



# Spatial acquisition, digital archiving, and interactive auralization of church acoustics

Toon van Waterschoot

**STADIUS** Center for Dynamical Systems, Signal  
Processing, and Data Analytics

**ESAT** – Department of Electrical Engineering

KU Leuven, Belgium



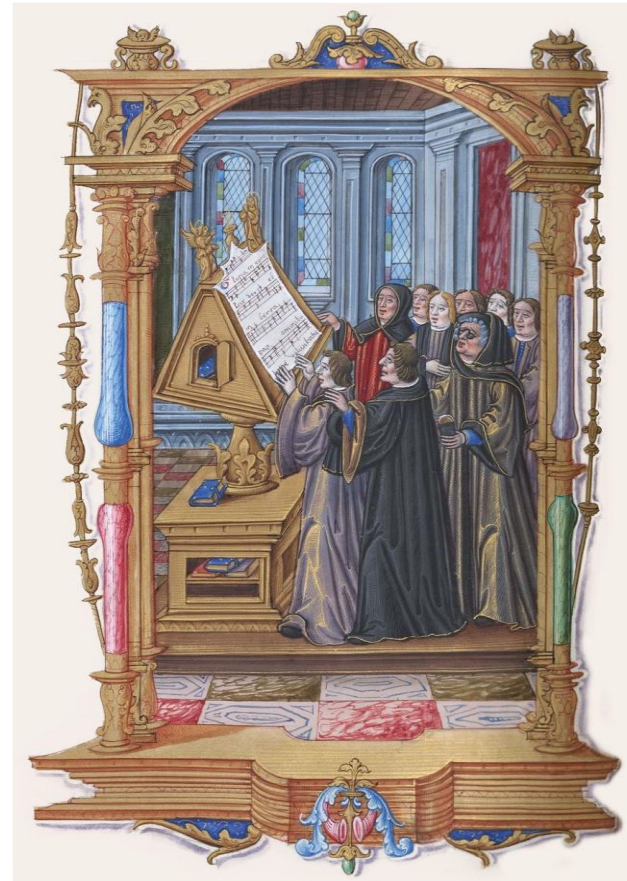
# Context & Motivation



Ghent Altarpiece (J. & H. van Eyck, 1432)

