KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Authors: Kathleen MacWilliam, Thomas Dietzen, Randall Ali, Toon van Waterschoot Title: State-space estimation of spatially dynamic room impulse responses using a room acoustic model-based prior Journal: Frontiers Signal Processing Vol: 4 Pages:
Archived version	☐ Submitted version ☐ Accepted (Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher) ☐ Published ■ Published (Open Access)
Acknowledgements	In this manuscript, the following funding is acknowledged: ■ EU H22020 ☐ EU Horizon Europe ■ FWO ☐ KULeuven Internal Funds ☐ VLAIO ☐ AI ☐ Other
Published version	`https://www.frontiersin.org/journals/signal-processing/articles/10.3389/frsip.2024.1426082/full#h11`
Copyright statement	© 2024 MacWilliam, Dietzen, Ali and van Waterschoot. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.
Author contact	kathleen.macwilliam@esat.kuleuven.be

frontiers | Frontiers in Signal Processing

TYPE Original Research
PUBLISHED 02 September 2024
DOI 10.3389/frsip.2024.1426082

Check for updates

OPEN ACCESS

EDITED BY
David V. Anderson,
Georgia Institute of Technology, United States

REVIEWED BY
Jelena Ćertić,
University of Belgrade, Serbia
Victor Lazzarini,
Maynooth University, Ireland

*CORRESPONDENCE
Kathleen MacWilliam,
kathleen.macwilliam@esat.kuleuven.be

RECEIVED 30 April 2024
ACCEPTED 01 August 2024
PUBLISHED 02 September 2024

CITATION
MacWilliam K, Dietzen T, Ali R and van Waterschoot T (2024) State-space estimation of spatially dynamic room impulse responses using a room acoustic model-based prior.
Front. Sig. Proc. 4:1426082.
doi: 10.3389/frsip.2024.1426082

COPYRIGHT
© 2024 MacWilliam, Dietzen, Ali and van Waterschoot. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

State-space estimation of spatially dynamic room impulse responses using a room acoustic model-based prior

Kathleen MacWilliam[1]*, Thomas Dietzen[1], Randall Ali[1,2] and Toon van Waterschoot[1]

[1]STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Department of Electrical Engineering (ESAT), KU Leuven, Leuven, Belgium, [2]Institute of Sound Recording, Department of Music and Media, University of Surrey, Guildford, United Kingdom

Room impulse responses (RIRs) between static loudspeaker and microphone locations can be estimated using a number of well-established measurement and inference procedures. While these procedures assume a time-invariant acoustic system, time variations need to be considered for the case of spatially dynamic scenarios where loudspeakers and microphones are subject to movement. If the RIR is modeled using image sources, then movement implies that the distance to each image source varies over time, making the estimation of the spatially dynamic RIR particularly challenging. In this paper, we propose a procedure to estimate the early part of the spatially dynamic RIR between a stationary source and a microphone moving on a linear trajectory at constant velocity. The procedure is built upon a state-space model, where the state to be estimated represents the early RIR, the observation corresponds to a microphone recording in a spatially dynamic scenario, and time-varying distances to the image sources are incorporated into the state transition matrix obtained from static RIRs at the start and end points of the trajectory. The performance of the proposed approach is evaluated against state-of-the-art RIR interpolation and state-space estimation methods using simulations, demonstrating the potential of the proposed state-space model.

KEYWORDS

state-space model, transition matrix, acoustic room impulse response interpolation, time-varying system, dynamic time warping

1 Introduction

A room impulse response (RIR) is the time-domain representation of the linear time-invariant (LTI) system that uniquely characterizes the cumulative impact of a room on sound waves between a specific static source and a microphone's position, effectively representing the room's acoustic environment. This concept is fundamental to various acoustic signal processing applications, including source localization (Evers et al., 2020), dereverberation (Naylor and Gaubitch, 2010), echo cancellation (Elko et al., 2003), and spatial audio reproduction (Schissler et al., 2017). Considerable research has been dedicated to developing robust measurement (Stan et al., 2002; Szöke et al., 2018) and estimation (Lin and Lee, 2006; Crocco and Bue, 2015; Ratnarajah et al., 2022) techniques for capturing RIRs, particularly in scenarios where the source and microphone remain static within the acoustic environment (Stan et al., 2002; Szöke et al., 2018). However, real-world situations often

involve spatially dynamic scenarios, where sources and microphones are subject to movement. This paper addresses the challenge of accurately estimating the early part of RIRs along a trajectory in a time-varying acoustic scenario with a stationary source and a moving microphone. This can be thought of as a time-varying system identification problem where the system to be identified may be referred to as a time-variant RIR.

In this context, it is important to consider the apparent contradiction between the previously defined RIR as an LTI system representation and the concept of a time-variant RIR. In this paper, the acoustic environment itself is assumed to be time-invariant, in which case an RIR between two static positions indeed corresponds to an LTI system. RIRs, however, are inherently location-variant—they depend on the locations of the source and the microphone. The time-variant RIR, as considered here, can be defined as the time-varying system representation (Cherniakov, 2003) relating the source signal to the microphone signal if the microphone moves—that is, has a time-varying location. In discrete time, this time-variant RIR can also be thought of as a collection of RIRs in the LTI sense over different time-instances: the location of the microphone changes at each time step, leading to a new discrete position in space associated with a time-invariant RIR. Time-variant RIR estimation, as defined above, has relevance across numerous acoustic signal processing applications, especially amid the growing interest in virtual acoustic environments. For instance, Ajdler et al. (2007) demonstrated the use of time-varying acoustic system models for enabling the rapid measurement of head-related impulse responses. Moreover, sub-optimal estimations of time-variant RIRs can impair the effectiveness of echo-cancellation systems in telepresence and communication technologies (Nophut et al., 2024).

In order to contextualize our contribution in estimating the early part of a time-variant RIR, it is necessary to provide a brief outline of the related literature. In this context, the concept of spatial RIR interpolation is highly relevant given that we can consider a time-varying RIR as a collection of RIRs over a discrete set of locations. RIR interpolation facilitates sound field rendering for dynamic source–microphone positions by filling spatial gaps in RIR data, as measurements or simulations are usually limited to sparse grids. Numerous approaches have been proposed for RIR interpolation, including compressed sensing methods (Mignot et al., 2013; Katzberg et al., 2018), spherical harmonics (Borra et al., 2019), physics-based models (Antonello et al., 2017; Hahmann and Fernandez-Grande, 2022), directional RIRs (Zhao et al., 2022), and neural networks (Pezzoli et al., 2022; Karakonstantis et al., 2024). Haneda et al. (1999) introduced a frequency-domain approach for interpolating room transfer functions (RTFs), which is particularly effective at lower frequencies; it was extended by Das et al. (2021). Additionally, several techniques for the interpolation of head-related transfer functions (HRTFs), which are critical for accurate binaural rendering and share some methodological approaches with RTF interpolation, have been explored (Carty, 2010). For the purpose of this paper, we focus on the interpolation of an RIR at a point between two microphone locations given their estimated RIRs for a common stationary source. Kearney et al. (2009) introduced Dynamic Time Warping (DTW)-based interpolation, dividing RIRs into early reflections and diffuse decay. This method temporally aligns and linearly interpolates early reflections while modeling the tail using the approach of Masterson et al. (2009) and laying the foundation for subsequent developments. Garcia-Gomez and Lopez (2018) expanded on Kearney's work, enhancing algorithm robustness and computational efficiency while retaining the core interpolation technique. Building on this, Bruschi et al. (2020) refined peak finding and matching aspects of the interpolation technique. Geldert et al. (2023) proposed a novel approach using partial optimal transport for interpolation, enabling non-bijective mapping of sound events between the early part of RIRs. It is important to highlight that these interpolation approaches operate outside of a conventional system identification framework. Instead, they use a limited set of measured RIRs and predominantly rely on a room acoustic sound propagation model such as the image source method (ISM) (Allen and Berkley, 1979). As opposed to interpolation strategies, fully data-driven approaches for estimating time-variant RIRs have also been investigated. In this case, the estimation relies directly on the source and microphone signals and can be framed as a system identification or adaptive filtering problem. This approach has been largely motivated by the need to obtain rapid measurements of head-related impulse responses (Enzner, 2008; Hahn and Spors, 2015) and for echo cancellation (Antweiler and Symanzik, 1995; Enzner, 2010), where a typical scenario involves an excitation signal being continuously captured by a moving microphone. Using carefully designed excitation signals (Hahn and Spors, 2015; Kuhl et al., 2018), time-variant RIRs can be estimated using a normalized least mean squares (NLMS) algorithm (Enzner, 2008; Antweiler et al., 2012) or more generally using a Kalman filter (Enzner, 2010). In the context of this paper, a Kalman filter is of particular interest as it is derived from a state-space model of the dynamic system where the state to be estimated is the time-variant RIR (Enzner, 2010). In such a state-space model, the evolution of the time-variant RIR is explicitly modeled using a first-order difference equation, which allows more modeling flexibility than other adaptive algorithms such as NLMS. One popular choice for the first-order difference model is to relate the states at two time instants by a transition factor and an additional process noise term. The transition coefficient and process noise covariance are typically set according to the expected variability of the state, influencing the convergence behavior of the Kalman filter. For instance, in Nophut et al. (2024), the transition coefficient was modeled as a function of the microphone velocity.

We here consider the problem of estimating the early part of the time-variant RIR between a stationary source and a microphone moving on a linear trajectory at constant velocity. We propose integrating RIR interpolation, derived from a room acoustic model, into a state-space-based framework for RIR estimation, thereby merging data-driven approaches with physical modeling. More specifically, rather than relying solely on a state transition factor within the state equation, we propose incorporating the ISM into a state transition matrix between the early segments of consecutive RIRs. We derive an analytical model for this transition matrix and subsequently propose estimating it from static RIRs at the trajectory's start and end points using a DTW-based algorithm. The proposed approach's performance is evaluated through simulations by comparing it to two alternatives: one that relies solely on the state equation for RIR estimation, resembling a pure interpolation method with the ISM-based transition matrix, and another that uses a conventional state-space estimation with a simple state transition factor. Our findings suggest

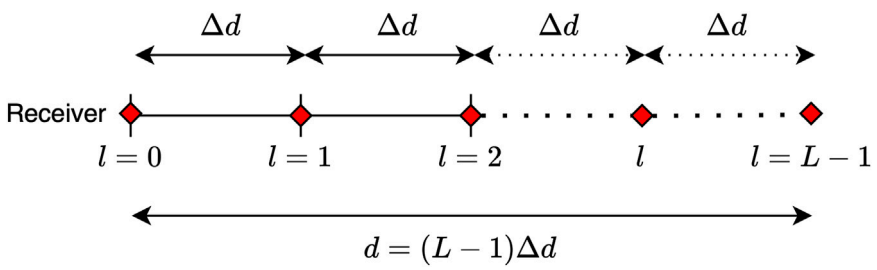

FIGURE 1
Linear microphone trajectory of length d divided into L equidistant locations spaced by Δd and indexed as $l = 0, \ldots, L-1$ such that $d = (L-1)\Delta d$.

that the proposed state-space model outperforms both alternatives in terms of normalized misalignment between the simulated "ground-truth" RIRs and estimated RIRs.

The subsequent sections of this paper are organized as follows. Section 2 introduces the signal model and provides an overview of the most pertinent state-of-the-art methods. Section 3 elaborates on the proposed RIR state-space model and outlines the update equations of the Kalman filter used to recursively estimate the defined state. Section 4 offers detailed derivations of the proposed room acoustic model-based state transition matrix. Finally, Section 5 presents experimental validation through simulations, followed by a discussion of the results.

2 Signal model, problem statement, and related state of the art

2.1 Signal model and problem statement

We first introduce the signal model to formally define the concept of an RIR as employed in this paper. When we henceforth use the term "RIR", it will specifically pertain to the early part of the RIRs, as we do not address the estimation of the late reverberant tail. We also assume that the source location remains static. Let k denote the time index, and let the location of the microphone relative to the linear trajectory be denoted by the one-dimensional location index l. If the microphone location is time-variant, an RIR defines the LTI system relating a source signal $x(k)$ to an observed signal $y(l,k)$. Let the RIR for microphone location l be denoted by $h(l,n)$, where n indexes the time-shift of the RIR samples. The signals $x(k)$ and $y(l,k)$ are then related by

$$y(l,k) = \sum_{n=0}^{N-1} h(l,n)x(k-n) + v(k), \tag{1}$$

where $v(k)$ is a noise term that we generally assume to be present in the observed signal and may include late reverberation. Now, defining the vectors $\mathbf{x}(k)$ and $\mathbf{h}(l)$ as

$$\mathbf{x}(k) = \begin{pmatrix} x(k) & x(k-1) & \ldots & x(k-N+1) \end{pmatrix}^T, \tag{2}$$

$$\mathbf{h}(l) = \begin{pmatrix} h(l,0) & h(l,1) & \ldots & h(l,N-1) \end{pmatrix}^T, \tag{3}$$

the convolution in Equation 1 can alternatively be written as

$$y(l,k) = \mathbf{x}^T(k)\mathbf{h}(l) + v(k). \tag{4}$$

Within the scope of this paper, we assume a linear microphone trajectory of length d, divided into L equidistant locations spaced by Δd and indexed as $l = 0, \ldots, L-1$, as illustrated in Figure 1, such that $L = d/\Delta d + 1$. We consider the problem of estimating $\mathbf{h}(l)$ using the source signal $x(k)$ and the observed signal $y(l,k)$ if the microphone moves along the trajectory at a constant velocity μ_{rx}—that is, if l varies over time. For this, we make the assumptions that $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$ are known, and that the time intervals which the individual reflections occupy within the time-variant RIR over the range of the trajectory do not overlap. An example of $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$ including only first-order reflections is shown in Figure 2, where peaks corresponding to the same source or reflection are labeled by the same number in each RIR.

