Joint Higher-Order Low-Rank Tensor Decompositions for Multichannel Acoustic System Compression

Matthias Blochberger, *Member, IEEE,* Filip Elvander, *Member, IEEE,* Jan Østergaard, *Senior Member, IEEE,* Toon van Waterschoot, *Member, IEEE*

***Abstract*—Multichannel room impulse responses (RIRs) in acoustics are often long and therefore the sample count becomes large if seen as finite impulse response (FIR) filters, making storage and real-time convolution costly. Pole-zero (PZ) models are an established way to represent RIRs compactly as infinite impulse response (IIR) transfer functions. We show that a PZ model implies a low-rank canonical polyadic decomposition (CPD) of tensorized RIRs: the early reflections form a sparse rank-1 sum, while the late reverberation admits a CPD with factors determined by the poles. This motivates higher-order-tensor-based compression of RIRs via the block term decomposition (BTD), a generalization of the CPD. We propose two multichannel compression schemes: (i) a stacked multichannel block term decomposition (SMCBTD) applied to a (D+1)-mode tensor of RIRs, and (ii) a joint multichannel block term decomposition (JMCBTD) that shares subsets of factors across channels discovered by spectral clustering on factor–subspace affinity graphs. Both are optimized by an alternating higher-order orthogonal iteration (HOOI) procedure. On simulated rooms as well as measured datasets, the proposed methods achieve consistently lower reconstruction error at equal compression ratios compared to separate single-channel decompositions and a generalized low-rank approximation of matrices (GLRAM) baseline. The largest gains occur with scalar cores (i.e., CPD), supporting the PZ-to-CPD modelling link. We also find D = 3 to be a practical sweet spot, offering most of the achievable gain at moderate compute/memory cost. The compressed models enable efficient downstream multichannel convolution.**

***Index Terms*—Room impulse response, tensor decomposition, block-term decomposition, canonical polyadic decomposition, spectral clustering**

I. Introduction

In room acoustic signal processing, the room impulse response (RIR) is a fundamental concept that describes the point-to-point acoustic path from a sound source to a microphone within a room [1]. It captures the effects of the room's geometry, materials, and potentially interior furnishings on the sound propagation, including reflections, diffractions, and absorptions. They are of great importance in various applications aiming to measure, evaluate, model or compensate for the acoustic properties of a room, such as in source localization [2], speech enhancement and source separation [3], room geometry estimation [4], or dereverberation [5].

The RIRs are typically modelled as either infinite impulse response (IIR) filters (e.g., [6], [7]) or finite impulse response (FIR) filters, depending on the application and the desired properties of the model. The FIR model is often preferred due to its simplicity, with the drawback of requiring a large number of parameters. For an average-sized room, the RIR can be several thousands of samples, while for large spaces, such as concert halls or churches, a RIR can be several tens to hundreds of thousands of samples long. This can lead to issues with storage and computational complexity, especially when multiple RIRs are considered. This is particularly relevant in applications that aim to reproduce the room in a virtual environment, such as virtual reality or augmented reality, where a potentially large number of RIRs measured within the acoustic space are needed. While IIR models are compact and can represent long RIRs with few parameters, they tend to be more difficult to work with, e.g., due to numerical stability issues. The pole-zero (PZ) model [6] is a common approach to model RIRs as IIR filters. It enables a mathematical representation of the room modes, the resonances of the room, through its poles. The room modes are shared by all RIRs in a room [8]. The method proposed in [9] takes advantage of common poles in a distributed estimation algorithm. Other literature uses the common-pole model for spatial interpolation of RIRs [10] by estimating modal parameters. In this paper, we show that there is a clear connection between the PZ model and low-rank tensor approximations by means of modal factorization of RIRs.

The need for compact representation of RIRs has led to the development of various compression techniques, with recent ones focusing on low-rank decompositions of RIRs. The truncated singular value decomposition (SVD) is a well-known method for low-rank approximation of matrices, which has been applied to matricized RIRs [11]. In [12], a generalized form of the SVD, the generalized low-rank approximation of matrices (GLRAM) [13] is used to jointly compress a set of matricized RIRs. We use the term "matricized" to describe the arrangement of consecutive RIR segments of appropriate length as columns of an appropriately sized matrix; the analogue term we use for higher-order tensors is "tensorized". Higher-order decompositions, such as the Kronecker product decomposition or the canonical polyadic decomposition (CPD) have been applied to individual RIRs arranged in tensor form [14], [15]. All these methods focus on compact representations

This research work was carried out at the ESAT Laboratory of KU Leuven, and has received funding from KU Leuven internal funds (C14/21/075), Research Foundation Flanders (FWO projects S005525N and G0A0424N), the Flemish Government (AI Research Program), and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369. The scientific responsibility is assumed by the authors.

Matthias Blochberger and Toon van Waterschoot are with the Department of Electrical Engineering (ESAT), STADIUS, KU Leuven, Leuven, Belgium (e-mail: matthias.blochberger@kuleuven.be)

Filip Elvander is with the Dept. of Information and Communications Engineering, Aalto University, Espoo, Finland

Jan Østergaard is with the Department of Electronic Systems, Aalborg University, Aalborg, Denmark

that allow efficient downstream uses of the compressed RIRs, mainly described in [14], [12] where the decompositions yield fast convolution algorithms. Other compression approaches for multichannel acoustic systems include [16] where neural audio codecs are applied. Although this approach is effective in terms of compression, the particular application does not consider joint compression of multiple RIRs nor does the compact representation yield more efficient downstream applications. While an effective approach, it does not consider the same requirements and is not considered in our comparisons.

We extend these low-rank compression approaches to a joint multichannel setting by introducing tensor-decomposition-based models for sets of RIRs. Our framework is based on the block term decomposition (BTD), which provides a convenient umbrella model for several standard low-rank tensor decompositions. In particular, the Tucker decomposition (TD), also known as low multilinear rank approximation (LMLRA), is the natural extension of matrix low-rank approximations such as GLRAM [12] to arbitrary D-mode tensors. The CPD can be viewed as a restricted TD model with a superdiagonal core whose dimensions are all equal to the CPD rank [17]. The block term decomposition (BTD) further generalizes these models by representing a tensor as a sum of block terms, where each block is itself a TD model with its own core and factor matrices. This hierarchy allows us to treat CPD, TD, and intermediate BTD models within a common framework. It is particularly useful in the present context because the PZ-based analysis suggests a low-rank tensor structure for tensorized RIRs, but does not determine a priori whether scalar-core, single-core, or multi-block structures are most parameter-efficient in practice. By comparing CPD, TD, and BTD models experimentally, we assess which member of this decomposition family best matches measured and simulated RIRs. As shown in Section V, scalar-core BTD, i.e., CPD, provides the most parameter-efficient representation in many of the considered scenarios, supporting the proposed connection between the PZ model and low-rank CPD structure.

The main contribution of this paper is the introduction of two joint compression methods. First, we propose multichannel compression by applying BTD to a $(D+1)$-mode tensor, i.e., a stacked tensor of D-mode tensorized RIRs, which we refer to as stacked multichannel block term decomposition (SMCBTD). Then we introduce a coupled BTD, named joint multichannel block term decomposition (JMCBTD), for which we formulate the problem as a joint multichannel tensor decomposition problem, where subsets of decomposition factors are shared across RIRs. The set of shared factors across subsets of channels is determined by a spectral clustering approach. The factors of individual RIRs are first computed using a standard BTD and then used to form graphs of decomposition factors within each mode and rank. The graphs have affinity edges between factors that are similar, i.e., have a small subspace distance. Spectral clustering then forms clusters, i.e. subgraphs, of factors that can be considered similar within a mode and rank. This differs from the coupled canonical polyadic decomposition (C-CPD) approach analyzed in [18], [19] which considers all rank factors of a mode to be shared across the decompositions of multiple tensors, while we consider an arbitrary subset of factors across modes and ranks to be shared, determined by spectral clustering. The optimization algorithm for BTD is based on alternating higher-order orthogonal iterations (HOOIs) [17], [20], [21], [22].

The paper is structured as follows: In Section II, we introduce the basic concepts of tensorization of RIRs and the BTD. In Section III, the connection between the PZ model of RIRs and the low-rank tensor representation is established. In Section IV, we propose the joint multichannel compression methods JMCBTD and SMCBTD. In Section V, we evaluate the proposed methods on simulated and measured data, showing that they outperform the state-of-the-art. In Section VI, we elaborate on a few points of discussion and future work, and we conclude in Section VII.

Notation: We use lowercase letters for scalar values (x), lowercase bold letters for vectors ($\mathbf{x} \in \mathbb{R}^I$), uppercase bold letters for matrices ($\mathbf{X} \in \mathbb{R}^{I_1 \times I_2}$), and calligraphic bold letters for D-mode tensors ($\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_D}$). Element-wise (multi-)indexing uses square brackets, e.g. $\mathbf{x}[i]$, $\mathbf{X}[i_1, i_2]$, and $\boldsymbol{\mathcal{X}}[i_1, ..., i_D]$ for the i-th, the (i_1, i_2)-th, and the $(i_1, ..., i_D)$-th element of the vector, matrix, and tensor, respectively. We use $\circ$ for the outer product of vectors, with multiple denoted as

$$\mathbf{x}_1 \circ \mathbf{x}_2 \circ \cdots \circ \mathbf{x}_D \equiv \bigcirc_{d=1}^{D} \mathbf{x}_d = \boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_D}$$

when $\mathbf{x}_d \in \mathbb{R}^{I_d}$ for $d \in [D] = \{1, 2, \ldots, D\}$. Mode-$d$ unfolding $\mathbf{X}_{(d)} \in \mathbb{R}^{I_d \times \prod_{\delta \neq d} I_\delta}$ of $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_D}$ stacks the mode-d fibres as columns. The d-mode product of a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_D}$ with a matrix $\mathbf{A} \in \mathbb{R}^{J \times I_d}$ is a contraction along mode d such that

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} \times_d \mathbf{A} \in \mathbb{R}^{I_1 \times \cdots \times I_{d-1} \times J \times I_{d+1} \times \cdots \times I_D}$$
$$\iff \mathbf{Y}_{(d)} = \mathbf{A}\mathbf{X}_{(d)} \in \mathbb{R}^{J \times \prod_{\delta \neq d} I_\delta}.$$

We denote the multiple-mode product on a tensor with $D \geq E$ modes as

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} \times_1 \mathbf{A}_1 \cdots \times_E \mathbf{A}_E \equiv \boldsymbol{\mathcal{X}} \times_{d \in [E]} \mathbf{A}_d.$$

We denote the multiple-mode product with all modes except d as

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} \times_{\delta \in [D] \setminus \{d\}} \mathbf{A}_\delta \equiv \boldsymbol{\mathcal{X}} \times_{\delta \neq d} \mathbf{A}_\delta.$$

More details on these tensor specific definitions can be found in [17]. Further we will make use of the following short-hand: For $\mathbf{A} \in \mathbb{R}^{I \times J}$ and $1 \leq K \leq \min\{I, J\}$, define $\mathrm{LSV}_K(\mathbf{A}) \in \mathbb{R}^{I \times K}$ to be the matrix whose columns are K leading left singular vectors of $\mathbf{A}$.

