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

Matthias Blochberger[1], Graduate Student Member, IEEE, Filip Elvander [2], Member, IEEE, and Toon van Waterschoot[1], Member, IEEE

[1]Department of Electrical Engineering, ESAT-STADIUS, KU Leuven, 3001 Leuven, Belgium
[2]Department of Information and Communications Engineering, Aalto University, 02150 Espoo, Finland

Corresponding author: Matthias Blochberger (email: matthias.blochberger@esat.kuleuven.be).

ABSTRACT Room acoustic systems, represented by multichannel room impulse responses (RIRs), have been considered in low-rank two-dimensional representations such as the singular value decomposition (SVD) or generalized low-rank approximation of matrices (GLRAM) with the goal of exploiting low-rank structure for compression. These methods are based on empirical and experimental findings which suggest that RIRs exhibit similarities in their low-rank decompositions. In this paper, we show that this approach can be extended to tensor decompositions of arbitrary D-mode tensors. We first show how the pole-zero (PZ) model of RIRs leads to a low-rank tensor representation of the RIRs with similar decomposition factor subspaces, which in turn leads to low-rank structures in multichannel room-acoustic systems. Then, we introduce methods for higher-order multichannel RIR compression based on the block term decomposition (BTD), a generalization of the canonical polyadic decomposition (CPD), Tucker decomposition (TD) and GLRAM. The first, straightforward, approach to multichannel compression is to apply BTD to a stacked $D+1$-mode tensor of D-mode tensorized RIRs. The second, more specialized, approach involves formulating a joint multichannel tensor decomposition problem where subsets of decomposition factors are shared across RIRs, leading to the joint multichannel block term decomposition (JMCBTD). We show in numerical experiments using simulated and measured RIRs that the proposed approaches lead to better reconstruction performance at equal compression ratios compared to individual decompositions and state-of-the-art methods.

INDEX TERMS Enter key words or phrases in alphabetical order, separated by commas. Using the *IEEE Thesaurus* can help you find the best standardized keywords to fit your article. Use the thesaurus access request form for free access to the *IEEE Thesaurus*.

I. INTRODUCTION

IN room acoustic signal processing, the 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 room equalization [?],

sound field reproduction [?], room geometry estimation [?], or dereverberation [?]. The finite impulse response (FIR) model is often preferred due to its simplicity, with the drawback of requiring a large number of parameters. For an averagely 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.

The RIRs are typically modelled as either infinite impulse responses (IIRs) (e.g., [2], [3]) or FIRs, depending on the application and the desired properties of the model. 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 PZ model [2] is a common approach to model RIRs as IIRs filters, because it allows for a mathematical representation of the room modes, which are the resonances of the room, through its poles. The room modes are shared by all RIRs in a room [4]. The method proposed in [5] takes advantage of common poles in a distributed estimation algorithm. In this paper, we show that there is a clear connection between the PZ model and low-rank tensor approximations of Kronecker/CPD by means of modal factorization of RIRs and arranging them in tensor form, closing a gap in literature.

The need for compact representation of finite RIRs has led to the development of various compression techniques, with recent ones focussing on low-rank decompositions of RIRs. The truncated SVD is a well known method for low-rank approximation of matrices, which has been applied to matricized RIRs [6]. In [7], a generalized form of the SVD, the GLRAM 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 CPD have been applied to individual RIRs arranged in tensor form [8], [9]. All these methods focus on compact representations that allow efficient down-stream uses of the compressed RIRs, mainly described in [8], [7] where the decompositions yield fast convolution algorithms. Other compression approaches for multichannel acoustic systems include [10] where neural audio codecs are applied. This approach is very effective in terms of compression, the particular application however, does not consider joint compression nor does the compact representation yield more efficient down-stream applications. While a very effective approach, it does not consider the same requirements and is not considered in our comparisons.

We extend and generalize these concepts by introducing a framework for joint compression of multiple RIRs based on the block term decomposition (BTD), which is a generalization of the CPD. It is further also a generalization of the TD, which is in turn the direct extension of GLRAM [7] from matrices to arbitrary D-mode tensors and is also known as low multilinear rank approximation (LMLRA). Further, the CPD is a special case of the TD where the core tensor's dimensions are all equal to the same rank and the core is superdiagonal [11]. The BTD is a sum of

Notation: We use lowercase letters for scalar values (x), lowercase bold letters for vectors $(\mathbf{x} \in \mathbb{R}^{I_x})$, uppercase bold letters for matrices $(\mathbf{X} \in \mathbb{R}^{I_1 \times I_2})$, and calligraphic bold letters for D-mode tensors $(\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_D})$. Addressing specific elements of a vector, matrix or tensor is denoted by square bracket (multi) indexing, 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 denote the outer product of vectors, matrices and tensors with $\circ$. Unfolding the tensor into a matrix $\mathbf{X}_{(d)} \in \mathbb{R}^{I_d \times \prod_{\delta \neq d} I_\delta}$ along mode d is defined as the matrix whose columns are the mode-d fibres of the tensor $\boldsymbol{\mathcal{X}}$. Further, the d-mode product of a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_D}$ with a matrix $\mathbf{A} \in \mathbb{R}^{K \times I_d}$ is a contraction along mode d such that

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} \times_d \mathbf{A} \in \mathbb{R}^{I_1 \times \cdots \times I_{d-1} \times K \times I_{d+1} \times \cdots \times I_D}$$

$$\iff \mathbf{Y}_{(d)} = \mathbf{A} \mathbf{X}_{(d)} \in \mathbb{R}^{K \times \prod_{\delta \neq d} I_\delta}.$$

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

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

More detail on these tensor specific definitions can be found in [11]. 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 $\text{LSV}_K(\mathbf{A}) \in \mathbb{R}^{I \times K}$ to be the matrix whose columns are K leading left singular vectors of $\mathbf{A}$.

multiple block terms, each of which is a TD with its own core and factor matrices. It is therefore able to capture more complex structures than the CPD or TD alone. By comparing the low-rank approximations using BTD and its edge cases CPD and TD, we evaluate how well the low-rank model based on the PZ model holds for RIRs in various scenarios. Numerical results will show that the most parameter-efficient decomposition of RIRs is BTD with scalar cores, i.e., CPD, strengthening the model connecting the PZ model with low-rank CPD.