2.2 Related state of the art

Before introducing the proposed approach, it is instructive to briefly introduce the most relevant concepts used in the state of the art. On the one hand, we consider RIR interpolation approaches that estimate $\mathbf{h}(l)$ from $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$ only based on a room acoustic sound propagation model, without making use of an observed signal $y(l,k)$. On the other hand, we consider data-driven approaches that estimate $\mathbf{h}(l)$ in an adaptive manner from a data set containing $x(k)$ and $y(l,k)$ over multiple time indexes k for a moving microphone, but without exploiting a room acoustic sound propagation model.

As previously mentioned, some RIR interpolation approaches (Kearney et al., 2009; Geldert et al., 2023) are motivated by the ISM for room acoustic sound propagation, which expresses an RIR as a sum of contributions from the original source and so-called image sources representing reflections from the boundaries of the room. Based on this model, RIR interpolation approaches infer the location-variant time of arrival (TOA) as well as the amplitude of the direct component and individual reflections (or equivalently, source and image source components) at a particular location l using the RIRs $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$. The general principle can be described as follows. Assuming that reflections corresponding to the same image source in $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$ can be identified, let $\mathbf{n}_r(0)$ and $\mathbf{n}_r(L-1)$ be vectors collecting the respective time-shift indices of the direct component and these reflections in $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$. The corresponding time-shift indices of $\mathbf{h}(l)$ can be interpolated as

$$\mathbf{n}_r(l) = \mathbf{n}_r(0) + \text{round}\big(\beta(l)\left(\mathbf{n}_r(L-1) - \mathbf{n}_r(0)\right)\big), \tag{5}$$

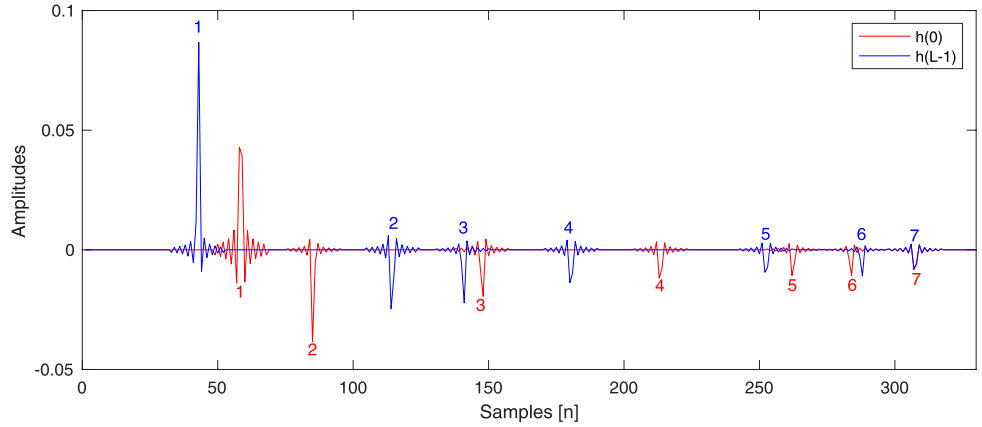

FIGURE 2
An example of simulated RIRs, $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$, including only first-order reflections. Peaks corresponding to the same source or reflection are labeled by the same number in each RIR.

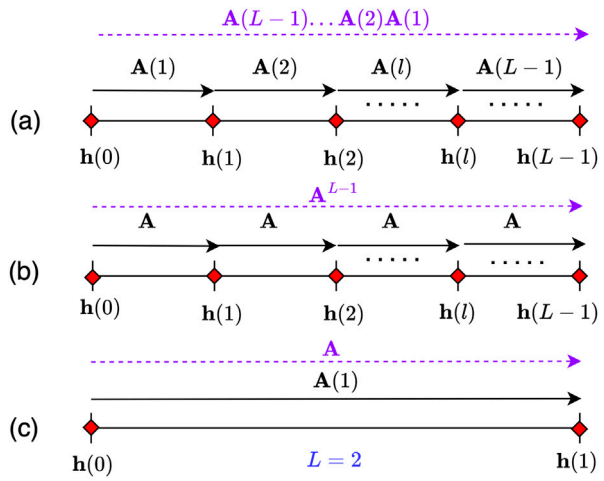

FIGURE 3
(a) Location-variant transition matrix model $\mathbf{A}(l)$ providing mapping between two RIRs $\mathbf{h}(l)$ and $\mathbf{h}(l-1)$ along a given trajectory. **(b)** Location-invariant transition matrix model $\mathbf{A}$ that approximates the mapping between two RIRs $\mathbf{h}(l)$ and $\mathbf{h}(l-1)$ along a given trajectory where the assumptions detailed in Section 4.2 hold. **(c)** Equivalent transition matrix models, $\mathbf{A}(1)$ and $\mathbf{A}$, for the special case of $L = 2$—that is, a transition matrix between the start and end points of the trajectory.

where $\mathbf{n}_r(L-1) - \mathbf{n}_r(0)$ can be conceived as sample-based time differences of arrival (TDOAs), $\beta(l) \in [0,1]$ with $\beta(0) = 0$ and $\beta(L-1) = 1$, and round$(\cdot)$ defines a rounding operation. The function $\beta(l)$ can be defined such that the interpolation is correct for the direct component (Kearney et al., 2009). Applying a similar interpolation rule for the amplitudes of the direct component and reflections, an estimate of $\mathbf{h}(l)$ can then be obtained. The collections of corresponding indices $\mathbf{n}_r(0)$ and $\mathbf{n}_r(L-1)$ can, for instance, be found by means of Dynamic Time Warping with $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$ as inputs (Kearney et al., 2009; Bruschi et al., 2020). Here, an accurate identification of all corresponding reflections requires that their order of arrival is the same in $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$.

Data-driven approaches (Enzner, 2010) to the estimation of location- or time-variant RIRs do not rely on explicit modeling of room acoustic sound propagation, but instead perform adaptive system identification given a data set containing $x(k)$ and $y(l,k)$ over multiple time indexes k. For the moment, let us assume that the moving microphone crosses one location l at each time instant k, such that the location-variant RIR $\mathbf{h}(l)$ can equivalently be thought of as a time-variant RIR $\mathbf{h}(k)$ with $k = l$. The convolution in Equation 1 then becomes

$$y(k) = \sum_{n=0}^{N-1} h(k,n)x(k-n) + v(k). \tag{6}$$

This can be expressed similarly to Equation 4 using the definitions in Equations 2 and 3. A popular approach to adaptively estimate $\mathbf{h}(k)$ is the Kalman filter, whose update equations are obtained from a state-space model where $\mathbf{h}(k)$ is the state. A commonly used state-space model can be defined as

$$\mathbf{h}(k) = \alpha\mathbf{h}(k-1) + \mathbf{w}(k), \tag{7}$$
$$y(k) = \mathbf{x}^T(k)\mathbf{h}(k) + v(k), \tag{8}$$

where the state equation in Equation 7 models the evolution of the state as a first-order difference equation, and the observation equation (Equation 8) relates the state to the observation $y(k)$ and is identical to Equation 6. The model commonly chosen in Equation 7 defines a factor α, referred to as the "forgetting" or "transition" factor, and a noise term $\mathbf{w}(k)$, referred to as "process noise", for which the covariance matrix is required in the Kalman filter. In practice, the choice of α and the covariance matrix of $\mathbf{w}(k)$ is typically not directly motivated by a physical model of the RIR evolution. Instead, they are often considered tuning parameters that can be used to control the convergence behavior of the Kalman filter, although they can be tuned depending on physical parameters such as the velocity of the microphone (Nophut et al., 2024).

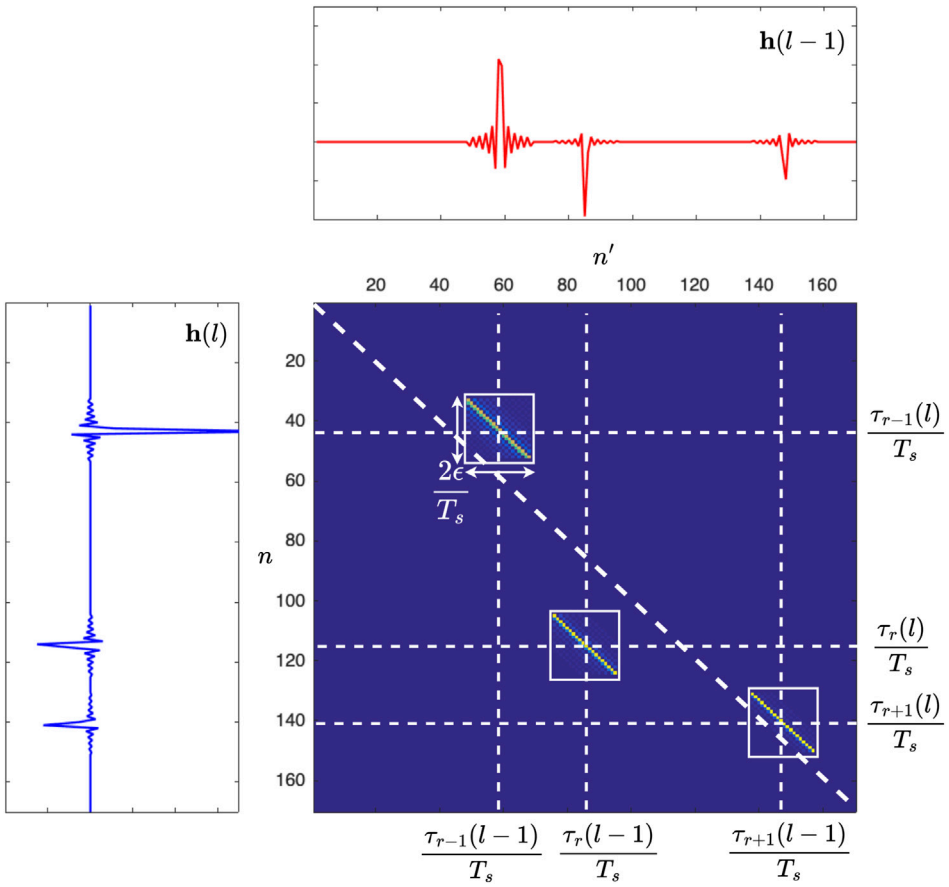

FIGURE 4
An example of an analytical location-variant transition matrix $\mathbf{A}(l)$ constructed from the early segments of two impulse responses, $\mathbf{h}(l)$ and $\mathbf{h}(l-1)$. The grid lines indicate the positions of the respective reflection TOAs.

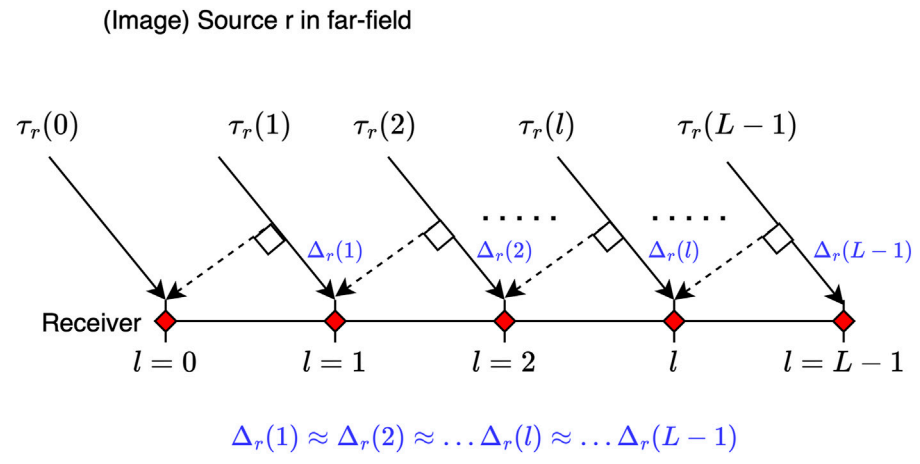

FIGURE 5
Illustration of the far-field assumption used to mitigate the dependence of $\Delta_r(l)$ on l.

