Low-Rank Tensor Modeling of Room Impulse Responses

Martin Jälmby
Dept. of Electrical Engineering
(ESAT-STADIUS) KU Leuven
Leuven, Belgium
martin.jalmby@esat.kuleuven.be

Filip Elvander
Dept. of Electrical Engineering
(ESAT-STADIUS) KU Leuven
Leuven, Belgium
filip.elvander@esat.kuleuven.be

Toon van Waterschoot
Dept. of Electrical Engineering
(ESAT-STADIUS) KU Leuven
Leuven, Belgium
toon.vanwaterschoot@esat.kuleuven.be

Abstract—In this paper, we showcase the benefit of low-rank modeling in the context of acoustic room impulse responses (RIRs). The RIR is modeled as a sum of exponentially decaying sinusoids, and by leveraging the inherent low-rank structure in tensorizations of such signals, we are able to achieve a very compact representation. The contribution of this paper is to provide a clear motivation for the use of low-rank models for the RIR, directly following from physical descriptions of room acoustics. Further, new light is shed on details of the derivation of the RIR model, in relation to solving the inhomogeneous Helmholtz equation. We demonstrate the very compact representations that can be achieved. Finally, the suggested model also displays good robustness properties with respect to the positions of the source and the receiver. Simulations performed on real-life room impulses responses support the theoretical findings.

Index Terms—Low-rank Modeling, Room Impulse Response, Tensor Decomposition

I. INTRODUCTION

Room impulse responses (RIRs) are essential to algorithms within a wide variety of acoustic signal processing tasks, such as speech dereverberation [1], source separation [2], source localization [3], echo cancellation [4], and artificial reverberation [5]. There are several different ways of modeling the RIR, including as an infinite impulse response (IIR) filter (see e.g. [6]), and as a finite impulse response (FIR) filter (see e.g. [7]). The advantage of an IIR filter model is that it offers a compact representation, but with the drawback of possible difficulties in estimating the filter parameters [8]. The FIR filter model, on the other hand, is appreciated for its simplicity, but has the downside that a large number of parameters is needed to fully represent the characteristics of the room [8]; for a regular office-sized room, such an FIR filter can be several thousands of taps long [9]. This may be prohibitive from a computational and memory requirement point of view, highlighting the need for compact representations of RIRs.

Low-rank models have received a lot of attention in signal processing literature as of late, and their usefulness has been demonstrated in areas such as data science [10], blind source

separation [11], blind system identification [12], and echo cancellation [13]. In works like [14] and [11] the authors demonstrate that several kinds of signals can be considered *low-rank*, in the sense that if the signal vector is reshaped into a matrix or tensor, that matrix or tensor will be of low-rank. These signals include polynomials, sinusoids and what we will use as model for the RIR, a sum of exponentially decaying sinusoids [15]. As we will see, by exploiting this inherent structure, one is able to dramatically reduce the number of parameters needed to represent the RIR.

The main contribution of this paper is to provide a clear, physical motivation for the use of low-rank matrix or tensor models of RIRs. We also shed new light on the approximations that are made throughout the derivation of the RIR model, all the way from the solution to the inhomogeneous Helmholtz equation. The suitability of these low-rank models is demonstrated with simulations on real-life recorded RIRs.

This paper is organized as follows: in Section II we introduce the signal model, along with the physical justification of the use of a sum of exponentially decaying sinusoids as RIR model, as well as the motivation for using a low-rank approximation. In Section III we present simulations that support the theoretical findings, and finally, Section IV summarizes and concludes this work and points out areas for possible future research.

