JASA EXPRESS LETTERS ARTICLE asa.scitation.org/journal/jel

Integrating atmospheric fluctuations into the image source method

Jeremy Lawrence,[1,a] Zeyu Xu,[1,b] Andreas Walther,[2,c] Nils Peters,[3,d] and Emanuël A. P. Habets[1,2,e]
[1]International Audio Laboratories Erlangen, Germany
[2]Fraunhofer Institute for Integrated Circuits (IIS), Erlangen, Germany
[3]Trinity College, The University of Dublin, Ireland

Abstract: Successive room impulse response (RIR) measurements exhibit subtle variations, mainly caused by atmospheric fluctuations due to spatiotemporal air-temperature differences. This work extends the image source method by incorporating temperature-induced variability through global deterministic drift and stochastic local volatility. Measurements in a shoebox room show that the proposed model reproduces key behaviors observed in practice, including the structure of differential RIRs, temporal variance of early reflections and coherence of late reverb, which fluctuation-free simulations fail to capture. The method increases realism without increasing the algorithmic complexity, enabling simulations that better reflect the variability encountered by algorithms in real measurement scenarios.

[Editor: Jonathan Andrew Hargreaves] https://doi.org/10.1121/10.0044473

Received: 23 March 2026 **Accepted:** 8 July 2026 **Published Online:** 27 July 2026

1. Introduction

Most acoustic simulators based on the image-source method (ISM) generate identical room impulse responses (RIRs) across repeated runs for identical source–receiver configurations and room geometries (Allen and Berkley, 1979). When variability is desired, additive white Gaussian noise (AWGN) is typically added to the synthesized RIR to emulate measurement noise. However, real-world measurements exhibit subtle variations between successive RIR estimates that increase with frequency and cannot be attributed solely to measurement noise, as demonstrated in several studies (Prawda, 2025; Prawda *et al.*, 2023, 2024). One major contributor to these differences is atmospheric variability within rooms, in particular, local air temperature fluctuations (Melikov *et al.*, 1997). These fluctuations introduce additional uncertainty in the RIRs and may therefore degrade beamforming, room geometry estimation, and especially acoustic-based human localization algorithms, which localize silent humans via minute RIR variations induced by the human body (Bhattacharjee *et al.*, 2025; Kim *et al.*, 2020; Lawrence *et al.*, 2025; Wang *et al.*, 2024; Yang *et al.*, 2022).

To systematically study the degradation induced by temperature fluctuations, a simple and flexible simulation model is desirable. In Prawda *et al.* (2024), these effects were quantified in terms of *drift*, describing global temperature changes, and *volatility*, describing local, spatially varying fluctuations. Atmospheric variability was investigated in simulation using stochastic reverberation models that are valid only for late reverberation (Badeau, 2018, 2019), and therefore do not capture the impact of these effects on early reflections and the direct sound. This limitation is particularly critical for applications in which the early reflections play a crucial role, e.g., the aforementioned acoustic-based human localization.

In this letter, we propose an extension of the ISM (Allen and Berkley, 1979) that incorporates both drift and volatility. The proposed model is implemented in the open-source diffraction-enhanced image source method (DEISM) simulator (Xu *et al.*, 2024). We validate the approach using real-world measurements, with a particular emphasis on early reflections. Beyond listener tracking, the proposed framework enables more realistic data augmentation for learning-based methods and improves simulation fidelity for spatial audio and beamforming applications, all without increasing the computational complexity.

[a]Corresponding author: jeremy.lawrence@fau.de
[b]Email: zeyu.xu@audiolabs-erlangen.de
[c]Email: andreas.walther@iis.fraunhofer.de
[d]Email: Nils.Peters@tcd.ie
[e]Email: emanuel.habets@audiolabs-erlangen.de

2. Background

2.1 Instantaneous coherence

In Secs. 2.2 and 4.3, we examine the similarity between successive RIR estimates $\hat{h}_1(t)$ and $\hat{h}_2(t)$, obtained with unchanged loudspeaker and microphone positions and with no modifications performed to the room, where t denotes time. The (real-valued) RIR estimates are modeled as

$$\hat{h}_1(t) = h_1(t) + n_1(t) \quad \text{and} \quad \hat{h}_2(t) = h_2(t) + n_2(t), \tag{1}$$

where $h_{1/2}(t)$ is the RIR and $n_{1/2}(t)$ is stationary additive noise. We assume $h_{1/2}(t)$ and $n_{1/2}(t)$ to be zero-mean and to have power spectral densities (PSDs) $S_{h_1h_1/h_2h_2}(t,f)$ and $S_{n_1n_1/n_2n_2}(t,f) = S_{nn}(f)$, where f indicates frequency. Furthermore, we assume that all terms are mutually uncorrelated except for $h_1(t)$ and $h_2(t)$, and that both RIRs exhibit identical PSDs, i.e., $S_{h_1h_1}(t,f) = S_{h_2h_2}(t,f)$. The latter holds under the assumption of a constant average room temperature and the fact that local temperature fluctuations constitute a random process that affects $h_1(t)$ and $h_2(t)$ equally, resulting in identical expected auto-correlations and PSDs. Following McCarthy (1997), the instantaneous coherence between $\hat{h}_1(t)$ and $\hat{h}_2(t)$ can be expressed as

$$\gamma^2_{\hat{h}_1\hat{h}_2}(t,f) = \frac{|S_{\hat{h}_1\hat{h}_2}(t,f)|^2}{S_{\hat{h}_1\hat{h}_1}(t,f)S_{\hat{h}_2\hat{h}_2}(t,f)}, \tag{2}$$