3 Proposed RIR state-space model

In the proposed approach, we aim to include prior knowledge on room acoustic sound propagation into a state-space model for time-variant RIR estimation. To this end, rather than resorting to a scalar factor as in Equation 7, we assume that the relation between two RIRs $\mathbf{h}(l)$ and $\mathbf{h}(l-1)$ on the trajectory can approximately be modeled by a location-invariant transition matrix $\mathbf{A}$ as

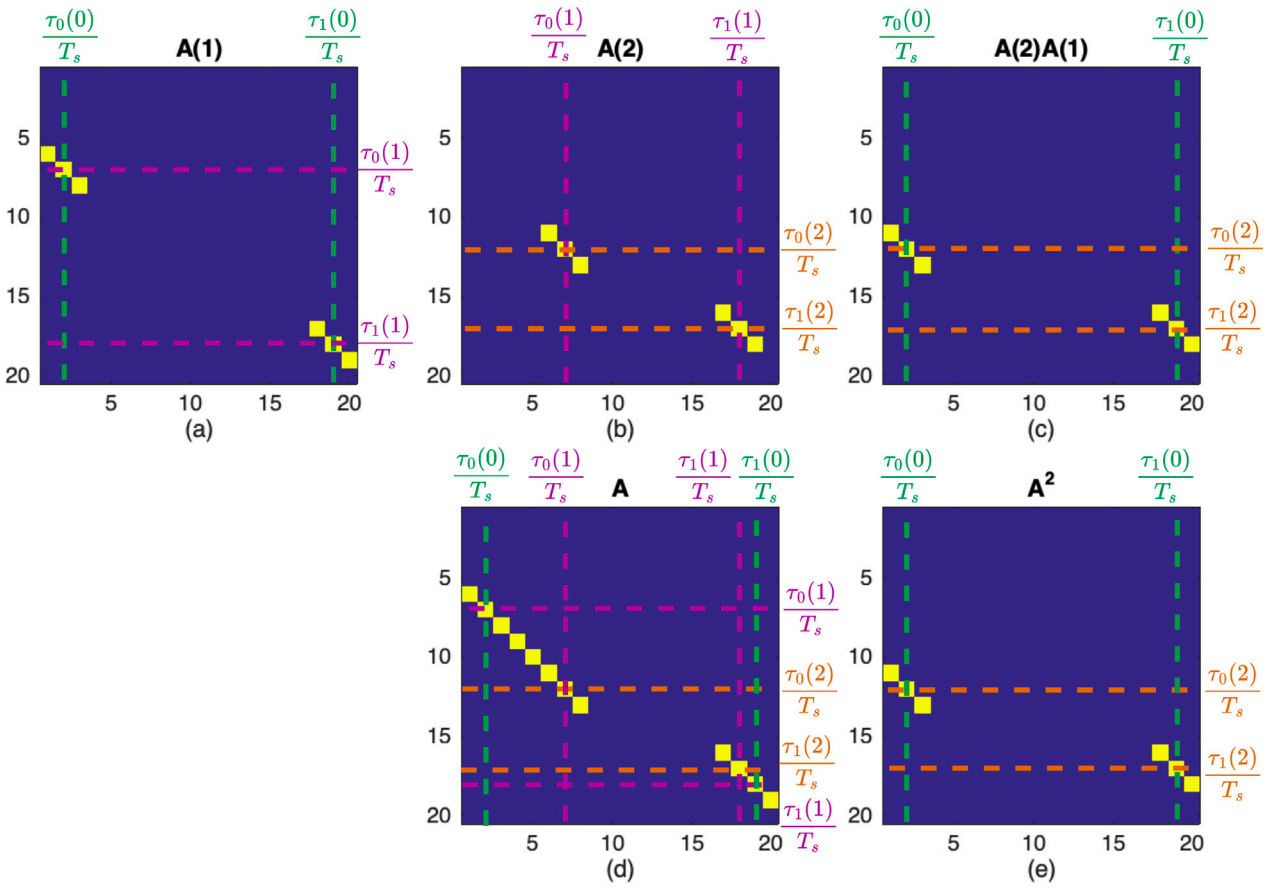

FIGURE 6
Illustration of the analytical transition matrices for a toy problem along a trajectory with $L = 3$, with $\frac{2\epsilon}{T_s} = 3$ and integer reflection TOAs. Subfigures **(a)** and **(b)** show the location-variant matrices $\mathbf{A}(l)$ for $l = 1$ and $l = 2$, respectively. Subfigure **(c)** shows their product, **A(2) A(1)**. Subfigure **(d)** shows the location-invariant matrix **A**, while **(e)** depicts $\mathbf{A}^2$ (i.e., $\mathbf{A}^{L-1}$ for L = 3).

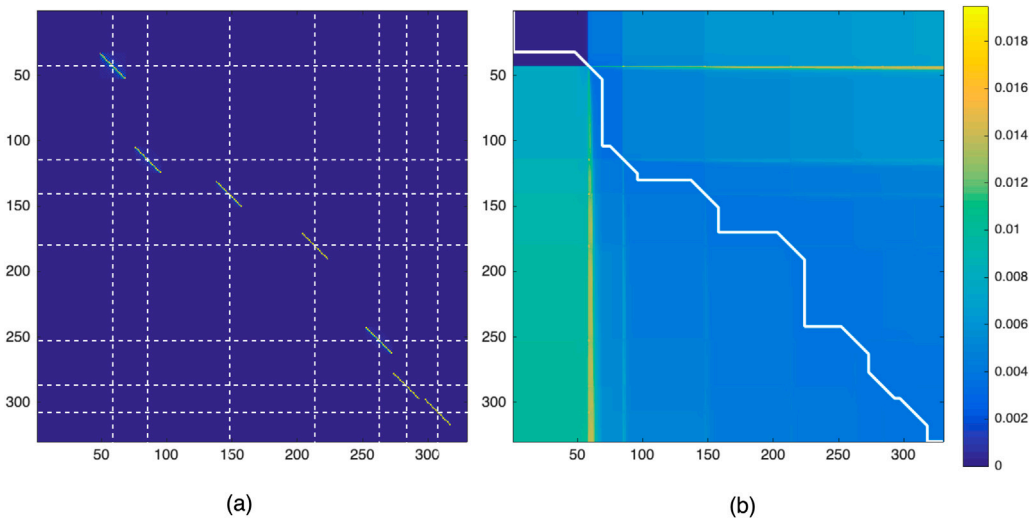

FIGURE 7
(a) Location-invariant matrix **A** for the special case of $L = 2$ and **(b)** accumulated distance matrix **D** constructed from $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$ overlaid with DTW warp path (white line).

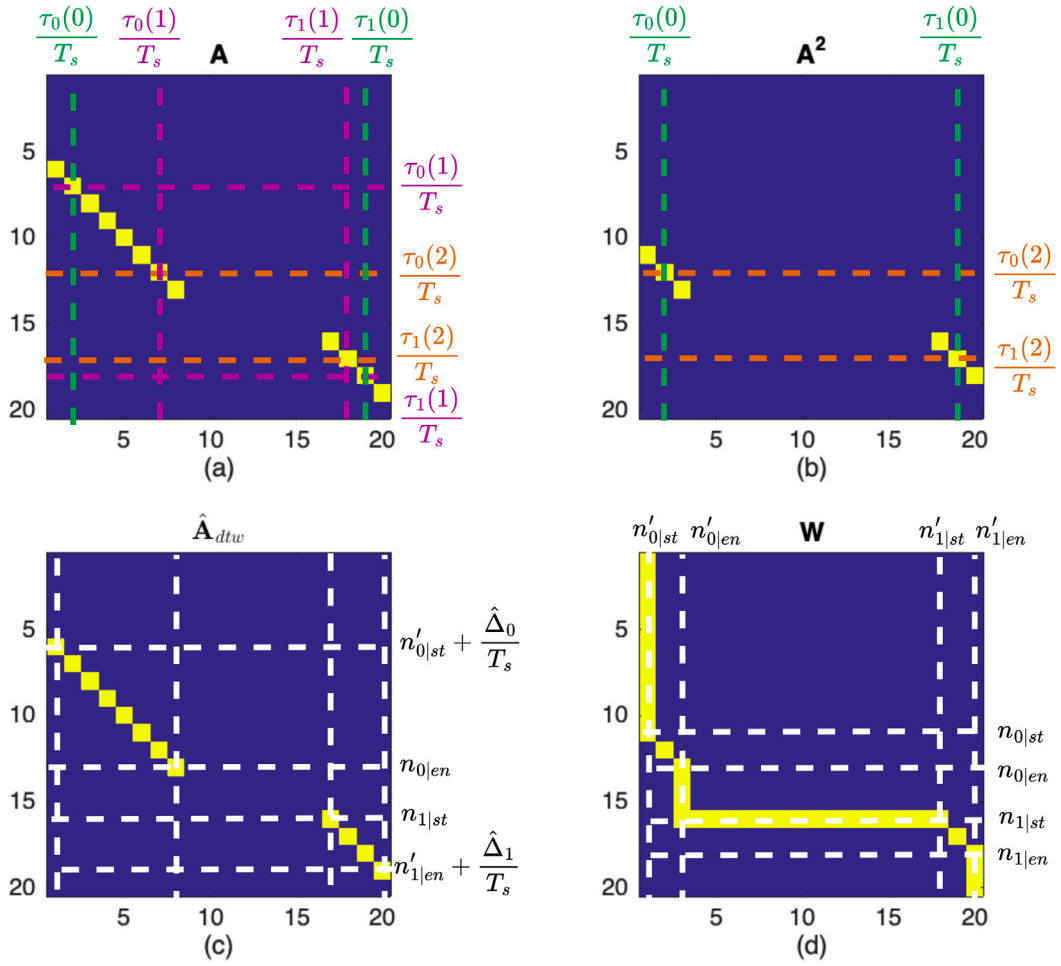

FIGURE 8
Illustration of the analytical transition matrices for a toy problem along a trajectory with $L = 3$, with $\frac{2\varepsilon}{T_s} = 3$ and integer reflection TOAs. Subfigure **(a)** shows the location-invariant matrix $\mathbf{A}$, while **(b)** depicts $\mathbf{A}^2$ (i.e., $\mathbf{A}^{L-1}$ for $L = 3$). Subfigures **(c)** and **(d)** respectively present the matrix $\mathbf{W}$, derived from the DTW warp path, and the estimated location-invariant matrix $\hat{\mathbf{A}}_{\text{dtw}}$.

TABLE 1 Description of the algorithms compared in the experimental results.

Algorithm name	Algorithm description
LI – A	Linear interpolation equivalent to Kalman filter update Equations 12–16 with the analytical transition matrix **A** derived in Section 4.2 and $y_\Omega(l) = 0$ and $\mathbf{x}_\Omega(l) = 0$
KF – α	Kalman filter update Equations 12–16 with **A** replaced with $\alpha = 1$
KF – A	Kalman filter update Equations 12–16 with the analytical transition matrix **A** derived in Section 4.2
KF – $\hat{\mathbf{A}}_{\text{dtw}}$	Kalman filter update Equations 12–16 with an estimate $\hat{\mathbf{A}}_{\text{dtw}}$ of the analytical matrix **A** as derived in Section 4.3

$$\mathbf{h}(l) \approx \mathbf{A}\mathbf{h}(l-1) \tag{9}$$

for $l = 1, \ldots, L-1$. The transition matrix $\mathbf{A}$ will be used to model changes in the TOA of the direct component and the early reflections between the neighboring locations l and $l-1$ and will be applied in the state equation of the proposed state-space model. As described in Section 4, the analytical definition of $\mathbf{A}$ will be based on the ISM, which also serves as the foundation of the RIR interpolation approaches discussed in Section 2.2. Furthermore, we will show that the transition matrix $\mathbf{A}$ can be modeled as location-invariant if the time intervals which the individual reflections occupy in the time-variant RIR over the range of the trajectory do not overlap, which is also implicitly assumed in Kearney et al. (2009) and Geldert et al. (2023). In practice, an estimate of $\mathbf{A}$ can be obtained from the presumed knowledge of the RIRs $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$ at the start and the end of the trajectory, as will be discussed in Section 4.3.

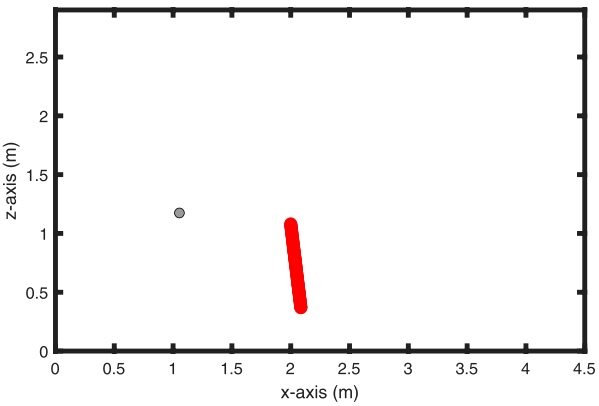

FIGURE 9
2-D illustration depicting the source position and trajectory of the microphone within a box-shaped room, as utilized for simulations.

While it is possible to interpolate between $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$ based on $\mathbf{A}$ only as $\mathbf{h}(l) \approx \mathbf{A}^l\mathbf{h}(0)$, a more accurate estimate can be obtained by using the signals recorded along the trajectory with a microphone moving at a constant velocity. Without loss of generality, let $k = 0$ denote the start of the recording at location $l = 0$. For a given velocity μ_{rx} of the microphone and temporal sampling period T_s, the smallest possible spatial period Δd that could be used is given by $\mu_{\text{rx}}T_s$. In this case, one RIR estimate can be obtained per time sample. In the proposed approach, we also consider spatial periods of

$$\Delta d = \Omega\mu_{\text{rx}}T_s,$$

where Ω is a positive integer that can be understood as a spatial down-sampling factor. If the microphone is at location with index l, we then observe sample