**Flemish primitives**

# Context & Motivation

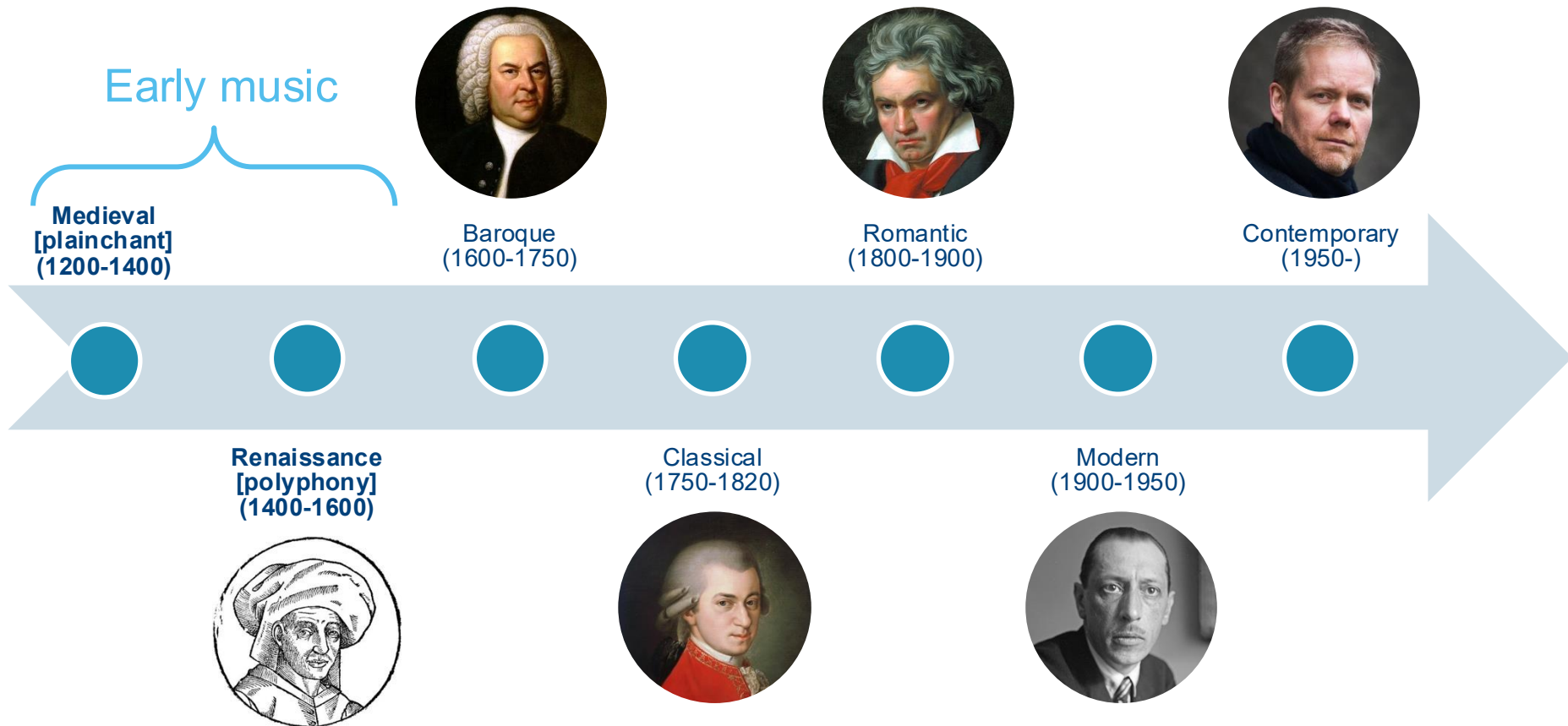


Nymphes des bois (J. des Prez, 1497, performed by Capella Pratensis)