II. Tensorized Acoustic Room Impulse Responses and Tensor Decompositions

Let L be the number of samples of an RIR $h_m(n)$ that are considered significant above the system noise floor, and assume this length to be the same for all m, since the RIRs share the same reverberation [1].

Tensor decompositions have been applied to RIRs in prior work: Considering matrices as 2-mode tensors, [11] uses the truncated SVD to compress multichannel RIRs arranged as a matrix. [12] uses GLRAM to jointly compress multiple matricized RIRs. Higher-order tensor decompositions have been applied to individual RIRs in [23], [14] where they are arranged in D-mode tensors and subsequently CPD is applied, as well as [15] in which the authors use the Kronecker decomposition, i.e., CPD, assuming that there are some decomposition factors that stay constant across channels

while others vary. The last example is similar in spirit to our proposed JMCBTD.

A. Tensorization

Choose positive integers $I_1, I_2, \ldots, I_D$ such that

$$L = \prod_{d=1}^{D} I_d. \tag{1}$$

Define the tensorization bijection between $n = 0, 1, \ldots, L-1$ and the multi-index $(i_1, i_2, \ldots, i_D)$ by

$$n = i_1 + i_2 I_1 + i_3 (I_1 I_2) + \cdots + i_D (I_1 I_2 \cdots I_{D-1}), \tag{2}$$

where $0 \le i_d \le I_d - 1$ for each d. Then define the D-mode tensorized RIR $\boldsymbol{\mathcal{H}}_m \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_D}$ by

$$\begin{aligned} &\boldsymbol{\mathcal{H}}_m[i_1, i_2, \ldots, i_D] \\ &\quad = h_m (i_1 + i_2 I_1 + i_3 I_1 I_2 + \cdots + i_D I_1 I_2 \cdots I_{D-1}). \end{aligned} \tag{3}$$

B. Block Term Decomposition

Let $\boldsymbol{\mathcal{H}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_D}$ denote a tensorized RIR. A BTD with R blocks approximates $\boldsymbol{\mathcal{H}}$ as [17], [20], [21], [22]

$$\boldsymbol{\mathcal{H}} \approx \widehat{\boldsymbol{\mathcal{H}}} = \sum_{r=1}^{R} \boldsymbol{\mathcal{G}}^{(r)} \times_{d \in [D]} \mathbf{A}^{(r,d)}, \tag{4}$$

where $\boldsymbol{\mathcal{G}}^{(r)} \in \mathbb{R}^{K_1^{(r)} \times \cdots \times K_D^{(r)}}$ is the core tensor of block r, and $\mathbf{A}^{(r,d)} \in \mathbb{R}^{I_d \times K_d^{(r)}}$ is its factor matrix in mode d. We consider the orthogonally constrained case with $K_d^{(r)} \le I_d$. For compact notation, we define the set of block-mode indices as

$$\mathcal{I} \triangleq \{1, \ldots, R\} \times \{1, \ldots, D\}. \tag{5}$$

The corresponding approximation problem is

$$\begin{aligned} \min_{\boldsymbol{\mathcal{G}}^{(r)}, \mathbf{A}^{(r,d)}} \quad & \left\| \boldsymbol{\mathcal{H}} - \sum_{r=1}^{R} \boldsymbol{\mathcal{G}}^{(r)} \times_{d \in [D]} \mathbf{A}^{(r,d)} \right\|_F^2 \\ \text{s.t.} \quad & \mathbf{A}^{(r,d)\top} \mathbf{A}^{(r,d)} = \mathbf{I} \quad \text{for all } (r,d) \in \mathcal{I}, \end{aligned} \tag{6}$$

The BTD generalizes several standard tensor decompositions. If each core is scalar, $\boldsymbol{\mathcal{G}}^{(r)} = g^{(r)} \in \mathbb{R}$, then (4) reduces to the rank-R CPD,

$$\widehat{\boldsymbol{\mathcal{H}}} = \sum_{r=1}^{R} g^{(r)} \mathbf{a}^{(r,1)} \circ \cdots \circ \mathbf{a}^{(r,D)}. \tag{7}$$

If $R = 1$, then (4) reduces to the Tucker decomposition,

$$\widehat{\boldsymbol{\mathcal{H}}} = \boldsymbol{\mathcal{G}} \times_{d \in [D]} \mathbf{A}^{(d)}. \tag{8}$$

The factors and cores in (4) can be computed using standard alternating HOOI-type algorithms for BTD; we refer the reader to [17], [20], [21], [22] for details. In Section IV, we adapt this principle to the proposed multichannel decompositions with shared factors.

III. Pole-Zero Model and Low-Rank Tensor Representation

To establish the connection between PZ models of RIRs and low-rank tensor representations, consider the causal rational model

$$y(n) = H(q)x(n) = \frac{B(q)}{A(q)} x(n), \tag{9}$$

where $H(q)$ represents the RIR $h(n)$, $B(q)$ and $A(q)$ are numerator and denominator polynomials of degrees P_B and P_A, respectively, and $q^{-l}x(n) = x(n-l)$. In the z-domain,

$$H(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z^{-1} + \cdots + b_{P_B} z^{-P_B}}{1 + a_1 z^{-1} + \cdots + a_{P_A} z^{-P_A}}, \tag{10}$$

with real coefficients a_k, b_k. We assume, for simplicity, that the poles z_r, $r = 1, \ldots, P_A$, are simple. Since $h(n) \in \mathbb{R}$, complex poles occur in conjugate pairs.

A. Modal Factorization of Impulse Responses

The decomposition used below is the discrete-time counterpart of the standard modal decomposition of finite-dimensional linear systems [24, Sec. 2.5.2]. In continuous time, modes associated with system eigenvalues λ_r appear as exponentials $e^{\lambda_r t}$. For the discrete-time rational transfer function $H(z)$, the analogous modal terms are powers z_r^n, where z_r are the poles of $H(z)$. Equivalently, this follows from the pole-residue expansion of the rational transfer function.

After separating the finite polynomial part of the transfer function, the impulse response can be decomposed as

$$h(n) = h_F(n) + h_R(n), \tag{11}$$

where $h_F(n)$ is a finite-length component and $h_R(n)$ is the recursive modal tail. For the improper case $P_B > P_A$, the finite component has length $P_F + 1$, with $P_F = P_B - P_A$, and can be written as

$$h_F(n) = \sum_{k=0}^{P_F} f_k \delta(n-k). \tag{12}$$

For a proper transfer function, this polynomial part is absent; nevertheless, in the acoustic interpretation below, we separately model the finite early-reflection component $h_F(n)$ as a sparse sum of delayed impulses.

Assuming simple poles z_r, $r = 1, \ldots, P_A$, the recursive tail is a sum of modal terms,

$$h_R(n) = \begin{cases} 0, & 0 \le n \le P_F, \\ \displaystyle\sum_{r=1}^{P_A} \beta^{(r)} z_r^{n-P_F-1}, & n \ge P_F + 1, \end{cases} \tag{13}$$

where $\beta^{(r)} \in \mathbb{C}$ are the pole residues. Since $h(n) \in \mathbb{R}$, non-real poles and residues occur in complex-conjugate pairs. In the following, we use (12)–(13) as the starting point for showing that the tensorized finite component and modal tail admit low-rank tensor representations.

B. Tensor Representation

Due to the infinite length of the tail, we further look at a truncated version of the impulse response with length L. Let us consider the head $h_F(n)$ for $n = 0, 1, \ldots, P_F$ which in practical terms consists of the early reflections of the RIR. This part of the RIR for the typical small-to-medium sized room is often considered on the order of 30 to 80 milliseconds, which corresponds to a few thousand samples depending on the sampling rate (e.g., at 48 kHz, this is 1500 to 4000 samples). Because this part of the RIR consists of a handful $N \ll P_F$ echoes,

$$h_F(n) = \sum_{k=1}^{N} \alpha_k \delta(n - \tau_k),$$

where α_k are the amplitudes of the echoes and $\tau_k \in \mathbb{N}$ are their delays. Let there be the multi-index $\tau_k = (i_1^{(k)}, i_2^{(k)}, \ldots, i_D^{(k)})$ defined in (2). We define the standard basis vectors

$$\mathbf{e}^{(k,d)} = [0, \ldots, 0, 1, 0, \ldots, 0]^\top \in \mathbb{R}^{I_d} \tag{14}$$

with 1 at the $i_d^{(k)}$-th position. Then the impulses $\delta(n-\tau_k)$ can be represented as rank-1 outer products with which we can write the tensorization of the head $h_F(n)$ as

$$\begin{aligned} \boldsymbol{\mathcal{H}}_F &= \sum_{k=1}^{N} \alpha_k \big(\mathbf{e}^{(k,1)} \circ \mathbf{e}^{(k,2)} \circ \cdots \circ \mathbf{e}^{(k,D)}\big) \\ &= \sum_{k=1}^{N} \alpha_k \times_{d\in[D]} \mathbf{e}^{(k,d)} \end{aligned} \tag{15}$$

with $\operatorname{rank}(\boldsymbol{\mathcal{H}}_F) \le N$.[1] Note that here the cores of the BTD are scalar with $K_d = 1$, i.e., a CPD.

Then, since for the tail $h_R(n)$ in (13),

$$h_R(n) = \sum_{r=1}^{P_A} \beta^{(r)} z_r^{n-P_F-1}, \quad n \ge P_F + 1, \tag{16}$$

one finds that using the tensorization bijection (2) for any $(i_1, \ldots, i_D)$ we have

$$\begin{aligned} \boldsymbol{\mathcal{H}}_R[i_1, \ldots, i_D] &= \sum_{r=1}^{P_A} \beta^{(r)} z_r^{(i_1 + i_2 I_1 + \cdots + i_D I_1 I_2 \cdots I_{D-1} - P_F - 1)} \\ &= \sum_{r=1}^{P_A} \widetilde{\beta}^{(r)} \left(z_r^{1}\right)^{i_1} \left(z_r^{I_1}\right)^{i_2} \left(z_r^{I_1 I_2}\right)^{i_3} \cdots \left(z_r^{I_1 I_2 \cdots I_{D-1}}\right)^{i_D}, \end{aligned} \tag{17}$$

where $\widetilde{\beta}^{(r)} = \beta^{(r)} z_r^{-P_F-1}$. So equivalently, we can write the tensor $\boldsymbol{\mathcal{H}}_R$ as the sum of the outer products

$$\begin{aligned} \boldsymbol{\mathcal{H}}_R &= \sum_{r=1}^{P_A} \widetilde{\beta}^{(r)} \left(\mathbf{a}^{(r,1)} \circ \mathbf{a}^{(r,2)} \circ \cdots \circ \mathbf{a}^{(r,D)}\right) \\ &= \sum_{r=1}^{P_A} \widetilde{\beta}^{(r)} \times_{d\in[D]} \mathbf{a}^{(r,d)} \end{aligned} \tag{18}$$