where $S_{\hat{h}_1\hat{h}_2}(t,f)$ is the cross-spectral density between the RIR estimates. Based on the modeled RIR estimates in Eq. (1) and the previously introduced assumptions regarding term-wise correlations, Eq. (2) can be reformulated as (Benesty *et al.*, 2008; Prawda *et al.*, 2024),

$$\gamma^2_{\hat{h}_1\hat{h}_2}(t,f) = \left(\frac{\mathrm{SNR}(t,f)}{1+\mathrm{SNR}(t,f)}\right)^2 \cdot \left(\frac{|S_{h_1h_2}(t,f)|}{S_{h_1h_1}(t,f)}\right)^2 = \gamma^2_{\mathrm{SNR}}(t,f) \cdot \gamma^2_{h_1h_2}(t,f), \tag{3}$$

where $\mathrm{SNR}(t,f) = S_{h_1h_1}(t,f)/S_{nn}(t,f)$. Here, $\gamma^2_{\mathrm{SNR}}(t,f)$ describes the coherence in terms of spectral signal-to-noise ratio (SNR) between $h_1(t,f)$ and $n_1(t,f)$, and $\gamma^2_{h_1h_2}(t,f)$ denotes the coherence in terms of similarity between $h_1(t,f)$ and $h_2(t,f)$ without any noise present.

2.2 Atmospheric fluctuations

To model the atmospheric fluctuations, we follow the derivation from Prawda *et al.* (2024) by discretizing the room volume into equally sized small voxels, where the (temperature-dependent) speed of sound within the k-th voxel is $c_k(t_1)$ and $c_k(t_2)$ at time instants t_1 and t_2. For a given acoustic path, the times of arrival (TOAs) $\tau(t_1)$ and $\tau(t_2)$ will slightly differ, resulting in a TOA difference of

$$\Delta\tau(t_1,t_2) = \tau(t_1) - \tau(t_2) = \sum_{k=1}^{K} l_k \left(\frac{1}{c_k(t_1)} - \frac{1}{c_k(t_2)}\right) = \sum_{k=1}^{K} l_k \Delta\zeta_k(t_1,t_2), \tag{4}$$

where K is the number of voxels traversed and l_k denotes the path length travelled in the k-th voxel. If we assume the change factors $\Delta\zeta_k(t_1,t_2)$ to be a set of K independent and identically distributed random variables with finite mean and variance, $\Delta\tau(t_1,t_2)$ becomes a Generalized Wiener Process for large K, i.e.,

$$\Delta\tau(t_1,t_2) \sim \mathcal{N}\left(\eta(t_1,t_2)\bar{\tau}, \left(\vartheta(t_1,t_2)\sqrt{\bar{\tau}}\right)^2\right), \tag{5}$$

where $\mathcal{N}(\mu,\sigma^2)$ is the normal distribution with mean μ and variance σ^2, $\bar{\tau} = \mathbb{E}[\tau(t_1)]$ denotes the TOA assuming a constant speed of sound (relative to the average room temperature at time t_1), and $\eta(t_1,t_2)$ and $\vartheta(t_1,t_2)$ are the drift and volatility (Durrett, 2019; Prawda *et al.*, 2024). Drift captures the change in the mean room temperature between t_1 and t_2, and volatility models the intensity of local temperature fluctuations over the same time interval. As the volume of the voxels tends towards zero, K will be large regardless of the acoustic path length; therefore, all TOA differences can be modeled according to Eq. (5). Note that the assumptions regarding independent and identical distribution regarding $\Delta\zeta_k(t_1,t_2)$ can be further weakened to different distributions (Petrov, 2000) and weak correlation (Sunklodas, 2000).

The drift between two real-world RIRs can be estimated by computing the windowed fractional delay between the signals over time as in Postma and Katz (2016). Since a change in average room temperature causes all propagation times to stretch by a fixed proportion, the fractional delay grows linearly over time. The (dimensionless) drift is equal to the slope of the delay (in seconds) per unit of time. The volatility, given in $s/\sqrt{s} = \sqrt{s}$, between two real-world RIRs can be quantified by modeling the expected coherence following (Prawda *et al.*, 2024). In particular, the expected cross-correlation [evaluated across the random $\Delta\tau(t_1,t_2)$ for a specific $\bar{\tau}$] between two signals $d_1(t) = \delta(t-\bar{\tau})$ and $d_2(t) = \delta(t-\bar{\tau}+\Delta\tau(t_1,t_2))$ is given by a normal density, i.e.,

$$R_{d_1d_2}(\bar{\tau},\Delta\tau(t_1,t_2)) = \frac{1}{\vartheta(t_1,t_2)\sqrt{\bar{\tau}2\pi}} e^{-1/2\left(\Delta\tau(t_1,t_2)-\eta(t_1,t_2)/\vartheta(t_1,t_2)\sqrt{\bar{\tau}}\right)^2}. \tag{6}$$

The expected cross-spectral density is computed as[1]

$$S_{d_1d_2}(\overline{\tau},f) = \mathcal{F}_{\Delta\tau(t_1,t_2)}\{R_{d_1d_2}(\overline{\tau},\Delta\tau(t_1,t_2))\} = e^{-i\eta(t_1,t_2)2\pi f} e^{-1/2\left[\overline{\tau}\vartheta(t_1,t_2)^2 4\pi^2 f^2\right]}, \tag{7}$$

where $\mathcal{F}.\{\cdot\}$ denotes the Fourier transform and i is the imaginary unit. If we assume each acoustic path of an RIR to be uncorrelated, the instantaneous coherence at time t between two RIRs affected by temperature fluctuations can be modeled with the instantaneous coherence between $d_1(t)$ and $d_2(t)$ for $\overline{\tau} = t$, i.e.,

$$\gamma^2_{h_1h_2}(t,f) \approx \frac{|S_{d_1d_2}(t,f)|^2}{S_{d_1d_1}(t,f)S_{d_2d_2}(t,f)} = |S_{d_1d_2}(t,f)|^2 = e^{-t\vartheta(t_1,t_2)^2 4\pi^2 f^2}. \tag{8}$$