## Franco-Flemish polyphony

KU LEUVEN

# Context & Motivation



A concise timeline of Western music

# Context & Motivation

## Early Music Research

### Source



### Performer



### Environment



# Context & Motivation

- **Research questions**

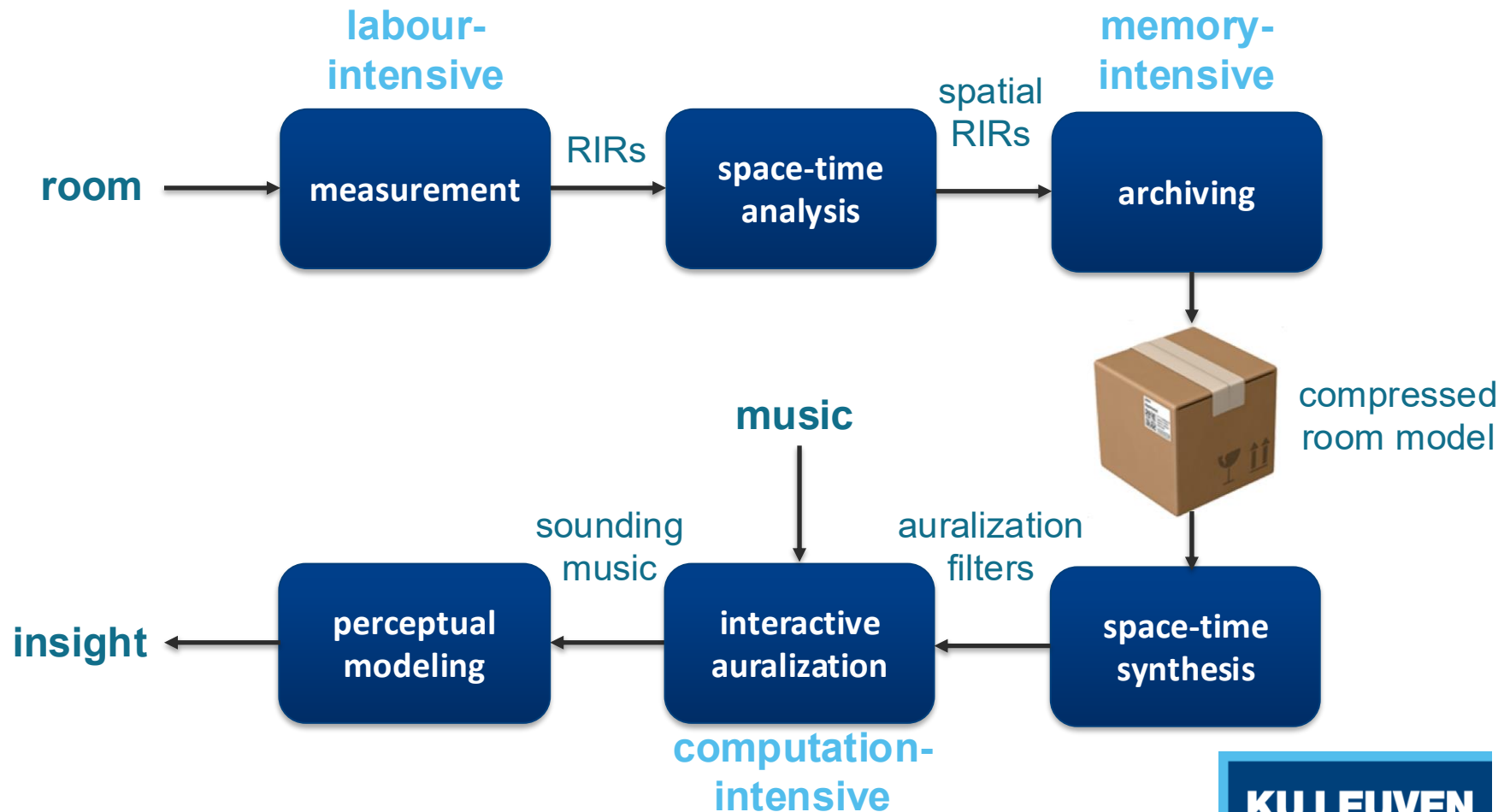
- How to **capture** the acoustics of a space to auralize/reproduce it at a different time and place?
- How to **archive** the acoustics of a historically valuable space for later reuse?
- How to achieve **interactive auralization** in a flexible and ecological manner?
- How to increase **understanding** of the relevance and quality of the acoustic characteristics of a space for musical performance?