[1]If early-reflection delays are fractional, the delayed impulses can be represented by truncated interpolation kernels, for example windowed sinc functions. Under the tensorization map, each non-zero sample of such a kernel corresponds to one standard-basis rank-one term. Therefore, if the k-th truncated kernel has T_k non-zero samples, the bound becomes $\operatorname{rank}(\boldsymbol{\mathcal{H}}_F) \le \min\{\sum_{k=1}^{N} T_k, P_F + 1\}$.

which represents a complex-valued rank-P_A CPD, with the dth mode-factor

$$\mathbf{a}^{(r,d)}[i_d] = z_r^{s_d i_d}, \quad i_d = 0, 1, \ldots, I_d - 1 \tag{19}$$

for each $r = 1, \ldots, P_A$ and $d = 1, \ldots, D$, where $s_d = \prod_{\delta=1}^{d-1} I_\delta$ is the stride for mode d. A complex-conjugate pole pair $z_r = \rho_r e^{j\omega_r}$ and $z_r^* = \rho_r e^{-j\omega_r}$, with conjugate residues $\beta^{(r)}$ and $\beta^{(r)*}$, can be written in real-valued form by expanding the complex modal factors into damped cosine and sine components. Define

$$\mathbf{c}_d[i_d] = \rho_r^{s_d i_d} \cos(\omega_r s_d i_d), \quad \mathbf{s}_d[i_d] = \rho_r^{s_d i_d} \sin(\omega_r s_d i_d). \tag{20}$$

Since

$$\mathbf{a}^{(r,d)} = \mathbf{c}_d + j\mathbf{s}_d,$$

the outer product of the complex modal factors expands as

$$\bigcirc_{d=1}^{D} (\mathbf{c}_d + j\mathbf{s}_d) = \sum_{\mathcal{S}\subseteq[D]} j^{|\mathcal{S}|} \bigcirc_{d=1}^{D} \mathbf{u}_d^{(\mathcal{S})}, \tag{21}$$

where

$$\mathbf{u}_d^{(\mathcal{S})} = \begin{cases} \mathbf{s}_d, & d \in \mathcal{S}, \\ \mathbf{c}_d, & d \notin \mathcal{S}. \end{cases} \tag{22}$$

Adding the conjugate pole contribution gives the real-valued tensor

$$\boldsymbol{\mathcal{H}}_R^{(r)} = \sum_{\mathcal{S}\subseteq[D]} 2\rho_r^{-P_F-1} \operatorname{Re}\big[\beta^{(r)} j^{|\mathcal{S}|}\big] \bigcirc_{d=1}^{D} \mathbf{u}_d^{(\mathcal{S})}. \tag{23}$$

Thus, each conjugate pole pair contributes at most 2^D real-valued CPD rank-one terms. Equivalently, by grouping $\mathbf{c}_d$ and $\mathbf{s}_d$ into $\mathbf{U}_d = [\mathbf{c}_d\ \mathbf{s}_d] \in \mathbb{R}^{I_d \times 2}$, the same contribution can be represented as one real-valued BTD block with multilinear ranks $K_d = 2$, $d = 1, \ldots, D$, whose $2 \times \cdots \times 2$ core contains the coefficients in (23). If some poles are real-valued, their contribution is a single real-valued rank-1 CPD term each, which we omit in our representation without loss of generality.

The tensorized RIR can therefore be written as the sum of the head and tail

$$\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{H}}_F + \boldsymbol{\mathcal{H}}_R \tag{24}$$

with $\operatorname{rank}(\boldsymbol{\mathcal{H}}) \le N + 2^{D-1} P_A$ in real-valued CPD form and as a real-valued BTD consisting of N scalar-core blocks for the finite component and $P_A/2$ rank-$(2, \ldots, 2)$ blocks for the conjugate pole pairs.

C. Factor Subspace Similarities

Let $m = 1, \ldots, M$ index RIRs measured at different positions in the same room, and assume that each tensorized RIR admits the representation $\boldsymbol{\mathcal{H}}_m = \boldsymbol{\mathcal{H}}_{mF} + \boldsymbol{\mathcal{H}}_{mR}$ in (24). For each mode d, define the factor subspace

$$\mathcal{S}_m^{(d)} = \operatorname{span}\left\{\mathbf{e}_m^{(1,d)}, \ldots, \mathbf{e}_m^{(N,d)}, \mathbf{a}_m^{(1,d)}, \ldots, \mathbf{a}_m^{(P_A,d)}\right\}. \tag{25}$$

Here, the vectors $\mathbf{e}_m^{(k,d)}$ are associated with the finite early-reflection component, while the vectors $\mathbf{a}_m^{(r,d)}$ are associated with the modal tail.

Under the common-pole model for RIRs measured in the same room [8], the pole locations z_r are shared across positions. Therefore, in the ideal modal-tail model, the factors $\mathbf{a}_m^{(r,d)}$ in (19) are shared across m, up to scaling and ordering ambiguities. The early-reflection factors, in contrast, are generally position-dependent. Consequently, the subspaces $\mathcal{S}_m^{(d)}$ are not expected to be identical across positions, but may exhibit similarities whenever common modal components, similar reflection delays, or other position-dependent acoustic regularities lead to aligned factor subspaces.

We quantify these similarities by comparing subspaces across positions. Let $d_S(\mathcal{S}_i^{(d)}, \mathcal{S}_j^{(d)})$ denote a standard subspace distance, such as the chordal or Grassmannian distance [25]. Figure 1 shows the resulting distances for RIRs simulated on a grid in a shoebox room. The observed patterns indicate that the factor subspaces contain non-trivial spatial structure. While part of this structure is consistent with the common-pole interpretation of the late reverberant tail, some patterns may also reflect early-reflection or delay-related structure induced by the room geometry and the tensorization. A detailed separation of these effects is beyond the scope of this work, but the observation motivates the data-driven sharing strategy used in the proposed JMCBTD: rather than imposing a fixed sharing pattern a priori, we infer shared factors from factor-subspace affinity graphs.

IV. Joint Multichannel RIR Compression using BTD

Based on the theoretical background of the shared factors subspaces, we can now propose a joint compression framework for multichannel RIRs using the BTD. We first introduce the SMCBTD and then the JMCBTD.

A. Stacked Multichannel BTD (SMCBTD)

We stack the M tensorized RIRs $\boldsymbol{\mathcal{H}}_m$ as defined in (3) into a $(D{+}1)$-mode tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_D \times M}$ such that

$$\boldsymbol{\mathcal{T}}[i_1, i_2, \ldots, i_D, m] = \boldsymbol{\mathcal{H}}_m[i_1, i_2, \ldots, i_D]. \tag{26}$$

Then we can apply BTD (6) to the tensor $\boldsymbol{\mathcal{T}}$ such that

$$\boldsymbol{\mathcal{T}} \approx \widehat{\boldsymbol{\mathcal{T}}} = \sum_{r=1}^{R} \boldsymbol{\mathcal{G}}^{(r)} \times_{d \in [D+1]} \mathbf{A}^{(r,d)} \tag{27}$$

with the analogous definitions as in Section II-B for a $(D{+}1)$-mode tensor. From Section III-C, we can expect that the added channel mode captures inter-RIR redundancy, while the remaining modes capture the intra-RIR low rank. The compression ratio for the SMCBTD is defined as

$$C_{\text{SMCBTD}} = 1 - \frac{\sum_{r=1}^{R} N_r}{ML} \tag{28}$$

with

$$N_r = \prod_{d=1}^{D+1} K_d^{(r)} + \sum_{d=1}^{D} I_d K_d^{(r)} + M K_{D+1}^{(r)}. \tag{29}$$

With high channel count M, large RIR length L, and large number of blocks R, the SMCBTD becomes computationally expensive and (temporarily) memory intensive. To alleviate this, we can consider a per-RIR decomposition with shared factors across the RIRs as described in the next section.

B. Joint Multichannel BTD (JMCBTD)

In addition to the SMCBTD, we can formulate a joint compression framework for multichannel RIRs using the JMCBTD with spectral clustering for improved factor sharing and compression efficiency. The similarity of the factor subspaces across RIRs implies that a reasonable approximation could be made by shared factors across RIRs, and thus reducing the number of parameters. For this, for each $(r,d) \in \mathcal{I}$, we need to define clusters of RIRs that share a factor matrix. Let such a cluster be denoted by

$$\mathcal{C}_c^{(r,d)} \subseteq \{1, \ldots, M\} \tag{30}$$

where $c = 1, \ldots, Q^{(r,d)}$ is the cluster index and $Q^{(r,d)}$ the number of clusters for the specific (r,d). The clusters are disjoint, i.e., $\mathcal{C}_c^{(r,d)} \cap \mathcal{C}_{c'}^{(r,d)} = \emptyset$ for $c \neq c'$ and all (r,d). Define the factor matrices as

$$\mathbf{A}_m^{(r,d)} = \begin{cases} \mathbf{U}_c^{(r,d)} & \text{if } m \in \mathcal{C}_c^{(r,d)}, \\ \mathbf{U}_m^{(r,d)} & \text{otherwise,} \end{cases} \tag{31}$$

where $\mathbf{U}_c^{(r,d)}$ denote the factors shared for cluster $\mathcal{C}_c^{(r,d)}$ and $\mathbf{U}_m^{(r,d)}$ the individual factors for each m. With this, we define the optimization problem for the joint multichannel RIR compression as

$$\begin{aligned} \min \quad & \sum_{m=1}^{M} \Big\| \boldsymbol{\mathcal{H}}_m - \sum_{r=1}^{R} \boldsymbol{\mathcal{G}}_m^{(r)} \times_{d \in [D]} \mathbf{A}_m^{(r,d)} \Big\|_F^2 \\ \text{s.t.} \quad & \mathbf{A}_m^{(r,d)\top} \mathbf{A}_m^{(r,d)} = \mathbf{I} \quad m = 1, \ldots, M,\ (r,d) \in \mathcal{I} \end{aligned} \tag{32}$$

As in the single-channel problem, the orthogonality constraint allows the application of an HOOI approach, again, limiting the multilinear core dimensions (multilinear rank) to $K_d^{(r)} \leq I_d$ for all d and r. The shared and individual factors are treated separately. While the update step for individual factors is unchanged, the shared factor update is performed using the SVD on stacked aggregated residuals from within the cluster such that

$$\mathbf{U}_c^{(r,d)} = \text{LSV}_{K_d^{(r)}}\big(\bar{\mathbf{Y}}_{c(d)}^{(r)}\big), \tag{33}$$

where

$$\bar{\mathbf{Y}}_{c(d)}^{(r)} = \big[\mathbf{Y}_{m(d)}^{(r)}\big]_{m \in \mathcal{C}_c^{(r,d)}} \tag{34}$$

is the horizontal concatenation of the residual unfoldings for all RIRs m in the cluster $\mathcal{C}_c^{(r,d)}$. The core tensors are updated as before. Algorithm 1 summarizes the overall procedure. Since the choice of the clusters is non-trivial, we now introduce a two-step approach using spectral clustering to determine data-driven sharing clusters in the next section.