Using Eq. (3), the volatility between two RIR estimates can be computed as

$$\vartheta(t_1,t_2) = \underset{\vartheta(t_1,t_2)}{\arg\min}\left|\int_0^{t_{\lim}}\int_{f_1}^{f_2}\left(\gamma^2_{\hat{h}_1\hat{h}_2}(t,f) - \gamma^2_{\mathrm{SNR}}(t,f)\,|S_{d_1d_2}(t,f)|^2\right)\mathrm{d}f\,\mathrm{d}t\right|, \tag{9}$$

where $t_{\lim}$ specifies the temporal integration limit, and $f_{1/2}$ define the frequency interval over which the volatility is evaluated. Note that Eq. (3) is valid only for $\eta(t_1,t_2) = 0$, as otherwise $S_{h_1h_1}(t,f) = S_{h_2h_2}(t,f)$ is no longer satisfied. Consequently, drift compensation (DComp), such as time-stretching through resampling (Postma and Katz, 2016), is required before volatility estimation.

2.3 Differential RIR

The differential RIR is the difference between two RIR estimates at different time instances and given by two studies (Lawrence *et al.*, 2025; Svensson and Nielsen, 1999),

$$\Delta\hat{h}_1^{(\Delta t)}(t) = \hat{h}_1(t) - \hat{h}_2(t) = h_1(t) - h_2(t) + n_1(t) - n_2(t), \tag{10}$$

where Δt indicates the time elapsed between the RIR estimates $\hat{h}_1(t)$ and $\hat{h}_2(t)$. Since $n_1(t)$ and $n_2(t)$ are uncorrelated, the noise power of $n_1(t) - n_2(t)$ is $2\sigma_n^2$, where σ_n^2 is the noise power of the individual noise terms. Although one might expect $|h_1(t) - h_2(t)| \ll |n_1(t) - n_2(t)|$ and therefore differential RIRs to consist of only noise in a static environment, measurements from real-world environments instead reveal clear structural correlations between the RIRs and their differential counterparts. In Sec. 4.3, we show that such correlations do not arise in simulated RIRs unless temperature fluctuations are modeled.

3. Extended image source method

The ISM simulates an RIR given the loudspeaker and microphone locations $\boldsymbol{r}_l = [x_l, y_l, z_l]$ and $\boldsymbol{r}_m = [x_m, y_m, z_m]$ under idealized conditions. The room is assumed to be cuboid-shaped with length, width, and height given by L_x, L_y, and L_z, and the coordinate origin is placed at the room corner, resulting in the room occupying the first octant. Each of the walls has reflection coefficients β_{x_1}, β_{x_2}, β_{y_1}, β_{y_2}, β_{z_1}, and β_{z_2}, where the subscripts a_1 and a_2 ($a \in \{x,y,z\}$) of $\beta_{(\cdot)}$ correspond to the walls at $a = 0$ and $a = L_a$. By defining triples $\boldsymbol{p} = (p_x, p_y, p_z)$ and $\boldsymbol{q} = (q_x, q_y, q_z)$, where each element of $\boldsymbol{p}$ and $\boldsymbol{q}$ takes on integer values [0, 1] and [−N, N], respectively, we obtain two sets $\mathcal{P}$ and $\mathcal{Q}$ of size 8 and $(2N+1)^3$. These indicate mirroring of the image source with respect to the room walls intersecting the origin and the number of wall-pair traversals along each axis, respectively. Note that the value of N bounds the maximum reflection order, but does not equal it. The RIR between the loudspeaker and the microphone can now be formulated, with both transducers modeled as omnidirectional, as (Habets, 2006; Xu *et al.*, 2024)

$$h(\boldsymbol{r}_m, \boldsymbol{r}_l, t) = \sum_{\boldsymbol{p}\in\mathcal{P}}\sum_{\boldsymbol{q}\in\mathcal{Q}} \beta_{\mathbf{p},\mathbf{q}} \frac{\delta(t - \tau_{\boldsymbol{p},\boldsymbol{q}}(\boldsymbol{r}_m, \boldsymbol{r}_l))}{4\pi||\boldsymbol{R}_{\boldsymbol{p},\boldsymbol{q}}(\boldsymbol{r}_m, \boldsymbol{r}_l)||_2}, \tag{11}$$

where $\delta(\cdot)$ is the Dirac delta function, $||\cdot||_2$ is the ℓ_2-norm, $\beta_{\mathbf{p},\mathbf{q}} = \beta_{x_1}^{|q_x-p_x|}\beta_{x_2}^{|q_x|}\beta_{y_1}^{|q_y-p_y|}\beta_{y_2}^{|q_y|}\beta_{z_1}^{|q_z-p_z|}\beta_{z_2}^{|q_z|}$ is the attenuation factor associated with the current image source, $\boldsymbol{R}_{\boldsymbol{p},\boldsymbol{q}}(\boldsymbol{r}_m, \boldsymbol{r}_l)$ is the vector from the current image source to the microphone, given by

$$\boldsymbol{R}_{\boldsymbol{p},\boldsymbol{q}}(\boldsymbol{r}_m, \boldsymbol{r}_l) = [x_m - x_l + 2p_x x_l - 2q_x L_x,\; y_m - y_l + 2p_y y_l - 2q_y L_y,\; z_m - z_l + 2p_z z_l - 2q_z L_z]^\top. \tag{12}$$