## Early Music Research



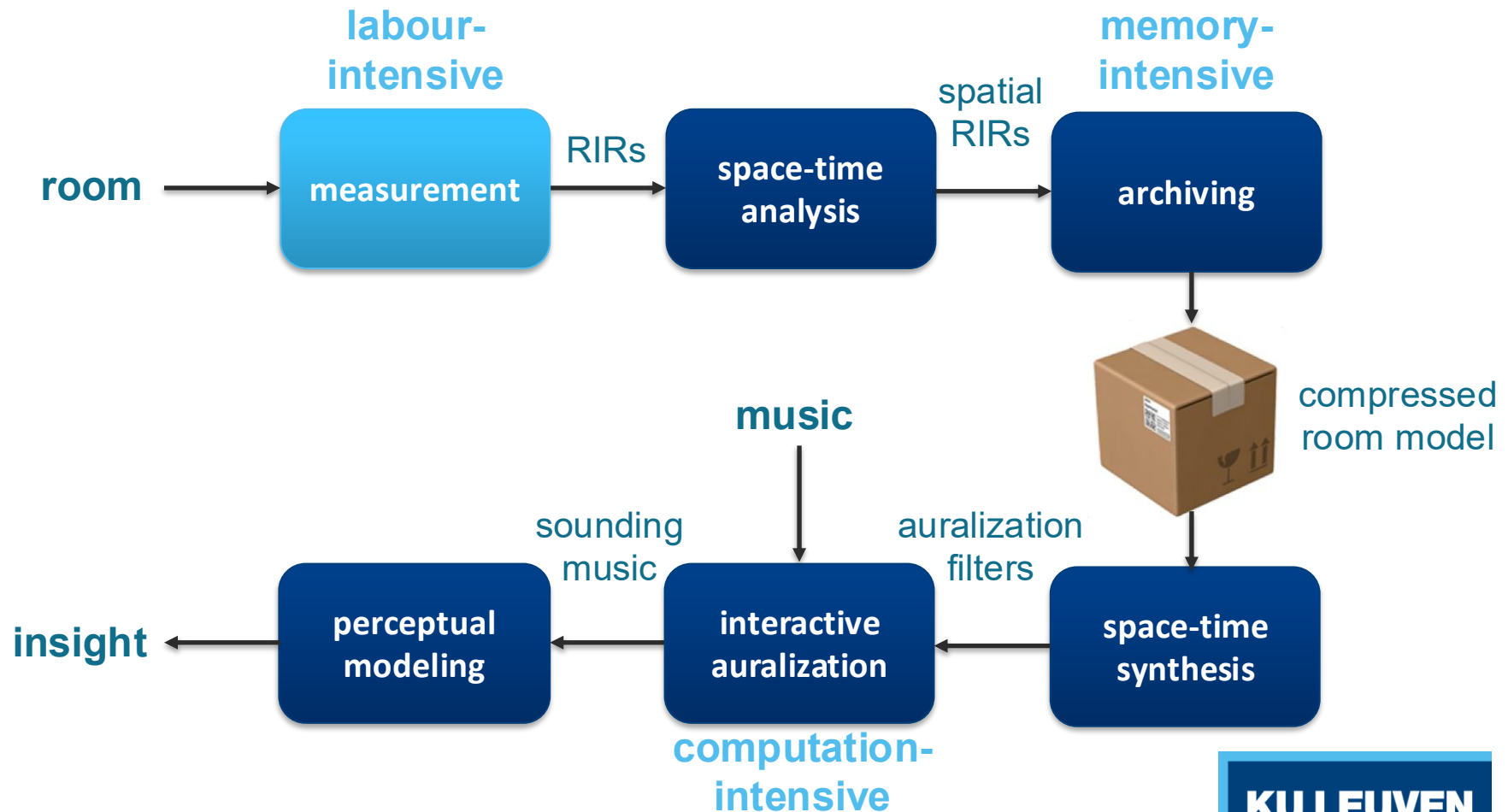
# Outline

- **System overview & challenges**



# Outline

- **System overview & challenges**



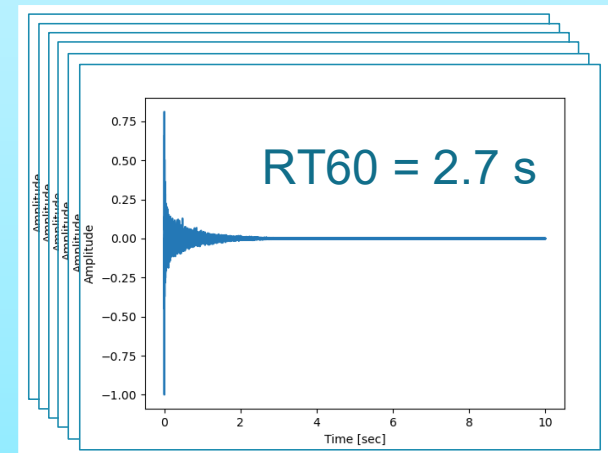
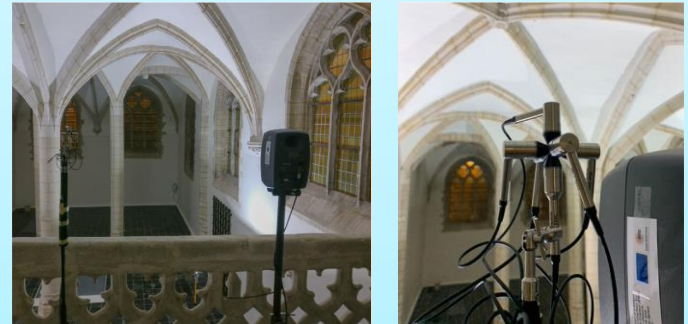
# Measurement