C. Spectral Clustering on Factor Subspace Affinity Graphs

The choice of channel clusters $\mathcal{C}_c^{(r,d)}$ is crucial for the performance of the proposed method. A fixed sharing pattern could be imposed manually, for example by sharing selected modes across all channels. Instead, we use a data-driven heuristic based on spectral clustering of factor-subspace affinity graphs. We define the factor subspace affinity graph for a (r,d) as

$$\mathcal{G}^{(r,d)} \triangleq \{\mathcal{V}^{(r,d)}, \mathcal{E}^{(r,d)}, w^{(r,d)}\} \tag{35}$$

Algorithm 1 Alternating HOOI for JMCBTD

Input: Tensors $\boldsymbol{\mathcal{H}}_m \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_D}$, number of blocks R, block ranks $(K_1^{(r)}, K_2^{(r)}, \ldots, K_D^{(r)})$, clusters $\mathcal{C}_c^{(r,d)}$, outer iteration count N_o, inner iteration count N_i
Output: Core tensors $\boldsymbol{\mathcal{G}}_m^{(r)} \in \mathbb{R}^{K_1^{(r)} \times K_2^{(r)} \times \ldots \times K_D^{(r)}}$, factor matrices $\mathbf{U}_m^{(r,d)} \in \mathbb{R}^{I_d \times K_d^{(r)}}$, $\mathbf{U}_c^{(r,d)} \in \mathbb{R}^{I_d \times K_d^{(r)}}$
Initialize: Random $\mathbf{U}_m^{(r,d)}$ for each r, d, m, random $\mathbf{U}_c^{(r,d)}$ for each r, d, c, random $\boldsymbol{\mathcal{G}}_m^{(r)}$ for each r, m
for N_o **do**
  for $r = 1, \ldots, R$ **do**
    for $m = 1, \ldots, M$ **do**
      $\boldsymbol{\mathcal{R}}_m^{(r)} = \boldsymbol{\mathcal{H}}_m - \sum_{l \neq r} \widehat{\boldsymbol{\mathcal{H}}}_m^{(l)}$
    end for
    for N_i **do**
      for $d = 1, \ldots, D$ **do**
        for $m = 1, \ldots, M$ **do**
          $\mathbf{Y}_{m(d)}^{(r)} \leftarrow \boldsymbol{\mathcal{R}}_m^{(r)} \times_{\delta \neq d} \mathbf{A}_m^{(r,\delta)\top}$
          if $m \notin \bigcup_{c=1}^{Q^{(r,d)}} \mathcal{C}_c^{(r,d)}$ **then**
            $\mathbf{U}_m^{(r,d)} \leftarrow \mathrm{LSV}_{K_d^{(r)}}\big(\mathbf{Y}_{m(d)}^{(r)}\big)$
          end if
        end for
        for $c = 1, \ldots, Q^{(r,d)}$ **do**
          $\bar{\mathbf{Y}}_{c(d)}^{(r)} = \big[\mathbf{Y}_{m(d)}^{(r)}\big]_{m \in \mathcal{C}_c^{(r,d)}}$
          $\mathbf{U}_c^{(r,d)} \leftarrow \mathrm{LSV}_{K_d^{(r)}}\big(\bar{\mathbf{Y}}_{c(d)}^{(r)}\big)$
        end for
      end for
    end for
    for $m = 1, \ldots, M$ **do**
      $\boldsymbol{\mathcal{G}}_m^{(r)} \leftarrow \boldsymbol{\mathcal{H}}_m \times_{d \in [D]} \mathbf{A}_m^{(r,d)\top}$
    end for
  end for
end for

where $\mathcal{V}^{(r,d)} = \{\mathbf{U}_1^{(r,d)}, \ldots, \mathbf{U}_M^{(r,d)}\}$ is the set of factors for all RIRs, $\mathcal{E}^{(r,d)} = \{(i,j) \mid w^{(r,d)}(i,j) > \tau\}$ is the set of edges, and $w^{(r,d)}$ is the weight function. The threshold $\tau \geq 0$ prunes the fully connected graph to a significant subgraph. The weight function $w^{(r,d)}(i,j)$ is defined as an affinity value based on the subspace distance by

$$w^{(r,d)}(i,j) = \exp\Big(-\frac{d_S(\mathbf{U}_i^{(r,d)}, \mathbf{U}_j^{(r,d)})}{\sigma^2}\Big), \tag{36}$$

where $d_S(\cdot,\cdot)$ is the Chordal distance and σ is a scaling factor. Note that we treat the factor matrices $\mathbf{U}_m^{(r,d)}$ as bases of the subspaces due to the orthogonality constraint.

To find the clusters, we employ spectral clustering on the graphs $\mathcal{G}^{(r,d)}$. The affinity matrix $\mathbf{W}^{(r,d)} \in \mathbb{R}^{M \times M}$ is constructed from the weight function $w^{(r,d)}(i,j)$ as follows:

$$\mathbf{W}^{(r,d)}[i,j] = \begin{cases} w^{(r,d)}(i,j) & \text{if } (i,j) \in \mathcal{E}^{(r,d)} \\ 0 & \text{otherwise} \end{cases} \tag{37}$$

We then truncate all affinity values below the threshold τ and remove the isolated nodes from the graph, i.e., rows and columns of $\mathbf{W}^{(r,d)}$ where the row and column-wise sum of elements is 0. From the truncated affinity matrix, denoted

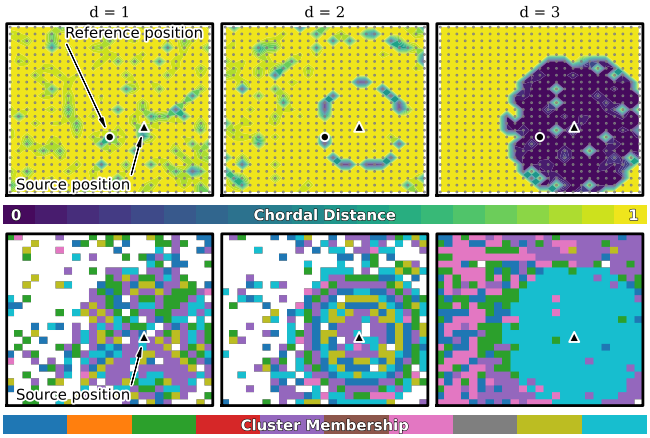

Fig. 1. **Top row:** Top views of a shoebox room with the Chordal distance between tensorized RIRs at a reference position and a grid of positions over the room. The distances are computed from the factor subspaces in mode $d = 1, 2, 3$, respectively. Darker colors indicate higher similarity (lower distance). **Bottom row:** Top views of the same room with spectral clustering applied. The clusters are indicated by different colors and show the spatial arrangement of the microphone positions within each cluster.

$\tilde{\mathbf{W}}^{(r,d)} \in \mathbb{R}^{\tilde{M} \times \tilde{M}}$ with $\tilde{M} \leq M$, we compute the normalized Laplacian matrix

$$\mathbf{L}^{(r,d)} = \mathbf{I} - \big(\mathbf{D}^{(r,d)}\big)^{-1/2} \tilde{\mathbf{W}}^{(r,d)} \big(\mathbf{D}^{(r,d)}\big)^{-1/2} \tag{38}$$

where the diagonal matrix $\mathbf{D}^{(r,d)}$ is the degree matrix defined by $\mathbf{D}^{(r,d)}[i,i] = \sum_j \tilde{\mathbf{W}}^{(r,d)}[i,j]$. We then compute the eigenpairs $(\lambda_n^{(r,d)}, \mathbf{v}_n^{(r,d)})$ of the Laplacian matrix $\mathbf{L}^{(r,d)}$ using eigenvalue decomposition (EVD), then select the $Q^{(r,d)}$ most significant eigenvalues and their corresponding eigenvectors through the maximum eigengap of the sorted sequence $\lambda_1^{(r,d)} \leq \lambda_2^{(r,d)} \leq \cdots \leq \lambda_{\tilde{M}}^{(r,d)}$ such that

$$Q^{(r,d)} = \min\{\arg\max_{1 \leq i < \tilde{M}} \big(\lambda_{i+1}^{(r,d)} - \lambda_i^{(r,d)}\big), Q_\mathrm{max}\}, \tag{39}$$

where Q_max is a safeguard maximum number of clusters. After row-normalization of the matrix formed by the eigenvectors

$$\tilde{\mathbf{V}}^{(r,d)} = \begin{bmatrix} \mathbf{v}_1^{(r,d)} & \cdots & \mathbf{v}_{Q^{(r,d)}}^{(r,d)} \end{bmatrix}, \tag{40}$$

we apply k-means clustering on the rows to obtain the final clusters $\mathcal{C}_c^{(r,d)}$ for $c = 1, \ldots, Q^{(r,d)}$. This partitions the M RIRs into $Q^{(r,d)}$ clusters, i.e., subgraphs of $\mathcal{G}^{(r,d)}$, for each (r,d) based on their similarity in the factor subspaces, and $M - \tilde{M}$ individual RIRs that do not belong to any cluster. While there is still some freedom in choosing the parameters τ, σ, and C_max, this approach is a more systematic data-driven method than manual choice of clusters. These clusters can then be used to initiate the joint JMCBTD compression algorithm. The overall procedure is summarized in Algorithm 2. The compression ratio for the JMCBTD is defined as

$$C_\mathrm{JMCBTD} = 1 - \frac{N_\mathrm{cores} + \sum_{(r,d) \in \mathcal{I}} \Big(N_\mathrm{shr}^{(r,d)} + N_\mathrm{ind}^{(r,d)}\Big)}{ML} \tag{41}$$

Algorithm 2 JMCBTD using spectral clustering

1: **for** $m = 1, \dots, M$ **do**
2: Compute individual BTDs for $\boldsymbol{\mathcal{H}}_m$
3: **end for**
4: **for** $(r, d) \in \mathcal{I}$ **do**
5: Construct graph $\mathcal{G}^{(r,d)}$
6: Apply spectral clustering to obtain $\mathcal{C}_c^{(r,d)}$
7: **end for**
8: Compute joint BTD (Algorithm 1) using clusters $\mathcal{C}_c^{(r,d)}$

where

$$N_{\text{cores}} = M \sum_{r=1}^{R} \prod_{d=1}^{D} K_d^{(r)}, \tag{42}$$

$$N_{\text{shr}}^{(r,d)} = Q^{(r,d)} I_d K_d^{(r)}, \text{and} \tag{43}$$

$$N_{\text{ind}}^{(r,d)} = \sum_{m \notin \bigcup_{c=1}^{Q^{(r,d)}} \mathcal{C}_c^{(r,d)}} I_d K_d^{(r)} \tag{44}$$

are the number of coefficients for the cores, shared factors, and individual factors, respectively.

Note that since the order of blocks r is arbitrary in the BTD, and therefore block contributions (r, d) across different channels m might not represent the same characteristics, it is advantageous to reorder them such that the subspace distances are minimized before clustering. This can be done for example using the Hungarian (Kuhn–Munkres) algorithm [26], [27] or the Auction algorithm [28]. We apply the latter using the Chordal distance as the cost metric, such that similar blocks are grouped together for better clustering and compression performance.

