forum acusticum 2026
8-12 September · Graz, Austria

STATE-SPACE MODELLING OF HEAD-RELATED TRANSFER FUNCTIONS

Art J. R. Pelling[1], Ennes Sarradj[1]

[1] *Department of Engineering Acoustics, TU Berlin, Einsteinufer 25, 10587 Berlin, Germany*

Abstract

State-space models can provide an efficient alternative to convolution-based rendering of head-related transfer functions (HRTFs), but directly identifying reduced state-space models from densely sampled datasets can be computationally infeasible. We present a loss-free projected Eigensystem Realization Algorithm (ERA) formulation for multiple-input multiple-output HRTF datasets, in which sampled directions are inputs and ears are outputs. The projection basis preserves the nonzero Hankel singular values exactly and therefore adds no approximation beyond dynamical order reduction. For finite impulse-response approximations, we prove that both the McMillan degree and the loss-free projection dimension are bounded by the product of the response length and number of outputs, independently of the number of sampled directions, and derive a projection-aware ERA error bound. For the 11 950-input, 256-sample FABIAN dataset, the input projection to 510 virtual inputs reduces the memory requirements of the economy-SVD ERA-step from 23.2 GiB to approximately 1.0 GiB, enabling computation of reduced-order state-space models on virtually any platform. We further show that interaural-time-delay removal accelerates singular-value decay and can greatly improve quality of identified models.

Keywords: *model order reduction, balanced truncation, head-related transfer function, auralization, data-driven modelling*

1. INTRODUCTION

Immersive virtual acoustic rendering requires listener-specific head-related transfer functions (HRTFs) that remain accurate as source and listener orientations change [1, 2]. Direction-dependent spectral changes and interaural time delays (ITDs) support sound source localization [2, 3]. HRTF models must therefore accommodate real-time source and head motion with adequate directional resolution [1].

Virtual acoustic rendering commonly selects directionally sampled head-related impulse responses (HRIRs) in the form of finite impulse response (FIR) filters and convolves them either in the time or frequency domain, e.g. using the uniformly partitioned overlap-save (UPOLS) method [4]. State-space models provide a less widespread alternative that can reduce real-time multiple-input multiple-output (MIMO) rendering cost, especially after order reduction [1, 5–7].

Balanced truncation and Eigensystem Realization Algorithm (ERA) have produced reduced order state-space HRTF models for individual responses and multi-direction datasets [1, 3, 5, 6, 8, 9]. They use a truncated singular value decomposition (SVD) of the Hankel matrix, whose cost grows with the number of measured directions and can become problematic for densely sampled HRTF datasets [10, 11]. Tangential projection, CUR factorization, and randomized SVD can reduce this cost by trading approximation quality for efficiency [7, 10–13]. Here, we exploit the special structure of HRTF datasets to compute a loss-free projected factorization that exactly preserves the Hankel singular values at a greatly reduced cost.

This paper identifies reduced-order MIMO state-space models from densely sampled HRTFs, with measurement directions as inputs and ears as outputs. We call ERA applied to the full, unprojected Hankel matrix direct ERA; loss-free projected ERA instead uses an exact virtual input basis without tangential truncation, reducing only the model's dynamical behaviour, not its spatial fidelity. We prove bounds on the McMillan degree, loss-free virtual input dimension, and reduced-model error, and evaluate the method using the FABIAN database [14]. We also compare original-delay and ITD-removed HRIRs to assess how delay handling affects singular-value decay and modelling error.

Corresponding author: *a.pelling@tu-berlin.de.*

2. PROJECTED ERA for HRTFs

In the following, we will consider MIMO discrete-time state-space systems with $m \in \mathbb{N}$ inputs and $p \in \mathbb{N}$ outputs. Such a system can be described by a so-called *realization* given by the matrix quadruple $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ with $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times m}$, $\mathbf{C} \in \mathbb{R}^{p\times n}$, and $\mathbf{D} \in \mathbb{R}^{p\times m}$, where $n \in \mathbb{N}$ is the order of the system. Its dynamics are governed by the discrete-time state equations

$$\begin{aligned} x_{k+1} &= \mathbf{A}x_k + \mathbf{B}u_k, \\ y_k &= \mathbf{C}x_k + \mathbf{D}u_k, \end{aligned}$$

for a causal input $u \in \ell_2^m$, output $y \in \ell_2^p$ and state $x \in \ell_2^n$, where ℓ_2^d denotes the square-summable sequences of d-dimensional vectors and $k \in \mathbb{N}_0$ is the discrete-time index. Assuming a zero initial state, a stable realization, and $u_k = 0$ for $k < 0$, the input-output behaviour is fully described by $y_k = \sum_{\ell\in\mathbb{N}_0} h_\ell u_{k-\ell}$, where the Markov parameters are

$$h_k = \begin{cases} \mathbf{D}, & k = 0, \\ \mathbf{C}\mathbf{A}^{k-1}\mathbf{B}, & k \geq 1. \end{cases} \tag{1}$$