II. SIGNAL MODEL

As previously indicated, in this work we consider modeling the RIR as a sum of exponentially decaying sinusoids. Assuming that the RIR is band-limited, and thereby can be sampled, the discrete time representation h is given by

$$h(\mathbf{r}_r, \mathbf{r}_s, n) = \sum_{m \in \mathbb{N}} \mu_m(\mathbf{r}_r, \mathbf{r}_s) e^{-\beta_m n} \cos(\omega_m n + \phi_m), \quad (1)$$

for $n = 1, 2, \ldots, N$. Here, μ_m denotes the initial amplitude, $\mathbf{r}_r \in \mathbb{R}^3$ the position of the receiver, $\mathbf{r}_s \in \mathbb{R}^3$ the position of the source, $\beta_m \in \mathbb{R}_+$ the exponential decay constant, $\omega_m \in [0, \pi]$ the angular frequency, and ϕ_m the phase. For ease of notation we will from hereon drop the dependence on $\mathbf{r}_r$ and $\mathbf{r}_s$ and simply refer to $h(\mathbf{r}_r, \mathbf{r}_s, n)$ as $h(n)$. As we will see, this model follows directly from simple physical descriptions of room acoustics, and directly allows for low-rank approximations.

Assuming that $q = \sqrt{N} \in \mathbb{N}$, the RIR vector $\boldsymbol{h} = [h(1), h(2), \ldots, h(N)]^T$ can be reshaped into a square matrix $\mathbf{H} \in \mathbb{R}^{q \times q}$,

$$\mathbf{H} = \begin{bmatrix} h(1) & h(q+1) & \ldots & h(q(q-1)+1) \\ \vdots & \vdots & & \vdots \\ h(q) & h(2q) & \ldots & h(N) \end{bmatrix}.$$

(2)

As will be shown, it follows from (1), together with some physical considerations of room acoustics, that $\mathbf{H}$ may be well approximated by a low-rank matrix $\hat{\mathbf{H}}$.

The idea of modeling an RIR as a sum of decaying sinusoids is well established in the literature. In works such as [16]–[18] this has been used for lower frequencies. The idea of low-rank modeling of RIRs is also anchored in previous work. This is done in e.g. [13], with support from the theoretical discussions in [11]. In this paper, however, we provide a previously missing link between said model, the actual physics of the problem and the use of low-rank tensor-based modeling. As is demonstrated, this allows for compact representations of RIRs.

A. Physical Motivation of Signal Model

The sound field in a source-free room is governed by the wave equation,

$$\nabla^2 p\left(\mathbf{r}_r, t\right) = \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2},$$

(3)

subject to initial and boundary conditions, where p denotes sound pressure, $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ is the Laplace operator, c the speed of sound, and t denotes time. Considering the frequency domain counterpart, we get the time-independent form of the wave equation, also known as the Helmholtz equation,

$$\left(\nabla^2 + k^2\right) P(\mathbf{r}_r, \omega) = 0,$$

(4)

where the wave number k and the angular frequency ω are related via $k = \frac{\omega}{c}$. We denote the eigenfunctions to the Laplace operator as ψ_m, with the corresponding eigenvalue k_m^2. For arbitrary room geometries, there are no analytical expressions for these eigenfunctions, or room modes, ψ_m. However, for simple geometries, such as shoe-box type rooms with rigid walls, there are closed-form expressions [15]. The eigenfunctions form a complete and orthogonal set [15] and any sound field can be expressed in terms of them [19]. This can be used when we look at a slightly modified version of (4), where we introduce a point source with volume velocity Q, at the source position $\mathbf{r}_s$, with a driving frequency ω. Further, denoting the value of the gas density by ρ_0, and the imaginary unit $i \triangleq \sqrt{-1}$, we get [15]

$$\left(\nabla^2 + k^2\right) G(\mathbf{r}_r, \mathbf{r}_s, \omega) = -i\omega\rho_0 Q\delta(\mathbf{r}_r - \mathbf{r}_s),$$

(5)

where $G(\mathbf{r}_r, \mathbf{r}_s, \omega)$ is the so called Green's function, or room transfer function. We can express it in terms of the aforementioned eigenfunctions ψ_m [15], yielding

$$G(\mathbf{r}_r, \mathbf{r}_s, \omega) = i\omega Q\rho_0 \sum_{m \in \mathbb{N}} \frac{\psi_m(\mathbf{r}_s)\psi_m(\mathbf{r}_r)}{K_m(k_m^2 - k^2)},$$