D. Computational Complexity and Memory Usage

Define $J_d^{(r)} = \prod_{\delta \neq d} K_\delta^{(r)}$. We use W for "work" to denote cost here, not to confuse with the compression ratio. One outer iteration with N_i inner iterations has cost

$$\sum_{r=1}^{R} \Big[W_{\text{rec}}^{(r)} + N_{\text{i}} \sum_{d=1}^{D} \big(W_Y^{(r,d)} + W_{\text{SVD}}^{(r,d)} \big) + W_{\text{core}}^{(r)} \Big], \tag{45}$$

where we use standard multi-mode product [17] and truncated-SVD costs [29],

$$W_{\text{rec}}^{(r)} = \sum_{d=1}^{D} I_d K_d^{(r)} \big(\prod_{\delta<d} I_\delta \big) \big(\prod_{\delta>d} K_\delta^{(r)} \big), \tag{46}$$

$$\begin{aligned} W_Y^{(r,d)} = &\sum_{\delta=1}^{d-1} I_\delta K_\delta^{(r)} \big(\prod_{\mu=1}^{\delta-1} K_\mu^{(r)} \big) \big(\prod_{\mu=\delta+1}^{d-1} I_\mu \big) I_d \big(\prod_{\mu=d+1}^{D} I_\mu \big) \\ &+ \sum_{\delta=d+1}^{D} I_\delta K_\delta^{(r)} \big(\prod_{\mu=1}^{d-1} K_\mu^{(r)} \big) I_d \big(\prod_{\mu=d+1}^{\delta-1} I_\mu \big) \big(\prod_{\mu=\delta+1}^{D} K_\mu^{(r)} \big), \end{aligned} \tag{47}$$

$$W_{\text{SVD}}^{(r,d)} = I_d J_d^{(r)} K_d^{(r)}, \tag{48}$$

$$W_{\text{core}}^{(r)} = \sum_{d=1}^{D} I_d K_d^{(r)} \big(\prod_{\delta<d} K_\delta^{(r)} \big) \big(\prod_{\delta>d} I_\delta \big). \tag{49}$$

While these expressions for the exact costs seem complex, their intuition is rather straightforward: for the reconstruction cost $W_{\text{rec}}^{(r)}$, the sequential multi-mode products over the sequence of modes d, i.e., a matrix vector product for each fiber, where the number of fibers change after each mode from I_δ to $K_\delta^{(r)}$, and the same applies for the $W_{\text{core}}^{(r)}$ cost in reverse. The cost $W_Y^{(r,d)}$ is a similar multi-mode product but skipping mode d, which separates the expression into two sums. Peak working memory is $L + \max_{r,d} I_d J_d^{(r)}$.

Stacking M RIRs for SMCBTD forms a $(D+1)$-mode tensor with a channel mode of size M and applies BTD (Section II-B). The channel-mode update performs a truncated SVD of size $M \times \prod_{d=1}^{D} K_d^{(r)}$, and all contractions traverse the full ML entries per each sweep. JMCBTD instead updates shared factors on per-cluster stacks (34), where the dominant SVD in mode d has size $I_d \times \sum_{m \in \mathcal{C}_c^{(r,d)}} J_d^{(r)}$, replacing M by the cluster size $|\mathcal{C}_c^{(r,d)}|$. Thus, the sweeps scale with $\max_c |\mathcal{C}_c^{(r,d)}|$. The peak memory shrinks accordingly. When $\max_c |\mathcal{C}_c^{(r,d)}| \ll M$ and L is large, JMCBTD reduces both computational cost and memory usage and the one-off cost of M individual BTDs and spectral clustering becomes negligible. This was confirmed by numerical experiments in Section V, where JMCBTD achieved faster computation times than SMCBTD, and at larger M and L the latter became infeasible due to memory constraints. By caching residuals, one can trade memory for computational effort (time), which becomes relevant for large datasets and when memory is the limiting factor.

V. Numerical Simulations

We conduct numerical experiments to evaluate the performance of the proposed method. For this, we first use simulated RIRs in shoebox rooms to show spatial clustering effects and subsequently use measured RIRs from the single- and multichannel audio recordings database (SMARD) [30] and shoebox room impulse response archive with varying absorption (SRIRACHA) [31] and church impulse response dataset (ChurchIR) dataset [32] to validate the applicability to real-world scenarios. Let the measure for the reconstruction error be the normalized misalignment (NM) defined as

$$\text{NM} = \frac{\|\boldsymbol{\mathcal{H}}_m - \widehat{\boldsymbol{\mathcal{H}}}_m\|_F}{\|\boldsymbol{\mathcal{H}}_m\|_F} = \frac{\|\mathbf{h}_m - \hat{\mathbf{h}}_m\|_2}{\|\mathbf{h}_m\|_2}, \tag{50}$$

where $\boldsymbol{\mathcal{H}}_m$ is the original tensor, $\widehat{\boldsymbol{\mathcal{H}}}_m$ is the reconstructed tensor, $\mathbf{h}_m$ is the original vectorized RIR and $\hat{\mathbf{h}}_m$ is the reconstructed vectorized RIR. Let the mean NM be defined as

$$\overline{\text{NM}} = \frac{1}{M} \sum_{m=1}^{M} \text{NM}, \tag{51}$$

and let

$$\overline{\text{NM}}_{\text{dB}} = 20 \log_{10}(\overline{\text{NM}}) \quad [\text{dB}] \tag{52}$$

be the NM in decibel scale.

For any set of RIRs, be it simulated or measured, we first estimate the reverberation time T_{60} and then truncate the RIRs to $L \geq L_{T_{60}} = T_{60} f_s$, where we define this as

Parameter	Simulated Array		SMARD 2X0X	SRIRACHA C1	ChurchIR
	Far	Near			
Room parameters					
Size $[\ell, w, h]$	8.00 m 8.00 m 3.00 m	8.00 m 8.00 m 3.00 m	2.87 m 7.34 m 8.09 m	6.22 m 3.85 m 3.07 m	51.08 m 24.89 m 23 m
T_{60}	0.53 s	0.53 s	0.15 s	0.5 s	$T_{30} = 7$ s (trunc.)
Critical distance d_c	1.12 m	1.12 m	1.92 m	0.69 m	2.81 m
Mixing time t_{mix}	14 ms	14 ms	13 ms	9 ms	171 ms
Source pos. $[\ell, w, h]$	4.0 m 4.0 m 1.2 m	1.0 m 1.0 m 1.2 m	see [30]	see [31]	see [32]
M	64	64	168	256	64
f_s	48 kHz	48 kHz	48 kHz	36 kHz	36 kHz
Decomposition parameters					
D	$3, 4, 5$	$3, 4, 5$	$3, 4, 5$	$3, 4, 5$	$3, 4, 5$
K_d	$1, 2, 4$	$1, 2, 4$	$1, 2, 4$	$1, 2, 4$	$1, 2$
σ	$0.2 \dots 1.2$	$0.2 \dots 1.2$	$0.6 \dots 1.8$	$0.4 \dots 6.0$	$0.4 \dots 0.6$
τ	0.4	0.4	0.4	0.4	0.4
Q_{max}	6	6	6	6	6
N_{o}	80	80	80	80	80
N_{i}	8	8	8	8	8

TABLE I

ROOM AND DECOMPOSITION PARAMETERS FOR THE NUMERICAL EXPERIMENTS. FOR THE MEASURED DATASETS, SOME PARAMETERS ARE REFERENCED TO THE ORIGINAL PUBLICATIONS. THE CRITICAL DISTANCE IS APPROXIMATED BY $d_c = 0.057\sqrt{V/T_{60}} = 1.12$ M AND THE MIXING TIME IN MS IS APPROXIMATED BY $t_{\text{MIX}} \approx \sqrt{V}$ [33], WHERE V IS THE ROOM VOLUME.

the smallest number of samples such that we have the most equidimensional tensor shape. For this, let $u = \lfloor \sqrt[D]{L_{T_{60}}} \rfloor$ be the largest integer such that $u^D \leq L_{T_{60}}$. Then choose the smallest $v \in \{1, ..., D\}$ such that $u^{D-v}(u+1)^v \geq L_{T_{60}}$. If necessary, we zero-pad the RIRs to length L. So then, we set the tensor dimensions as

$$I_d = \begin{cases} u+1 & \text{if } d \leq v, \\ u & \text{otherwise} \end{cases} \tag{53}$$

for $d = 1, ..., D$. We aim for an equidimensional tensor shape, due to the better compression performance as discussed in [14].

The methods in our comparison are the proposed SMCBTD and JMCBTD as well as the GLRAM method [12] as a baseline. The GLRAM method is essentially a 2-mode Tucker decomposition, i.e., BTD with $R = 1$. In this comparison, we consider the left matrices of the GLRAM as common and do a refinement step with individual right matrices. The compression ratio for the GLRAM method is varied by changing the dimensions of the core matrix.

A. Simulated Room Impulse Responses

We simulate a shoebox room using the randomized image source method (RISM) [34], placing a compact $M = 8 \times 8 = 64$ square grid microphone array with a grid spacing of 0.1 m centered in the room of dimensions 8.00 m$\times 8.00$ m$\times 3.00$ m. The absorption coefficients of the walls are set to achieve an average reverberation time of $T_{60} \approx 0.5$ s. Two source positions are considered: a close position within the array's near field and a far position in the array's far field, as illustrated in Figure 2. The room dimensions and reverberation time are listed in Table I. The RIRs are simulated at a sampling frequency of $f_s = 48$ kHz and truncated as described above. We compare the proposed methods JMCBTD and SMCBTD with the GLRAM-based method in [12]. We vary the tensorization order D and the compression rate to evaluate the performance. The decomposition parameters used are listed in Table I.

Figure 2 shows how the reconstruction error varies spatially across the array for a specific compression rate $C = 0.8$ and tensorization order $D = 4$. We can observe that in this scenario, while all compared methods perform similarly for the near-field position at a minimum reconstruction error at the closest microphone, the advantage of the proposed JMCBTD method is significant in the far-field scenario. We can observe the error increasing with distance from the center of the array for all methods, but the JMCBTD method maintains a lower error rate compared to the others with -16 dB mean NM compared to -13 dB for GLRAM and -12 dB for SMCBTD at the far-field position. The SMCBTD shows similar performance to the GLRAM-based method in [12] in this scenario. Further, Figure 2 shows example RIRs and their reconstructions for the far-field position for a compression rate of $C = 0.8$ comparing the proposed JMCBTD and SMCBTD methods with the GLRAM-based method with tensorization order $D = 4$.

Figure 3 shows the mean reconstruction error versus compression ratio for the simulated compact microphone array at the near-field position with tensorization order $D = 3$. We compare the proposed joint JMCBTD with individual BTD for core ranks $K_d \in \{1, 4\}$ and clustering parameters $\sigma \in \{0.4, 0.5, 0.6\}$. We can observe that by clustering decomposition factors and sharing them across multiple channels, yields lower reconstruction error at the same compression rate.