$\tau_{\boldsymbol{p},\boldsymbol{q}}(\boldsymbol{r}_m, \boldsymbol{r}_l)$ is the corresponding TOA, computed as

$$\tau_{\boldsymbol{p},\boldsymbol{q}}(\boldsymbol{r}_m, \boldsymbol{r}_l) = \frac{||\boldsymbol{R}_{\boldsymbol{p},\boldsymbol{q}}(\boldsymbol{r}_m, \boldsymbol{r}_l)||_2}{c}, \tag{13}$$

and c is the (predefined and constant) speed of sound.

We now extend the standard ISM to simulate an RIR with given drift and volatility values η and ϑ. Specifically, for each image source $(\boldsymbol{p},\boldsymbol{q})$, we replace the deterministic TOA $\tau_{\boldsymbol{p},\boldsymbol{q}}(\boldsymbol{r}_m, \boldsymbol{r}_l)$ in Eq. (11) with a perturbed TOA

$\tau_{p,q}(\boldsymbol{r}_m, \boldsymbol{r}_l) + \varepsilon(\tau_{p,q}(\boldsymbol{r}_m, \boldsymbol{r}_l), \eta, \vartheta)$, where ε is an independent random perturbation drawn according to Eq. (5) with $\overline{\tau} = \tau_{p,q}(\boldsymbol{r}_m, \boldsymbol{r}_l)$. This yields

$$\tilde{h}(\boldsymbol{r}_m, \boldsymbol{r}_l, t) = \sum_{p \in \mathcal{P}} \sum_{q \in \mathcal{Q}} \beta_{\mathbf{p},\mathbf{q}} \frac{\delta(t - \tau_{p,q}(\boldsymbol{r}_m, \boldsymbol{r}_l) - \varepsilon(\tau_{p,q}(\boldsymbol{r}_m, \boldsymbol{r}_l), \eta, \vartheta))}{4\pi||\boldsymbol{R}_{p,q}(\boldsymbol{r}_m, \boldsymbol{r}_l)||_2}, \tag{14}$$

where

$$\varepsilon(\tau_{p,q}(\boldsymbol{r}_m, \boldsymbol{r}_l), \eta, \vartheta) \sim \mathcal{N}\left(\eta \tau_{p,q}(\boldsymbol{r}_m, \boldsymbol{r}_l), \left(\vartheta \sqrt{\tau_{p,q}(\boldsymbol{r}_m, \boldsymbol{r}_l)}\right)^2\right). \tag{15}$$

Note that the perturbations are drawn independently across image sources, corresponding to the assumption that distinct acoustic paths traverse independent sets of voxels. This approximation becomes less accurate for paths that share (long) common segments. As an example, the proposed extension was incorporated into DEISM, which is publicly available at https://github.com/audiolabs/DEISM. The extension requires only $\mathcal{O}(1)$ additional computation per path, leaving the overall algorithmic complexity unchanged.

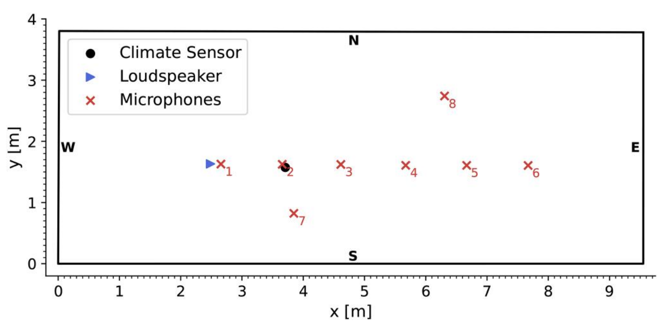
If two RIRs are simulated with the same volatility ϑ, the perturbations ε_1 and ε_2 are independent, so the variance of their difference $\varepsilon_1 - \varepsilon_2$ is $2\vartheta^2 \tau_{p,q}(\boldsymbol{r}_m, \boldsymbol{r}_l)$. To obtain a desired inter-simulation volatility $\vartheta(t_1, t_2)$, the per-simulation volatility must therefore be set to $\vartheta = \vartheta(t_1, t_2)/\sqrt{2}$. Alternatively, one may assign $\vartheta = 0$ to one simulation and $\vartheta = \vartheta(t_1, t_2)$ to the other. This yields the same inter-simulation volatility but renders one RIR deterministic, which may introduce artifacts when multiple pairs share an identical, unperturbed reference, e.g., in data augmentation scenarios. Similarly, to simulate a given drift $\eta(t_1, t_2)$ between two RIRs, we can assign $\eta = 0$ to one simulation and $\eta = \eta(t_1, t_2)$ to the other. Equivalently, the speed of sound from Eq. (13) may be modified to $c = \tilde{c}/[1 + \eta(t_1, t_2)]$ for one simulation, where $\tilde{c}$ is the speed of sound used for the other simulation.

In practical use, $\eta(t_1, t_2)$ can be selected to be proportional to the expected global temperature variation, and thus to the change in the environment's speed of sound, between times t_1 and t_2. Since empirical data on typical volatility values in real-world scenarios remain limited, $\vartheta(t_1, t_2)$ can be chosen either from a short calibration sequence in the real-world target environment or arbitrarily set to increase with the growing time difference between t_1 and t_2. The values we observed, together with those reported in Prawda *et al.* (2024), range from 2×10^{-6} to $2 \times 10^{-5}\ \sqrt{\mathrm{s}}$, though further measurements across diverse environments may reveal different characteristic levels.