$$k = l\Omega.$$

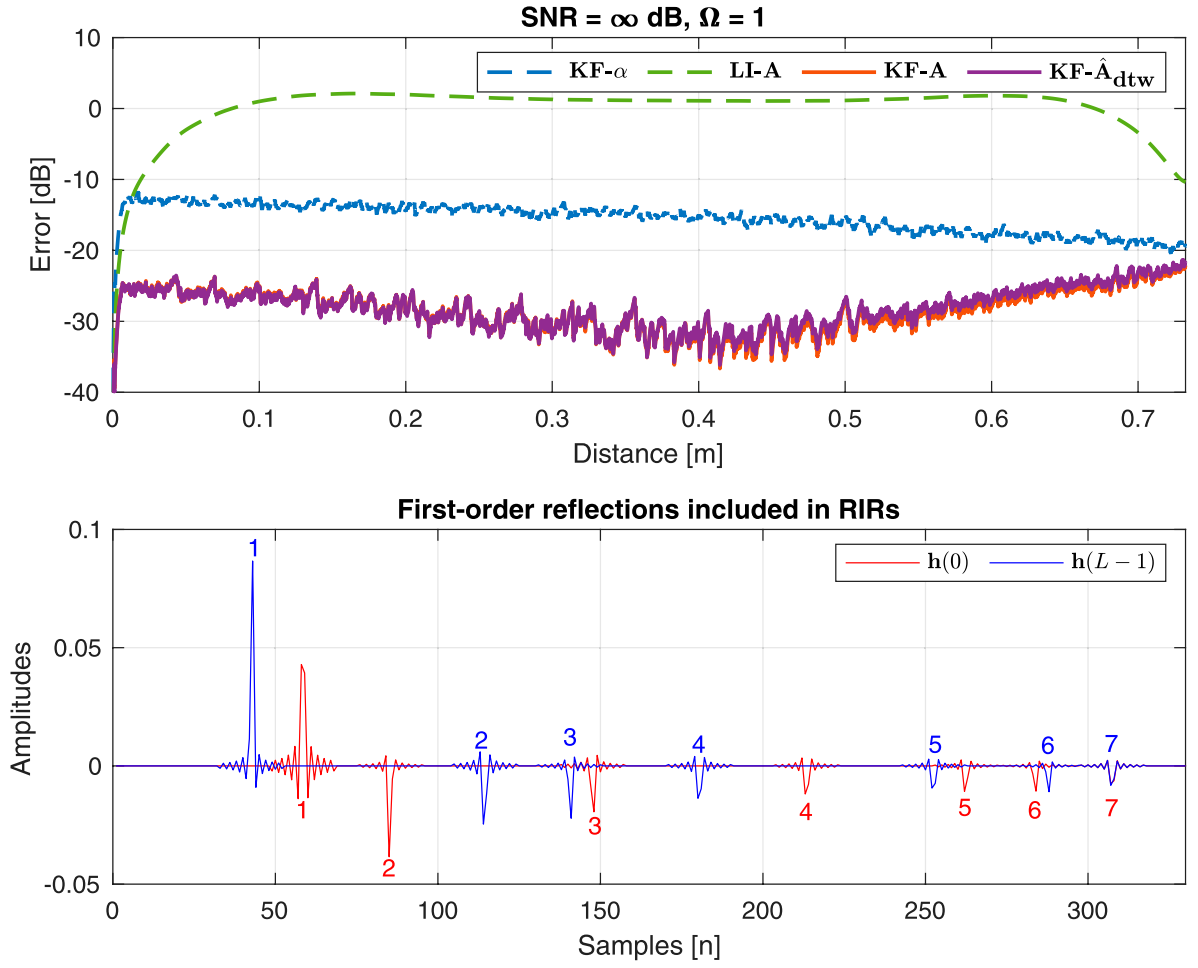

FIGURE 10
(Top) Result 1: Normalized misalignment along the trajectory between the "ground-truth" RIR $\mathbf{h}(l)$ and the estimated RIR $\hat{\mathbf{h}}(l)$ using algorithms detailed in Table 1. This is for the case of a noiseless observed output signal $y_\Omega(l)$. (Bottom) Simulated RIRs corresponding to the start and end of the trajectory, denoted as $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$, respectively. Peaks corresponding to the same source or reflection are labeled by the same number in each RIR.

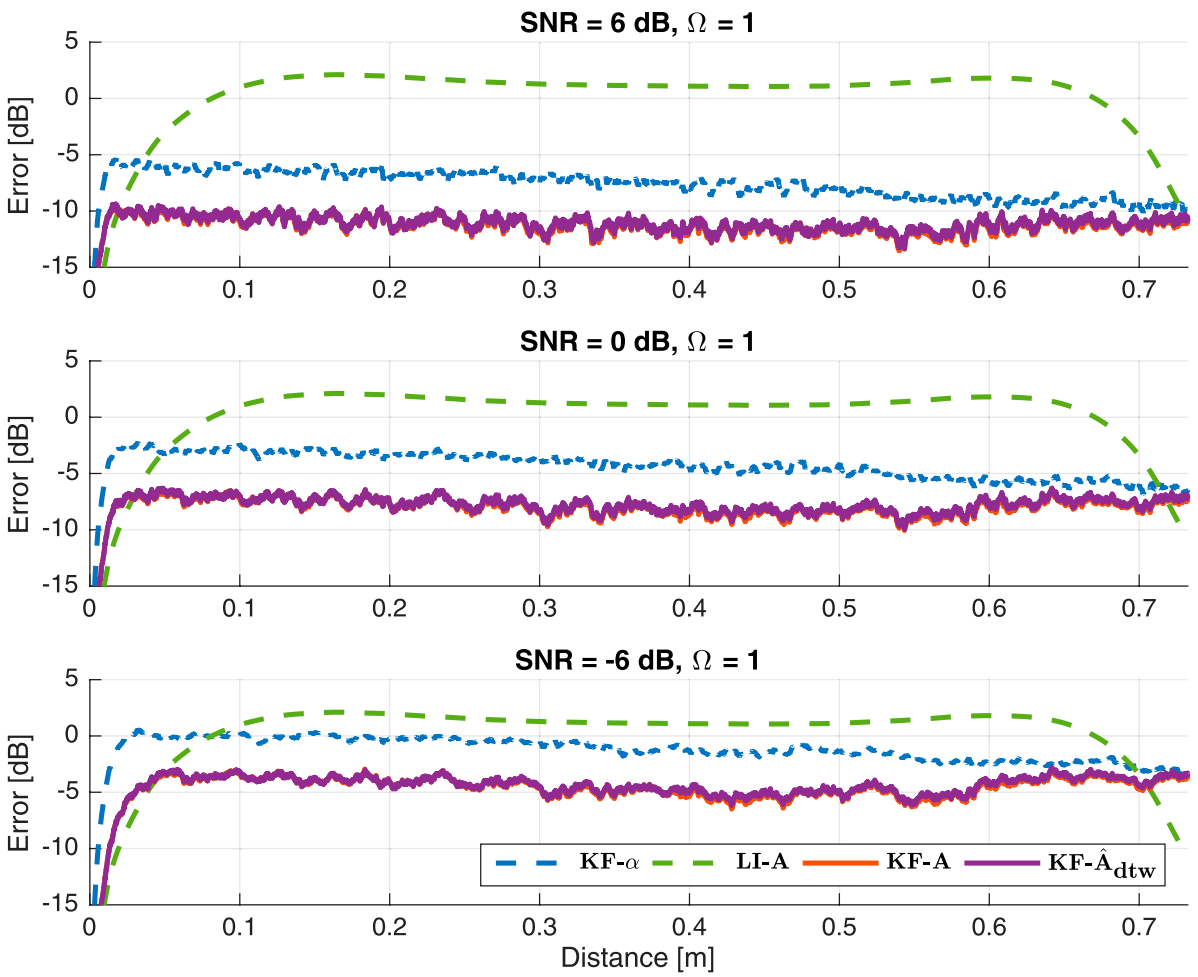

FIGURE 11
Result 2: Normalized misalignment along the trajectory between the "ground-truth" RIR $\mathbf{h}(l)$ and the estimated RIR $\hat{\mathbf{h}}(l)$ using algorithms detailed in Table 1. This assessment is conducted for different levels of noise added to the observed output signal $y_\Omega(l)$.

With Equation 9 and defining $y_\Omega(l) = y(l\Omega)$, $\mathbf{x}_\Omega(l) = \mathbf{x}(l\Omega)$, and $v_\Omega(l) = v(l\Omega)$ for ease of presentation, we can define the state-space model of the proposed approach as

$$\mathbf{h}(l) = \mathbf{A}\mathbf{h}(l-1) + \mathbf{w}(l), \tag{10}$$

$$y_\Omega(l) = \mathbf{x}_\Omega^T(l)\mathbf{h}(l) + v_\Omega(l), \tag{11}$$

where $\mathbf{w}(l)$ is the process noise accounting for modeling errors in the definition of $\mathbf{A}$. From the state-space model, an estimate $\hat{\mathbf{h}}(l)$ of $\mathbf{h}(l)$ can be obtained using the Kalman filter with one recursion per location l. The Kalman filter update equations are outlined in the next section.

3.1 Update equations of the Kalman filter

The Kalman filter (Simon, 2006) can be used to recursively estimate the state defined by a state-space model. For the proposed state-space model in Equations 10 and 11, let $\mathbf{Q} = \mathrm{E}[\mathbf{w}(l)\mathbf{w}^T(l)]$ and $R = \mathrm{E}[v_\Omega^2(l)]$ be the presumed location-invariant covariance matrix of $\mathbf{w}(l)$ and variance of $v_\Omega(l)$, respectively, where $\mathrm{E}[\cdot]$ denotes the expectation operation. The update equations of the Kalman filter for the model in Equations 10 and 11 are then given by

$$\hat{\mathbf{h}}(l) = \mathbf{A}\hat{\mathbf{h}}^+(l-1),. \tag{12}$$

$$\mathbf{P}(l) = \mathbf{A}\mathbf{P}^+(l-1)\mathbf{A}^T + \mathbf{Q},. \tag{13}$$

$$\mathbf{k}(l) = \frac{\mathbf{P}(l)\mathbf{x}_\Omega(l)}{\mathbf{x}_\Omega^T(l)\mathbf{P}(l)\mathbf{x}_\Omega(l) + R}, \tag{14}$$

$$\hat{\mathbf{h}}^+(l) = \hat{\mathbf{h}}(l) + \mathbf{k}(l)\left(y_\Omega(l) - \mathbf{x}_\Omega^T(l)\hat{\mathbf{h}}(l)\right),. \tag{15}$$

$$\mathbf{P}^+(l) = \left(\mathbf{I} - \mathbf{k}(l)\mathbf{x}_\Omega^T(l)\right)\mathbf{P}(l), \tag{16}$$

where Equations 12 and 13 are commonly referred to as the prediction step producing prior estimates, and Equations 14–16 are referred to as the update step producing posterior estimates. In these equations, $\hat{\mathbf{h}}(l)$ and $\hat{\mathbf{h}}^+(l)$ are state estimates, $\mathbf{P}(l)$ and $\mathbf{P}^+(l)$ are estimates of the state-estimation error covariance matrix, $\mathbf{k}(l)$ is the Kalman gain, $\mathbf{I}$ the identity matrix, and the superscript $^+$ distinguishes posterior from prior estimates. In the prediction step Equations 12 and 13, the previously acquired posterior estimates $\hat{\mathbf{h}}^+(l-1)$ and $\mathbf{P}^+(l-1)$ are propagated based on the state Equation 10 only, without taking into account the

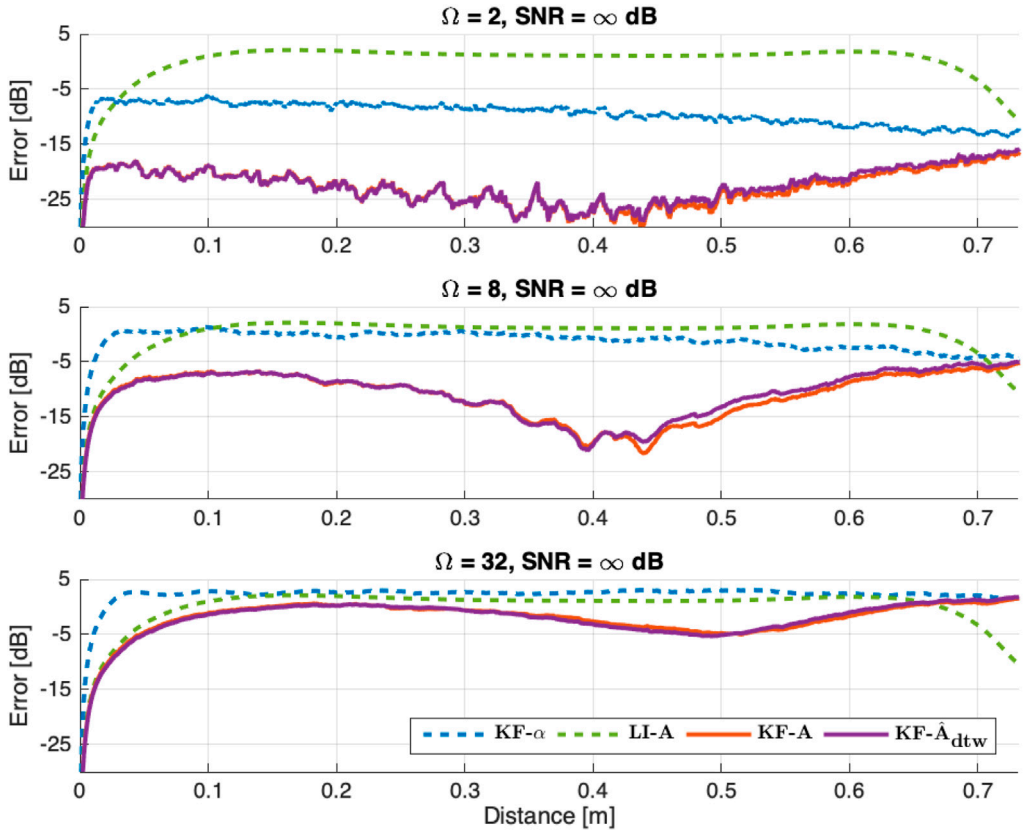

FIGURE 12
Result 3: Normalized misalignment along the trajectory between the "ground-truth" RIR $\mathbf{h}(l)$ and the estimated RIR $\hat{\mathbf{h}}(l)$ using algorithms detailed in Table 1. This is for the case of a noiseless observed output signal $y_\Omega(l)$ with different spatial sampling along the trajectory as a function of Ω.

observation at recursion l, yielding the prior estimates $\hat{\mathbf{h}}(l)$ and $\mathbf{P}(l)$. The update step Equations 14–16 is based on the observation Equation 11. Here, the Kalman gain $\mathbf{k}(l)$ and the error term $y_\Omega(l) - \mathbf{x}_\Omega^T(l)\hat{\mathbf{h}}(l)$ are computed based on $y_\Omega(l)$ and $\mathbf{x}_\Omega(l)$ and are used to update the prior estimates $\hat{\mathbf{h}}(l)$ and $\mathbf{P}(l)$, yielding the posterior estimates $\hat{\mathbf{h}}^+(l)$ and $\mathbf{P}^+(l)$.