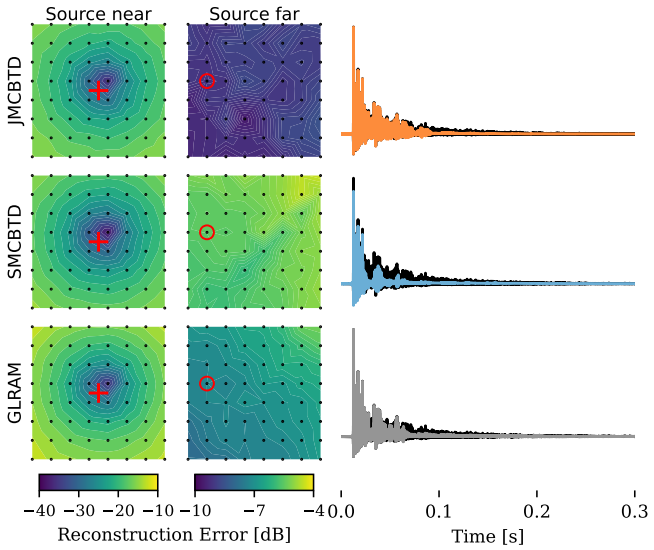

Fig. 2. This figure illustrates the reconstruction error NM over the compact microphone array geometry (dots indicate microphones) for the two source positions. For the near source position (left), it is marked +. We observe the reconstruction error for the proposed JMCBTD method is significantly lower across the array at the far-field position. Additionally, example RIRs (black) and their reconstructions comparing JMCBTD and SMCBTD with tensorization order $D = 4$, and GLRAM, all at compression rate $C = 0.8$ for a selected microphone position (circled in red ○) are shown for the far-field position.

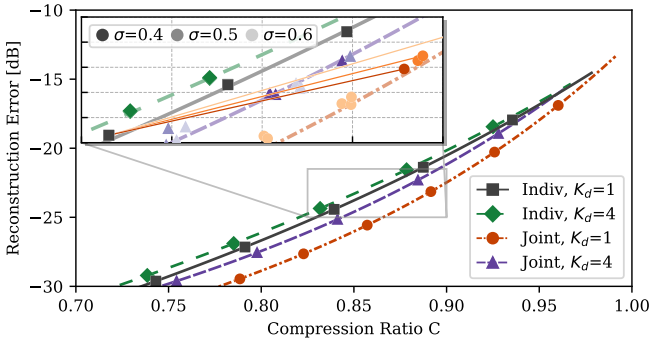

Fig. 3. This figure shows the mean reconstruction error NM versus compression rate comparing individual BTD and the proposed JMCBTD for different core ranks $K_d \in \{1, 4\}$ and clustering parameters $\sigma \in \{0.4, 0.5, 0.6\}$. The data is from the simulated compact microphone array at the near-field position, and tensorization order $D = 3$. It shows that by clustering decomposition factors and sharing them across multiple channels, yields lower reconstruction error at the same compression rate. We can see that the performance gain is more pronounced for lower core ranks, especially $K_d = 1$. Curves are cubic spline fits on the data for readability. In the inset figure, we indicate the effect of varying σ by drawing connecting lines between data points with the same block number R compressed individually and jointly.

We can see that the performance gain is more pronounced for lower core ranks, especially $K_d = 1$.

Figure 4 and Figure 5 show the mean reconstruction error versus compression rate for different tensorization orders $D \in \{3, 4, 5\}$ for a simulated compact microphone array at a near- and far-field scenario. Additionally, the core rank is varied as $K_d \in \{1, 2, 4\}$ for all modes d. We can observe that at constant compression rate, increased core rank leads to worse performance, with this effect being more pronounced for higher tensorization orders. This simulation provides numerical evidence for the low-rank CPD model of tensorized RIRs that was introduced in Section III.

B. Measured Room Impulse Responses

The second set of experiments is conducted using measured RIRs from the SMARD, which contains RIRs measured in a small room using compact microphone arrays and SRIRACHA, which contains RIRs measured in a shoebox room with densely placed source positions over a grid covering large part of the room. Further, we use ChurchIR, to evaluate the performance in a highly reverberant environment where RIRs exhibit a very dense modal structure [1]. We use the parameters listed in Table I for the proposed algorithm.

1) SMARD: The SMARD dataset [30] contains RIRs measured for different setups of compact microphone arrays and loudspeakers in a small room. For this, we select the subset of configurations using the directional loudspeaker and group the RIRs. The grouping is done such that each group contains RIRs measured using the same loudspeaker position but different microphone array setups. By the SMARD naming convention, we denote these grouped configurations as 2X0X grouping all configurations where X is any digit $\{0, 1, 2\}$. This results in $M = 168$ channels. All RIRs are measured at a sampling frequency of $f_s = 48$ kHz and truncated as described at the beginning of Section V. The average reverberation time is $T_{60} \approx 0.15$ s.

Figure 6 shows the mean reconstruction error versus compression ratios for different tensorization orders $D \in \{3, 4, 5\}$. The proposed method's performance gain is larger in this dataset compared to the simulated scenario across all tensorization orders, especially at $K_d = 1$ with a reconstruction error gain of around 8 to 10 dB over the GLRAM-based method. Again, we can observe that at constant compression rate, increased core rank leads to decreased performance, with this effect being more pronounced for higher tensorization orders. It is worth noting that the JMCBTD method's compression rate is computed after spectral clustering which leads to a smaller range of compression rates visualized in the plots.

2) SRIRACHA: The SRIRACHA dataset [31] contains a large set of measured RIRs in a shoebox room with densely placed source and receiver positions. We select two microphone positions, one at the center of the source position grid and one on the edge of the grid, and use an equally spaced subset of $M = 256$ source positions for each microphone position. All RIRs are provided at a sampling frequency of $f_s = 36$ kHz and are already truncated to the average $T_{60} \approx 0.5$ s at 16,000 samples. The truncation process described at the beginning of Section V therefore may pad the RIRs with zeros to reach a more equidimensional tensor shape.

Figure 7 shows the mean reconstruction error versus compression ratios for different tensorization orders $D \in \{3, 4, 5\}$. The proposed JMCBTD and SMCBTD methods outperform the GLRAM-based method across all tensorization orders, especially at core rank $K_d = 1$ with a reconstruction error gain of around 5 to 8 dB. The effect of increased core rank leading to decreased performance at constant compression rate is not as pronounced in this dataset compared to the previous

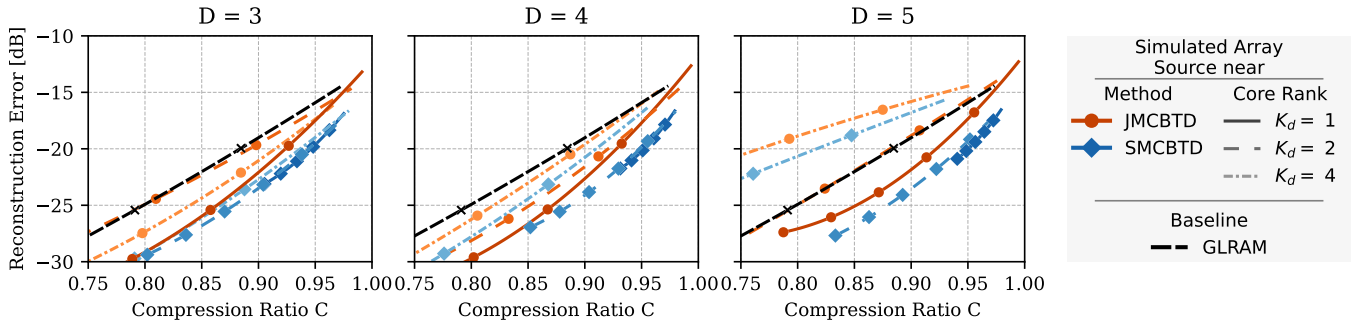

Fig. 4. Compression rate vs. mean reconstruction error for different tensorization orders D for a simulated compact microphone array at a near-field source position, see Figure 2. We compare the proposed JMCBTD and SMCBTD with the GLRAM-based method in [12].

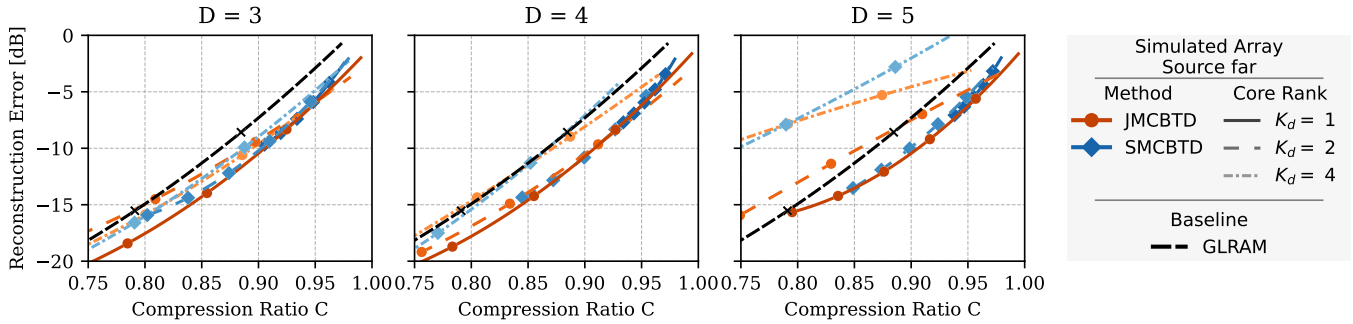

Fig. 5. Compression rate vs. mean reconstruction error for different tensorization orders D for a simulated compact microphone array at a far-field source position, see Figure 2. We compare the proposed JMCBTD and SMCBTD with the GLRAM-based method in [12].

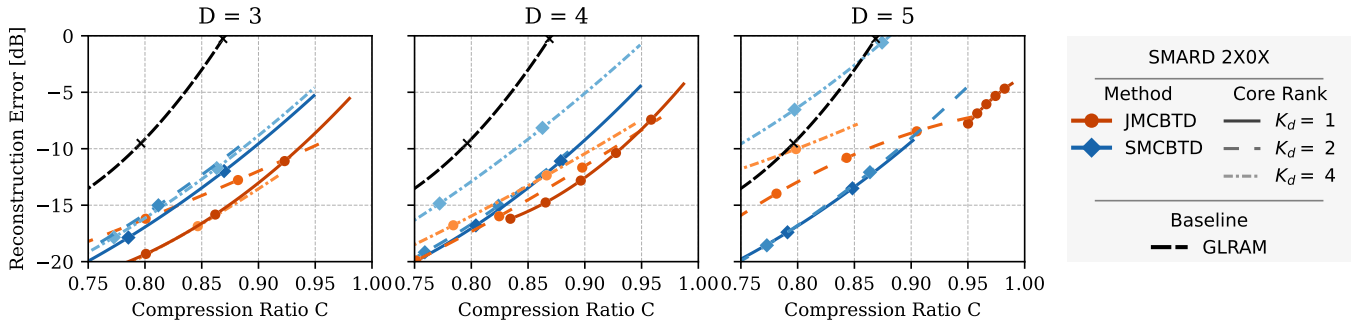

Fig. 6. Compression rate vs. mean reconstruction error for different tensorization orders D on the set of impulse responses in SMARD [30] using all RIRs in the 2X0X configuration. The performance gain of the proposed methods is substantial especially at $K_d = 1$, across all tensorization orders $D = 3, 4, 5$.