4. Evaluation

4.1 Experimental setup

We estimated real-world RIRs as a direct comparison with the extended ISM. The measurements were taken in a shoebox-shaped room with dimensions 9.54 m × 3.82 m × 2.56 m with an estimated reverberation time (T_{60}) of approximately 1.9 s. The T_{60} was estimated using backward integration (Schroeder, 1965) and subsequent estimation of the slope of the linear part of the RIR's energy decay. A Neumann KH 80 DSP loudspeaker (Neumann, Berlin, Germany) and eight dbx RTA-M measurement microphones (dbx, Sandy, UT) were placed in the room according to the floor plan in Fig. 1, with the loudspeaker oriented toward the east side of the room. The positions of the loudspeaker, microphones, and room boundaries were measured using a laser rangefinder, and the measurement setup was deliberately configured in an asymmetric layout to increase temporal separation between early reflections. The microphone recording levels were level-matched within the audio interface (MOTU Stage B-16; MOTU, Cambridge, MA) to ±0.25 dB at 250 Hz and ±0.35 dB at 1 kHz using a GRAS 42AG sound calibrator (GRAS, Holte, Denmark). The recordings exhibited a SNR of approximately 40 dB and the room temperature was constantly monitored during the experiment using a SensorTec UFT75-ST3 temperature and humidity sensor (Meltec Systems, Netphen, Germany). During the experiment, the measured temperature slowly drifted from 20.2°C to 20.1 °C and the relative humidity from 39.0 % to 39.2 %. The absolute humidity stayed



	x [m]	y [m]	z [m]
●	3.707	1.574	0.890
▶	2.487	1.631	1.145
$\times_1$	2.657	1.628	1.158
$\times_2$	3.658	1.624	1.161
$\times_3$	4.614	1.622	1.158
$\times_4$	5.673	1.608	1.157
$\times_5$	6.668	1.607	1.152
$\times_6$	7.671	1.606	1.150
$\times_7$	3.847	0.823	1.155
$\times_8$	6.307	2.741	1.156

Fig. 1. Experimental setup for both real-world and simulated RIRs.

constant at $6.8\,\mathrm{g/m^3}$. The room's heating system was switched off for the entire measurement session, and no disturbances such as ventilation, door movements, or open windows occurred. Before performing the first measurement, the room was allowed to settle undisturbed for 15 min.

The RIRs between the loudspeaker and each microphone were simultaneously estimated using 1 s exponential sine sweeps spanning 60 Hz–20 kHz, incorporating a 50–60 Hz fade-in and a 20–24 kHz fade-out. Each sweep was followed by 3 s of silence, and the RIRs were estimated via deconvolution of the recorded signals (Farina, 2000). Additionally, the RIR estimates were bandpass-filtered after deconvolution using a zero-phase, second-order-sections 5th-order Butterworth filter. First, a high-pass filter with cutoff 60 Hz was applied, followed by a low-pass filter with cutoff 20 kHz. Both filtering stages used forward-backward processing (`scipy.signal.sosfiltfilt` in Python) to eliminate phase distortion and ensure a maximally flat passband. In total, 400 RIRs were captured at a sampling rate of 48 kHz, 50 for each microphone, over the course of approximately 8.6 min. The measurement sequence consisted of 10 back-to-back sweeps, followed by 80 s of silence, repeated five times. Both the extended pause and the 10-sweep grouping were introduced to provide a trade-off between maintaining continuous RIR estimation and reducing loudspeaker heating from biasing the results.

4.2 Simulation parameters

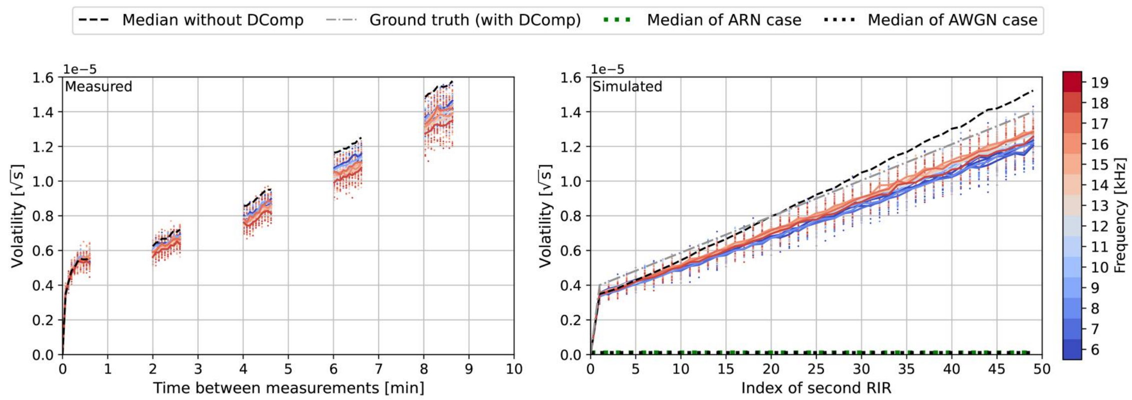
We generate simulated RIRs using the DEISM introduced in Xu *et al.* (2024), noting that any ISM-based simulator could be used instead. Among other features, DEISM includes the possibility to use complex impedances and angle-dependent reflection coefficients. The room was modeled as an ideal shoebox with the same dimensions as the real environment, and both the loudspeaker and microphones were treated as ideal omnidirectional sources and receivers. The wall acoustic impedances were set to the real-world values inferred from the measured reverberation time, following the procedure in Prinn and Badeau (2026). For each loudspeaker-microphone pair, we simulated 50 RIRs of length 1 s with a maximum reflection order of 75, using the real microphone coordinates. The sampling rate was 48 kHz, and the speed of sound was fixed at 343.5 m/s, matching the value estimated from the first measured RIRs. During the simulations, the drift parameter η from Eq. (14) was set to increase linearly from 0 to 3×10^{-4}. The volatility parameter ϑ was set to $0\sqrt{\mathrm{s}}$ for the first simulation and then increased linearly from 4×10^{-6} to $1.4 \times 10^{-5}\sqrt{\mathrm{s}}$. These parameter ranges were selected such that the simulations mimic the observed real-world late reverberation behavior.