Equivalently, in the frequency domain, the output $\mathbf{Y}(z) \in \mathbb{C}^p$ is given by a multiplication of the input $\mathbf{U}(z) \in \mathbb{C}^m$ with the transfer function

$$\mathbf{H}(z) = \mathbf{C}(z\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D},$$

where z is the frequency and $\mathbf{I}_n \in \mathbb{R}^{n\times n}$ is the identity.

For the HRTF dataset considered below, each measured source direction is an input and the two ears are outputs. Thus, one realization shares a state matrix across every direction-ear transfer function, making it a common-acoustical-pole (CAP) model in the representational sense of [15].

2.1. ERA

In discrete time, the impulse response is the sequence of Markov parameters (1). This relationship allows the identification of a matching state-space model from an impulse response sequence $(h_0, h_1, \ldots, h_{2s-1})$, $s \in \mathbb{N}$, via the Hankel matrix of Markov parameters

$$\mathcal{H} = \begin{bmatrix} \mathbf{CB} & \mathbf{CAB} & \cdots & \mathbf{CA}^{s-1}\mathbf{B} \\ \mathbf{CAB} & \mathbf{CA}^2\mathbf{B} & \cdots & \mathbf{CA}^{s}\mathbf{B} \\ \vdots & \vdots & \iddots & \vdots \\ \mathbf{CA}^{s-1}\mathbf{B} & \mathbf{CA}^{s}\mathbf{B} & \cdots & \mathbf{CA}^{2s-2}\mathbf{B} \end{bmatrix}. \tag{2}$$

The Hankel matrix can be factored into the observability and controllability matrices $\mathcal{O} \in \mathbb{R}^{ps\times n}$ and $\mathcal{C} \in \mathbb{R}^{n\times ms}$:

$$\mathcal{H} = \underbrace{\begin{bmatrix} \mathbf{C} \\ \vdots \\ \mathbf{CA}^{s-1} \end{bmatrix}}_{=:\mathcal{O}} \underbrace{\begin{bmatrix} \mathbf{B} & \cdots & \mathbf{A}^{s-1}\mathbf{B} \end{bmatrix}}_{=:\mathcal{C}} \in \mathbb{R}^{ps\times ms}.$$

From this factorization, a realization can be constructed via

$$\mathbf{A} = \mathcal{O}_f^\dagger \mathcal{O}_l,\ \mathbf{B} = \mathcal{C}\begin{bmatrix} \mathbf{I}_m \\ 0 \end{bmatrix},\ \mathbf{C} = [\mathbf{I}_p\ \ 0]\mathcal{O},\ \mathbf{D} = h_0, \tag{3}$$

where $\mathcal{O}_f$ and $\mathcal{O}_l$ denote the first and last $p(s-1)$ rows of the observability matrix $\mathcal{O} \in \mathbb{R}^{ps\times n}$, respectively, and $\dagger$ denotes the Moore–Penrose pseudoinverse [16, 17]. Classically, the factorization is computed from a (truncated) SVD, i.e., for a truncation order $r \leq \min\{ms, ps\}$ write the SVD as

$$\mathcal{H} = \begin{bmatrix} \mathbf{U}_r & * \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_r & 0 \\ 0 & * \end{bmatrix} \begin{bmatrix} \mathbf{V}_r^\top \\ * \end{bmatrix},$$

where $\mathbf{U}_r \in \mathbb{R}^{ps\times r}$ comprises the first r left singular vectors as columns, $\mathbf{\Sigma}_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$ the first r singular values and $\mathbf{V}_r \in \mathbb{R}^{ms\times r}$ the first r right singular vectors as columns. By choosing

$$\mathcal{O} = \mathbf{U}_r\mathbf{\Sigma}_r^{1/2} \quad \text{and} \quad \mathcal{C} = \mathbf{\Sigma}_r^{1/2}\mathbf{V}_r^\top, \tag{4}$$