The main contribution of this paper is the introduction of two joint compression methods. The straightforward approach to multichannel compression in a tensor decomposition context is to apply 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 variant of BTD, aptly named JMCBTD, and a corresponding optimization strategy. 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 low distance in the factor space. Spectral clustering then forms clusters, i.e. subgraphs, of factors that can be considered similar within

<Society logo(s) and publication title will appear here.>

a mode and rank. This differs from the coupled canonical polyadic decomposition (C-CPD) approach analyzed in [12], [13] 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) [11], [14], [15], [16].

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 it outperforms 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.

II. Tensorized Acoustic Room Impulse Responses and Tensor Decompositions

Let us consider M RIRs $h_1(n), h_2(n), \ldots, h_M(n)$ measured in the same room. Let L be the number of samples of a RIR $h_m(n)$ that we consider significant.

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

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

B. Block term decomposition (BTD)

Let us consider a tensor $\mathcal{H} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_D}$ and a BTD with R blocks, where each block r has its own core tensor $\mathcal{G}^{(r)} \in \mathbb{R}^{K_1^{(r)} \times K_2^{(r)} \times \cdots \times K_D^{(r)}}$ and factor matrices $\mathbf{A}^{(r,d)} \in \mathbb{R}^{I_d \times K_d^{(r)}}$ for each mode d. Then the BTD is given by

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

Let us say that $K_d^{(r)} \leq I_d$ for all r such that we can use the constrained decomposition where $\mathbf{A}^{(r,d)\top} \mathbf{A}^{(r,d)} = \mathbf{I}$. The BTD [16], [15], [14] is a generalization of the CPD and TD. If the core dimensions are chosen such that $\mathcal{G}^{(r)} = g^{(r)} \in \mathbb{R}$ is a scalar, then this reduces to the rank-R CPD

$$\widehat{\mathcal{H}} = \sum_{r=1}^{R} g^{(r)} \mathbf{u}^{(r,1)} \circ \cdots \circ \mathbf{u}^{(r,D)}, \tag{5}$$

while if $R = 1$ and $K_d > 1$ for some d, it reduces to the TD

$$\widehat{\mathcal{H}} = \mathcal{G} \times_{d \in [D]} \mathbf{A}^{(d)}. \tag{6}$$

Computing the BTD can be done using an alternating HOOI algorithm. For convenience, let us define the set of all block-mode indices (r, d) as

$$\mathcal{I} \triangleq \{1, \ldots, R\} \times \{1, \ldots, D\}. \tag{7}$$

The procedure for computing the BTD using alternating HOOI [11] is as follows: The minimization problem is given by

$$\min_{\mathbf{A}^{(r,d)}} \quad \left\| \mathcal{H} - \sum_{r=1}^{R} \mathcal{G}^{(r)} \times_{d \in [D]} \mathbf{A}^{(r,d)} \right\|_F^2 \\ \text{s.t.} \quad \mathbf{A}^{(r,d)\top} \mathbf{A}^{(r,d)} = \mathbf{I} \quad \text{for all } (r,d) \in \mathcal{I} \tag{8}$$

using tensor algebra, we can write this as

$$\max_{\mathbf{A}^{(r,d)}} \quad \left\| \sum_{r=1}^{R} \mathcal{H} \times_{d \in [D]} \mathbf{A}^{(r,d)\top} \right\|_F^2 \tag{9}$$

which further is also

$$\max_{\mathbf{A}^{(r,d)}} \quad \left\| \mathbf{A}^{(r,d)\top} \mathbf{Y}_{(d)}^{(r)} \right\| \tag{10}$$

were $\mathbf{Y}_{(d),m}^{(r)}$ is the mode-d unfolding of the tensor

$$\mathcal{Y}^{(r,d)} = \left[\mathcal{H} - \sum_{s \neq r} \widehat{\mathcal{H}}^{(s)} \right] \times_{\delta \neq d} \mathbf{A}^{(r,\delta)\top} \tag{11}$$

where

$$\widehat{\mathcal{H}}^{(s)} = \mathcal{G}^{(s)} \times_{d \in [D]} \mathbf{A}^{(s,d)} \tag{12}$$

We end up with a nested loop in which, for each block r we compute the residual tensor (11) and then within we iterate over the modes d (repeatedly) and compute the SVD of the mode-d unfolding $\mathbf{Y}_{(d)}^{(r)}$ to get the leading left singular vectors

$$\mathbf{A}^{(r,d)} = \mathrm{LSV}_{K_d^{(r)}} \left(\mathbf{Y}_{(d)}^{(r)} \right). \tag{13}$$

Algorithm 1 summarizes the overall procedure.

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

To establish the connection between the PZ model of RIRs and the low-rank tensor representation, let us consider the model

$$y(n) = H(q)x(n) = \frac{B(q)}{A(q)} x(n) \tag{14}$$

where $H(q)$ is the filter representing the causal RIR $h(n)$ of the room, $B(q)$ and $A(q)$ are the numerator and denominator polynomials of the filter with degrees P_B and P_A, respectively, and $x(n)$ and $y(n)$ are the input and output signal, respectively. The operator q is the shift operator, defined by $q^{-l}x(n) = x(n-l)$. In the z-domain, this is equivalent to

$$H(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z^{-1} + \cdots + b_{P_B} z^{-P_B}}{a_0 + a_1 z^{-1} + \cdots + a_{P_A} z^{-P_A}}, \tag{15}$$

with real coefficients a_k, b_k and let us denote the simple poles of it as z_r for $r = 1, \ldots, P_A$.