(6)

where K_m is a normalizing constant. This is also known as the modal decomposition. In general, the eigenfunctions and eigenvalues are complex-valued, although for the special case of perfectly reflecting room boundaries, they are real-valued [15]. The eigenfunctions and eigenvalues depend only on the dimensions of the room, and not on the source and receiver positions [20]. Furthermore, the eigenvalue k_m^2, corresponding to mode m, satisfies $k_m = \frac{\tilde{\omega}_m}{c} + i\frac{\beta_m}{c}$, where $\tilde{\omega}_m$ is the resonance frequency of the mth mode, and β_m is its decay constant [15]. Inserting this in (6), along with $k = \frac{\omega}{c}$, and expanding the denominator yields

$$G(\mathbf{r}_r, \mathbf{r}_s, \omega) = \sum_{m \in \mathbb{N}} \frac{A_m i\omega}{-\omega^2 + \tilde{\omega}_m^2 + 2i\tilde{\omega}_m\beta_m - \beta_m^2},$$

(7)

where, for the sake of compactness, we collected all the, for our purposes, less interesting variables in $A_m \triangleq A_m(\mathbf{r}_s, \mathbf{r}_r) = Q\rho_0 c^2 \psi_m(\mathbf{r}_s)\psi_m(\mathbf{r}_r)/K_m$.

Further, we are going to assume that the decay constants are very small compared to the eigenfrequencies, i.e. $\beta_m \ll \tilde{\omega}_m$ [15]. This means that each mode in the sum of (7) is impactful only when the driving frequency ω is very close to the resonance frequency $\tilde{\omega}_m$. This allows us to, for each term in the sum, replace $\tilde{\omega}_m$ with ω in the third term of the denominator. With this,

$$\begin{aligned} G(\mathbf{r}_r, \mathbf{r}_s, \omega) &\approx \sum_{m \in \mathbb{N}} \frac{A_m i\omega}{-\omega^2 + \tilde{\omega}_m^2 + 2i\omega\beta_m - \beta_m^2} \\ &= \sum_{m \in \mathbb{N}} \frac{A_m i\omega}{-\omega^2 + \beta_m^2 + 2i\beta_m\omega + \omega_m^2} \\ &= \sum_{m \in \mathbb{N}} A_m \frac{i\omega + \beta_m}{-\omega^2 + \beta_m^2 + 2i\beta_m\omega + \omega_m^2} \\ &\quad - \sum_{m \in \mathbb{N}} \frac{A_m\beta_m}{\omega_m} \frac{\omega_m}{-\omega^2 + \beta_m^2 + 2i\beta_m\omega + \omega_m^2} \\ &= \sum_{m \in \mathbb{N}} A_m \mathcal{F}\left(\theta(t)e^{-\beta_m t}\cos(\omega_m t)\right) \\ &\quad - \sum_{m \in \mathbb{N}} \frac{A_m\beta_m}{\omega_m} \mathcal{F}\left(\theta(t)e^{-\beta_m t}\sin(\omega_m t)\right) \end{aligned}$$

(8)

where $\theta(t)$ denotes the Heaviside step function, $\mathcal{F}$ the Fourier transform and $\omega_m \triangleq \sqrt{\tilde{\omega}_m^2 - 2\beta_m^2} > 0$ (see also [21] for a time-domain analog). This manoeuvre is something that has been done implicitly before [6], [15], but to the best of the authors' knowledge, this is the first time the necessity of it is made explicit. We see that the adjusted resonance frequency ω_m is a slight shift towards zero, compared to the modal

frequency $\tilde{\omega}_m$. With this, the impulse response $h(t)$ is obtained by taking the inverse Fourier transform of (8), yielding