the resulting realization is approximately internally balanced [16, 18]. Assuming that the impulse response has decayed to negligible magnitude by index s, i.e., $h_k \approx 0$ for $k > s$, the resulting realization is stable [10, 13, 16]. Additionally, the realization given by (3), (4) satisfies the a priori error bound

$$\sum_{k=1}^{2s-1} \|\mathbf{C}_r\mathbf{A}_r^{k-1}\mathbf{B}_r - h_k\|_{\mathrm{F}}^2 \leq (r+m+p)\cdot\sigma_{r+1}(\mathcal{H})^2, \tag{5}$$

where $\sigma_{r+1}(\mathcal{H})$ is the first neglected Hankel singular value and $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm [13, 16].

2.2. Loss-Free Projected ERA

By Section 2.1, we can construct reduced order models from HRIR datasets[1]. Recall that $\mathcal{H} \in \mathbb{R}^{ps\times ms}$. For datasets with many input directions, m is large, which can make the block-Hankel matrix infeasibly wide for a direct SVD factorization. A natural approach would be to apply tangential projections with a reduced number of inputs in the Tangential Eigensystem Realization Algorithm (TERA) framework [10]. As mentioned, this can accelerate the computations albeit yielding a suboptimal reduced order model. As we will show now, the special structure of HRIRs allows us to omit tangential truncation and use an exact basis for the row space of $\mathcal{H}$ for projection instead.

HRIRs typically decay within a few milliseconds, so it is common practice to consider FIR approximations where the HRIR is set to zero after some time index s. The following theorem bounds the McMillan degree of an FIR approximation with s nonzero samples:

Theorem 2.1 (McMillan degree bound)**.** *Let $h_k \in \mathbb{R}^{p\times m}$, $k = 0, \ldots, s$, denote an FIR-approximated impulse response sequence with $h_k = 0$ for every $k > s$.*

[1] HRTFs with equidistant frequency bins can be transformed into HRIRs with an FFT.

Then, the rational transfer function given by $\mathbf{H}_s(z) = \sum_{k=0}^{s} h_k z^{-k}$ *has McMillan degree at most* ps. *In particular, for a MIMO-HRTF,* $p = 2$ *(the two ears), so its McMillan degree is at most* $2s$*, independently of the number* m *of sampled directions.*

Proof. Let $\mathcal{H}_\infty = [h_{i+j-1}]_{i,j\geq 1}$ be the infinite Hankel matrix of $\mathbf{H}_s$. Because the sequence is zero for $k > s$, $\mathcal{H}_\infty$ is zero outside its leading $s \times s$ block, which is $\mathcal{H}$ in (2). Thus $\operatorname{rank}(\mathcal{H}_\infty) = \operatorname{rank}(\mathcal{H}) \leq ps$. The McMillan degree of a rational function equals the rank of its infinite Hankel matrix, proving the result. □

This bound is a property of finite HRIR, independent of its realization. Thus, it applies whether or not a state-space model is constructed. Since (2) captures the Hankel singular values of an s-sample MIMO-HRIR entirely, retaining its $r = \operatorname{rank}(\mathcal{H}) \leq 2s$ nonzero singular components in ERA exactly realizes the corresponding zero-extended HRTF. Thus, $2s$ is a worst-case order bound rather than necessarily the minimal order. Following input-side TERA [10], let

$$\Theta_R = \begin{bmatrix} h_1^\top & \cdots & h_s^\top \end{bmatrix}^\top = \bar{\mathbf{U}}\bar{\mathbf{\Sigma}}\bar{\mathbf{V}}^\top \in \mathbb{R}^{ps\times m}$$

be its economy SVD[2]. The direct term is retained without projection, so $\mathbf{D}_r = \mathbf{D} = h_0$. Unlike standard TERA, we omit tangential truncation: with $q = \operatorname{rank}(\Theta_R)$, set $\mathbf{W}_2 = \bar{\mathbf{V}} \in \mathbb{R}^{m\times q}$, $\mathbf{P}_2 = \mathbf{W}_2\mathbf{W}_2^\top$, and define the projected Markov parameters $\hat{h}_k = h_k\mathbf{W}_2$.