Algorithm 1 Alternating HOOI for BTD

1: **Input:** Tensor $\mathcal{H} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_D}$, number of blocks R, block ranks $(K_1^{(r)}, K_2^{(r)}, \ldots, K_D^{(r)})$, outer iteration count N_o, inner iteration count N_i

2: **Output:** Core tensors $\mathcal{G}^{(r)} \in \mathbb{R}^{K_1^{(r)} \times K_2^{(r)} \times \ldots \times K_D^{(r)}}$ for each r and factor matrices $\mathbf{A}^{(r,d)} \in \mathbb{R}^{I_d \times K_d^{(r)}}$ for each d and r

3: **Initialize:** Random $\mathbf{A}^{(r,d)}$ for each r, d, random $\mathcal{G}^{(r)}$ for each r

4: **for** N_o **do**

5:     **for** $r = 1, \ldots, R$ **do**

6:         $\mathcal{H}^{(l \neq r)} = \mathcal{H} - \sum_{l \neq r} \widehat{\mathcal{H}}^{(l)}$

7:         **for** inner iteration count N_i **do**

8:             **for** $d = 1, \ldots, D$ **do**

9:                 $\mathcal{Y}^{(r,d)} = \mathcal{H}^{(l \neq r)} \times_{\delta \neq d} \mathbf{A}^{(r,\delta)\top}$

10:                $\mathbf{A}^{(r,d)} \leftarrow \mathrm{LSV}_{K_d^{(r)}}\left(\mathbf{Y}_{(d)}^{(r)}\right)$

11:             **end for**

12:         **end for**

13:         $\mathcal{G}^{(r)} \leftarrow \mathcal{H} \times_{d \in [D]} \mathbf{A}^{(r,d)\top}$

14:     **end for**

15: **end for**

A. Modal factorization of impulse responses

We aim to derive the impulse response $h(n)$ of the filter $H(q)$ in terms of its poles and zeros, which is known as the modal factorization. Let us make the assumption that

$$P_B > P_A, \tag{16}$$

so the filter

$$H(q) = \frac{B(q)}{A(q)} \tag{17}$$

is *improper*. Let us define the monic polynomials

$$\tilde{A}(z) = z^{P_A} A(z) = a_0 z^{P_A} + a_1 z^{P_A - 1} + \cdots + a_{P_A}, \tag{18}$$
$$\tilde{B}(z) = z^{P_B} B(z) = b_0 z^{P_B} + b_1 z^{P_B - 1} + \cdots + b_{P_B}. \tag{19}$$

Since the filter is improper, $\deg \tilde{B} = P_B > P_A = \deg \tilde{A}$, we apply polynomial long division to write:

$$\tilde{B}(z) = F(z)\tilde{A}(z) + R(z), \tag{20}$$

where $F(z)$ is a polynomial of degree $P_F = P_B - P_A$ and $R(z)$ is a polynomial of degree $P_A - 1$ or less, i.e., $\deg R < P_A$. Write

$$F(z) = f_0 z^{P_F} + f_1 z^{P_F - 1} + \cdots + f_d, \tag{21}$$
$$R(z) = r_0 z^{P_A - 1} + r_1 z^{P_A - 2} + \cdots + r_{P_A - 1}. \tag{22}$$

Then

$$\frac{\tilde{B}(z)}{\tilde{A}(z)} = F(z) + \frac{R(z)}{\tilde{A}(z)}. \tag{23}$$

Because $H(q) = B(q)/A(q)$ in z-notation equals

$$H(z) = z^{-P_F}\left[F(z) + R(z)/\tilde{A}(z)\right], \tag{24}$$

the polynomial $F(z)$ is preceded by z^{-P_F}. Concretely, if

$$F(z) = \sum_{k=0}^{P_F} f_k z^{P_F - k}, \tag{25}$$

then

$$z^{-P_F} F(z) = \sum_{k=0}^{P_F} f_k z^{-k}. \tag{26}$$

Applying that to an impulse $\delta(n)$ yields

$$h_F(n) = \sum_{k=0}^{P_F} f_k \delta(n - k). \tag{27}$$

Thus

$$h_F(n) = \begin{cases} f_k, & n = k, \quad 0 \le k \le P_F, \\ 0, & n > P_F, \end{cases} \tag{28}$$

where $f_0 = b_0/a_0$. This is the *head* of the impulse response, which is FIR with length P_F.

Now, let us consider the remainder term $R(z)/\tilde{A}(z)$. The polynomial $\tilde{A}(z)$ has degree P_A and can be written as

$$\tilde{A}(z) = a_0 \prod_{r=1}^{P_A} (z - z_r), \tag{29}$$

with simple poles z_r. Then with

$$\frac{R(z)}{\tilde{A}(z)} = \sum_{r=1}^{P_A} \frac{\beta^{(r)}}{1 - z_r z^{-1}}, \tag{30}$$

and

$$\beta^{(r)} = \frac{R(z_r)}{\tilde{A}'(z_r)}, \tag{31}$$

where $\tilde{A}'(z_r)$ is the derivative of $\tilde{A}(z)$ evaluated at z_r. Each term $\beta^{(r)}/(1 - z_r z^{-1})$ has impulse $\beta^{(r)} z_r^n, n \ge 0$. Summed over the number of poles, the "undelayed" impulse is

$$\sum_{r=1}^{P_A} \beta^{(r)} z_r^n, \quad n \ge 0. \tag{32}$$

This is the *tail* of the impulse response, which is infinite in length. Because $H(z)$ includes z^{-d} in front, the actual remainder's contribution is

$$h_R(n) = \begin{cases} 0, & 0 \le n \le P_F - 1, \\ \sum_{m=1}^{P_A} \beta^{(r)} z_r^{n - P_F}, & n \ge P_F. \end{cases} \tag{33}$$

By combining head and tail

$$h(n) = h_F(n) + h_R(n), \tag{34}$$

we obtain the final formula

$$h(n) = \begin{cases} f_n, & 0 \le n \le P_F - 1, \\ f_{P_F} + \sum_{m=1}^{P_A} \beta^{(r)}, & n = P_F, \\ \sum_{m=1}^{P_A} \beta^{(r)} z_r^{n - P_F}, & n \ge P_F + 1, \\ 0, & n < 0, \end{cases} \tag{35}$$

<Society logo(s) and publication title will appear here.>

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,

$$h_F(n) = \sum_{k=0}^{N} \alpha_k \delta(n - \tau_k),$$

where α_k are the amplitudes of the echoes and τ_k are their delays. Let there be the multi-index $\tau_k \leftrightarrow (i_1^{(k)}, i_2^{(k)}, \ldots, i_D^{(k)})$ defined in (2). Then let's define the standard basis vectors

$$\mathbf{e}^{(k,d)} = [0, \ldots, 0, 1, 0, \ldots, 0]^\top \in \mathbb{R}^{I_d \times 1} \tag{36}$$