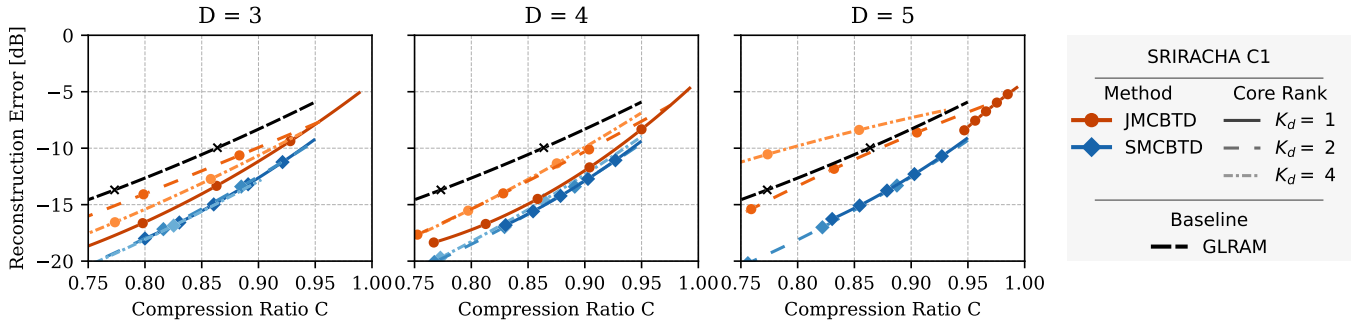

Fig. 7. Compression rate vs. mean reconstruction error for different tensorization orders D on the SRIRACHA C1 dataset, spatially downsampled to $M = 256$ positions. The performance gain of the proposed JMCBTD and SMCBTD methods is substantial especially at $K_d = 1$, across all tensorization orders $D = 3, 4, 5$. Here SMCBTD performs slightly better than JMCBTD.

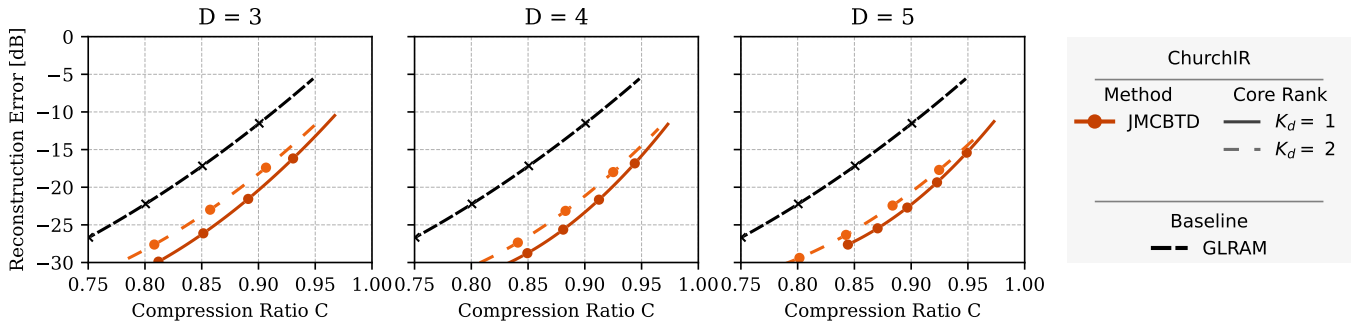

Fig. 8. Compression rate vs. mean reconstruction error for different tensorization orders D on a selection of $M = 64$ RIRs from the ChurchIR dataset. The performance gain of the proposed JMCBTD method is substantial especially at $K_d = 1$, across all tensorization orders $D = 3, 4, 5$. Here, the advantage of the JMCBTD method over SMCBTD is substantial as the temporary memory requirements of SMCBTD make it infeasible to be applied to such long RIRs.

ones. Also, the clustering effect in the shared factors found by the proposed JMCBTD method is less pronounced compared to the previous datasets. A potential reason for this is that this room, while being a shoebox, exhibits a denser modal structure than the room used in SMARD. It is worth noting that the JMCBTD method's compression rate is computed after spectral clustering which leads to a smaller range of compression rates visualized in the plots.

3) ChurchIR: The ChurchIR dataset contains RIRs measured in a church using higher-order spherical microphone arrays. We select two higher order arrays (Eigenmike) of 32 channels each and jointly compress this set of $M = 64$ RIRs. Specifically, we select source position "SC" and receiver positions "E1" and "E2" as described in [32]. The room being a church, the reverberation time is very high with an average of $RT_{30} = 4.12$ s according to [32]. The authors of the dataset provide truncated RIRs to 7.0 s. All RIRs are provided at a sampling frequency of $f_s = 48$ kHz, which we downsample to $f_s = 36$ kHz. Figure 8 shows the mean reconstruction error versus compression ratios for different tensorization orders $D \in \{3, 4, 5\}$. The proposed JMCBTD method outperforms the GLRAM-based method across all tensorization orders, again, especially at core rank $K_d = 1$ with a reconstruction error gain of around 10 dB. The proposed SMCBTD method becomes infeasible to apply in this scenario due to the temporary memory requirements growing beyond available working memory, see Section IV-D. The effectiveness of the methods in this highly reverberant scenario is more pronounced than in the previous datasets. So despite the denser modal structure of the RIRs in this dataset, its relationship to the much larger reverberation time and therefore higher sample counts, still allows for a low-rank tensor representation after tensorization. Further, here, increasing the number of modes D seems to have a more pronounced effect on the performance compared to the previous datasets, where a gain of around 2 dB can be observed when going from $D = 3$ to $D = 4$ at core rank $K_d = 1$.

C. Tail Reconstruction

To complement the standard reconstruction error, which is dominated by the high-energy direct sound and early reflections, we additionally evaluate the late reverberant part of the impulse response using a log-envelope error measure. Let n_{tail} denote the starting sample of the tail, corresponding to the mixing time $t_{\text{mix}} \approx \sqrt{V}$ in ms [33], used as a reasonable approximation of the transition from early reflections to late reverberation. For an impulse response $h[n]$, we first compute a smoothed local-energy envelope

$$e_h[n] = (h^2[n]) * w[n] \tag{54}$$

where $w[n]$ is a short smoothing window. The corresponding tail-start normalized envelope is

$$\tilde{E}_h[n] = 10 \log_{10}(e_h[n]) - E_h[n_{\text{tail}}], \quad \forall n. \tag{55}$$

This is computed analogously for the reconstructed RIR $\hat{h}[n]$, and we then define the tail envelope error (TEE) as

$$\text{TEE} = \frac{1}{L - n_{\text{tail}}} \sum_{n=n_{\text{tail}}}^{L-1} |\tilde{E}_h[n] - \tilde{E}_{\hat{h}}[n]|. \tag{56}$$

This metric is expressed in dB and specifically quantifies how well the local decay profile of the late reverberant tail is preserved, which corresponds better to diffuse late reverberation than sample-wise misalignment.

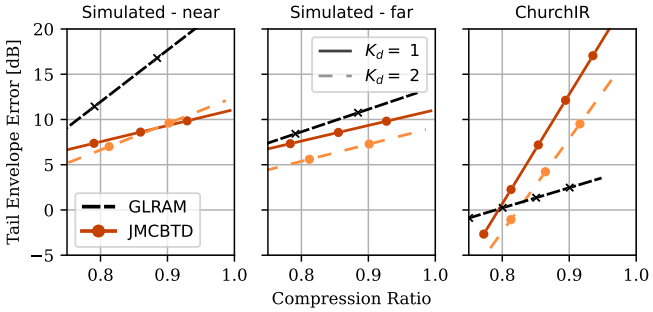

Fig. 9. Compression rate vs. TEE on the simulated and ChurchIR datasets representing short and long RIRs, taking into account the tail from the mixing time at 14 ms and 171 ms, respectively. Lower values represent better tail reconstruction.

Figure 9 shows the TEE for the simulated dataset and ChurchIR plotted over compression ratio. We use a rectangular smoothing window of 10 ms length and use the mixing times as listed in Table I. We limit the investigation to two datasets and a comparison between the baseline and JMCBTD with core sizes $K_d \in 1, 2$ for the simulated data and ChurchIR

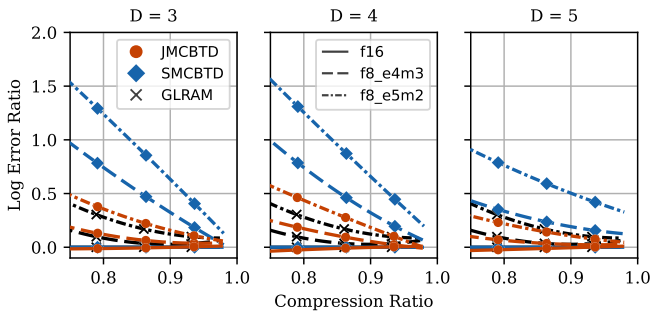

Fig. 10. The LER (57) for tensorization orders $D \in \{3, 4, 5\}$ over compression ratio C for the algorithms JMCBTD, SMCBTD and the baseline GLRAM.

to represent short and long RIRs, respectively. Compared to the reconstruction error, where scalar cores show the best performance, here the trend is that larger cores enable better tail reconstruction. Further, for ChurchIR, the baseline GLRAM performs better at higher compression ratios, while the proposed method does so for lower compression ratios.

D. Floating-Point Precision

We perform a simple check to test the decomposition's sensitivity to decreasing floating-point precision. For this, we convert the tensor decompositions for JMCBTD, SMCBTD, and GLRAM from the original 32-bit floating-point representation into 16- and 8-bit formats and compute the reconstruction error (50) w.r.t the original RIRs. We compute the log error ratio (LER)

$$\mathrm{LER}_{\{16,8\}} = \log_2 \frac{\overline{\mathrm{NM}}_{\{16,8\}}}{\overline{\mathrm{NM}}_{32}} \tag{57}$$

for the following PyTorch [35] floating-point types: `float16`, `float8_e4m3` (4-bit exponent; 3-bit mantissa), and `float8_e5m2` (5-bit exponent; 2-bit mantissa), with the `float32` type as the baseline. This ratio, computed for decompositions at various compression ratios, shows how large the error induced by the decreased floating point precision is compared to the original high-precision decomposition. The logarithm symmetrizes the values w.r.t. the unit ratio. A LER of 0 indicates no reconstruction loss due to reduced floating-point precision, while a value of 1 indicates the reconstruction error induced by precision and decomposition are equally large. In Figure 10, we show the LERs for the proposed and baseline methods for the simulated RIR scenario. We can observe the LER decreases with higher compression ratios. This implies that the error induced by the decreased precision becomes less significant at higher compression ratios compared to the error induced by the compressed representation itself. Generally, 16-bit precision does not introduce additional error for all three methods, while for 8-bit precision, for all methods, `float8_e4m3` shows larger LER than `float8_e5m2`. SMCBTD is affected more significantly than JMCBTD and the baseline GLRAM.