$$h(t) = \sum_{m \in \mathbb{N}} \theta(t) e^{-\beta_m t} \left(A_m \cos(\omega_m t) - \frac{A_m \beta_m}{\omega_m} \sin(\omega_m t) \right)$$
$$= \sum_{m \in \mathbb{N}} \theta(t) \mu_m e^{-\beta_m t} \cos(\omega_m t + \phi_m) \qquad (9)$$

where $\mu_m = A_m \sqrt{1 + \left(\frac{\beta_m}{\omega_m} \right)^2}$ and $\phi_m = \arctan\left(\frac{\beta_m}{\omega_m} \right)$. This is the continuous-time counterpart of (1).

Considering that the number of modes, N_f, below a certain physical frequency f, satisfies $N_f \approx \frac{4\pi}{3c^3} f^3$ [19], one would expect to have to use a very large number of terms in (9) in order to get a good approximation of the RIR. However, as pointed out in [22], because of the large number of resonant frequencies with overlapping modal curves, the number of distinguishable peaks in the frequency response is in practice considerably smaller than N_f. That is, (9) may be well approximated by the truncated sum

$$h(t) \approx \sum_{m=1}^{M} \mu_m e^{-\beta_m t} \cos(\omega_m t + \phi_m) \theta(t), \qquad (10)$$

where M is on the order of 10^2 [22]. Given the finite number of terms, the approximated RIR in (10) is band-limited [1] and can be sampled, yielding the discrete-time representation in (1), after truncation to N discrete-time samples.

B. Low-Rank Properties

Here we will see how truncation of the modal decomposition (9), at some finite number of terms M, corresponds to finding a rank-$2M$ approximation of the reshaped signal vector.

It is well established, see e.g. [11] and references therein, that when a signal vector consisting of a sum of M causal, exponentially decaying sinusoids, is reshaped into a matrix, that matrix will have rank $2M$. This becomes clear if we decompose $\mathbf{H}$ in the following fashion. Let $\mathbf{H} = \mathbf{H}_1 + \mathbf{H}_2 + \cdots + \mathbf{H}_M$, where $\mathbf{H}_m$ corresponds to the mth decaying sinusoid,

$$\mathbf{H}_m = \begin{bmatrix} h_m(1) & h_m(q+1) & \ldots & h_m(q(q-1)+1) \\ \vdots & \vdots & & \vdots \\ h_m(q) & h_m(2q) & \ldots & h_m(N) \end{bmatrix}$$
$$= \frac{\mu_m e^{i\phi_m}}{2} \begin{bmatrix} z_m \\ z_m^2 \\ \vdots \\ z_m^q \end{bmatrix} \begin{bmatrix} 1 & z_m^q & \cdots & z_m^{(q-1)q} \end{bmatrix} \qquad (11)$$
$$+ \frac{\mu_m e^{-i\phi_m}}{2} \begin{bmatrix} z_m^* \\ z_m^{*2} \\ \vdots \\ z_m^{*q} \end{bmatrix} \begin{bmatrix} 1 & z_m^{*q} & \cdots & z_m^{*(q-1)q} \end{bmatrix}.$$

Here, $z_m = e^{i\omega_m - \beta_m}$ and $*$ denotes complex conjugation. We see that each of the M terms composing $\mathbf{H}$ can be written as a sum of two rank-1 matrices, i.e. as a rank-2 matrix. As long as z_m are distinct, the matrix $\mathbf{H}$ will have rank $2M$ [14]. For this type of signal, this low-rank structure persists also for higher-order tensorization. The rank of a D-dimensional tensor $\mathcal{G} \in \mathbb{R}^{S_1 \times S_2 \times \cdots \times S_D}$, where S_d denotes the size of the dth dimension, is defined as the smallest number of rank-1 tensors that generate the tensor as their sum. Furthermore, a D-dimensional tensor is said to be rank-1 if it can be written as the outer product of D vectors [23]. Consequently, a rank-R tensor $\mathcal{G}$ can be written as