To ensure comparable SNR between the simulations and the measurements, we added the measured noise from each real-world RIR to its corresponding simulated RIR at the same SNR. The noise was extracted from the late tail of each measured RIR, where the signal had already decayed below the noise floor. Finally, we applied the same bandpass filter used in the real-world processing chain. For comparison, we also generated two fluctuation-free simulation sets: one augmented with additive real-world noise (ARN) and one with AWGN, both added at identical SNR and followed by the same bandpass filtering. We refer to these as the ARN and AWGN cases.

4.3 Comparison and discussion of results

The comparison between real-world and simulated RIRs is carried out by examining the volatility estimates, the behavior of the differential RIRs, and the variance of the early reflections. To compare the volatility estimates, we first apply DComp following Postma and Katz (2016), reducing the residual drift to $\eta(t_1, t_2) < 5 \times 10^{-7}$, corresponding to a reduction in maximum drift from 14.4 down to 0.024 samples per second. Volatility is then estimated using Eq. (9), where each estimate is formed between the first measurement and every subsequent one. Since the instantaneous coherence is highly susceptible to noise, we approximate it using short-time coherence evaluated on 1024-sample windows with a hop size of 512. Within each window, the PSDs are estimated using Welch's method (Welch, 1967), employing 128-sample segments with a 64-sample hop and a fast Fourier transform (FFT) length of 1024. Following Prawda *et al.* (2024), t_{lim} is set such that the volatility is evaluated only while the RIR maintains at least 30 dB SNR, since lower SNRs tend to produce unreliable coherence estimates. The volatility is individually evaluated at frequencies from 6 to 19 kHz in 1 kHz increments, using the FFT bin closest to each target frequency. Frequencies below 6 kHz were omitted, as the volatility estimation is less accurate in this range since the fluctuations predominantly affect higher frequencies. For completeness, we also computed the median volatility estimates across all frequencies without applying DComp, as well as on the ARN and AWGN cases.

The results in Fig. 2 show that the real-world volatility estimates steadily increase with the time difference between measurements, with a noticeable initial jump, the cause of which warrants further investigation. The volatility estimates of the simulations that include atmospheric fluctuations behave similarly, whereas the ARN and AWGN cases remain at zero. Moreover, the gap between the DComp and non-DComp estimates widens as the drift increases, underscoring the need to correct even small drifts before estimating volatility. Despite the improved alignment, the estimated volatility values from the simulation are slightly lower than the parameters used for the simulation. A plausible explanation is the finite reflection order used during the simulation, which may lead to slightly higher-than-expected coherence. Note that the apparent divergence of the volatility estimates across frequencies over time is a visual effect of the increasing volatility. Their relative differences remain within ±5% at all times.

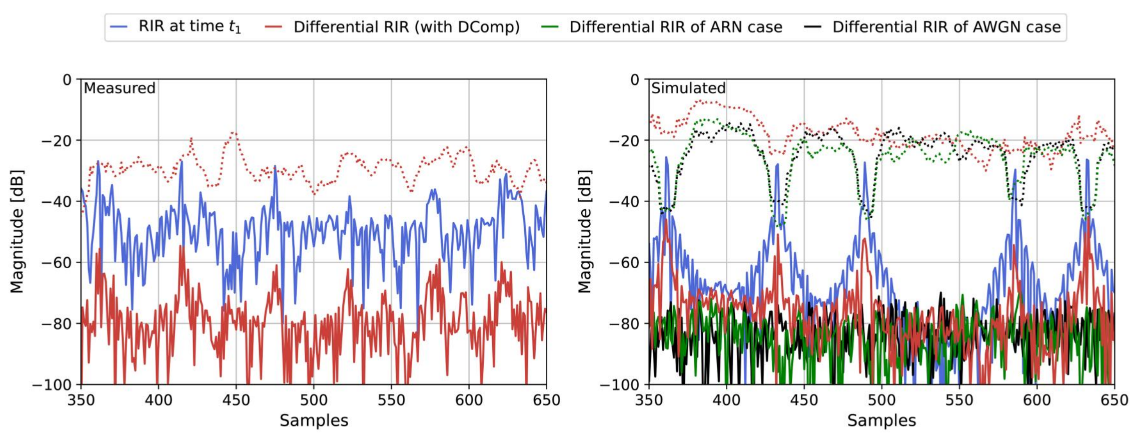


(a) Volatility estimated on real-world RIRs. (b) Volatility estimated on simulated RIRs.

Fig. 2. Comparison of real-world and simulated volatilities. Results are averaged across all eight microphones. Simulations that include atmospheric fluctuations (colored lines) align more closely with real-world estimates than the ARN and AWGN cases. Note that the simulation index is aligned with the measurement timeline, although the simulation itself does not evolve in physical time.

