Towards Faster Continuous Multi-Channel HRTF Measurements Based on Learning System Models

iKS Institute of Communication Systems Prof. Dr.-Ing. Peter Jax
RWTH AACHEN UNIVERSITY

Tobias Kabzinski and Peter Jax

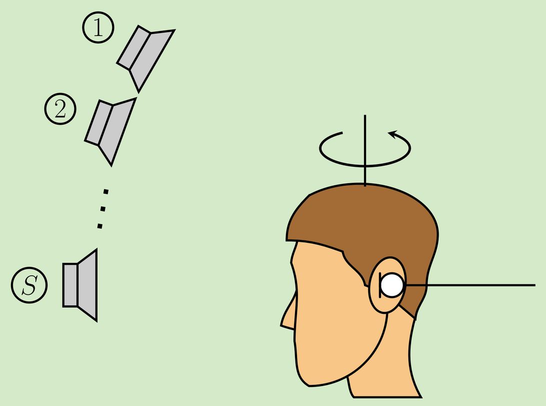
1 Introduction

Head-related transfer functions (HRTFs) / impulse responses (HRIRs):

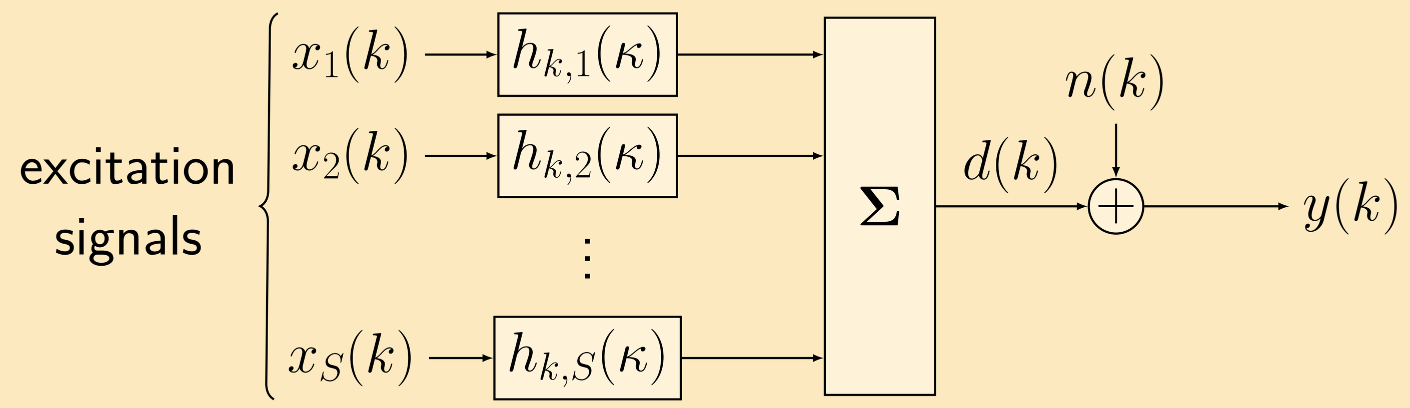
- Describe acoustic transmission between sound source and human ear
- Applications in binaural audio
- Individualization beneficial: binaural synthesis, crosstalk cancellation

Measurement of HRTFs:

- Two approaches: stop-and-go or continuous
- Goal: **faster continuous** measurements
  - Learning a system model
  - Offline processing



2 System and Signal Model



Task: estimate S time-variant HRIRs $h_{k,s}$ of length $\tilde{L}$ at time k from excitation signals $x_s(k)$ and noisy microphone signal

$$y(k) = \sum_{s=1}^{S} \sum_{\kappa=0}^{\tilde{L}-1} x_s(k-\kappa)\, h_{k,s}(\kappa) + n(k)$$

▶ Traditional solution: NLMS algorithm [1]

State-space problem

$\mathbf{z}_k = [\mathbf{h}_{k,1}^{\mathrm{T}}, \ldots, \mathbf{h}_{k,S}^{\mathrm{T}}]^{\mathrm{T}}$ $\mathbf{q}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma})$

$$\boxed{\begin{aligned} \mathbf{z}_k &= \mathbf{A}\mathbf{z}_{k-1} + \mathbf{q}_k && \text{(state equation)} \\ y(k) &= \mathbf{C}_k \mathbf{z}_k + n(k) && \text{(observation equation)} \end{aligned}}$$