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

$$
\begin{aligned}
\mathcal{H}_F &= \sum_{k=0}^{N} \alpha_k \left(\mathbf{e}^{(k,1)} \circ \mathbf{e}^{(k,2)} \circ \cdots \circ \mathbf{e}^{(k,D)} \right) \\
&= \sum_{k=0}^{N} \alpha_k \times_{d \in [D]} \mathbf{e}^{(k,d)}
\end{aligned}
\tag{37}
$$

with $\mathrm{rank}(\mathcal{H}_F) \leq N$ Note that here the core in BTD form is scalar $K_d = 1$. Since for the tail $h_R(N)$ for $n \geq P_F$,

$$h_R(n) = \sum_{r=1}^{P_A} \beta^{(r)} z_r^{n - P_F}, \tag{38}$$

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

$$
\begin{aligned}
\mathcal{H}_R[i_1, \ldots, i_D] &= \sum_{r=1}^{P_A} \beta^{(r)} z_r^{(i_1 + i_2 I_1 + \cdots + i_D I_1 I_2 \cdots I_{D-1} - P_F)} \\
&= \sum_{r=1}^{P_A} \widetilde{\beta}^{(r)} \left(z_r^1 \right)^{i_1} \left(z_r^{I_1} \right)^{i_2} \left(z_r^{I_1 I_2} \right)^{i_3} \cdots \left(z_r^{I_2 I_3 \cdots I_{D-1}} \right)^{i_D},
\end{aligned}
\tag{39}
$$

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

$$
\begin{aligned}
\mathcal{H}_R &= \sum_{r=1}^{P_A} \widetilde{\beta}^{(r)} \left(\mathbf{a}^{(r,1)} \circ \mathbf{a}^{(r,2)} \circ \cdots \circ \mathbf{a}^{(r,D)} \right) \\
&= \sum_{r=1}^{P_A} \widetilde{\beta}^{(r)} \times_{d \in [D]} \mathbf{a}^{(r,d)}
\end{aligned}
\tag{40}
$$

a rank-P_A CPD, with the dth mode-factor

$$\mathbf{a}^{(r,d)}[i_d] = z_r^{\left(i_d \prod_{\delta=1}^{d-1} I_\delta \right)}, \quad i_d = 0, 1, \ldots, I_d - 1 \tag{41}$$

for each $r = 1, \ldots, P_A$ and $d = 1, \ldots, D$.

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

$$\mathcal{H} = \mathcal{H}_F + \mathcal{H}_R \tag{42}$$

with $\mathrm{rank}(\mathcal{H}) \leq N + P_A$. This shows that the PZ model of RIRs leads to a low-rank tensor representation of the RIRs in the form of a CPD, i.e., a BTD with scalar cores $K_d = 1$ for all d.

C. Factor subspace similarities

Let there be a set of positions $m = 1, \ldots, M$ in a room for each of which we assume that the PZ model and its tensorization $\mathcal{H}_m = \mathcal{H}_{mF} + \mathcal{H}_{mR}$ (42) applies to. Let the subspace for mode d be spanned by the set of factors,

$$\mathcal{S}_m^{(d)} = \mathrm{span}\{\mathbf{e}_m^{(1,d)}, \ldots, \mathbf{e}_m^{(K,d)}, \mathbf{a}_m^{(1,d)}, \ldots, \mathbf{a}_m^{(P_A,d)}\}. \tag{43}$$

It is easy to see that if the RIRs share the same poles, i.e., the same $\mathbf{a}_m^{(r,d)}$ across m for $r = 1, \ldots, P_A$, then the subspaces $\mathcal{S}_m^{(d)}$ are similar across the positions m particularly in the "later" modes, i.e., larger d. This is due to the improved separation of basis vectors from larger exponents for $|z_r| < 1$ in (41) and therefore better conditioning of the basis which leads to better identification in the decomposition. While the head factors $\mathbf{e}_m^{(k,d)}$ are not generally shared across the RIRs, they may be similar across the positions m if the delays of the early reflections are similar or have integer-multiple relations relative to the tensorization's mode lengths.

We can verify this by computing distances between the subspaces. Then define the distance between two subspaces $d_S(\mathcal{S}_i^{(d)}, \mathcal{S}_j^{(d)})$ as one of the common metrics for subspace distances [17], such as the Chordal or Grassmannian distance. Figure 1 shows the Chordal distance plotted over the top view of a shoebox room with RIRs simulated at a grid of positions. It is evident that the subspaces are highly similar in the "last" tensor mode $d = 3$ and show less or no similarity for $d = 2$ and $d = 1$, respectively.

IV. Joint Multichannel RIR Compression using Block term decomposition (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 block term decomposition (SMCBTD)

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

$$\mathcal{T}[m, i_1, i_2, \ldots, i_D] = \mathcal{H}_m[i_1, i_2, \ldots, i_D]. \tag{44}$$

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

$$\mathcal{T} \approx \widehat{\mathcal{T}} = \sum_{r=1}^{R} \mathcal{G}^{(r)} \times_{d \in [D+1]} \mathbf{A}^{(r,d)} \tag{45}$$

with the analogous definitions as in Section B for a $(D+1)$-mode tensor. From the theoretical background in Section C,

we can expect that in the channel mode $d = 1$ the high similarity of the subspaces across the RIRs causes low-rank structure as discussed in Section C. The later modes $d = 2, ..., D+1$ will show the low-rank structure of the individual RIRs as discussed in Section B. The compression ratio for the SMCBTD is defined as

$$C_{\text{SMCBTD}} = 1 - \frac{\sum_{r=1}^{R} \left(\prod_{d=1}^{D+1} K_d^{(r)} + \sum_{d=1}^{D+1} I_d K_d^{(r)} \right)}{ML}. \tag{46}$$

B. Joint multichannel block term decomposition (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 hard-shared factors across RIRs. Thus, the goal is to reduce the number of coefficients by considering shared factors. For this, for each $(r, d) \in \mathcal{I}$, we need to define sets/clusters of RIRs that share a factor matrix. Let such a cluster be denoted as

$$\mathcal{C}_c^{(r,d)} \subseteq \{1, \ldots, M\} \tag{47}$$

where $c = 1, \ldots, C^{(r,d)}$ is the cluster index and $C^{(r,d)}$ the number of clusters for the specific (r, d). Let us define the factor matrices as

$$\mathbf{A}_m^{(r,d)} = \begin{cases} \mathbf{U}_c^{(r,d)} & \text{if } m \in \mathcal{C}_c^{(r,d)}, \\ \mathbf{U}_m^{(r,d)} & \text{if } m \notin \bigcup_{c=1}^{C^{(r,d)}} \mathcal{C}_c^{(r,d)} \end{cases}, \tag{48}$$

which replace the individual factor matrices at the specified RIR m and block-mode index (r, d). With this, we define the optimization problem for the joint multichannel RIR compression as

$$\min \quad \sum_{m=1}^{M} \left\| \mathcal{H}_m - \sum_{r=1}^{R} \mathcal{G}_m^{(r)} \times_{d \in [D]} \mathbf{A}_m^{(r,d)} \right\|_F^2 \tag{49}$$

$$\text{s.t.} \quad \mathbf{A}_m^{(r,d)\top} \mathbf{A}_m^{(r,d)} = \mathbf{I} \quad m = 1, \ldots, M, \ (r, d) \in \mathcal{I}$$