- **RIR measurement**
  - = well-established procedure
- **Challenges:**
  - (Nightly) access to venue
  - Ownership of measurements

- Labour-intensive
- RIR = point-to-point model

→ *how to choose source/mic positions?*

## Case study: Nassau Chapel, Royal Library Brussels



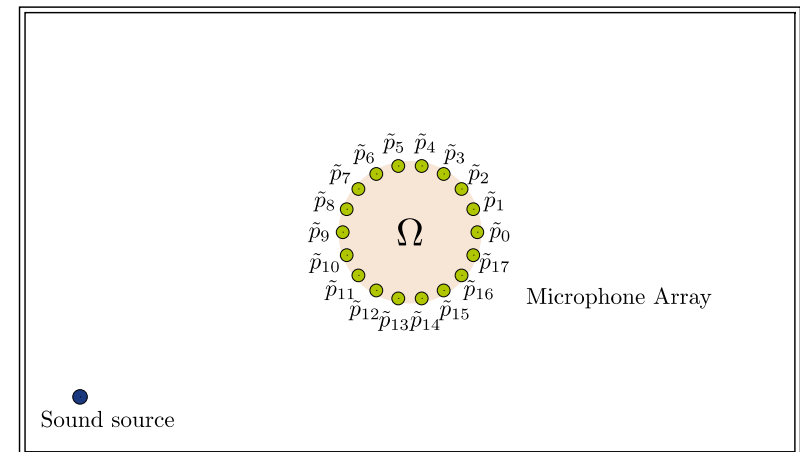
# Measurement

- **Key idea: RIR measurement at “few” positions + RIR interpolation**
  - Reduce labor-intensity
  - Expand potential use to other positions than those measured
  - Reduce memory required for archiving
  - Allow for auralization with moving sources/listeners
- **RIR interpolation problem:**

*given* RIR measurements

$$\tilde{\mathbf{P}} = [\tilde{\mathbf{p}}_0 \dots \tilde{\mathbf{p}}_{N_m-1}]$$

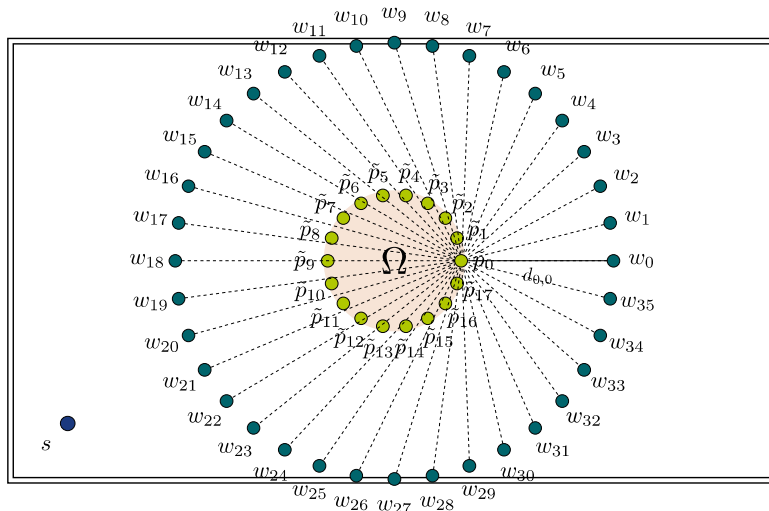
*find* RIR for same source  
at any mic position  $\mathbf{r}$  in  $\Omega$



# Measurement

- **Time-domain equivalent source method (TESM)**
  - Model RIR  $h(t, \mathbf{r})$  for mic position  $\mathbf{r}$  in  $\Omega$  as sum of propagating equivalent source signals  $w_l(t) \neq \delta(t)$

$$h(t, \mathbf{r}) = \sum_{l=0}^{N_w-1} \frac{1}{4\pi d_l(\mathbf{r})} \delta(t - d_l(\mathbf{r})/c) * w_l(t) = \mathcal{D}(\mathbf{r})[\mathbf{W}]$$



independent of  $\mathbf{r}$  !