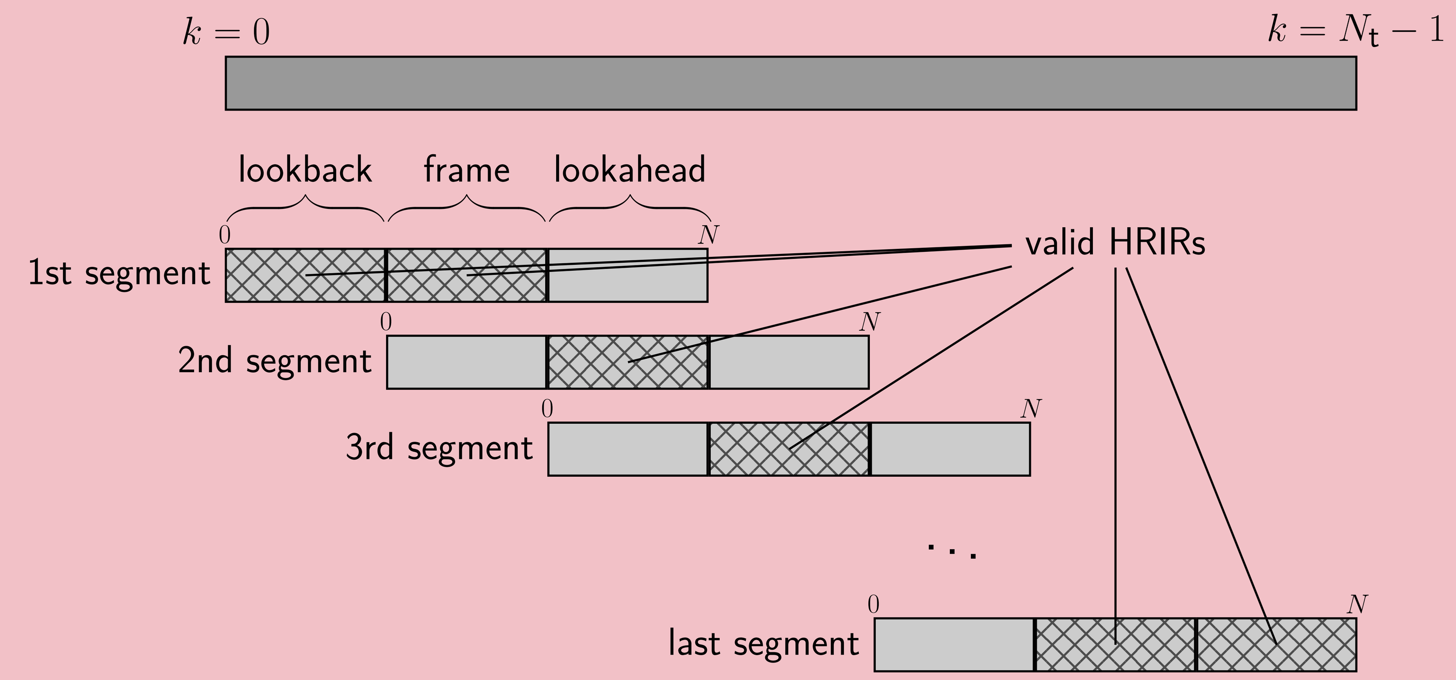
observation matrix $\mathbf{C}_k$ depends on excitation signals $n(k) \sim \mathcal{N}(0, \sigma_n^2)$

- State estimation requires parameters $\boldsymbol{\theta} = \{\mathbf{A}, \boldsymbol{\Gamma}, \sigma_n^2, \boldsymbol{\mu}_0, \mathbf{P}_0\}$, e.g.,
  - Initial state $\boldsymbol{\mu}_0$, initial error covariance $\mathbf{P}_0$ heuristically chosen
  - Scalar fading factor $\mathbf{A} = \gamma \mathbf{I}$ with $0 \ll \gamma \leq 1$
  - Online-estimate of diagonal process noise covariance $\boldsymbol{\Gamma}$ and σ_n^2
- Kalman *Filter*: $\boldsymbol{\mu}_k = \mathbb{E}\left\{ \mathbf{z}_k \middle| \{\mathbf{x}_\ell\}_{\ell=0}^{k}, \{y(\ell)\}_{\ell=0}^{k} \right\}$