Theorem 2.2 (Loss-free virtual inputs)**.** *Assume* $h_k = 0$ *for every* $k > s$. *Then,* $q \leq \operatorname{rank}(\mathcal{H}) \leq ps$ *and* $h_k = h_k\mathbf{P}_2$ *for every* $k \geq 1$. *Consequently, the projected Hankel matrix constructed from* $\hat{h}_k = h_k\mathbf{W}_2$ *has the same nonzero Hankel singular values as (2). For every reduced order, the Markov parameters of the projected and unprojected ERA models are identical.*

Proof. Θ_R is the first block column of $\mathcal{H}$, and $\mathbf{W}_2$ spans range$(\Theta_R^\top)$; hence $q \leq \operatorname{rank}(\mathcal{H})$ and $h_k = h_k\mathbf{P}_2$. With $\mathcal{W} = \mathbf{I}_s \otimes \mathbf{W}_2$, $\mathcal{H} = \mathcal{H}\mathcal{W}\mathcal{W}^\top$, so $\hat{\mathcal{H}} = \mathcal{H}\mathcal{W}$ has the same nonzero singular values. Corresponding truncated SVDs differ only in their input coordinates; thus (4) gives the same reduced Markov parameters after backprojection. □

Theorems 2.1 and 2.2 give a projected ERA workflow that replaces a large SVD of size $(2s \times ms)$ by two smaller SVDs of sizes $(2s \times m)$ and $(2s \times qs)$ that can lead to substantial gains when $q \ll m$: First, compute $\mathbf{P}_2$ from the measured data, identify a reduced realization $(\mathbf{A}_r, \hat{\mathbf{B}}_r, \mathbf{C}_r, h_0)$ from $\hat{h}$ with ERA, then backproject by $\mathbf{B}_r = \hat{\mathbf{B}}_r\mathbf{W}_2^\top$. The factor $\mathbf{W}_2$ of $\mathbf{P}_2$ maps a physical input direction e_j to virtual input coefficients $\mathbf{W}_2^\top e_j$, encoding each virtual input as a linear combination of input directions. For a HRTF dataset, $q \leq 2s$, even when additional measured directions are added, so this strategy can be applied to HRTF datasets of practically any level of spatial fidelity. Finally, the previous rank considerations can be used to sharpen the a priori error bound for ERA:

[2] An economy-size/rank-revealing LQ factorization can also be used here.

Theorem 2.3 (Projection-aware error bound)**.** *Under the assumptions of Theorem 2.2, let* $(\mathbf{A}_r, \hat{\mathbf{B}}_r, \mathbf{C}_r, h_0)$ *be the order-*r *ERA realization of* $\hat{h}$. *Its backprojected realization* $(\mathbf{A}_r, \hat{\mathbf{B}}_r\mathbf{W}_2^\top, \mathbf{C}_r, h_0)$ *satisfies*

$$\sum_{k=1}^{2s-1} \|\mathbf{C}_r\mathbf{A}_r^{k-1}\hat{\mathbf{B}}_r\mathbf{W}_2^\top - h_k\|_{\mathrm{F}}^2 \leq (r+q+p)\cdot\sigma_{r+1}(\mathcal{H})^2.$$

Proof. Apply (5) to $\hat{h}$, which has q inputs. By Theorem 2.2, the singular values agree and backprojection preserves the error norm because $\mathbf{W}_2$ has orthonormal columns. □

3. EXPERIMENTAL SETUP

3.1. FABIAN Dataset and Preprocessing

We use the 256-sample, 44 100 Hz HRIRs for the neutral HATO in the FABIAN HRTF database [14] with $m = 11\,950$ source directions at 2° elevation and azimuth increments. We treat them as $h_0, \ldots, h_{255}$ of a MIMO discrete-time system with $p = 2$ outputs (the two ears), so $s = 255$ non-direct Markov parameters are $h_1, \ldots, h_{255}$. We zero-extend the sequence by $h_k = 0$ for $k > s$. Let $h^{(i,j)} \in \mathbb{R}^{2s}$ denote the resulting HRIR from the i-th source to the j-th receiver. The k-th Markov parameter contains the k-th sample of every HRIR, i.e. $\{h_k\}_{i,j} = h_k^{(i,j)} \in \mathbb{R}^{p\times m}$. Before system identification, the HRIRs are normalized so that the mean energy of single-channel HRTFs is one.