$$\mathcal{G} = \sum_{r=1}^{R} \mathbf{h}_r^{(1)} \circ \mathbf{h}_r^{(2)} \circ \cdots \circ \mathbf{h}_r^{(D)}, \qquad (12)$$

where $\circ$ denotes outer product, and $\mathbf{h}_r^{(d)} \in \mathbb{R}^{S_d}$. Assuming [2] that $\sqrt[D]{N} \in \mathbb{N}$, where N is the length of $\mathbf{h}$, the RIR may be reshaped into a D-dimensional tensor $\mathcal{H} \in \mathbb{R}^{S \times S \times \cdots \times S}$. Throughout this paper we are going to assume that $S_1 = S_2 = \cdots = S_D = S = \sqrt[D]{N}$ and that $\sqrt[D]{N} \in \mathbb{N}$. Then we can reshape the RIR $\mathbf{h}$ into a D-dimensional tensor , that will be of rank $2M$ as well. The upside of this higher-order tensorization is the possibility of further decreasing the number of coefficients needed to represent the RIR, turning what is commonly called the *curse of dimensionality*, into a blessing instead.

A tensor can be decomposed in several different ways. The purpose of this paper is not to showcase the strength and weaknesses of different tensor decompositions, for that the reader is referred to works such as [23] and [24]. The aim is rather to demonstrate that the low-rank structure obtained in (2), with $D = 2$, persists for $D > 2$, and the possible compactness of representation that follows. Therefore, we have chosen to use the polyadic decomposition, where the tensor is written as a sum of rank-one tensors [24]. For the numerical computation of these tensor decompositions we use the high-level function *cpd* of the Matlab toolbox Tensorlab [25].

---

[1] In practice, given that the human hearing is limited to the range of approximately 20 Hz - 20 kHz [19], the impulse response may, using appropriate filtering, be considered band-limited to this interval. This also serves as justification for removing the cavity mode [19], corresponding to $m = 0$, with $\tilde{\omega}_m = 0$.

[2] In general, arbitrary tensor dimensions may be utilized, as they all gives rise to a rank-$2M$ tensor. The square case is considered here, as to simplify the exposition.
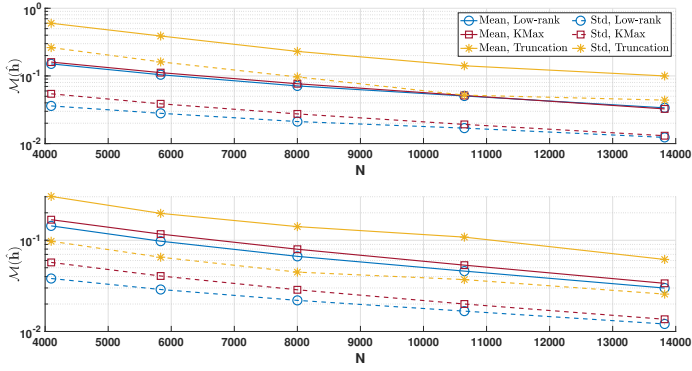
Fig. 1. Averaged normalized misalignments, and corresponding standard deviations, initial delay kept (top), and removed (bottom) respectively, as a function of N.

Finally, it should be noted that the physical motivation given here serves only as a justification of the low-rank modeling employed in this paper. The method described is purely data-driven.

III. NUMERICAL RESULTS

In order to illustrate the usefulness of the considered low-rank framework, and to demonstrate its applicability to actual measured RIRs, we apply it to the *single- and multichannel audio recordings database* (SMARD) [3]. These recordings are sampled at 48 kHz, for 12 seconds, yielding impulse responses of $576 \cdot 10^3$ taps. These are recorded at various source and receiver positions, and with varying equipment, in a room of size 7.34 m $\times 8.09$ m $\times 2.87$ m, with a reverberation time of approximately 0.15 s. In total, the dataset contains 1008 RIRs. It should be noted that in all numerical simulations performed here, the RIRs are truncated at some discrete-time index $N \ll 576 \cdot 10^3$.