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 aggregated residuals from within the cluster for the SVD step such that

$$\mathbf{U}_c^{(r,d)} = \text{LSV}_{K_d^{(r)}} \left(\sum_{m \in \mathcal{C}_c^{(r,d)}} \mathbf{Y}_{m,(d)}^{(r)} \right).$$

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 find optimal clusters in the next section.

Algorithm 2 Alternating HOOI for joint BTD

1: **Input:** Tensors $\mathcal{H}_m \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_D}$, 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

2: **Output:** Core tensors $\mathcal{G}_m^{(r)} \in \mathbb{R}^{K_1^{(r)} \times K_2^{(r)} \times \ldots \times K_D^{(r)}}$, factor matrices $\mathbf{U}_m^{(r,d)} \in \mathbb{R}^{I_d \times K_d^{(r)}}$, $\bar{\mathbf{U}}_c^{(r,d)} \in \mathbb{R}^{I_d \times K_d^{(r)}}$

3: **Initialize:** Random $\mathbf{U}_m^{(r,d)}$ for each r, d, m, random $\bar{\mathbf{U}}_c^{(r,d)}$ for each r, d, c, random $\mathcal{G}_m^{(r)}$ for each r, m

4: **for** N_o **do**
5:     **for** $r = 1, \ldots, R$ **do**
6:         **for** $m = 1, \ldots, M$ **do**
7:             $\mathcal{H}_m^{(l \neq r)} = \mathcal{H}_m - \sum_{l \neq r} \widehat{\mathcal{H}}_m^{(l)}$
8:         **end for**
9:         **for** N_i **do**
10:             **for** $d = 1, \ldots, D$ **do**
11:                 **for** $c = 1, \ldots, C^{(r,d)}, m \in \mathcal{C}_c^{(r,d)}$ **do**
12:                     $\mathcal{Y}_c^{(r,d)} = \mathcal{H}_m^{(l \neq r)} \times_{\delta \neq d} \bar{\mathbf{U}}_c^{(r,\delta)\top}$
13:                     $\bar{\mathbf{U}}_c^{(r,d)} \leftarrow \text{LSV}_{K_d^{(r)}} \left(\sum_{m \in \mathcal{C}_c^{(r,d)}} \mathbf{Y}_{c(d)}^{(r,d)} \right)$
14:                 **end for**
15:                 **for** $m = 1, \ldots, M$ **do**
16:                     $\mathcal{Y}_m^{(r,d)} = \mathcal{H}_m^{(l \neq r)} \times_{\delta \neq d} \mathbf{U}_m^{(r,\delta)\top}$
17:                     $\mathbf{U}_m^{(r,d)} \leftarrow \text{LSV}_{K_d^{(r)}} \left(\mathbf{Y}_{m(d)}^{(r,d)} \right)$
18:                 **end for**
19:             **end for**
20:         **end for**
21:         **for** $m = 1, \ldots, M$ **do**
22:             $\mathcal{G}_m^{(r)} \leftarrow \mathcal{H}_m \times_{d \in [D]} \mathbf{A}_m^{(r,d)\top}$
23:         **end for**
24:     **end for**
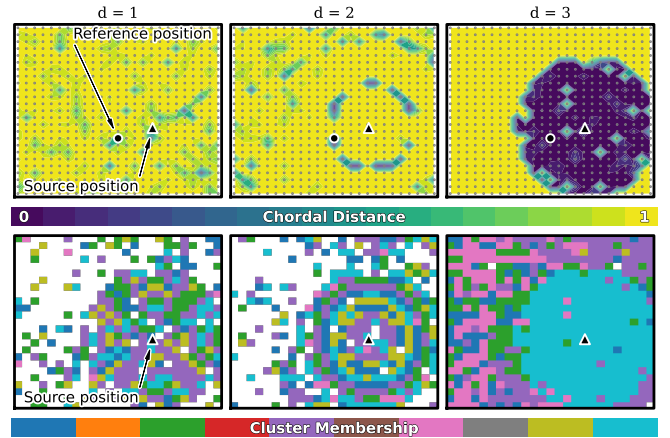25: **end for**

FIGURE 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.

<Society logo(s) and publication title will appear here.>

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 heuristic choice is possible choosing more shared factors in modes where higher similarity is expected. However, to find optimal clusters, we employ spectral clustering on the factor subspace affinity graphs constructed from the factor matrices $\bar{\mathbf{U}}_c^{(r,d)}$. Let us define the factor subspace affinity graph for a (r,d) as

$$\mathcal{G}^{(r,d)} \triangleq \{\mathcal{V}^{(r,d)}, \mathcal{E}^{(r,d)}, w^{(r,d)}\} \tag{50}$$

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 ubspace distance by

$$w^{(r,d)}(i,j) = \exp\left(-\frac{d_S(\mathbf{U}_i^{(r,d)}, \mathbf{U}_j^{(r,d)})}{\sigma^2}\right), \tag{51}$$

where $d_S(\cdot,\cdot)$ is the subspace distance function 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 optimal 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:

$$\mathbf{W}^{(r,d)}[i,j] = \begin{cases} w^{(r,d)}(i,j) & \text{if } (i,j) \in \mathcal{E}^{(r,d)} \\ 0 & \text{otherwise} \end{cases} \tag{52}$$