3.2. Delay Handling

We use two delay-handling conditions. *Original-delay* HRIRs are normalized, zero-extended responses without shifts. For *ITD-removed* HRIRs, we estimate each response onset as the last sample below 20 dB relative to its maximum and shift the response by this estimate[3]. For the FABIAN dataset, this retains $s = 237$ nonzero samples.

3.3. Implementation and Timing

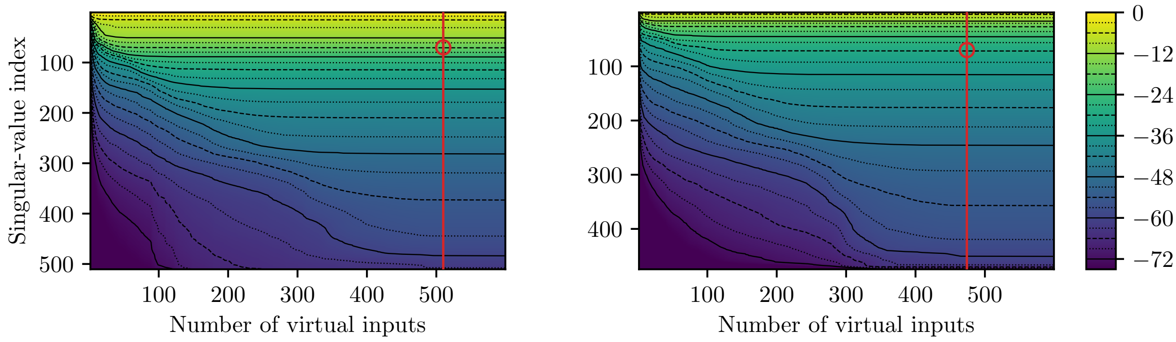
Timings cover HRIR preprocessing, projected ERA factorization, and reduced-realization construction, excluding data loading. Three uncached trials per case ran on an AMD Ryzen 7 8845HS laptop (8 cores, 16 threads; 14 GiB memory) under Arch Linux, Python 3.13, NumPy 2.4.6, and SciPy 1.18.0 with 16 OpenBLAS threads.

4. RESULTS

4.1. Loss-Free Virtual Input Projection

For the original-delay FABIAN responses, the number of loss-free virtual inputs is 510; after individual ITD removal it is $2\cdot 237 = 474$. Figure 1 depicts the Hankel singular values decay for a varying (truncating) number of virtual inputs in the TERA framework. The red

[3] See `find_impulse_response_start` (pyFAR v0.8.0).



(a) Original-delay HRIRs; colors show singular values.

(b) After individual ITD removal.

Figure 1: Hankel singular-value maps for neutral-orientation FABIAN HRIRs. Colors are dB relative to the maximum; red lines mark loss-free virtual input caps and circles $r = 70$.

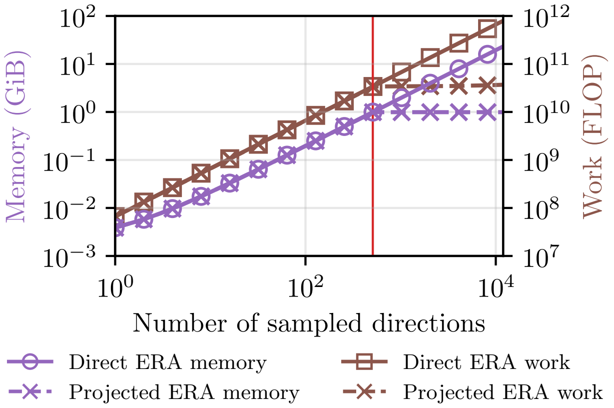


Figure 2: Economy-SVD memory and full-SVD work versus sampled directions, including dense Hankel and economy-size factors. The red line marks the loss-free virtual input cap $q = 510$.

