to degrade results by decreasing the number of windows containing isolated echoes. The opportunity to consider adaptive window sizes is left for future work.

Subset	σ_{geo}	$Q = 1$		$Q = 2$	
		MAE	CE (%)	MAE	CE (%)
$\mathcal{J}_{\boldsymbol{k}}^{*}$	0 cm	0.095	79.7	0.128	71.1
	2 cm	0.108	74.7	0.132	69.6
$\mathcal{J}_{\boldsymbol{k}}^{\text{top}}$	2 cm	0.128	71.7	0.119	69.7
$\mathcal{J}_{\boldsymbol{k}}^{\text{ransac}}$	2 cm	0.088	82.9	**0.082**	**84.6**

Table I: Mean absolute error (MAE) and percentage of correct estimates (CE) obtained by minimizing (7) over different RIR window subsets and over image sources up to order Q.

Two metrics are used to evaluate the estimation of absorption coefficients over all frequencies, surfaces and rooms: the mean absolute error (MAE) and the total percentage of *correctly estimated* (CE) coefficients, *i.e.* with an error smaller than 0.1 (recall that absorption coefficients take values between 0 and 1). This threshold is meant to reflect what would be an appropriate tolerance for acoustic diagnosis purpose.

The first experiment compares the results obtained by minimizing the proposed objective (7) over the three presented RIR window subsets, namely, $\mathcal{J}_{\boldsymbol{k}}^{*}$, $\mathcal{J}_{\boldsymbol{k}}^{\text{top}}$ and $\mathcal{J}_{\boldsymbol{k}}^{\text{ransac}}$. Here, $S = 3$ sources and $M = 3$ microphones (9 RIRs) are used in each of the 500 rooms, and the PSNR is fixed to 50 dB. Note that in some rare cases ($< 1\%$ for $\mathcal{J}^{\text{ransac}}$ using 4 RIRs or more) a RIR subset $\mathcal{J}_{(k)}$ can be empty, making the estimation of α_k impossible or very poor. In such cases, the estimate $\hat{\alpha}_k(f)$ is arbitrarily set to the middle value 0.5.

The first row of table I shows results obtained using $\mathcal{J}_{\boldsymbol{k}}^{*}$ and assuming the geometry is exactly known, *i.e.*, $\sigma_{\text{geo}} = 0$ cm. Using echoes of order 1 only ($Q = 1$), a MAE just below 0.1 is obtained and nearly 80% of absorption coefficients are correctly estimated, demonstrating the viability of the proposed optimization scheme (7). However, as expected, both metrics are significantly worsened when geometrical perturbations (σ_{geo} = 2 cm) are added. In addition, in both cases, the method is not able to leverage RIR windows containing echoes of order 2 ($Q = 2$), which only degrade results. This owes to the fact that higher order echoes are likelier to interfere with those of order 2, which cannot be taken into account by this direct, deterministic approach. Using only the most likely window for each echo according to the proposed probabilistic model ($\mathcal{J}_{\boldsymbol{k}}^{\text{top}}$) slightly improve matters when $Q = 2$, but degrades results for $Q = 1$ due to the ensuing reduction of available observations. Finally, the proposed RANSAC-based uncertainty-aware solution outperformed the two others in all cases, lowering the MAE to nearly 0.08 and increasing the correct estimation rate to nearly 85% using echoes up to order 2. This shows that the proposed scheme of robustly selecting from a larger pool of RIR windows can successfully exclude spurious observations affected by mis-modeled interference between echoes and measurement noise.

The second experiment focuses on the best performing method in the first one, namely, $\mathcal{J}_{\boldsymbol{k}}^{\text{ransac}}$, $Q = 2$. It jointly studies the influence of the number of sources S and microphones M on its performances and its robustness to measurement noise and geometrical error. As expected, Fig. 1 reveals that reducing the PSNR and increasing σ_{geo} systematically decreases performance. On the other hand, increasing the number of available RIRs per room consistently improves both the MAE and CE metrics, eventually compensating the degradation. This suggests that the proposed approach succeeds in selecting the most relevant RIR windows despite their increasing number. Note that with $PSNR = 50$ dB and $\sigma_{geo} = 2$ cm, using only 4 RIRs with $(S, M) = (2, 2)$ suffices to reach a satisfying MAE close to 0.1, with 78% of correct estimates. Interestingly, similar results are obtained for a somewhat more practical setup consisting of $S = 1$ source and $M = 5$ microphones.