The rows and columns of isolated nodes, i.e., nodes with low or no similarity to any other nodes, need to be removed. This implies removing the rows and columns where the row and column-wise sum of elements is below the threshold τ. From the truncated affinity matrix, denoted $\tilde{\mathbf{W}}^{(r,d)} \in \mathbb{R}^{\tilde{M} \times \tilde{M}}$ with $\tilde{M} \leq M$, we compute the normalized Laplacian matrix

$$\mathbf{L}^{(r,d)} = \mathbf{I} - \left(\mathbf{D}^{(r,d)}\right)^{-1/2} \tilde{\mathbf{W}}^{(r,d)} \left(\mathbf{D}^{(r,d)}\right)^{-1/2} \tag{53}$$

where $\mathbf{D}^{(r,d)}$ 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 $C^{(r,d)}$ most significant eigenvalues and their corresponding eigenvectors through the maximum eigengap

$$C^{(r,d)} = \arg \max_{1 \leq i \leq M} \left(\lambda_{i+1}^{(r,d)} - \lambda_i^{(r,d)}\right). \tag{54}$$

We additionally limit the number of clusters

$$C^{(r,d)} \leftarrow \min\{C^{(r,d)}, C_{\max}\}. \tag{55}$$

After row-normalization of the matrix formed by the eigenvectors

$$\tilde{\mathbf{V}}^{(r,d)} = \begin{bmatrix} \mathbf{v}_1^{(r,d)} & \cdots & \mathbf{v}_{C^{(r,d)}}^{(r,d)} \end{bmatrix}, \tag{56}$$

we apply k-means clustering on the rows to obtain the final clusters $\mathcal{C}_c^{(r,d)}$ for $c = 1, \ldots, C^{(r,d)}$. This partitions the M RIRs into $C^{(r,d)}$ clusters, i.e., subgraphs of $\mathcal{G}^{(r,d)}$, for

each (r,d) based on their similarity, i.e., closeness on the Grassmannian manifold of each block-mode space. 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 a heuristic choice of clusters. These clusters can then be used to initiate the joint JMCBTD compression algorithm.

Algorithm 3 Joint BTD using spectral clustering

1: **for** $m = 1, \ldots, M$ **do**
2:     Compute individual BTDs (Algorithm 1) for $\mathcal{H}_m$
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 2) using clusters $\mathcal{C}_c^{(r,d)}$

The compression ratio for the JMCBTD is defined as

$$C_{\text{C-BTD}} = 1 - \frac{N_{\text{cores}} + \sum_{(r,d)\in\mathcal{I}}\left(N_{\text{shr}}^{(r,d)} + N_{\text{ind}}^{(r,d)}\right)}{ML} \tag{57}$$

where

$$N_{\text{cores}} = M \sum_{r=1}^{R} \prod_{d=1}^{D} K_d^{(r)}, \tag{58}$$

$$N_{\text{shr}}^{(r,d)} = C^{(r,d)} I_d K_d^{(r)}, \text{ and} \tag{59}$$

$$N_{\text{ind}}^{(r,d)} = \sum_{m \notin \bigcup_{c=1}^{C^{(r,d)}} \mathcal{C}_c^{(r,d)}} I_d K_d^{(r)} \tag{60}$$

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 [18], [19] or the Auction algorithm [20], which we employ here, such that similar blocks are grouped together for better clustering and compression performance.

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) [21] and shoebox room impulse response archive with varying absorption (SRIRACHA) [22] and church impulse response dataset (ChurchIR) [23] to validate the applicability to real-world scenarios. Let the error measure for the reconstruction be the normalized misalignment (NM) defined as

$$\text{NM}(\mathcal{H}_m, \widehat{\mathcal{H}}_m) = \frac{\|\mathcal{H}_m - \widehat{\mathcal{H}}_m\|_F}{\|\mathcal{H}_m\|_F} = \frac{\|\mathbf{h}_m - \hat{\mathbf{h}}_m\|}{\|\mathbf{h}_m\|}, \tag{61}$$

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

$$\overline{\text{NM}} = \frac{1}{M} \sum_{m=1}^{M} \text{NM}(\boldsymbol{\mathcal{H}}_m, \widehat{\boldsymbol{\mathcal{H}}}_m) \qquad (62)$$

and let

$$\overline{\text{NM}}_{\text{dB}} = 20 \log_{10}(\overline{\text{NM}}) \qquad (63)$$

be the NM in decibel scale.

For any set of RIRs, be it simulated or measured, we first estimate the reverberation time RT_{60} and then truncate the RIRs to $L \geq I_{RT_{60}} = RT_{60} f_s$, where we define this as the smallest number of samples such that we have the most equidimensional tensor shape. For this, let $u = \lfloor \sqrt[D]{I_{RT_{60}}} \rfloor$ be the largest integer such that $u^D \leq I_{RT_{60}}$. Then choose the smallest $v \in \{1, ..., D\}$ such that $u^{D-v}(u+1)^v \geq I_{RT_{60}}$. We aim for an equidimensional tensor shape, due to the better compression performance as discussed in [8].

The methods in our comparison are the proposed SM-CBTD and JMCBTD as well as the GLRAM method [7] 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) [24], 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 $RT_{60} \approx 0.5$ 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 1. 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 [7]. We vary the tensorization order D and the compression rate to evaluate the performance. The decomposition parameters used are listed in Table 1.

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 close source position at a minimum reconstruction error at the closest microphone, the advantage of the proposed JMCBTD method is significant in the far source position. 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
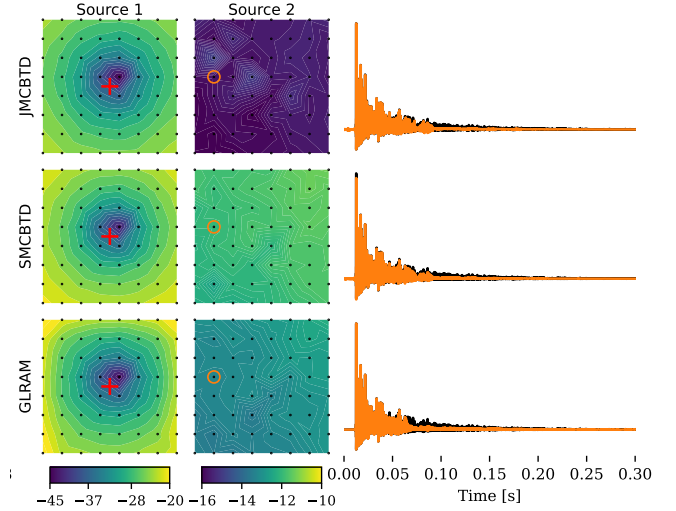

FIGURE 2. This figure illustrates the reconstruction error NM over the compact microphone array geometry for the two source position (near and far field). We can observe that the reconstruction error for the proposed JMCBTD method is significantly lower across the array at the far field position. Additionally, example RIRs and their reconstructions using the proposed JMCBTD method with tensorization order $D = 4$ at compression rate $C = 0.8$ for a selected microphone position (circled in orange) are shown for the far field position (differences would be difficult to see for the near field position).