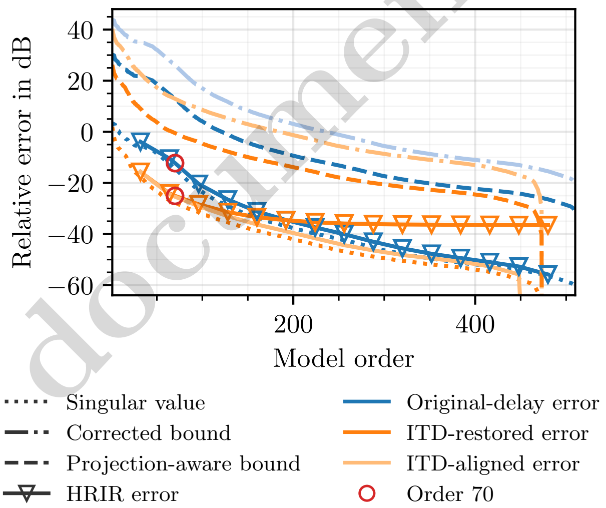


Figure 3: Hankel singular-value decay, a priori bounds, and relative HRIR reconstruction errors for neutral-orientation FABIAN data. Colors distinguish delay handling; line styles distinguish quantities; and circles mark restored-response errors at $r = 70$; the dark curve uses the aligned reference. For ITD-removed responses, the bounds apply to the aligned system, so comparison with restored-response errors is empirical.

vertical lines mark the respective values for q based on Theorem 2.2 at which the input projection becomes loss-free. To the right of these, the contours remain horizontal, showing that further virtual inputs leave the Hankel singular values unchanged. Several contours are already nearly horizontal to the left of these lines, suggesting that a lossy tangential truncation could further reduce the input dimension for a chosen accuracy target. We stick to q based on Theorem 2.2 because the computational gains are already sufficient and an exact preservation of the HRTFs is preferable. Figure 2 shows that direct economy-SVD storage would require 23.2 GiB. Capping the virtual input dimension at $q = 510$ reduces storage to approximately 1.0 GiB and full-SVD work by a factor of about 23. These savings are theoretical, not measured speedups, because the direct factorization exceeded the available memory of our platform.

4.2. Effect of ITD-Alignment

Figure 3 compares singular-value decay, the two a priori bounds, and measured relative HRIR reconstruction errors. As can be seen from Figs. 1 and 3, individual ITD removal produces a faster singular value decay and reduces the relative error at a fixed reduced order. However, the relative error of ITD-restored reduced order models does not decrease below −36.5 dB. At $r = 224$, the ITD-restored models perform worse than the original-delay models. A further increase of model order does not improve the error substantially. This, however, does not stem from the projected ERA routine but is due to the truncation of nonzero values of the HRIR when removing the estimated ITDs. The ITD-aligned error also depicted in Fig. 3 continues to decay to −281.7 dB at order 480, showing that the high-order plateau is not caused by a lack of convergence against the aligned reference. An increase in model order does in fact improve the approximation of the ITD-removed HRIR but at a certain point, the error in truncation during removal becomes larger than the error in the model reduction.

Earlier HRTF work likewise found that alignment or phase correction can reduce effective spatial order [19], and that ITD can degrade reduced state-space models [1, 13]. The present result is consistent with these observations, but it does not establish a causal mechanism.

4.3. Order Selection

The common reduced order $r = 70$ was selected as the lowest order whose original-delay reconstruction error is below $-12\,\mathrm{dB}$ ($-12.3\,\mathrm{dB}$). This threshold is set arbitrarily and not directly motivated by a perceptual study. Earlier HRTF state-space studies reported and perceptual order thresholds between 7 and 24 for an array of 50 HRTFs [6], but these conditions are not directly comparable with the present 11 950-direction dataset. The literature therefore only contextualizes the order scale; perceptual and localization validation is still required.

Figure 4 examines the individual left and right responses at azimuth 45° and elevation 15°. The time-domain panel compares the measured responses with the delay-restored, ITD-removed order-70 model. The magnitude panel compares the measured responses, the original-delay and ITD-removed models, and the corresponding ear-specific errors. Thus, the individual-response comparison complements the aggregate errors in Fig. 3 at the selected common order.

5. CONCLUSIONS

We presented a projected ERA formulation for densely sampled MIMO HRTFs. For finite length-s FIR-HRIRs, the McMillan degree and loss-free virtual input dimension are bounded by $2s$, independently of the number of sampled directions. This data-driven bound preserves the nonzero Hankel singular values and can also guide investigation of other Hankel- or SVD-based HRTF representations.

For the 11 950-direction FABIAN dataset, loss-free projection reduces estimated economy SVD storage from 23.2 GiB to approximately 1.0 GiB, lifting the memory limit of the evaluated laptop platform. The projection-aware error bound is correspondingly tighter than the classical ERA bound because it uses $q = 510$ rather than all $m = 11\,950$ input directions. Individual ITD removal accelerated singular-value decay and reduced the relative reconstruction error from $-12.3\,\mathrm{dB}$ to $-25.1\,\mathrm{dB}$. The study considers one neutral head-above-torso orientation (HATO) of the FABIAN head. Future work should test further HATOs and directional samplings, assess perceptual and localization implications, and evaluate dead-time-splitting extraction [13] for improved real-time rendering.