In order to evaluate the accuracy of the proposed low-rank model, we consider three different scenarios. Specifically we consider the accuracy of the model as a function of the length N of the RIR, tensorization dimension D, and memory saving. Here, the memory saving $C(\hat{\boldsymbol{h}})$ corresponding to an approximation $\hat{\boldsymbol{h}}$ is defined as

$$C(\hat{\boldsymbol{h}}) = 1 - \frac{\Upsilon(\hat{\boldsymbol{h}})}{N}, \qquad (13)$$

where $\Upsilon(\hat{\boldsymbol{h}})$ is the number of coefficients of $\hat{\boldsymbol{h}}$ and $C(\hat{\boldsymbol{h}}) \in [0, 1)$. For $C(\hat{\boldsymbol{h}}) = 0$ there is no memory saving, whereas for $C(\hat{\boldsymbol{h}})$ closer to 1, the degree of memory saving is larger. For the proposed method, $\Upsilon(\hat{\boldsymbol{h}}) = RD \sqrt[D]{N}$, where R is the rank of the approximation.

In the first scenario, where N is varied, $C(\hat{\boldsymbol{h}})$ and D are kept fixed at 0.8 and 3, respectively. In the second scenario, the tensor dimension D is varied, with $N = 7000$ and $C(\hat{\boldsymbol{h}}) = 0.8$ are being fixed. [3] In the third scenario, $C(\hat{\boldsymbol{h}})$ is varied and $N = 8000$ and $D = 3$ are kept fixed. It may here be noted that in all three scenarios, the tensor rank R varies with N, D, and $C(\hat{\boldsymbol{h}})$, respectively.

[3] In order for $\sqrt[D]{N} \in \mathbb{N}$, N must be adjusted as D is changed
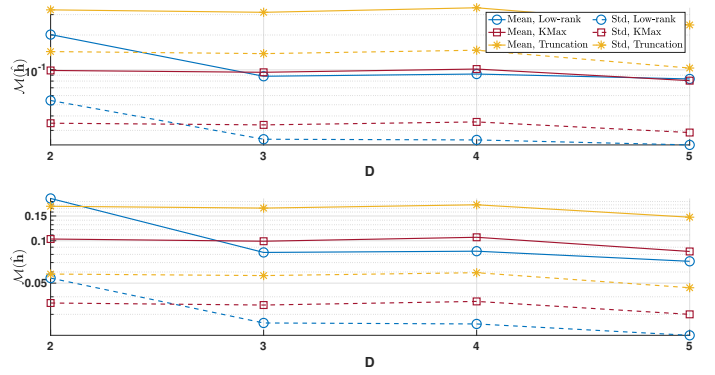

Fig. 2. Averaged normalized misalignments, and corresponding standard deviations, initial delay kept (top), and removed (bottom) respectively, as a function tensorization dimension D.
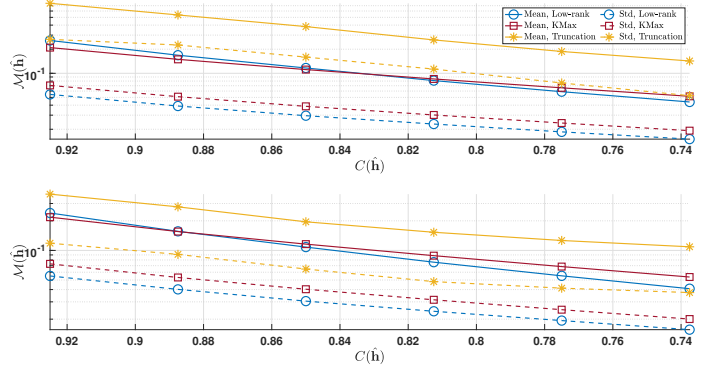

Fig. 3. Averaged normalized misalignments, and corresponding standard deviations, initial delay kept (top), and removed (bottom) respectively, as a function of space saving $C(\hat{\boldsymbol{h}})$.