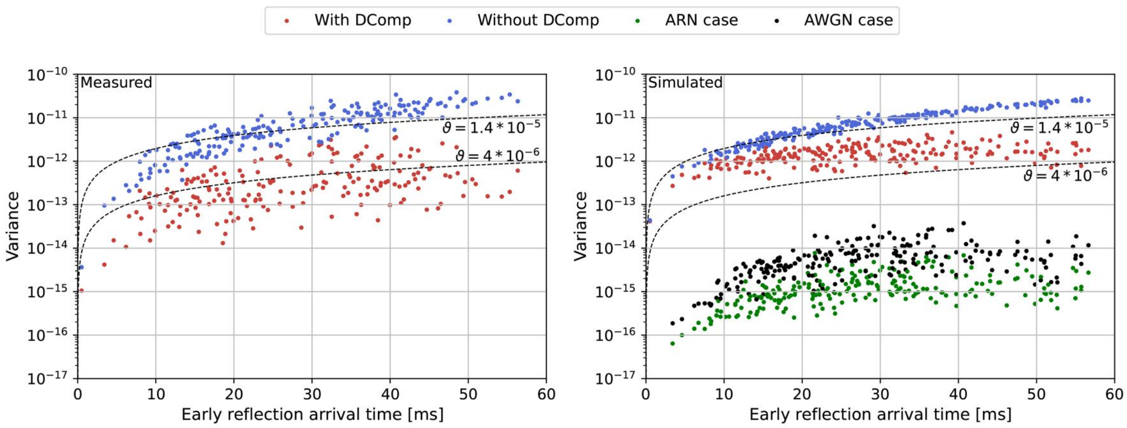
A qualitative comparison is performed using one exemplary differential RIR. Figure 3(a) shows the first RIR measured at microphone 2, together with the differential RIR computed from the first and last measurements with DComp. As can be observed, the differential RIR broadly follows the shape of the original RIR, but at a level roughly 30 dB lower. The simulated counterpart with atmospheric volatility and DComp in Fig. 3(b) exhibits similar behavior, albeit at only approximately 20 dB level difference. This larger discrepancy is consistent with the early reflection variance analysis in the following. In contrast, the ARN and AWGN cases yield a differential RIR that consists only of uniform noise, resulting in large deviations from the differential RIR that includes atmospheric fluctuations during high-energy peaks of the RIR.

Finally, we investigate the variance of the early reflections. Prominent peaks within the first 60 ms of each RIR were manually identified, after which the RIRs were upsampled by a factor of 1000 to enable precise peak tracking. The temporal variance of these peaks was then computed across all 50 RIRs for each microphone. This analysis was performed with and without DComp, and additionally on the ARN and AWGN cases. The results are shown in Fig. 4. Note that for a sampling rate of 48 kHz, a variance of 10^{-11} corresponds to a standard deviation of the TOA jitter of approximately $\sqrt{10^{-11}} \cdot 48\,000 \approx 0.15$ samples. The range of the real-world variances with DComp extends approximately one order of magnitude below the lower bound expected from the lowest volatility estimate, indicating that longer acoustic paths are needed for temperature-induced fluctuations to fully develop in the RIRs. In contrast, the variances of the



(a) Real-world differential RIR. (b) Simulated differential RIRs.

Fig. 3. Qualitative comparison of simulated and real-world RIR using a differential RIR at microphone 2. When atmospheric fluctuations are incorporated, the simulated differential RIR follows the shape of the underlying RIR, closely matching the behavior of the real-world differential RIR. In contrast, the ARN and AWGN cases produce only noise. The dotted lines represent the difference between the differential RIRs and the RIR after applying a 10-sample moving average smoothing.



(a) Early reflection variances on real-world RIRs. (b) Early reflection variances on simulated RIRs.

Fig. 4. Early reflection variance across all microphones for DComp, non-DComp, ARN and AWGN cases. Simulations that include atmospheric fluctuations more closely match real-world observations than the ARN and AWGN cases, though they exhibit a smaller spread than observed in the measurements. The dashed lines bound the variance range predicted by the volatility parameters used.

simulations with fluctuations and DComp fall well within the range defined by the volatility parameters, which also explains the discrepancy between the level differences in Figs. 3(a) and 3(b). The variances obtained without DComp appear more similar because they are dominated by the drift component. These observations suggest that the model assumptions made in Sec. 3 are better satisfied by late reverberation than by early reflections. Nevertheless, the ARN and AWGN cases differ from the measurements by approximately two orders of magnitude, with the ARN case producing slightly lower variance than the AWGN case. Follow-up analysis suggests that this stems from the real-world noise exhibiting a non-flat spectrum with elevated low-frequency energy, which reduces its impact on the early reflection TOA estimates. Additionally, the ARN and AWGN cases exhibit a slight time-dependence, which can be attributed to the gradual reduction in early reflection amplitudes over time, decreasing the SNR. These results indicate that, despite the proposed model's limitations, it still yields much more realistic behavior than the standard ISM, even when enhanced with noise.

5. Conclusion

We introduced an extension of the ISM that incorporates temperature fluctuations as a deterministic temperature drift and stochastic local volatility to more realistically simulate repeated RIR estimates. A comprehensive comparison between real-world measurements and simulations in a shoebox room demonstrated that the model captures several characteristic behaviors observed in practice. The simulated atmospheric volatility estimates closely follow those derived from measurements, and the differential RIRs exhibit the same qualitative structure, in contrast to fluctuation-free simulations, where differential RIRs reduce to uncorrelated noise. The early reflection analysis further revealed similar variance patterns between measurements and simulations, with the real-world variances displaying a larger spread. This suggests that the specific acoustic path of each early reflection has a greater impact on the variance, whereas the late reverb is less sensitive to the exact path, motivating further model refinement. We also observed that the current model produces RIRs with slightly lower volatility than the chosen values, warranting further investigation. Nonetheless, the proposed model yields a substantially more realistic representation of repeated RIR measurements than the standard ISM, even when augmented with AWGN or real-world noise, without increasing the computational complexity.