Source Code Availability

The source code used to compute the results is available at `doi:10.5281/zenodo.21727073` under the MIT license.

Acknowledgements

The work of Art J. R. Pelling was funded by DFG (German Research Foundation) – project no.: 504367810.

References

[1] N. H. Adams and G. H. Wakefield, "State-Space Synthesis of Virtual Auditory Space," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 5, pp. 881–890, Jul. 2008. DOI: `10.1109/TASL.2008.924151`

[2] H. Møller, "Fundamentals of binaural technology," *Applied Acoustics*, vol. 36, no. 3, pp. 171–218, Jan. 1992. DOI: `10.1016/0003-682X(92)90046-U`

[3] P. Georgiou and C. Kyriakakis, "Modeling of head related transfer functions for immersive audio using a state-space approach," in *Conference Record of the Thirty-Third Asilomar Conference on Signals, Systems, and Computers (Cat. No.CH37020)*, vol. 1, Pacific Grove, CA, USA: IEEE, 1999, pp. 720–724, ISBN: 978-0-7803-5700-6.

[4] F. Wefers, *Partitioned Convolution Algorithms for Real-Time Auralization* (Aachener Beiträge Zur Technischen Akustik Band 20). Berlin: Logos Verlag Berlin GmbH, 2015, ISBN: 978-3-8325-3943-6.

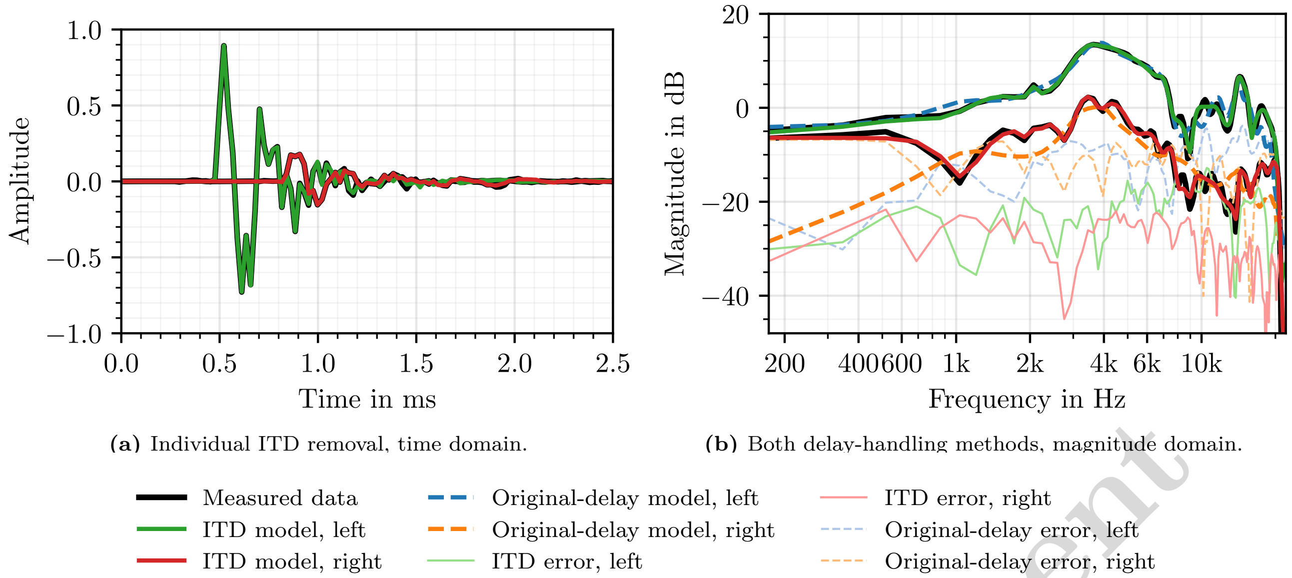
[5] N. Adams and G. Wakefield, "Efficient Binaural Display using MIMO State-Space Systems," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 1, Apr. 2007, pp. I-169–I–172. DOI: `10.1109/ICASSP.2007.366643`