In all three scenarios, accuracy is measured by the normalized misalignment, defined as

$$\mathcal{M}\left(\hat{\boldsymbol{h}}\right) = \frac{\|\hat{\boldsymbol{h}} - \boldsymbol{h}\|_2}{\|\boldsymbol{h}\|_2}, \qquad (14)$$

where $\boldsymbol{h}$ is the RIR we are aiming to approximate. Furthermore, we consider two different starting points for the RIR; the first one being when the impulse is emitted from the source (i.e. the RIR contains an initial delay), the second one when the direct component arrives at the microphone (i.e. the RIR does not contain an initial delay). The results presented here are averaged over all the 1008 RIRs of the dataset.

For comparison, we consider two alternative approaches to RIR approximation. In the first alternative approach, we consider simply truncating the RIR some finite discrete-time index K, setting the last $N - K$ values to 0. In the second alternative approach, $\boldsymbol{h}$ is approximated by keeping only its K largest entries (in absolute value), and setting the rest to 0. We refer to these approximations as *Truncation* and *KMax* respectively. As to be comparable to the proposed method, K is set equal to the number of coefficients used by the proposed low-rank model, varying with the different scenarios, i.e., all these methods under comparison yields the same value for the memory saving $C(\hat{\boldsymbol{h}})$.

Figure 1 shows the average normalized misalignment and standard deviation for the discussed approximations, as N is

varied. As may be noted, the proposed method outperforms both comparison methods in terms of normalized misalignment when the initial delay is removed, and on par with *KMax* when the initial delay is kept. Furthermore, the standard deviation is lower, indicating a robustness to factors such as distance to walls, and distance between source and receiver.

Figure 2 displays the accuracy of the proposed model as a function of the dimension of the tensorization. The clear drop in normalized misalignment from $D = 2$ to $D = 3$ displays the clear benefit of performing a higher-order tensorization, as opposed to only a matricization. The low standard deviation once again points to robustness to source and receiver placement.

Figure 3 depicts the the proposed method as a function of memory saving. Below a certain compression rate the suggested method outperforms *KMax* and, yet again, the standard deviation is consistently lower.

The systematically lower standard deviation is an encouraging discovery. In light of the derivation of the time-domain signal model from the modal decomposition in the frequency domain, this is to be expected, since the eigenfunctions do not depend on source and receiver position.

It can also be observed that the suggested model performs better when the initial delay of the RIR is removed, particularly in comparison to the performance of *KMax*. This is also reasonable; the delay is ignored by *KMax*, whereas it has to be explicitly handled by the proposed model, preventing it from more accurately modeling the more information-dense parts of the RIR.

IV. CONCLUSIONS AND FUTURE WORK

In this paper we have given a glimpse of the potential of low-rank modeling of room impulse responses, as well as provided a physical justification for the use of low-rank approximations. We have also clarified the importance of certain approximations made throughout the derivation of the RIR model, which in previous works have been vague. We notice that the low-rank modeling clearly outperforms simple truncation and performs on par with using the largest entries in the RIR, in terms of normalized misalignment. The standard deviation, however, is consistently lower, suggesting that the proposed method is robust to the placement of source and receiver throughout the room.

In this paper many design choices had to be made, for the sake of the brevity of the presentation. Future research will focus on how the choice of these parameters, as well as properties of the room, impacts the success of the low-rank model described here. We also intend to further investigate the error introduced by the suggested approximation. If this approximation distorts the output of a convolution in a systematic way, with respect to more perceptually oriented evaluation, that could be an indication of what applications this method would be particularly useful in.

REFERENCES

[1] J. Y. C. Wen, N. D. Gaubitch, E. A. P. Habets, T. Myatt, and P. A. Naylor, "Evaluation of speech dereverberation algorithms using the MARDY database," in *Proc. 2006 Int. Workshop Acoustic Echo Noise Control (IWAENC '06)*, Paris, France, Sept. 2006.