To implement the Kalman filter, $\mathbf{Q}$ and R are required as inputs. While R can potentially be obtained from background noise recordings, $\mathbf{Q}$ can be considered a tuning parameter as the variance of $\mathbf{w}(l) = \mathbf{h}(l) - \mathbf{A}\mathbf{h}(l-1)$ will not be known exactly in practice. The Kalman filter also needs to be initialized by defining $\hat{\mathbf{h}}^+(0)$ and $\mathbf{P}^+(0)$. As we assume that $\mathbf{h}(0)$ is known, we set $\hat{\mathbf{h}}^+(0) = \mathbf{h}(0)$. This choice of $\hat{\mathbf{h}}^+(0)$ corresponds to the assumption that the initial state estimation error is 0, such that $\mathbf{P}^+(0)$ can be chosen to have a small norm. Finally, a practical implementation requires an estimate of $\mathbf{A}$, which will be discussed in more detail in Sections 4.2 and 4.3.

At this point, it is instructive to interpret Equations 12–16 in relation to the state of the art as discussed in Section 2. In the interpolation approaches in Kearney et al. (2009) and Geldert et al. (2023) on the one hand, recorded signals are not available, which corresponds to the assumption that $y_\Omega(l) = 0$ and $\mathbf{x}_\Omega(l) = 0$. In this case, we find that $\hat{\mathbf{h}}^+(l) = \hat{\mathbf{h}}(l)$ and $\mathbf{P}^+(l) = \mathbf{P}(l)$ in Equations 15 and 16, i.e. $\hat{\mathbf{h}}(l)$ no longer depends on $\mathbf{P}(l)$ and the Kalman filter update equations effectively reduce to $\hat{\mathbf{h}}(l) = \mathbf{A}\hat{\mathbf{h}}(l-1) = \mathbf{A}^l\mathbf{h}(0)$. As $\mathbf{A}$ is defined by interpolating TOAs of reflections between $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$ based on a room acoustic model, this case is conceptually similar to the aforementioned interpolation approaches and can be referred to as "linear interpolation." The state-space space approaches in Enzner (2010), on the other hand, are obtained if $\Omega = 1$ and the state transition matrix $\mathbf{A}$ is replaced by a scalar α.

4 Proposed room acoustic model-based transition matrix

This section provides a detailed explanation of the methodology followed to obtain a suitable room acoustic model-based transition matrix for use in Equation 10 and is organized as follows. In Section 4.1, we derive an analytical expression for a location-variant transition matrix model $\mathbf{A}(l)$ based on the ISM. A suitable modification of the analytical transition matrix for use in the state-space model is then proposed in Section 4.2 in order to obtain a location-invariant transition matrix model $\mathbf{A}$ valid within the limits of the trajectory. The models presented are parameterized by the exact reflection TOAs, which are not inherently available in practice. Therefore, Section 4.3 introduces an approach based on DTW to obtain an estimate of $\mathbf{A}$ from $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$ by TDOA interpolation.

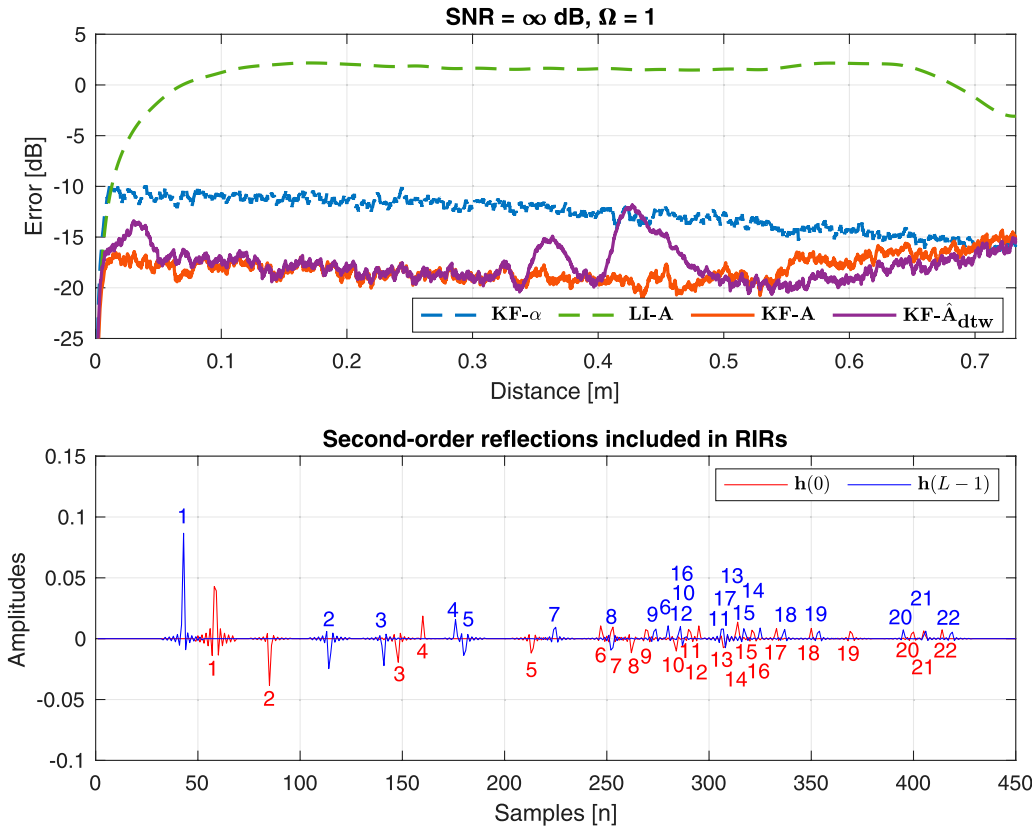

FIGURE 13
(Top) Result 4: Normalized misalignment along the trajectory between the "ground-truth" RIR $\mathbf{h}(l)$ and the estimated RIR $\hat{\mathbf{h}}(l)$ using algorithms detailed in Table 1. This is for the case of a noiseless observed output signal $y_\Omega(l)$ with simulated RIRs inclusive of second-order reflections. (Bottom) Simulated RIRs corresponding to the start and end of the trajectory, denoted as $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$, respectively. Peaks corresponding to the same source or reflection are labeled by the same number in each RIR.

4.1 Analytical location-variant transition matrix model

The objective of this section is to derive a transition matrix model $\mathbf{A}(l) \in \mathbb{R}^{N\times N}$ that approximates the mapping between two RIRs $\mathbf{h}(l)$ and $\mathbf{h}(l-1)$:

$$\mathbf{h}(l) \approx \mathbf{A}(l)\mathbf{h}(l-1). \tag{17}$$

An illustration of this relation along a given trajectory is shown in Figure 3a.

To achieve this objective, we make use of the ISM. In this model, the RIR is expressed as the sum of contributions from the original sound source and additional image sources, which represent reflections within the room. Let the RIR at location l with continuous time-shift t be denoted by $h(l,t)$. With $\delta(t)$ denoting the Dirac delta function, the ISM models $h(l,t)$ as

$$h(l,t) = \sum_{r\in\mathcal{R}} a_r(l)\delta(t-\tau_r(l)), \tag{18}$$

where $r \in \mathcal{R} = \{0, 1, ..R-1\}$ is the (image) source or reflection index, and $\tau_r(l)$ and $a_r(l)$, respectively, denote the TOA and amplitude of the (image) source r observed at location l. Ultimately, we aim to express a discretized version of $h(l,t)$ in terms of a discretized version of $h(l-1,t)$ as in Equation 17. To this end, we start by introducing parameters of $h(l-1,t)$ into $h(l,t)$ as follows.

$$h(l,t) = \sum_{r\in\mathcal{R}} a_r(l-1)\frac{a_r(l)}{a_r(l-1)}\delta(t-\tau_r(l)+\tau_r(l-1)-\tau_r(l-1)). \tag{19}$$

The inclusion of $a_r(l-1)/a_r(l-1)$ and $\tau_r(l-1)-\tau_r(l-1)$ preserves the definition of $h(l,t)$ in Equation 18, while allowing us to find a representation that more closely resembles $h(l-1,t)$. By incorporating $\tau_r(l-1)-\tau_r(l-1)$, we can isolate a Dirac delta function that only includes a time shift of $\tau_r(l-1)$ using the following identity derived from the sifting property (Hoskins, 2009):

$$\begin{aligned}&\delta(t-\tau_r(l)+\tau_r(l-1)-\tau_r(l-1))\\ &=\int_{-\infty}^{+\infty}\delta(t-\tau_r(l)+\tau_r(l-1)-t')\delta(t'-\tau_r(l-1))dt'\end{aligned}$$

Therefore, Equation 19 can alternatively be expressed as

$$h(l,t) = \sum_{r\in\mathcal{R}}\int_{t'}\frac{a_r(l)}{a_r(l-1)}\delta(t-\tau_r(l)+\tau_r(l-1)-t')a_r(l-1)\delta(t'-\tau_r(l-1))\,dt'. \tag{20}$$

To obtain a discrete time-shift representation in terms of the time-shift index n, we impose a band limit with critical sampling period T_s. Band-limiting a Dirac delta function results in a sinc-function (Hoskins, 2009), which will be sampled at discrete time indices. Defining

$$\Delta_r(l) = \tau_r(l) - \tau_r(l-1) \tag{21}$$

as the TDOA of reflection r between locations l and $l-1$, we can now write Equation 20 as

$$h(l,n) = \sum_{r\in\mathcal{R}} \sum_{n'} \frac{a_r(l)}{a_r(l-1)} \mathrm{sinc}\left(n - \frac{\Delta_r(l)}{T_s} - n'\right) a_r(l-1) \mathrm{sinc}\left(n' - \frac{\tau_r(l-1)}{T_s}\right). \tag{22}$$

At this point, it is advantageous to introduce a time-shift-dependent approximation of Equation 22. At any n, the RIR sample $h(l,n)$ is dominated by a subset of reflections only, which we denote as $\tilde{\mathcal{R}}(l,n)$. Due to the decaying nature of the sinc-function, the product $\mathrm{sinc}(n - \Delta_r(l)/T_s - n')\mathrm{sinc}(n' - \tau_r(l-1)/T_s)$ in Equation 22 will have a prominent peak only if the two sinc-functions are closely aligned—that is, if nT_s is near $\Delta_r(l) + \tau_r(l-1) = \tau_r(l)$. In the following, the product is considered negligible for $|nT_s - \tau_r(l)| > \epsilon$, where ϵ is a misalignment threshold and 2ϵ can be understood as the effective temporal width of reflection r. For ease of presentation, we define the interval of width 2ϵ around $\tau_r(l)$ as

$$\mathcal{T}_r(l) = [\tau_r(l) - \epsilon,\ \tau_r(l) + \epsilon], \tag{23}$$

which contains the TOAs of the non-negligible components of reflection r in the RIR at location l. With Equation 23, the misalignment condition $|nT_s - \tau_r(l)| > \epsilon$ can alternatively be expressed as $n \notin \mathcal{T}_r(l)/T_s$. Based on this, we define $\tilde{\mathcal{R}}(l,n)$ as[1]

$$\tilde{\mathcal{R}}(l,n) = \left\{ r \in \mathcal{R} \;\middle|\; n \in \frac{\mathcal{T}_r(l)}{T_s} \right\}. \tag{24}$$

With Equation 24, we can therefore approximate Equation 22 as

$$h(l,n) \approx \sum_{n'} \sum_{r\in\tilde{\mathcal{R}}(l,n)} \frac{a_r(l)}{a_r(l-1)} \mathrm{sinc}\left(n - \frac{\Delta_r(l)}{T_s} - n'\right) a_r(l-1) \mathrm{sinc}\left(n' - \frac{\tau_r(l-1)}{T_s}\right), \tag{25}$$

where we have swapped the order of summation over n and r. Analogous to Equation 18, we introduce the coefficients $h(l-1,n')$ defined by the ISM as

$$h(l-1,n') = \sum_{r'\in\mathcal{R}} a_{r'}(l-1) \mathrm{sinc}\left(n' - \frac{\tau_{r'}(l-1)}{T_s}\right)$$

and rewrite Equation 25 as

$$h(l,n) \approx \sum_{n'} \sum_{r\in\tilde{\mathcal{R}}(l,n)} \frac{a_r(l)}{a_r(l-1)} \mathrm{sinc}\left(n - \frac{\Delta_r(l)}{T_s} - n'\right) h(l-1,n') - e(l,n) \tag{26}$$

with the error term $e(l,n)$ defined as

$$e(l,n) = \sum_{n'} \sum_{\substack{r\in\tilde{\mathcal{R}}(l,n) \\ r'\in\tilde{\mathcal{R}} \\ r'\neq r}} \frac{a_r(l)}{a_r(l-1)} \mathrm{sinc}\left(n - \frac{\Delta_r(l)}{T_s} - n'\right) a_{r'}(l-1) \mathrm{sinc}\left(n' - \frac{\tau_{r'}(l-1)}{T_s}\right). \tag{27}$$