3 Proposed Measurement Method

1. Idea: exploit that **future samples** are available (offline processing)

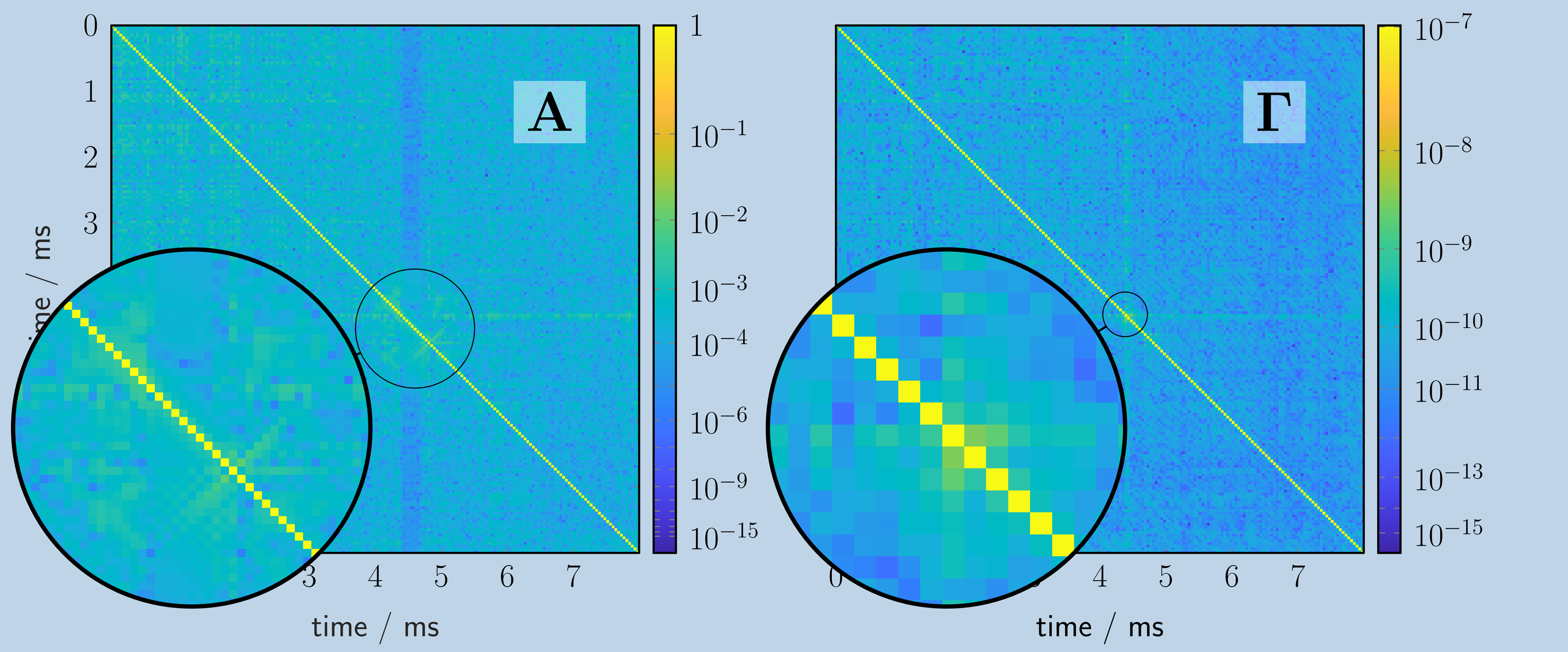
   → Kalman *Smoother*: $\hat{\boldsymbol{\mu}}_k = \mathbb{E}\left\{ \mathbf{z}_k \middle| \{\mathbf{x}_\ell\}_{\ell=0}^{N}, \{y(\ell)\}_{\ell=0}^{N} \right\}$
2. Idea: **learn parameters** $\boldsymbol{\theta} = \{\mathbf{A}, \boldsymbol{\Gamma}, \sigma_n^2, \boldsymbol{\mu}_0, \mathbf{P}_0\}$ with Expectation Maximization (EM) algorithm [2]
   - Generic state transition matrix $\mathbf{A}$ → better state prediction
   - E-step: compute expected log-likelihood } repeat for $\mathcal{I}$ iterations
   - M-step: update parameters
   - Memory problem: sampling rate of 24 kHz, $S = 3$ loudspeakers, 8 ms estimated HRIR lengths, 20-second recording → 1.2 TB memory ↯
3. Idea: process **small segments** to limit memory requirements
   - Divide total recording of length N_{t} into small segments of length N
   - Learn *independent* set of parameters per segment