-16 dB mean NM compared to -13 dB for GLRAM and -12 dB for SMCBTD at the far position. The SMCBTD shows similar performance to the GLRAM based method in [7] in this scenario. Further, Figure 2 shows example RIRs and their reconstructions for the far field position for a compression rate of $C = 0.8$ using the proposed JMCBTD method with tensorization order $D = 4$. Figure 3 shows the mean reconstruction error versus compression rate for different tensorization orders $D \in \{3, 4, 5\}$ for a simulated compact microphone array at a close and far source position. 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 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 do not exhibit modal behaviour. We use the parameters listed in Table 1 for the proposed algorithm.

<Society logo(s) and publication title will appear here.>

Experiment	Room Parameters					Decomposition Parameters						
	Size [l, w, h]	RT_{60}	Source Pos.	M	f_s	D	K_d	σ	τ	$C_{\max}$	N_o	N_i
Simulated Array	8.00 m, 8.00 m, 3.00 m	0.53 s	4.0, 4.0, 1.2 m	64	48 kHz	3, 4, 5	1, 2, 4	0.2 . . . 1.2	0.4	6	80	8
			1.0, 1.0, 1.2 m									
SMARD 2X0X	7.34 m, 8.09 m, 2.87 m	0.15 s	see [21]	168	48 kHz			0.6 . . . 1.8				
SRIRACHA C1	6.22 m, 3.85 m, 3.07 m	0.5 s	see [22]	256	36 kHz			0.8 . . . 2.0				

TABLE 1. Room and decomposition parameters for the numerical experiments. The source positions for the simulated array experiments are illustrated in Figure 2.
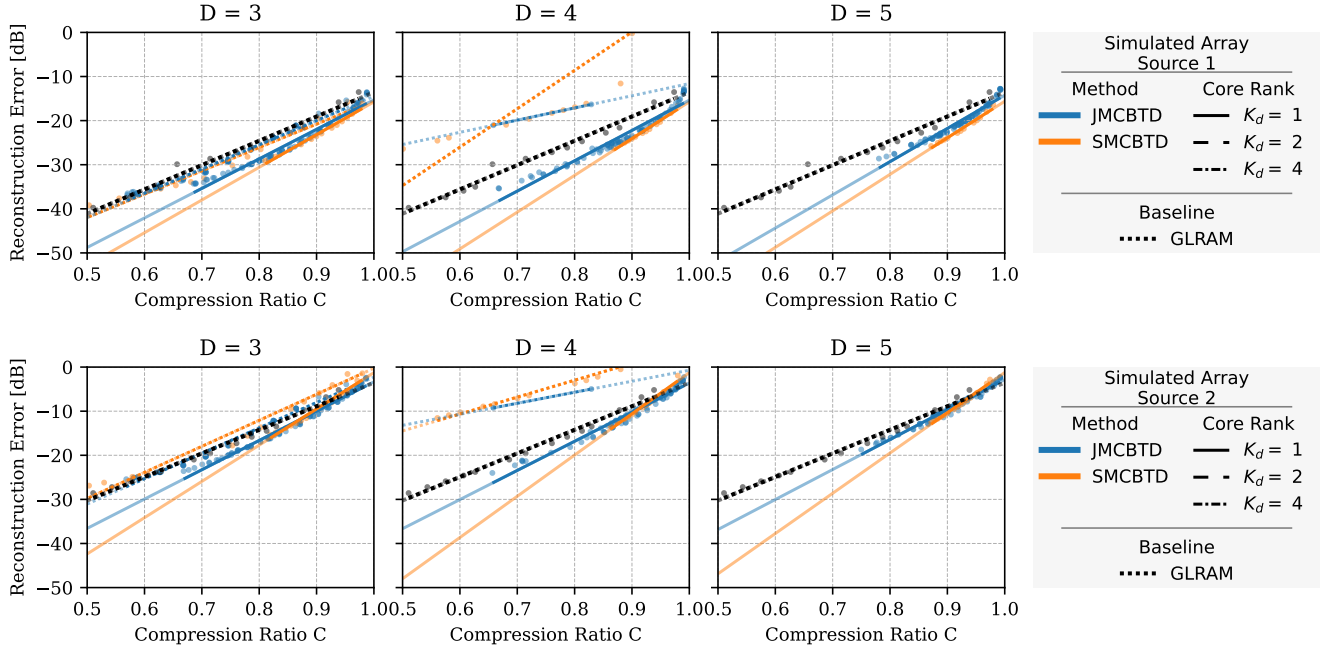

FIGURE 3. Compression rate vs. mean reconstruction error for different tensorization orders D for a simulated compact microphone array at a close and far source position, see Figure 2. We compare the proposed JMCBTD and SMCBTD with the GLRAM based method in [7]. In simulated rooms with low reverberation time, the proposed methods achieve lower reconstruction error across a range of compression rates, when the core rank is kept at 1, where the performance gain is less pronounced at the far position.

1) SMARD

The SMARD dataset [21] 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 $RT_{60} \approx 0.15$ s.

Figure 4 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 constance compression rate, increased core rank leads to decreased performance, with this effect being more pronounced for higher tensorization orders. Further, we can observe that there is a clear clustering effect in the shared factors found by the proposed JMCBTD method, which aligns well with the microphone arrays of the configurations. This can be seen by the affinity measures within and across the different microphone arrays. The average affinity within the same microphone array is around TODO while the average affinity across different microphone arrays is around TODO.