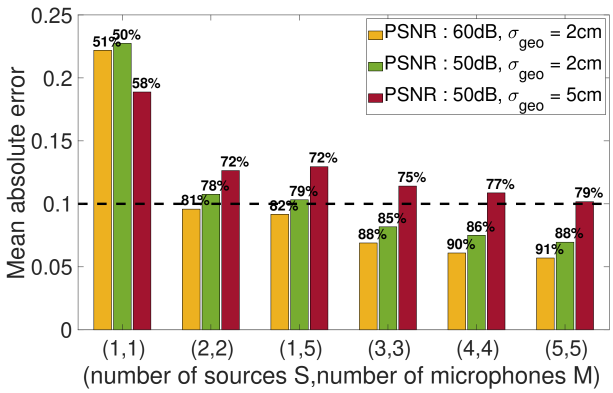

Fig. 1: Mean absolute error and percentage of correct estimates achieved with the proposed RANSAC selection procedure with $Q = 2$ for different numbers of sources and microphones and various levels of noise and geometrical error.

V. Conclusion

In this work, the inverse problem of estimating the surface absorption profiles of a room using multiple impulse responses was investigated. This estimation was achieved by optimizing an objective function in the magnitude time-frequency domain on a subset of relevant room impulse response windows selected with a geometrically-guided approach. Results showed that this approach performs accurate estimations of the absorption coefficients above 400 Hz, even when geometrical errors and ambient noise are included. The probabilistic and robust version of this approach based on the RANSAC algorithm was shown to improve estimation by leveraging additional information provided by second-order reflections. Increasing the number of RIRs was shown to significantly reduce errors and to improve the robustness of the approach.

Future work will include applications of the presented approach to real data, which may be difficult due to longer source-microphone responses $\widetilde{d}[t]$. Deep generative models optimizing an objective function similar to (7) will be developed to improve the generalization to real data using both labelled simulated data and unlabelled real measurements during training. Ultimately, the aim is to perform estimation *blindly* from RIRs, *i.e.*, with partial or no geometrical knowledge, by jointly estimating the absorption profiles and the geometrical parameters.

References

[1] C. Foy, A. Deleforge, and D. Di Carlo, "Mean absorption estimation from room impulse responses using virtually supervised learning," *The Journal of the Acoustical Society of America*, vol. 150, no. 2, pp. 1286–1299, 2021.

[2] E. Brandão, A. Lenzi, and S. Paul, "A review of the in situ impedance and sound absorption measurement techniques," *Acta Acustica united with Acustica*, vol. 101, no. 3, pp. 443–463, 2015.

[3] J. Hald, W. Song, K. Haddad, C.-H. Jeong, and A. Richard, "In-situ impedance and absorption coefficient measurements using a double-layer microphone array," *Applied Acoustics*, vol. 143, pp. 74–83, 2019.

[4] G. P. Nava, Y. Yasuda, Y. Sato, and S. Sakamoto, "On the in situ estimation of surface acoustic impedance in interiors of arbitrary shape by acoustical inverse methods," *Acoustical science and technology*, vol. 30, no. 2, pp. 100–109, 2009.

[5] N. Antonello, T. van Waterschoot, M. Moonen, and P. Naylor, "Evaluation of a numerical method for identifying surface acoustic impedances in a reverberant room," in *Proc. of the10th European Congress and Exposition on Noise Control Engineering*, 2015, pp. 1–6.

[6] N. Bertin, S. Kitić, and R. Gribonval, "Joint estimation of sound source location and boundary impedance with physics-driven cosparse regularization," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 6340–6344.

[7] Y. Okawa, Y. Watanabe, Y. Ikeda, and Y. Oikawa, "Estimation of acoustic impedances in a room using multiple sound intensities and fdtd method," in *Advances in Acoustics, Noise and Vibration - Proceedings of the 27th International Congress on Sound and Vibration, ICSV 2021*. Silesian University Press, 2021.

[8] W. Yu and W. B. Kleijn, "Room acoustical parameter estimation from room impulse responses using deep neural networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 436–447, 2020.

[9] S. Dilungana, A. Deleforge, C. Foy, and S. Faisan, "Learning-based estimation of individual absorption profiles from a single room impulse response with known positions of source, sensor and surfaces," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 263, no. 1. Institute of Noise Control Engineering, 2021, pp. 5623–5630.

[10] 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.

[11] K. G. Derpanis, "Overview of the RANSAC algorithm," *Image Rochester NY*, vol. 4, no. 1, pp. 2–3, 2010.

[12] J. S. Abel and P. Huang, "A simple, robust measure of reverberation echo density," in *Audio Engineering Society Convention 121*. Audio Engineering Society, 2006.

[13] S. M. Schimmel, M. F. Muller, and N. Dillier, "A fast and accurate "shoebox" room acoustics simulator," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp. 241–244.