[2] E. Hadad, F. Heese, P. Vary, and S. Gannot, "Multichannel audio database in various acoustic environments," in *Proc. 2014 Int. Workshop Acoustic Signal Enhancement (IWAENC '14)*, Antibes, France, Sept. 2014.

[3] 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)*, Antibes, France, Sept. 2014.

[4] G. W. Elko, E. Diethorn, and T. Gaensler, "Room impulse response variation due to thermal fluctuation and its impact on acoustic echo cancellation," in *Proc. 2003 Int. Workshop Acoustic Echo Noise Control (IWAENC '03)*, Kyoto, Japan, Sept. 2003.

[5] J. S. Abel, S. Coffin, and K. Spratt, "A modal architecture for artificial reverberation with application to room acoustics modeling," in *Preprints AES 137th Convention*, Los Angeles, CA, USA, Oct. 2014, AES Preprint 9208.

[6] 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, 2017.

[7] C. Huszty, N. Bukuli, Á. Torma, and F. Augusztinovicz, "Effects of filtering of room impulse responses on room acoustics parameters by using different filter structures," *J. Acoust. Soc. Amer.*, vol. 123, pp. 3617, 2008.

[8] G. Vairetti, *Efficient parametric modeling, identification and equalization of room acoustics*, Ph.D. thesis, KU Leuven, 2018.

[9] P. A. Naylor and N. D. Gaubitch, *Speech Dereverberation*, Springer, 2010.

[10] M. Udell and A. Townsend, "Why are big data matrices approximately low rank?," *SIAM J. Math. Data Sci.*, vol. 1, no. 1, pp. 144–160, 2019.

[11] M. Boussé, O. Debals, and L. De Lathauwer, "A tensor-based method for large-scale blind source separation using segmentation," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 346–358, 2017.

[12] M. Boussé, O. Debals, and L. De Lathauwer, "Tensor-based large-scale blind system identification using segmentation," *IEEE Trans. Signal Process.*, vol. 65, no. 21, pp. 5770–5784, 2017.

[13] C. Paleologu, J. Benesty, and S. Ciochină, "Linear system identification based on a Kronecker product decomposition," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 10, pp. 1793–1808, 2018.

[14] L. De Lathauwer, "Blind separation of exponential polynomials and the decomposition of a tensor in rank-$(L_r,L_r,1)$ terms," *SIAM J. Matrix Anal. Appl.*, vol. 32, no. 4, pp. 1451–1474, 2011.

[15] H. Kuttruff, *Room Acoustics*, Spon Press, London, 2009.

[16] R. Mignot, G. Chardon, and L. Daudet, "Low frequency interpolation of room impulse responses using compressed sensing," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 1, pp. 205–216, 2014.

[17] T. P. Vu and H. Lissek, "Sound field reconstruction in a room using low-rank approximation of mode shapes: validation and application," in *Proc. 26th Int. Congr. Sound Vib. (ICSV '19)*, Montréal, QC, Canada, July 2019.

[18] H. P. Tukuljac, T. P. Vu, H. Lissek, and P. Vandergheynst, "Joint estimation of the room geometry and modes with compressed sensing," in *Proc. 2018 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '18)*, Calgary, AB, Canada, Apr. 2018, pp. 6882–6886.

[19] F. Jacobsen and P. M. Juhl, *Fundamentals of General Linear Acoustics*, Wiley, 2013.

[20] 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, 1994.

[21] S. Bilbao, *Numerical Sound Synthesis*, Wiley, 2009.

[22] L. S. H. Ngia, "Recursive identification of acoustic echo systems using orthonormal basis functions," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 3, pp. 278–293, 2003.

[23] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[24] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, 2015.

[25] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, "Tensorlab 3.0," Mar. 2016, Available online. https://www.tensorlab.net.