2) SRIRACHA

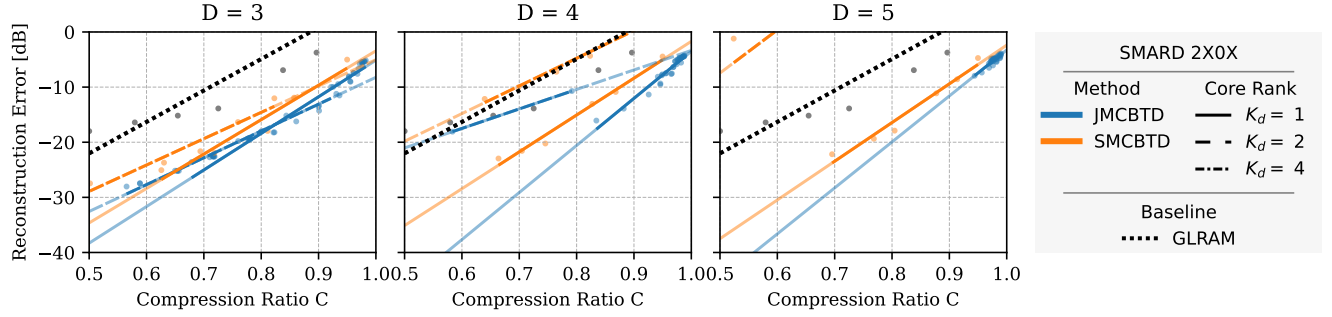The SRIRACHA dataset [22] contains a large set of measured RIRs in a shoebox room with densely placed source

FIGURE 4. Compression rate vs. mean reconstruction error for different tensorization orders D on the set of impulse responses in SMARD [21] 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 $D = 3, 4, 5$.

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 $f_s = 32$ kHz and are already truncated to the average $RT_{60} \approx 0.5$ s at 16,000 samples. The truncation process described at the beginning of Section V therefore might pad the RIRs with zeros to reach a more equidimensional tensor shape.

Figure 5 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, 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 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 less modal behaviour than the room used in SMARD.

3) ChurchIR
The ChurchIR dataset contains RIRs measured in a church using higher-order spherical microphone arrays. The spatial sampling is relatively wide compared to the room dimensions.

> **Todo: [25]**
>
> relevant comparison for Ambi RIRs? c

VI. Discussion
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. Similar im-

provements were observed for the SMCBTD method. An important discussion point is the difference between the two methods in terms of their feasibility applied to large datasets. The SMCBTD method requires the construction of potentially very large temporary tensors in intermediate steps of the alternating least squares (ALS) algorithm, specifically in the computation of per-block reconstructed tensors for larger number of blocks R. As our implementation favours vectorization/parallelization over memory efficiency, this can lead to high temporary memory consumption for large datasets, which is the reason for the omission of the SMCBTD results in some comparisons.

We would like to point to the fact that the GLRAM approach with shared left matrices, is a special case of the decomposition introduced in this paper. It is a $D = 2$, $R = 1$ JMCBTD where all channels are in one cluster. An extension of the GLRAM method, essentially an in-between step towards the method proposed here, is applying a clustering approach to the left matrices. We did however skip this in our comparisons as it does not add additional information.

Future extensions of the proposed methods are a single step clustering approach for the JMCBTD avoiding the two-step approach, and data-driven multilinear rank detection for both JMCBTD and SMCBTD, which was considered a chosen hyperparameter for this work. Despite the results showing that a scalar core having the best compression performance, it might still yield valuable insight into multichannel RIR tensor modelling.

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 literature. We then proposed two methods for joint compression of multichannel RIRs using the BTD. The first method SMCBTD
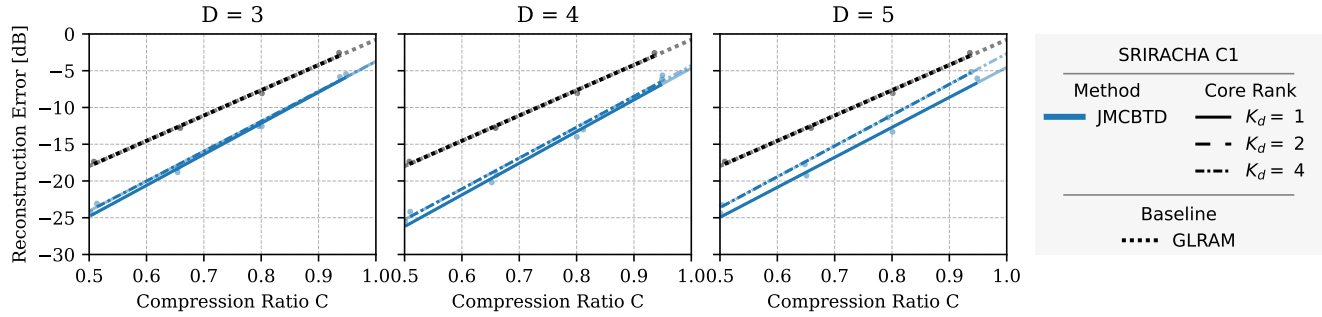
<Society logo(s) and publication title will appear here.>

FIGURE 5. Compression rate vs. mean reconstruction error for different tensorization orders D **on the SRIRACHA C1 dataset, spatially downsampled to** $M = 256$ **positions. This comparison does not include the SMCBTD method, as its computation was infeasible for this high-channel-count dataset due to memory constraints. The performance gain of the proposed JMCBTD method is substantial especially at** $K_d = 1$**, across all tensorization orders** $D = 3, 4, 5$**.**

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 the improved performance of the proposed methods over the baseline method, of which the JMCBTD method is superior.

REFERENCES

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

[2] 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.

[3] 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.

[4] Y. Haneda, S. Makino, and Y. Kaneda, "Multiple-point equalization of room transfer functions by using common acoustical poles," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 4, pp. 325–333, Jul. 1997.

[5] 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.

[6] 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.

[7] 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.

[8] ——, "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.

[9] G. Huang, J. Benesty, J. Chen, C. Paleologu, S. Ciochină, W. Kellermann, and I. Cohen, "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.

[10] 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.

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

[12] 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.

[13] 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.

[14] 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.

[15] L. De Lathauwer, "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.

[16] ——, "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.

[17] K. Ye and L.-H. Lim, "Schubert varieties and distances between subspaces of different dimensions," Jun. 2016.

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

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

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

[21] 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.

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

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

[24] 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.

[25] S. Kitić and J. Daniel, "Blind identification of ambisonic reduced room impulse response," *IEEE Trans. Audio Speech Lang. Process.*, vol. 32, pp. 443–458, 2024.

Matthias Blochberger (Graduate Student Member IEEE) photograph and biography not available at the time of publication.

Filip Elvander (Member, IEEE) photograph and biography not available at the time of publication.

Toon van Waterschoot (Member, IEEE), photograph and biography not available at the time of publication.