Given the relation between $h(l,n)$ and $h(l-1,n')$ in Equation 26, we can define the elements of $\mathbf{A}(l)$ in Equation 17 assuming that $e(l,n)$ as defined in Equation 27 is negligible. Before turning to the elements of $\mathbf{A}(l)$, let us first verify the conditions under which this assumption holds. As before, we can argue that the product $\mathrm{sinc}(n - \Delta_r(l)/T_s - n')\mathrm{sinc}(n' - \tau_{r'}(l-1)/T_s)$ will be negligible if the two sinc-functions are misaligned, where $|nT_s - \Delta_r(l) - \tau_{r'}(l-1)| > \epsilon$. As shown in Supplementary Appendix SA, this condition is equivalent to

$$\mathcal{T}_{r'}(l-1) \cap \mathcal{T}_r(l-1) = \varnothing, \quad r' \neq r, \tag{28}$$

which essentially says that the reflections in $h(l-1,n)$ should be well-separated in time. Neglecting $e(l,n)$ and referring back to the vector representation of RIRs in Equation 3, it can be seen that Equation 26 can be expressed in the form $\mathbf{h}(l) \approx \mathbf{A}(l)\mathbf{h}(l-1)$ as anticipated in Equation 17, where the element of $\mathbf{A}(l)$ at index $(n+1, n'+1)$ is given by

$$A_{n+1,n'+1}(l) = \sum_{r\in\tilde{\mathcal{R}}(l,n)} \frac{a_r(l)}{a_r(l-1)} \mathrm{sinc}\left(n - \frac{\Delta_r(l)}{T_s} - n'\right) \tag{29}$$

if $\tilde{\mathcal{R}}(l,n) \neq \varnothing$ and otherwise $A_{n+1,n'+1}(l) = 0$—that is, at indices where $h(l,n)$ does not contain reflections. In the actual implementation of such a matrix, one may choose to truncate the sinc-functions in Equation 29 for $|n'T_s - \tau_r(l-1)| > \epsilon$. As illustrated in Figure 4, this results in a matrix $\mathbf{A}(l)$ composed of Toeplitz-structured submatrices of size $\lfloor \frac{2\epsilon}{T_s} \rfloor \times \lfloor \frac{2\epsilon}{T_s} \rfloor$ containing sampled sinc-functions, with one such submatrix per reflection r. These sampled sinc-functions are vertically and horizontally centered around $\tau_r(l)/T_s$ and $\tau_r(l-1)/T_s$, which are non-integer in general.[2] Practically speaking, we note that submatrices centered above the main diagonal of $\mathbf{A}(l)$ shift the corresponding reflection in $\mathbf{h}(l-1)$ toward smaller TOAs, while entries below shift it toward larger TOAs.

4.2 Analytical location-invariant transition matrix model

For the application at hand, we anticipate employing a location-invariant transition matrix model as described in Equation 9 and illustrated in Figure 3b. We show that by using a location-invariant matrix, in which TOA intervals are defined to span the entire trajectory, we can reduce the necessity for accurate TOA estimates. This efficiency is achieved because these TOA

1 For notational convenience when working with an interval $[a,b]$, we introduce $\frac{[a,b]}{c}$ to represent $[\frac{a}{c}, \frac{b}{c}]$ (see Equations 24 and 34), and $[a,b] - c$ to represent $[a-c, b-c]$ (Equation 38).

2 In the special case that $\tau_r(l)/T_s$ and $\tau_r(l-1)/T_s$ are integers, the sinc-functions are sampled symmetrically around their peak, which implies that all but the peak sample lie at the zero crossings of the sinc-functions, resulting in diagonal submatrices.

intervals, along with TDOA estimates, can be obtained directly from $\mathbf{h}(0)$ to $\mathbf{h}(L-1)$ using the DTW-based method outlined in Section 4.3. In this section, we therefore derive a location-invariant transition matrix $\mathbf{A}$ from the definition of $\mathbf{A}(l)$ in Equation 29. As will be shown, location-invariance can be achieved within the limits of the trajectory under some assumptions.

In Equation 29, we observe two dependencies on l: in the TDOA $\Delta_r(l)$ and in the set $\tilde{\mathcal{R}}(l,n)$. The dependency of $\Delta_r(l)$ on l disappears under the following conditions. As illustrated in Figure 5, under the assumptions of a far-field scenario where the (image) sources are far away in relation to the length of the trajectory, $a_r(l)/a_r(l-1)=1$ and $\Delta_r(l)=\Delta_r$, meaning corresponding reflection TOAs are shifted by the same amount at each step along the trajectory. Consequently, the reflection TOA $\tau_r(l)$ can be approximated by the linear relation

$$\tau_r(l)=\tau_r(0)+l\Delta_r. \tag{30}$$

The dependence of $\tilde{\mathcal{R}}(l,n)$ on l as defined in Equation 24 can be resolved as follows. The model in Equation 9 is required to hold for $l=1,\ldots,L-1$. Over this range of l, the components of reflection r occupy the interval

$$\mathcal{T}_r=\left[\tau_{r|\min},\,\tau_{r|\max}\right] \tag{31}$$

$$\text{with}\quad \tau_{r|\min}=\min\left(\tau_r(1),\,\tau_r(L-1)\right)-\epsilon, \tag{32}$$

$$\tau_{r|\max}=\max\left(\tau_r(1),\,\tau_r(L-1)\right)+\epsilon, \tag{33}$$

which can be understood as an extension of Equation 23.[3]

Based on Equation 31, we define the set similar to Equation 24 without the dependency on l:

$$\tilde{\mathcal{R}}(n)=\left\{r\in\mathcal{R}\;\middle|\;n\in\frac{\mathcal{T}_r}{T_s}\right\} \tag{34}$$

The approximation in Equation 26 can then be replaced by

$$h(l,n)\approx\sum_{n'}\sum_{r\in\tilde{\mathcal{R}}(n)}\operatorname{sinc}\left(n-\frac{\Delta_r}{T_s}-n'\right)h(l-1,n')-e(l,n), \tag{35}$$

with the error term $e(l,n)$ redefined as

$$e(l,n)=\sum_{n'}\sum_{\substack{r\in\tilde{\mathcal{R}}(n)\\ r'\in\mathcal{R}\\ r'\neq r}}\operatorname{sinc}\left(n-\frac{\Delta_r}{T_s}-n'\right)\operatorname{sinc}\left(n'-\frac{\tau_{r'}(l-1)}{T_s}\right). \tag{36}$$

Again, the product $\operatorname{sinc}(n-\Delta_r/T_s-n')\operatorname{sinc}(n'-(\tau_{r'}(l-1)/T_s)$ in Equation 36 will be negligible if the two sinc-functions are misaligned, where $|nT_s-\Delta_r-\tau_{r'}(l-1)|>\epsilon$. As shown in the Supplementary Appendix SB, this condition is equivalent to

$$\mathcal{T}_{r'}^{-}\cap\mathcal{T}_r^{-}=\varnothing,\qquad r'\neq r \tag{37}$$

where

$$\begin{aligned}\mathcal{T}_r^{-}&=\mathcal{T}_r-\Delta_r\\&=\left[\tau_{r|\min}^{-},\,\tau_{r|\max}^{-}\right],\end{aligned} \tag{38}$$

$$\text{with}\quad \tau_{r|\min}^{-}=\tau_{r|\min}-\Delta_r=\min\left(\tau_r(0),\,\tau_r(L-2)\right)-\epsilon-\Delta_r, \tag{39}$$

$$\tau_{r|\max}^{-}=\tau_{r|\max}-\Delta_r=\max\left(\tau_r(0),\,\tau_r(L-2)\right)+\epsilon-\Delta_r, \tag{40}$$

and the limits in Equations 39, 40 are obtained from Equations 32, 33, and $\tau_r(1)-\Delta_r=\tau_r(0)$ and $\tau_r(L-1)-\Delta_r=\tau_r(L-2)$ according to Equation 30. The condition in Equation 37 essentially states that the intervals that the individual reflections occupy over the range of the trajectory (with exclusion of the end point $l=L-1$) may not overlap. This can be understood as an extension of condition Equation 28 derived for the location-variant transition matrix. Neglecting $e(l,n)$ and referring back to the vector representation of RIRs in Equation 3, it can be seen that Equation 35 can be expressed in the form $\mathbf{h}(l)\approx\mathbf{A}\mathbf{h}(l-1)$ as anticipated in Equation 9, where the element of $\mathbf{A}$ at index $(n+1,n'+1)$ is given by

$$A_{n+1,n'+1}=\sum_{r\in\tilde{\mathcal{R}}(n)}\operatorname{sinc}\left(n-\frac{\Delta_r}{T_s}-n'\right). \tag{41}$$

if $\tilde{\mathcal{R}}(n)\neq\varnothing$ and otherwise $A_{n+1,n'+1}(l)=0$.

Given the assumption of access to exact reflection TOAs $\tau_r(0)$ and $\tau_r(L-1)$, we estimate Δ_r using Equation 30 as follows:

$$\Delta_r=\frac{\tau_r(L-1)-\tau_r(0)}{L-1}, \tag{42}$$

Note that from a conceptual point of view, the terms $\tau_r(l)$, $\tau_r(0)$, $\tau_r(L-1)-\tau_r(0)$, and $l/(L-1)$ in Equations 30 and 42 are comparable to $\mathbf{n}_r(l)$, $\mathbf{n}_r(0)$, $\mathbf{n}_r(L-1)-\mathbf{n}_r(0)$, and $\beta(l)$ in Equation 5, respectively. Note, however, that the latter are sample-based, while the former are defined in continuous time-shifts and can indeed be used to implement non-integer shifts by using the sinc-functions in Equation 29.

Keeping in mind Figures 3a and b, for an intuitive understanding of the relationship between $\mathbf{A}(l)$ and $\mathbf{A}$, we refer the reader to the toy example illustrated in Figure 6. Here, a trajectory with $L=3$ is considered, and matrices $\mathbf{A}(l)$ and $\mathbf{A}$ are, respectively, formed using Equations 29 or 41 with integer TOAs and $2\epsilon/T_s=3$. It can be seen that the result of taking $\mathbf{A}^{L-1}$ yields an identical matrix to that of the multiplication of $\mathbf{A}(L-1)\cdots\mathbf{A}(2)\mathbf{A}(1)$ (Figures 6c, e). Furthermore, we note that the resulting matrix is equivalent to the matrix constructed using either Equations 29 or 41 for the case of $L=2$: a transition matrix between the start and end points of the trajectory as shown in Figure 3c.

4.3 Dynamic Time Warping transition matrices

As it stands, the analytical solution proposed in Section 4.2 cannot be applied directly to RIRs and requires inherent knowledge of the TOAs in $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$. Therefore, we propose an approach to estimate the matrix $\mathbf{A}$ using a DTW algorithm to temporally align elements in $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$. DTW achieves temporal alignment by flexibly warping time, ensuring that similar patterns in the two sequences are matched, despite variations in timing. As discussed in Section 2.2, prior applications of DTW have involved linear interpolation between two impulse responses. However, the method proposed here builds upon the observation that the warp path derived from DTW can be linked to both the set $\tilde{\mathcal{R}}(n)$ and $\tau_r(L-1)-\tau_r(0)$, the latter of which is necessary for

3 Note that if $\Delta_r<2\epsilon$, we find that $\mathcal{T}_r=\bigcup_{l=1}^{L-1}\mathcal{T}_r(l)$.

estimating Δ_r via Equation 42. Leveraging these parameters enables the construction of an appropriate analytical location-invariant transition matrix $\mathbf{A}$, denoted as $\hat{\mathbf{A}}_{\text{dtw}}$.

The DTW algorithm (Müller, 2007) involves computing an accumulated cost matrix $\mathbf{D} \in \mathbb{R}^{(N+1)\times(N+1)}$, representing the cumulative costs of aligning each element in $\mathbf{h}(L-1)$ with each element in $\mathbf{h}(0)$, respectively, indexed by n and n'. Here, the cost is defined as the Euclidean distance $\|h(L-1,n) - h(0,n')\|_2$, and $\mathbf{D}$ is constructed using the recurrence relation

$$D_{n+2,n'+2} = \|h(L-1,n) - h(0,n')\|_2 + \min\{D_{n+1,n'+2}, D_{n+2,n'+1}, D_{n+1,n'+1}\}, \tag{43}$$

which is initialized such that $D_{1,1} = 0$, while $D_{n+2,1}$ and $D_{1,n'+2}$ are set to ∞ for all n and n', respectively[4]. Owing to the prominent reflection peaks in the structure of early RIRs, the matrix $\mathbf{D}$ will be formed with grid-like entries (Figure 7b)

A warp path, denoted by pairs of indices $(n+2, n'+2)$ representing optimal alignment in the sense of smallest accumulated cost, is established by backtracking from $D_{N+2,N+2}$ to $D_{2,2}$. The recurrence relation in Equation 43 guarantees a monotonically increasing order, thus ensuring that peaks in sequences are mapped in the same order they appear, preserving temporal correspondence during alignment. Note that a single point in one sequence can correspond to multiple points in the other, and vice versa. An illustration of the warp path through $\mathbf{D}$ can be seen in Figure 7b.

The elements along the horizontal and vertical "gridlines" in $\mathbf{D}$ contain local maxima values, whose row and column indices should respectively correspond to integer estimates of reflection TOA time-shift indices $\frac{\tau_r(L-1)}{T_s}$ and $\frac{\tau_r(0)}{T_s}$. The elements at the intersections of these "gridlines" represent local minima in $\mathbf{D}$, signifying potential alignment points between reflection peaks in $\mathbf{h}(L-1)$ and $\mathbf{h}(0)$. Consequently, the warp path tends to traverse these intersections in a diagonal manner, guided by the minimization of alignment costs. Here, a clear link between the shape of the warp path through $\mathbf{D}$ (Figure 7b) and the analytical matrix $\mathbf{A}$ constructed using Equation 41 for $L = 2$ (Figure 7a) becomes apparent. Recall that the case of $L = 2$ results in a matrix $\mathbf{A}$ that maps the RIR at the start of the trajectory to the RIR at its end (Figure 3c).