[6] N. H. Adams and G. H. Wakefield, "State-space models of head-related transfer functions for virtual auditory scene synthesis," *The Journal of the Acoustical Society of America*, vol. 125, no. 6, pp. 3894–3902, Jun. 2009. DOI: `10.1121/1.3124778`

[7] A. J. R. Pelling and E. Sarradj, "Efficient Forced Response Computations of Acoustical Systems with a State-Space Approach," *Acoustics*, vol. 3, no. 3, pp. 581–593, 2021. DOI: `10.3390/acoustics3030037`

[8] J. Mackenzie, J. Huopaniemi, V. Valimaki, and I. Kale, "Low-order modeling of head-related transfer functions using balanced model truncation," *IEEE Signal Processing Letters*, vol. 4, no. 2, pp. 39–41, Feb. 1997. DOI: `10.1109/97.554467`

[9] D. W. Grantham, J. A. Willhite, K. D. Frampton, and D. H. Ashmead, "Reduced order modeling of head related impulse responses for virtual acoustic displays," *The Journal of the Acoustical Society of America*, vol. 117, no. 5, pp. 3116–3125, Apr. 2005. DOI: `10.1121/1.1882944`



(a) Individual ITD removal, time domain.

(b) Both delay-handling methods, magnitude domain.

Figure 4: Measured and order-70 FABIAN responses at azimuth 45° and elevation 15°: (a) the individually ITD-removed model in time and (b) both delay-handling methods in magnitude. Black curves are measurements; normal and lighter shades identify reduced models and errors; dashed blue/orange curves are original-delay left/right results.

[10] B. Kramer and S. Gugercin, "Tangential interpolation-based eigensystem realization algorithm for MIMO systems," *Mathematical and Computer Modelling of Dynamical Systems*, vol. 22, no. 4, pp. 282–306, Jul. 2016. DOI: 10.1080/13873954.2016.1198389

[11] R. Minster, A. K. Saibaba, J. Kar, and A. Chakrabortty, "Efficient Algorithms for Eigensystem Realization Using Randomized SVD," *SIAM Journal on Matrix Analysis and Applications*, vol. 42, no. 2, pp. 1045–1072, Jan. 2021. DOI: 10.1137/20M1327616

[12] B. Kramer and A. A. Gorodetsky, "System identification via CUR-factored hankel approximation," *SIAM Journal on Scientific Computing*, vol. 40, no. 2, A848–A866, Jan. 2018. DOI: 10.1137/17M1137632

[13] A. J. R. Pelling and E. Sarradj, "Adaptive reduced order modelling of acoustic LTI systems with input–output dead time," *Mechanical Systems and Signal Processing*, vol. 242, p. 113 613, Jan. 2026. DOI: 10.1016/j.ymssp.2025.113613

[14] F. Brinkmann et al., "A High Resolution and Full-Spherical Head-Related Transfer Function Database for Different Head-Above-Torso Orientations," *Journal of the Audio Engineering Society*, vol. 65, no. 10, pp. 841–848, Oct. 2017. DOI: 10.17743/jaes.2017.0033

[15] Y. Haneda, S. Makino, Y. Kaneda, and N. Kitawaki, "Common-acoustical-pole and zero modeling of head-related transfer functions," *Speech and Audio Processing, IEEE Transactions on*, vol. 7, pp. 188–196, Apr. 1999. DOI: 10.1109/89.748123

[16] S. Kung, "A new identification and model reduction algorithm via singular value decomposition," in *Proceedings of the 12th Asilomar Conference on Circuits, Systems and Computers*, 1978, pp. 705–714.

[17] J.-N. Juang and R. S. Pappa, "An eigensystem realization algorithm for modal parameter identification and model reduction," *Journal of Guidance, Control, and Dynamics*, vol. 8, no. 5, pp. 620–627, Sep. 1985. DOI: 10.2514/3.20031

[18] B. D. Schutter, "Minimal state-space realization in linear system theory: An overview," *Journal of Computational and Applied Mathematics*, vol. 121, no. 1–2, pp. 331–354, Sep. 2000. DOI: 10.1016/S0377-0427(00)00341-1

[19] Z. Ben-Hur, D. L. Alon, R. Mehra, and B. Rafaely, "Efficient Representation and Sparse Sampling of Head-Related Transfer Functions Using Phase-Correction Based on Ear Alignment," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 2249–2262, Nov. 2019. DOI: 10.1109/TASLP.2019.2945479