VI. Discussion

One important aspect of the proposed methods is that the compression not just reduces the number of parameters needed to represent the RIRs, but also allows for efficient downstream applications such as convolution with anechoic signals. The efficiency of such convolution operations is described in detail in [12] for the CPD and GLRAM methods. Since BTD generalizes CPD, similar computational savings can be achieved.

The floating-point precision experiment suggests that the proposed decompositions can be stored in 16-bit precision with negligible additional reconstruction error, making reduced-precision storage a promising complement to tensor-based compression. More aggressive 8-bit quantization appears more sensitive to the chosen format and is left for future optimization.

The numerical experiments in Section V show that the proposed JMCBTD method outperforms existing methods for joint compression of multichannel RIRs across a range of compression rates and tensorization orders. While the performance gain is significant from $D = 2$ to $D = 3$ in all experiments, the gain from $D > 3$ is marginal to none in most scenarios. This indicates that a tensorization order of $D = 3$ is sufficient to capture the multilinear structure of RIRs for compression purposes. Similar improvements were observed for the SMCBTD method. The exception is the ChurchIR dataset, where increasing the tensorization order from $D = 3$ to $D = 4$ shows a more pronounced performance gain. This leads to the hypothesis that the method's effectiveness increases with increasing RIR lengths.

The numerical results underline the theoretical correspondence between PZ model and low-rank tensor representations, and point to the CPD as the most efficient representation in terms of reconstruction error. Nonetheless, the more general BTD shows better tail reconstruction in some cases. This, combined with the potential to further investigate multilinear low-rank structures, i.e., cores that are larger than scalar, for data-driven rank estimation instead of rank being a fixed hyperparameter keep the BTD a valuable representation.

In implementation, there is a trade-off between computation time and temporary memory requirement to make: intermediate residuals can either be kept in memory or recomputed on-demand. This number grows with both increasing tensorization order D and number of blocks R. In all experiments, the range of compression ratios covered shrinks with increasing tensorization order D due to this effect as the feasibility of the computations is limited by available memory (and time). Here, the difference between JMCBTD and SMCBTD is that the former requires smaller intermediate residuals due to the smaller working sets in each block update. This makes the JMCBTD method more scalable for higher channel counts M, RIR lengths L, and number of blocks R. The fact of increasing computation time and temporary memory requirement with increasing tensorization order D, combined with the marginal performance gain in terms of reconstruction error observed in the experiments, suggests that $D = 3$ is a good choice in practice.

Future extensions of the proposed methods are a single step clustering approach for the JMCBTD avoiding the two-step approach, and data-driven detection of the multilinear rank for both JMCBTD and SMCBTD, which was considered a fixed hyperparameter for this work. Although scalar cores

yield the best compression performance in our experiments, they may still provide valuable insight into multichannel RIR tensor modelling. Further, while the clustering approach used in the JMCBTD method shows good performance gains, it is currently not possible to predict the clustering parameters for a certain target compression rate and reconstruction error, therefore further research is needed to address this challenge.

VII. Conclusion

In this paper, we first introduced BTD for the compression of acoustic RIRs, generalizing previous work on CPD decomposition and GLRAM-based methods. We showed how the PZ model leads to a low-rank tensor when reshaping a single RIR into a higher-order tensor, which confirms empirical findings of existing work, bridging a gap in the literature. We then proposed two methods for joint compression of multichannel RIRs using the BTD. The first method SMCBTD compresses a stacked tensor of RIRs, while the second JMCBTD method considers shared factors across RIRs. We proposed a two-step approach using spectral clustering to find optimal clusters of RIRs that share factors. Numerical experiments using simulated and measured RIRs showed improved performance of the proposed methods over the baseline method.

References

[1] H. Kuttruff, *Room Acoustics*, 5th ed. Boca Raton, Fla.: CRC Press [u.a.], 2009.

[2] C. Evers *et al.*, "The LOCATA Challenge: Acoustic Source Localization and Tracking," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 28, pp. 1620–1643, 2020.

[3] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A Consolidated Perspective on Multimicrophone Speech Enhancement and Source Separation," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 4, pp. 692–730, Apr. 2017.

[4] K. MacWilliam, F. Elvander, and T. van Waterschoot, "Simultaneous Acoustic Echo Sorting and 3-D Room Geometry Inference," in *Proc. 2023 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '23)*. Rhodes Island, Greece: IEEE, 2023, pp. 1–5.

[5] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*, ser. Signals and Communication Technology. Springer Science & Business Media, 2010.

[6] J. Mourjopoulos and M. Paraskevas, "Pole and zero modeling of room transfer functions," *J. Sound Vib.*, vol. 146, no. 2, pp. 281–302, Apr. 1991.

[7] G. Vairetti, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. van Waterschoot, "A Scalable Algorithm for Physically Motivated and Sparse Approximation of Room Impulse Responses With Orthonormal Basis Functions," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 7, pp. 1547–1561, Jul. 2017.

[8] Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 2, pp. 320–328, Apr. 1994.

[9] T. van Waterschoot and M. Moonen, "Distributed estimation and equalization of room acoustics in a wireless acoustic sensor network," in *Proc. 20th European Signal Process. Conf. (EUSIPCO '12)*, Aug. 2012, pp. 2709–2713.

[10] O. Das, P. Calamia, and S. V. Amengual Gari, "Room Impulse Response Interpolation from a Sparse Set of Measurements Using a Modal Architecture," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jun. 2021, pp. 960–964.

[11] G. Huang, J. Benesty, and J. Chen, "Dimensionality Reduction of Room Acoustic Impulse Responses and Applications to System Identification," *IEEE Signal Process. Lett.*, vol. 30, pp. 1107–1111, 2023.

[12] M. Jälmby, F. Elvander, and T. van Waterschoot, "Multi-Channel Low-Rank Convolution of Jointly Compressed Room Impulse Responses," *IEEE Open J. Signal Process.*, vol. 5, pp. 850–857, 2024.

[13] J. Ye, "Generalized Low Rank Approximations of Matrices," in *Proc. Twenty-First Int. Conf. Mach. Learn.*, ser. Icml '04. New York, NY, USA: Association for Computing Machinery, 2004, p. 112.

[14] M. Jälmby, F. Elvander, and T. van Waterschoot, "Compression of room impulse responses for compact storage and fast low-latency convolution," *EURASIP J. Audio, Speech, Music Process.*, vol. 2024, no. 1, p. 45, Sep. 2024.

[15] G. Huang *et al.*, "Acoustic System Identification with Partially Time-Varying Models Based on Tensor Decompositions," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC '22)*, 2022, pp. 1–5.

[16] A. I. Mezza, A. Bernardini, and F. Antonacci, "Large-Scale Room Impulse Response Dataset Compression with Neural Audio Codecs," in *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*, Sep. 2024, pp. 1–8.

[17] T. G. Kolda and B. W. Bader, "Tensor Decompositions and Applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Aug. 2009.

[18] M. Sørensen and L. D. De Lathauwer, "Coupled Canonical Polyadic Decompositions and (Coupled) Decompositions in Multilinear Rank-$(L_{r,n}, L_{r,n}, 1)$ Terms—Part I: Uniqueness," *SIAM J. Matrix Anal. & Appl.*, vol. 36, no. 2, pp. 496–522, Jan. 2015.

[19] M. Sørensen, I. Domanov, and L. De Lathauwer, "Coupled Canonical Polyadic Decompositions and (Coupled) Decompositions in Multilinear Rank-$(L_{r,n}, L_{r,n}, 1)$ Terms—Part II: Algorithms," *SIAM J. Matrix Anal. & Appl.*, vol. 36, no. 3, pp. 1015–1045, Jan. 2015.

[20] L. De Lathauwer, "Decompositions of a Higher-Order Tensor in Block Terms—Part I: Lemmas for Partitioned Matrices," *SIAM J. Matrix Anal. & Appl.*, vol. 30, no. 3, pp. 1022–1032, Jan. 2008.

[21] ——, "Decompositions of a Higher-Order Tensor in Block Terms—Part II: Definitions and Uniqueness," *SIAM J. Matrix Anal. & Appl.*, vol. 30, no. 3, pp. 1033–1066, Jan. 2008.

[22] L. De Lathauwer and D. Nion, "Decompositions of a Higher-Order Tensor in Block Terms—Part III: Alternating Least Squares Algorithms," *SIAM J. Matrix Anal. & Appl.*, vol. 30, no. 3, pp. 1067–1083, Jan. 2008.

[23] M. Jälmby, F. Elvander, and T. van Waterschoot, "Low-Rank Room Impulse Response Estimation," *IEEE/ACM Trans. Audio Speech Lang. Process.*, pp. 1–13, 2023.

[24] T. Kailath, *Linear Systems*. Prentice-hall Englewood Cliffs, NJ, 1980, vol. 156.

[25] K. Ye and L.-H. Lim, "Schubert varieties and distances between subspaces of different dimensions," *SIAM J. Matrix Anal. & Appl.*, vol. 37, no. 3, pp. 1176–1197, 2016.

[26] H. W. Kuhn, "The Hungarian method for the assignment problem," *Nav. Res. Logist.*, vol. 2, no. 1-2, pp. 83–97, 1955.

[27] J. Munkres, "Algorithms for the assignment and transportation problems," *SIAM J. Appl. Math.*, vol. 5, no. 1, pp. 32–38, 1957.

[28] D. P. Bertsekas, "A distributed algorithm for the assignment problem," *Lab. for Information and Decision Systems Working Paper, MIT*, vol. 3, 1979.

[29] G. H. Golub and C. F. Van Loan, *Matrix Computations*, fourth edition ed., ser. Johns Hopkins Studies in the Mathematical Sciences. Baltimore: The Johns Hopkins University Press, 2013.

[30] J. K. Nielsen, J. R. Jensen, S. H. Jensen, and M. G. Christensen, "The single- and multichannel audio recordings database (SMARD)," in *Proc. 2014 Int. Workshop Acoustic Signal Enhancement (IWAENC '14)*, Sep. 2014, pp. 40–44.

[31] A. J. R. Pelling, A. Kujawski, and E. Sarradj, "SRIRACHA: Shoebox Room Impulse Response Archive with Varying Absorption," Jul. 2025.

[32] R. Giampiccolo, S. Parrinelli, and F. Antonacci, "ChurchIR: A Dataset of Multichannel Church Impulse Responses for Spatial Audio Applications," in *Proc. 33rd European Signal Process. Conf. (EUSIPCO '25)*, 2025, pp. 8–12.

[33] G. Defrance and J.-D. Polack, "Measuring the mixing time in auditoria," *J. Acoust. Soc. Am.*, vol. 123, no. 5, p. 3499, 2008.

[34] E. De Sena, N. Antonello, M. Moonen, and T. van Waterschoot, "On the modeling of rectangular geometries in room acoustic simulations," *IEEE Trans. Audio Speech Lang. Process.*, vol. 23, no. 4, pp. 774–786, 2015.

[35] J. Ansel *et al.*, "PyTorch 2: Faster machine learning through dynamic python bytecode transformation and graph compilation," in *29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24)*. ACM, Apr. 2024.