4a Evaluation: Learned Parameters

Simulated measurements with rigid sphere "HRTFs" for different velocities

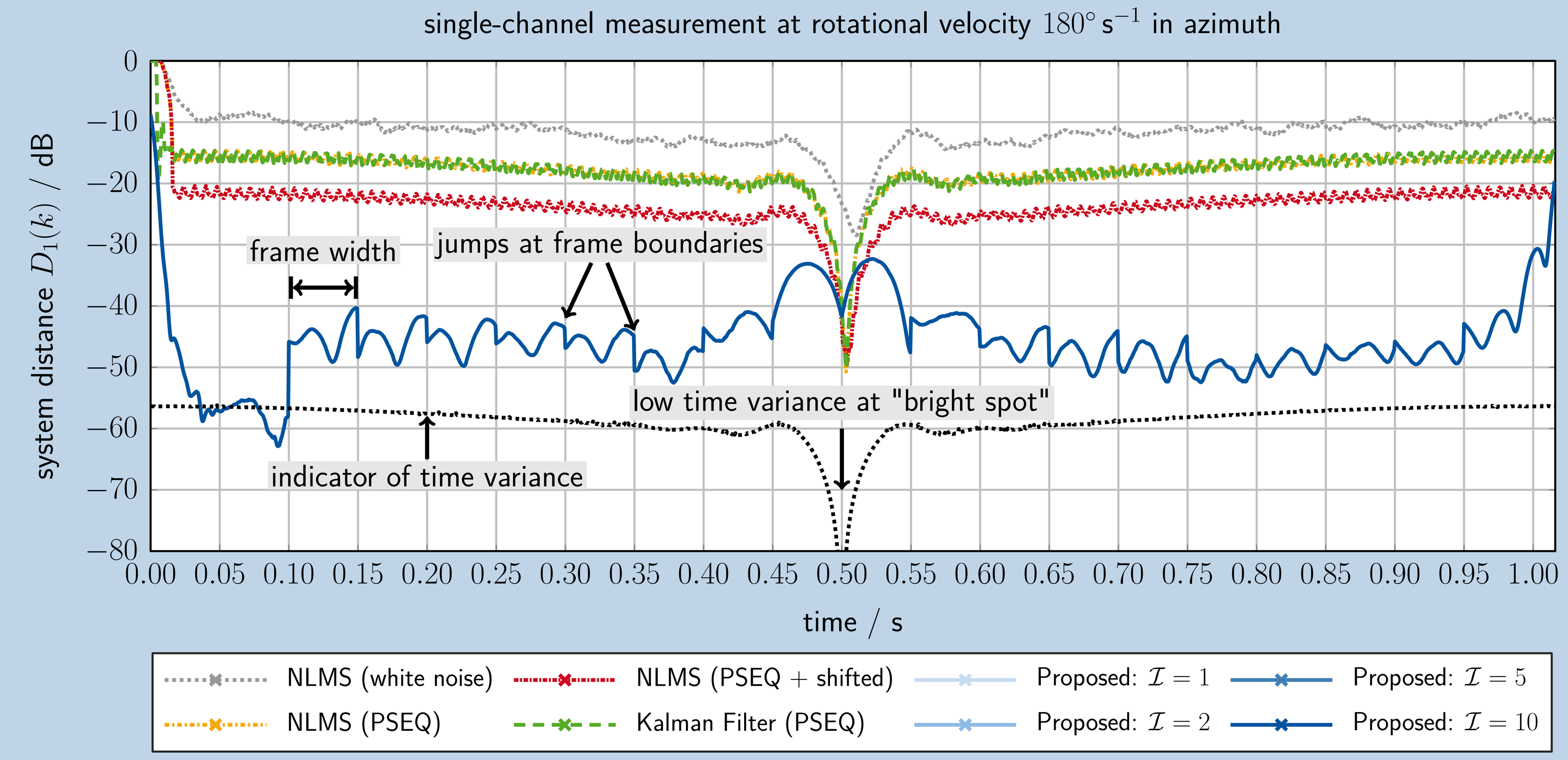
- Exemplary results (1st segment) for single-channel and $\mathcal{I} = 10$ iterations:

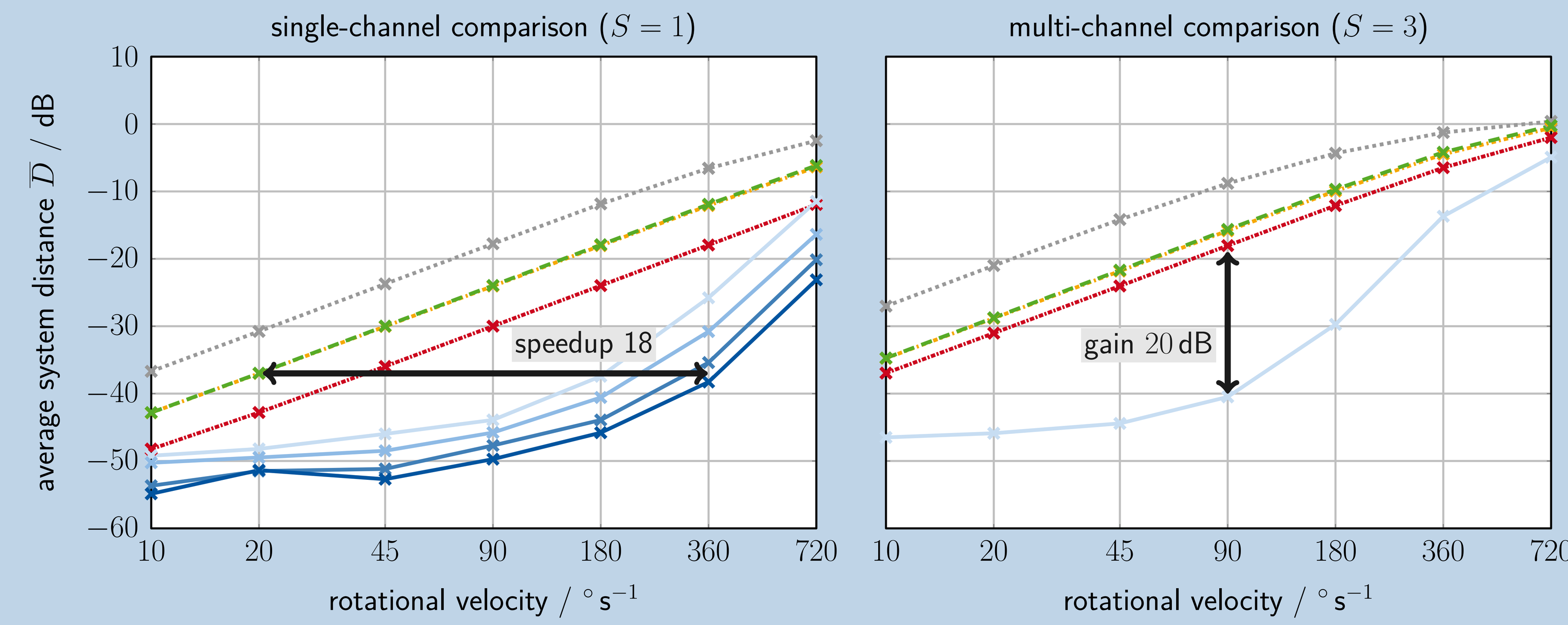


▶ Larger weights and covariances near base delay of 4.4 ms

4b Evaluation: Identification Performance

- White noise or perfect sequence (PSEQ) excitation [3]





- Open: evaluation in real HRTF measurement setting

5 Conclusions

▶ Learning system model with EM very beneficial for spherical head model

▶ **Increase quality** of HRTFs or **reduce duration** of measurement

References

[1] G. Enzner, "3D-continuous-azimuth acquisition of head-related impulse responses using multi-channel adaptive filtering," in *IEEE Workshop on Appl. of Signal Process. to Audio and Acoust. (WASPAA)*, 2009, pp. 325–328.

[2] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[3] C. Antweiler and G. Enzner, "Perfect sequence LMS for rapid acquisition of continuous-azimuth head related impulse responses," in *IEEE Workshop on Appl. of Signal Process. to Audio and Acoust. (WASPAA)*, 2009, pp. 281–284.