In order to use the shape of the warp path through $\mathbf{D}$, we need to construct a matrix $\mathbf{W} \in \mathbb{R}^{N\times N}$, with ones placed at the indices corresponding to the pairs in the warp path and zeros elsewhere:

$$W_{n+1,n'+1} = \begin{cases} 1 & \text{if } (n+2, n'+2) \text{ is in the warp path} \\ 0 & \text{otherwise} \end{cases}$$

Such a matrix $\mathbf{W}$ is shown in Figure 8d for the previous toy problem with integer TOAs and $\frac{2\epsilon}{T_s} = 3$. It can be seen that the diagonal segments of $\mathbf{W}$ are consistent with the toy problem matrix $\mathbf{A}^{L-1}$, which is again displayed in Figure 8b. We therefore propose to leverage the positions of the diagonal segments in $\mathbf{W}$ to obtain the estimates of Δ_r and $\tilde{\mathcal{R}}(n)$ needed to construct $\mathbf{A}$ as follows.

Firstly, integer estimates of $(\tau_r(L-1) - \tau_r(0))/T_s$ are obtained by measuring the "distances" (number of elements) from the diagonal segments in $\mathbf{W}$ to the main diagonal. These estimates can in turn be used to find estimates of Δ_r using the relation in Equation 42 and the known sampling period T_s. We employ $\hat{\Delta}_r$ and $\hat{\tau}_r$ to distinguish the estimates found in this manner:

$$\hat{\Delta}_r = \frac{\hat{\tau}_r(L-1) - \hat{\tau}_r(0)}{L-1}, \tag{44}$$

Secondly the start and end indices of the diagonal segments, denoted, respectively, as $(n_{r|\text{st}}, n'_{r|\text{st}})$ and $(n_{r|\text{en}}, n'_{r|\text{en}})$, are retrieved. These indices are then used to find a suitable estimate of the range $\mathcal{T}_r$ for the set $\tilde{\mathcal{R}}(n)$ in Equation 34 as

$$\hat{\mathcal{T}}_r = [\hat{\tau}_{r|\min}, \hat{\tau}_{r|\max}] \tag{45}$$

$$\hat{\tau}_{r|\min} = T_s \cdot \min\left(n'_{r|\text{st}} + \frac{\hat{\Delta}_r}{T_s}, n_{r|\text{st}}\right), \tag{46}$$

$$\hat{\tau}_{r|\max} = T_s \cdot \max\left(n_{r|\text{en}}, n'_{r|\text{en}} + \frac{\hat{\Delta}_r}{T_s}\right), \tag{47}$$

It should be ensured that the estimated ranges still result in the condition in Equation 37 being met—that is, that corresponding column ranges do not overlap. Finally, we construct the matrix $\hat{\mathbf{A}}_{\text{dtw}}$ using Equations 41 and 34 with $\Delta_r = \hat{\Delta}_r$ and $\mathcal{T}_r = \hat{\mathcal{T}}_r$ where $\hat{\Delta}_r$ and $\hat{\mathcal{T}}_r$ are obtained using Equations 44 to 47. Figure 8c illustrates an example of a matrix obtained using this method for the previous toy problem. It is apparent that $\hat{\mathbf{A}}_{\text{dtw}}$ closely aligns with the desired $\mathbf{A}$ matrix shown in Figure 8a. Importantly, the described method allows us to derive $\mathbf{A}$ without needing to explicitly estimate individual TOAs.

5 Simulations

In this section, we systematically evaluate the performance of transition matrices $\mathbf{A}$ and $\hat{\mathbf{A}}_{\text{dtw}}$ in state-space RIR estimation. The algorithms considered are summarized in Table 1, with **LI** – $\mathbf{A}$ and **KF** – α serving as reference methods, and **KF** – $\mathbf{A}$ and **KF** – $\hat{\mathbf{A}}_{\text{dtw}}$ representing our primary contributions from the proposed methods outlined in Section 3. In Section 5.1, we provide details on the experimental setup, including the simulated acoustic environment, Kalman filter parameters, and the performance measure used. The results of the experiments are presented in Section 6.

5.1 Acoustic environment

For simplicity, we consider a box-shaped room with dimensions bounded by $[0, 4.50] \times [0, 5.80] \times [0, 2.90]$ m. We place a fixed-position sound source at $[1.05, 2.98, 1.17]$ m, and define a linear trajectory for the microphone spanning approximately 0.7 m, starting at $[1.94, 3.10, 1.09]$ m and ending at $[1.99, 2.95, 0.37]$ m. Along this trajectory, we have established that all first-order reflections arrive in the same order. The set-up is shown in Figure 9.

4 In this paper, matrices are indexed from (*1,1*), following the convention of many programming languages. However, our RIR sequences are indexed from 0. To align with the zero-based indexing of our RIRs and accommodate the additional row and column used solely for initialization, we index the entries of the accumulated cost matrix as $D_{n+2,n'+2}$

At a standard speech sampling rate of 16 kHz (Nophut et al., 2024) and with a microphone velocity of 0.25 m/s, the ISM (Allen and Berkley, 1979) is used to simulate the ground-truth early RIRs $\mathbf{h}(l)$, including reflections up to the second order, at various points along the trajectory. Subsequently, the observed output signal $y_\Omega(l)$ is obtained using Equation 11 for a given input $x_\Omega(l)$ and measurement noise $v_\Omega(l) \sim \mathcal{N}(0, \sigma_y^2)$, where the variance σ_v^2 is adjusted based on the experimental conditions. White Gaussian noise is chosen as the input signal, denoted as $x(l) \sim \mathcal{N}(0, \sigma_x^2)$, with the variance σ_x^2 set as $10\log_{10}(\sigma_x^2) = -20$ dB.

5.2 Kalman filter parameters

As outlined in Section 3.1, the Kalman filter relies on specific input parameters for its operation. These parameters include the process noise covariance matrix $\mathbf{Q}$ and the variance of the measurement noise R. Additionally, the filter requires initialization with a state estimate $\hat{\mathbf{h}}^+(0)$ and an initial guess for the state estimation error covariance matrix $\mathbf{P}^+(0)$.

Given the assumption of knowledge of the RIR at the start of the trajectory $\mathbf{h}(0)$, the filter is initialized with $\hat{\mathbf{h}}^+(0) = \mathbf{h}(0)$. Since we are simulating measurement noise with variance σ_v^2, we set $R = \sigma_v^2$. As previously mentioned, R can in practice potentially be estimated using background noise recordings. The appropriate $\mathbf{A}$ matrix is selected for each algorithm, as outlined in Table 1, with the considered width 2ϵ of the sinc functions in the proposed analytical transition matrices set to $\frac{2\epsilon}{T_s} = 20$. It was found that for the reference Kalman filter, a transition factor of $\alpha = 1$ was suitable, and varying this did not change the results in a significant manner. After some initial testing of the Kalman filter, the tuning parameter $\mathbf{Q}$ was chosen as $\mathbf{Q} = \sigma_w^2\mathbf{I}$, where $\mathbf{I}$ is an identity matrix and $10\log(\sigma_w^2) = -30$ dB.

5.3 Performance measure

In assessing the accuracy of our estimated RIRs $\hat{\mathbf{h}}^+(l)$, we employ the normalized misalignment $\mathcal{M}_{\text{dB}}(l)$. This metric is defined thus:

$$\mathcal{M}_{\text{dB}}(l) = 20\log_{10}\left(\frac{\|\hat{\mathbf{h}}^+(l) - \mathbf{h}(l)\|_2}{\|\mathbf{h}(l)\|_2}\right).$$

Here, $\mathbf{h}(l)$ represents the ground truth early RIRs.

5.4 Experiments

We categorize experiments into the following four sets to evaluate algorithm performance.

- **Experiment 1: Ideal case.** In the first set of experiments, we examine the "ideal" scenario where no noise is present on the observed output signal $y_\Omega(l)$ ($\sigma_v^2 = 0$) and every sample along the trajectory is considered ($\Omega = 1$). Simulated RIRs exclusively include first-order reflections.
- **Experiment 2: Noise sensitivity**. Building upon the setup in Experiment 1 with $\Omega = 1$ and only first-order reflections included, the second set of experiments evaluates algorithm robustness when different noise levels are added to the observed output signal—that is, varying SNR = $\{6, 0, -6\}$ dB.
- **Experiment 3: Spatial sampling effects.** The third set of experiments employs different spatial downsampling factors Ω to assess the impact of using fewer data points along the trajectory. In this case, we once again consider $\sigma_v^2 = 0$ and only first-order reflections, while varying $\Omega = \{2, 8, 32\}$. Note that $\Omega = 2$ corresponds to 50% of the samples along the trajectory being used.
- **Experiment 4: Second-order reflections included.** The final set of experiments investigates the inclusion of second-order reflections in the RIRs and the corresponding simulated observed signal. We once again examine the "ideal" scenario where no noise is present on the observed output signal $y_\Omega(l)$ ($\sigma_v^2 = 0$) and every sample along the trajectory is considered ($\Omega = 1$). Please note that the analytical transition matrix $\mathbf{A}$ is still constructed solely using first-order TOAs, as not all second-order reflections meet the necessary assumptions along the given trajectory. As a robustness measure, after constructing $\mathbf{A}$, ones are added along the main diagonal in any part of the matrix with empty rows. This step ensures that second-order reflections are not entirely discarded. On the other hand, the DTW algorithm operates directly on the RIRs, and thus, when constructing $\hat{\mathbf{A}}_{\text{dtw}}$, it inherently assumes that all reflections arrive in the same order in $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$. In other words, it may incorrectly associate some reflections in $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$ as belonging to the same image source.

6 Results

We present the results corresponding to the experiments outlined above. The reader is urged to take careful note of the different y-axis scaling in the presented results.

6.1 Result 1: ideal case

The results of Experiment 1 are presented at the top of Figure 10, where the normalized misalignment $\mathcal{M}_{\text{dB}}(l)$ between the ground-truth RIR $\mathbf{h}(l)$ and the estimated RIR $\hat{\mathbf{h}}(l)$ is displayed for the various algorithms at different positions along the trajectory. The bottom of Figure 10 shows the start and end point RIRs used to obtain our transition matrices. This is provided to give the reader an idea of how the source and reflection TOAs change over the trajectory. Furthermore, it can be seen from the labeled TOAs that there are no overlapping intervals, which aligns with our assumptions.

At the beginning of the trajectory, all algorithms exhibit zero error owing to their initialization with a known RIR $\mathbf{h}(0)$. Following a subsequent steep increase in error, distinctive properties of the curves resulting from the different algorithms become evident. The reference Kalman filter algorithm ($\mathbf{KF} - \alpha$), which utilizes a transition factor of $\alpha = 1$ and relies solely on microphone observations, demonstrates a gradual decrease in error throughout the remainder of the trajectory. Excluding the start and end segments, it can be observed that Algorithm $\mathbf{LI} - \mathbf{A}$,

whose performance solely relies on the model of **A**, results in an error curve with a subtle dip in it. This non-monotonic behavior can be attributed to the position of the direct source relative to the linear microphone trajectory (Figure 9) and the fact that the source is not sufficiently in the far-field for our assumption that TDOAs remain approximately constant to hold. Toward the end of the trajectory, Algorithm **LI** − **A** exhibits a rapid decrease in error. This is because the matrix **A** is derived from accurate TOAs at the start and end of the trajectory, and therefore estimates should most closely align with the ground-truth RIRs around these positions.

In general, changes in the error curve of the Kalman filter algorithms using a transition matrix (**KF** − **A** and **KF** − $\hat{\mathbf{A}}_{\text{dtw}}$) correlate with changes in the error curve of Algorithm **LI** − **A**. However, a more significant dip in error is observed, and the curves nearly converge with those of Algorithm **KF** − α at the end of the trajectory. This behavior may be a result of the rapid convergence of **KF** − **A** and **KF** − $\hat{\mathbf{A}}_{\text{dtw}}$ under ideal conditions (i.e., no noise), making the Kalman filter more sensitive to the degradation of the transition matrix model accuracy.

If we now consider the relative performance of the algorithms, we note that the Kalman filter employing "ideal" analytical transition matrix Algorithm **KF** − **A** showcases a substantial improvement which is, on average, approximately −15 dB compared to the reference Kalman filter, Algorithm **KF** − α. Importantly, similar improvement extends to the case where the analytical transition matrix approximated using DTW is used—Algorithm **KF** − $\hat{\mathbf{A}}_{\text{dtw}}$. Moreover, the advantage of employing a state-space model over a linear interpolation method becomes evident from the comparatively poor overall results of Algorithm **LI** − **A**. The observed enhancements in accuracy support our preliminary hypothesis: the combination of room acoustic model-based interpolation and a data-driven state-space model provides a more accurate approach to early RIR estimation.

6.2 Result 2: noise sensitivity

The results of Experiment 2 are presented in Figure 11 and once again exhibit error curve shapes consistent with those observed in Experiment 1. However, as anticipated, overall performance is worse, except for Algorithm **LI** − **A**, which remains unaffected since it does not rely on the noisy microphone measurements. Additionally, we observe slower convergence and less pronounced dips in the error curves of the algorithms **KF** − **A** and **KF** − $\hat{\mathbf{A}}_{\text{dtw}}$ compared to the noiseless case. This is likely because the less reliable observations prevent convergence to such low error levels in the first place, making the dips less noticeable.

It is evident that even under poor SNR conditions, our methods consistently outperform both reference algorithms. Notably, at an SNR of −6 dB, an average of approximately −5 dB difference persists, further underscoring the advantages of the proposed approach, even when measurements contain high levels of noise.