Acknowledgments

The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg and Fraunhofer IIS.

Author Declarations

Conflict of Interest

The authors have no conflicts to disclose.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1]Note that Eqs. (14) and (15) of Prawda *et al.* (2024) are missing a factor of $4\pi^2$ (confirmed with the authors); the expression here includes it.

Allen, J. B., and Berkley, D. A. (**1979**). "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**(4), 943–950.

Badeau, R. (**2018**). "Unified stochastic reverberation modeling," in *Proceedings of the 26th European Signal Processing Conference (EUSIPCO)* (IEEE, Piscataway, NJ), pp. 2175–2179.

Badeau, R. (**2019**). "Common mathematical framework for stochastic reverberation models," J. Acoust. Soc. Am. **145**(4), 2733–2745.

Benesty, J., Chen, J., and Huang, Y. (**2008**). "On the importance of the Pearson correlation coefficient in noise reduction," IEEE Trans. Audio. Speech Lang. Process. **16**(4), 757–765.

Bhattacharjee, S. S., Jonas Fuglsig, A., Christensen, F., Jensen, J. R., and Graesbùll Christensen, M. (**2025**). "Robust fixed-filter sound zone control with audio-based position tracking," in *Proceedings of the ICASSP 2025* (IEEE, Piscataway, NJ), pp. 1–5.

Durrett, R. (**2019**). *Cambridge Series in Statistical and Probabilistic Mathematics Probability: Theory and Examples*, 5th ed. (Cambridge University Press, Cambridge, MA).

Farina, A. (**2000**). "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proceedings of the 108th Audio Engineering Society Convention*, Paris, France.

Habets, E. A. (**2006**). "Room impulse response generator," Technische Universiteit Eindhoven, Tech. Rep. **2**, 1–21.

Kim, K., Wang, S., Ryu, H., and Lee, S. Q. (**2020**). "Acoustic-based position estimation of an object and a person using active localization and sound field analysis," Appl. Sci. **10**(24), 9090.

Lawrence, J., Tuna, C., Walther, A., and Peters, N. (**2025**). "Acoustic position estimation of a silent listener," in *Proceedings of the ICASSP 2025* (IEEE, Piscataway, NJ), pp. 1–5.

McCarthy, D. (**1997**). "Short-time coherence functions: Instantaneous and tuned," Mech. Syst. Sig. Process. **11**(2), 169–185.

Melikov, A., Krüger, U., Zhou, G., Madsen, T., and Langkilde, G. (**1997**). "Air temperature fluctuations in rooms," Build. Environ. **32**(2), 101–114.

Petrov, V. V. (**2000**). *Classical-Type Limit Theorems for Sums of Independent Random Variables, 1–24* (Springer Berlin Heidelberg, Berlin, Heidelberg)

Postma, B. N. J., and Katz, B. F. G. (**2016**). "Correction method for averaging slowly time-variant room impulse response measurements," J. Acoust. Soc. Am. **140**(1), EL38–EL43.

Prawda, K. (**2025**). "Sensitivity of room impulse responses in changing acoustic environment," in *Proceedings of ICASSP 2025* (IEEE, Piscataway, NJ), pp. 1–5.

Prawda, K., Schlecht, S. J., and Välimäki, V. (**2023**). "Time variance in measured room impulse responses," in *Proceedings of the 10th Convention of the European Acoustics Association, Forum Acusticum 2023* (European Acoustics Association, Torino, Italy), pp. 1639–1646.

Prawda, K., Schlecht, S. J., and Välimäki, V. (**2024**). "Short-time coherence between repeated room impulse response measurements," J. Acoust. Soc. Am. **156**(2), 1017–1028.

Prinn, A. G., and Badeau, R. (**2026**). "Verification of statistical wave field theory reverberation time predictions," Appl. Acoust. **250**, 111337.

Schroeder, M. R. (**1965**). "New method of measuring reverberation time," J. Acoust. Soc. Am. **37**(3), 409–412.

Sunklodas, J. (**2000**). *Approximation of Distributions of Sums of Weakly Dependent Random Variables by the Normal Distribution* (Springer, Berlin-Heidelberg).

Svensson, P., and Nielsen, J. L. (**1999**). "Errors in MLS measurements caused by time variance in acoustic systems," J. Audio Eng. Soc. **47**(11), 907–927

Wang, M., Clarke, S., Wang, J.-H., Gao, R., and Wu, J. (**2024**). "Soundcam: A dataset for finding humans using room acoustics," Adv. Neural Inf. Process. Syst. **36**, pp. 52238–52264.

Welch, P. (**1967**). "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," IEEE Trans. Audio Electroacoust. **15**(2), 70–73.

Xu, Z., Herzog, A., Lodermeyer, A., Habets, E. A. P., and Prinn, A. G. (**2024**). "Simulating room transfer functions between transducers mounted on audio devices using a modified image source method," J. Acoust. Soc. Am. **155**(1), 343–357.

Yang, Z., Fan, X., Isler, V., and Park, H. S. (**2022**). "PoseKernelLifter: Metric lifting of 3D human pose using sound," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE Computer Society, New Orleans, LA), pp. 13169–13179.