6.3 Result 3: spatial sampling effects

The results of Experiment 3 are presented in Figure 12. As fewer measurements are used, the disparity between the results obtained using the linear interpolation method and those obtained using a Kalman filter naturally diminishes. Additionally, the performance of the reference Kalman filter (**KF** − α) deteriorates more rapidly than that of our proposed methods (**KF** − **A** and **KF** − $\hat{\mathbf{A}}_{\text{dtw}}$). We further note an apparent shift in the position of the dip in the error curve of Algorithms **KF** − **A** and **KF** − $\hat{\mathbf{A}}_{\text{dtw}}$ as fewer samples are used. This is likely a consequence of the change in how often the state equation is propagated.

For $\Omega = 8$, equivalent to approximately 12.5% of the available sampling points, Algorithms **KF** − α and **LI** − **A** exhibit similar performance, while our proposed algorithms consistently outperform both by an average of approximately −12 dB. Despite the trivial choice between the two reference methods when only a small number of measurements are available, as demonstrated by the case of $\Omega = 32$ (equivalent to 3.13% of sampling points), our proposed method still demonstrates an overall improvement in performance.

6.4 Result 4: second-order reflections included

The results of Experiment 4 are shown at the top of Figure 13. It is important to recall that in Experiment 4, the analytical transition matrix **A** used in Algorithm **KF** − **A** is still constructed using only first-order TOAs. This is because not all second-order reflections meet the necessary assumptions along the given trajectory, as indicated by the reflection TOA labels toward the end of the RIR at the bottom of Figure 13—for example, the interval over which reflection 6 moves overlaps with reflections 7, 8, and 9. However, by adding ones along the diagonal of **A** in any empty rows, we ensure that second-order reflections are not entirely discarded. This allows us to leverage prior knowledge, even if incomplete, and explains the strong convergence with Algorithm **KF** − α toward the end of the trajectory. That being said, Algorithm **KF** − **A** still demonstrates an average improvement of approximately −7 dB over Algorithm **KF** − α.

It is important to reiterate that in estimating the matrix $\hat{\mathbf{A}}_{\text{dtw}}$, we directly utilize the RIRs $\mathbf{h}(0)$ and $\mathbf{h}(L-1)$, thus implicitly incorporating second-order reflection information. However, in this case, not all second-order reflections adhere to the assumption that the time intervals' individual reflections occupy over the trajectory do not overlap. Given that DTW inherently maps sequences in a monotonic manner, assuming consistent order among significant points like peaks, certain reflection peaks might be incorrectly mapped between RIRs. Consequently, the points along the trajectory where second-order reflections in the ground-truth RIRs begin to overlap likely correspond to the jumps observed in the error curve of Algorithm **KF** − $\hat{\mathbf{A}}_{\text{dtw}}$. Nonetheless, this serves as a robust test demonstrating that despite deviations from certain assumptions regarding all reflections, the proposed estimated matrix still yields an overall enhancement over Algorithm **KF** − α.

7 Conclusion

This paper investigates the estimation of early segments of time-varying RIRs through a state-space model incorporating the ISM within the state transition matrix. Simulation results indicate that this approach outperforms both RIR interpolation and a purely

data-driven state-space model using a transition factor. Moreover, a practical method for estimating such a matrix has been proposed and has a similar performance to the "ideal" analytical transition matrix derived. It is important to acknowledge that certain assumptions inherent to our method limit its application within specific areas of a room. This necessitates further research in order to improve the robustness of the approach, potentially through adaptive estimation of the transition matrix. Furthermore, experimental validation using real measurement data is required to assess the effectiveness of the proposed approach in real-world scenarios.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

KM: conceptualization, formal analysis, investigation, methodology, validation, writing–original draft, and writing–review and editing. TD: conceptualization, methodology, supervision, writing–original draft, and writing–review and editing. RA: conceptualization, supervision, writing–original draft, and writing–review and editing. TV: conceptualization, funding acquisition, supervision, and writing–review and editing.

Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by

- The European Research Council under the European Union's Horizon 2020 research and innovation program/ERC Consolidator Grant: SONORA (no. 773268). This paper reflects only the authors' views, and the EU is not liable for any use that may be made of the contained information.
- KU Leuven Internal Funds C14/21/075 "A holistic approach to the design of integrated and distributed digital signal processing algorithms for audio and speech communication devices"
- FWO Research Project: "The Boundary Element Method as a State-Space Realization Problem" (G0A0424N)

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frsip.2024.1426082/full#supplementary-material

References

Ajdler, T., Sbaiz, L., and Vetterli, M. (2007). Dynamic measurement of room impulse responses using a moving microphone. *J. Acoust. Soc. Amer. (JASA)* 122, 1636–1645. doi:10.1121/1.2766776

Allen, J. B., and Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Amer. (JASA)* 65, 943–950. doi:10.1121/1.382599

Antonello, N., De Sena, E., Moonen, M., Naylor, P. A., and Van Waterschoot, T. (2017). Room impulse response interpolation using a sparse spatio-temporal representation of the sound field. *IEEE/ACM Trans. Audio Speech Lang. Process. (TransASLP_ACM)* 25, 1929–1941. doi:10.1109/taslp.2017.2730284

Antweiler, C., and Symanzik, H. G. (1995). "Simulation of time variant room impulse responses," in Proc. 1995 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '95), Detroit, MI, May 08-12, 1995, Vol. 5, 3031–3034. doi:10.1109/icassp.1995.479484

Antweiler, C., Telle, A., Vary, P., and Enzner, G. (2012). "Perfect-sweep NLMS for time-variant acoustic system identification," in Proc. 2012 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '12), Kyoto, Japan, 25-30 March 2012, 517–520. doi:10.1109/ICASSP.2012.6287930

Borra, F., Gebru, I. D., and Markovic, D. (2019). "Soundfield reconstruction in reverberant environments using higher-order microphones and impulse response measurements," in Proc. 2021 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '19), Brighton, Great Britain, 12-17 May 2019, 281–285. doi:10.1109/ICASSP.2019.8682961

Bruschi, V., Nobili, S., Cecchi, S., and Piazza, F. (2020). "An innovative method for binaural room impulse responses interpolation," in Audio Engineering Society Convention 148 (AES148Conv).

Carty, B. (2010). "Movements in binaural space: issues in HRTF interpolation and reverberation, with applications to computer music,". Ph.D. thesis (Maynooth, Ireland: National University of Ireland).

Cherniakov, M. (2003). *An introduction to parametric digital filters and oscillators.* John Wiley & Sons.

Crocco, M., and Bue, A. D. (2015). "Room impulse response estimation by iterative weighted l1-norm," in Proc. 23rd European Signal Process. Conf. (EUSIPCO '15), Nice, France, August 31-September 4, 2015, 1895–1899.

Das, O., Calamia, P., and Gari, S. V. A. (2021). "Room impulse response interpolation from a sparse set of measurements using a modal architecture," in Proc. 2021 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '21), Toronto, Canada, 6-11 June 2021, 960–964. doi:10.1109/ICASSP39728.2021.9414399

Elko, G. W., Diethorn, E., and Gänsler, T. (2003). "Room impulse response variation due to temperature fluctuations and its impact on acoustic echo cancellation," in Proc. 2003 Int. Workshop Acoustic Echo Noise Control (IWAENC '03), Kyoto, Japan, September 8-11, 2003, 67–70.

Enzner, G. (2008). "Analysis and optimal control of lms-type adaptive filtering for continuous-azimuth acquisition of head related impulse responses," in Proc. 2008 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '08), Las Vegas, Nevada, 31 March-4 April 2008, 393–396. doi:10.1109/ICASSP.2008.4517629

Enzner, G. (2010). "Bayesian inference model for applications of time-varying acoustic system identification," in Proc. 18th European Signal Process. Conf. (EUSIPCO '10), Aalborg, Denmark, August 23-27, 2010, 2126–2130.

Evers, C., Löllmann, H. W., Mellmann, H., Schmidt, A., Barfuss, H., Naylor, P. A., et al. (2020). The locata challenge: Acoustic source localization and tracking. *IEEE/ACM*

Trans. Audio Speech Lang. Process. (TransASLP_ACM) 28, 1620–1643. doi:10.1109/taslp.2020.2990485

Garcia-Gomez, V., and Lopez, J. J. (2018). "Binaural room impulse responses interpolation for multimedia real-time applications," in AES 144th Convention (AES148Conv), Milan, Italy, May 23-26, 2018.

Geldert, A., Meyer-Kahlen, N., and Schlecht, S. J. (2023). "Interpolation of spatial room impulse responses using partial optimal transport," in Proc. 2023 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '23), Rhodes Island, Greece, June 4-June 10, 2023, 1–5. doi:10.1109/ICASSP49357.2023.10095452

Hahmann, M., and Fernandez-Grande, E. (2022). A convolutional plane wave model for sound field reconstruction. *J. Acoust. Soc. Amer. (JASA)* 152 (5), 3059–3068. doi:10.1121/10.0015227

Hahn, N., and Spors, S. (2015). "Continuous measurement of impulse responses on a circle using a uniformly moving microphone," in Proc. 23rd European Signal Process. Conf. (EUSIPCO '15), Nice, France, August 31-September 4, 2015, 2536–2540. doi:10.1109/EUSIPCO.2015.7362842

Haneda, Y., Kaneda, Y., and Kitawaki, N. (1999). Common-acoustical-pole and residue model and its application to spatial interpolation and extrapolation of a room transfer function. *IEEE Trans. Speech, Audio Process (TransASSP)* 7, 709–717. doi:10.1109/89.799696

Hoskins, R. F. (2009). *Delta functions: Introduction to generalised functions*. Horwood Publishing.

Karakonstantis, X., Caviedes-Nozal, D., Richard, A., and Fernandez-Grande, E. (2024). Room impulse response reconstruction with physics-informed deep learning. *J. Acoust. Soc. Amer. (JASA)* 155 (2), 1048–1059. doi:10.1121/10.0024750

Katzberg, F., Mazur, R., Maass, M., Böhme, M., and Mertins, A. (2018). "Spatial interpolation of room impulse responses using compressed sensing," in Proc. 2018 Int. Workshop Acoustic Signal Enhancement (IWAENC '18), Tokyo, Japan, September 17–20, 2018, 426–430. doi:10.1109/IWAENC.2018.8521390

Kearney, G., Masterson, C., Adams, S., and Boland, F. (2009). "Dynamic time warping for acoustic response interpolation: Possibilities and limitations," in Proc. 17th European Signal Process. Conf. (EUSIPCO '09), Glasgow, Scotland, 24-28 August 2009, 705–709.

Kuhl, S., Nagel, S., Kabzinski, T., Antweiler, C., and Jax, P. (2018). "Tracking of time-variant linear systems: Influence of group delay for different excitation signals," in Proc. 2018 Int. Workshop Acoustic Signal Enhancement (IWAENC '18), Tokyo, Japan, September 17–20, 2018, 131–135. doi:10.1109/IWAENC.2018.8521372

Lin, Y., and Lee, D. (2006). Bayesian regularization and nonnegative deconvolution for room impulse response estimation. *IEEE Trans. Signal Process (TransSP)* 54, 839–847. doi:10.1109/tsp.2005.863030

Masterson, C., Kearney, G., and Boland, F. (2009). "Acoustic impulse response interpolation for multichannel systems using dynamic time warping," in Proc. AES 35th Int. Conf. Audio for Games (AES35Conf), London, United Kingdome, February 11-13, 2009.

Mignot, R., Daudet, L., and Ollivier, F. (2013). Room reverberation reconstruction: Interpolation of the early part using compressed sensing. *IEEE Trans. Audio Speech Lang. Process. (TransASLP)* 21, 2301–2312. doi:10.1109/tasl.2013.2273662

Müller, M. (2007). "Dynamic time warping," in *Information retrieval for music and motion*. Berlin, German: Springer, 69–84.

Naylor, P. A., and Gaubitch, N. D. (2010). *Speech dereverberation*. Vol. 2. Berlin, Germany: Springer.

Nophut, M., Preihs, S., and Peissig, J. (2024). Velocity-controlled Kalman filter for an improved echo cancellation with continuously moving microphones. *J. Audio Eng. Soc. (JAES)* 72, 33–43. doi:10.17743/jaes.2022.0116

Pezzoli, M., Perini, D., Bernardini, A., Borra, F., Antonacci, F., and Sarti, A. (2022). Deep prior approach for room impulse response reconstruction. *Sensors* 22, 2710. doi:10.3390/s22072710

Ratnarajah, A., Ananthabhotla, I., Ithapu, V. K., Hoffmann, P. F., Manocha, D., and Calamia, P. T. (2022). "Towards improved room impulse response estimation for speech recognition," in Proc. 2023 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '23), Rhodes Island, Greece, 4-10 June 2023, 1–5. doi:10.1109/ICASSP49357.2023.10094770

Schissler, C., Stirling, P., and Mehra, R. (2017). "Efficient construction of the spatial room impulse response," in 2017 IEEE Virtual Reality (VR), Los Angeles, CA, 18-22 March 2017, 122–130. doi:10.1109/VR.2017.7892239

Simon, D. (2006). *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. Hoboken, NJ: John Wiley & Sons.

Stan, G.-B., Embrechts, J.-J., and Archambeau, D. (2002). Comparison of different impulse response measurement techniques. *J. Audio Eng. Soc. (JAES)* 50, 249–262.

Szöke, I., Skácel, M., Mošner, L., Paliesek, J., and Černocký, J. H. (2018). Building and evaluation of a real room impulse response dataset. *IEEE J. Sel. Topics Signal Process. (JSTSP)* 13, 863–876. doi:10.1109/JSTSP.2019.2917582

Zhao, J., Zheng, X., Ritz, C., and Jang, D. (2022). Interpolating the directional room impulse response for dynamic spatial audio reproduction. *Appl. Sci.* 12, 2061. doi:10.3390/app12042061