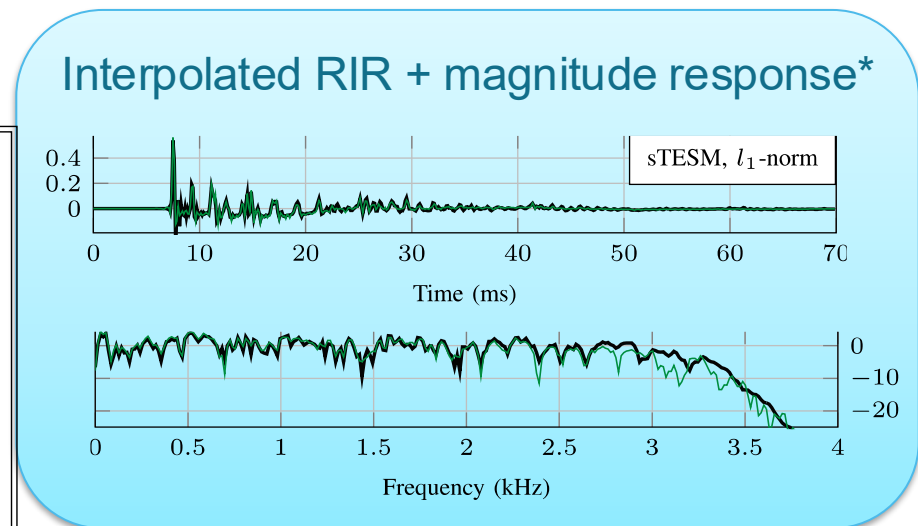
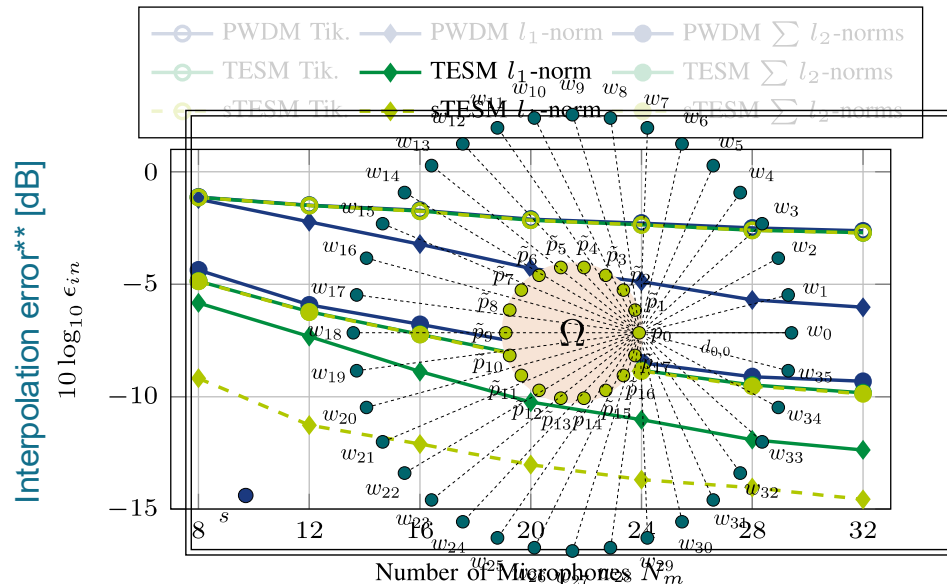
$\mathbf{W} = [\mathbf{w}_0 \dots \mathbf{w}_{N_w-1}]$   
 $\mathcal{D}(\mathbf{r})[\cdot]$  = linear operator  
 (delays + attenuations)

# Measurement

- **Time-domain equivalent source method (TESM)**

- Finding TESM signals  $w_l(t)$  by matching pressure signals  $\tilde{p}_m(t)$  (= measured RIR) at microphones = **underdetermined problem**
- Regularization by **imposing spatio-temporal sparsity**

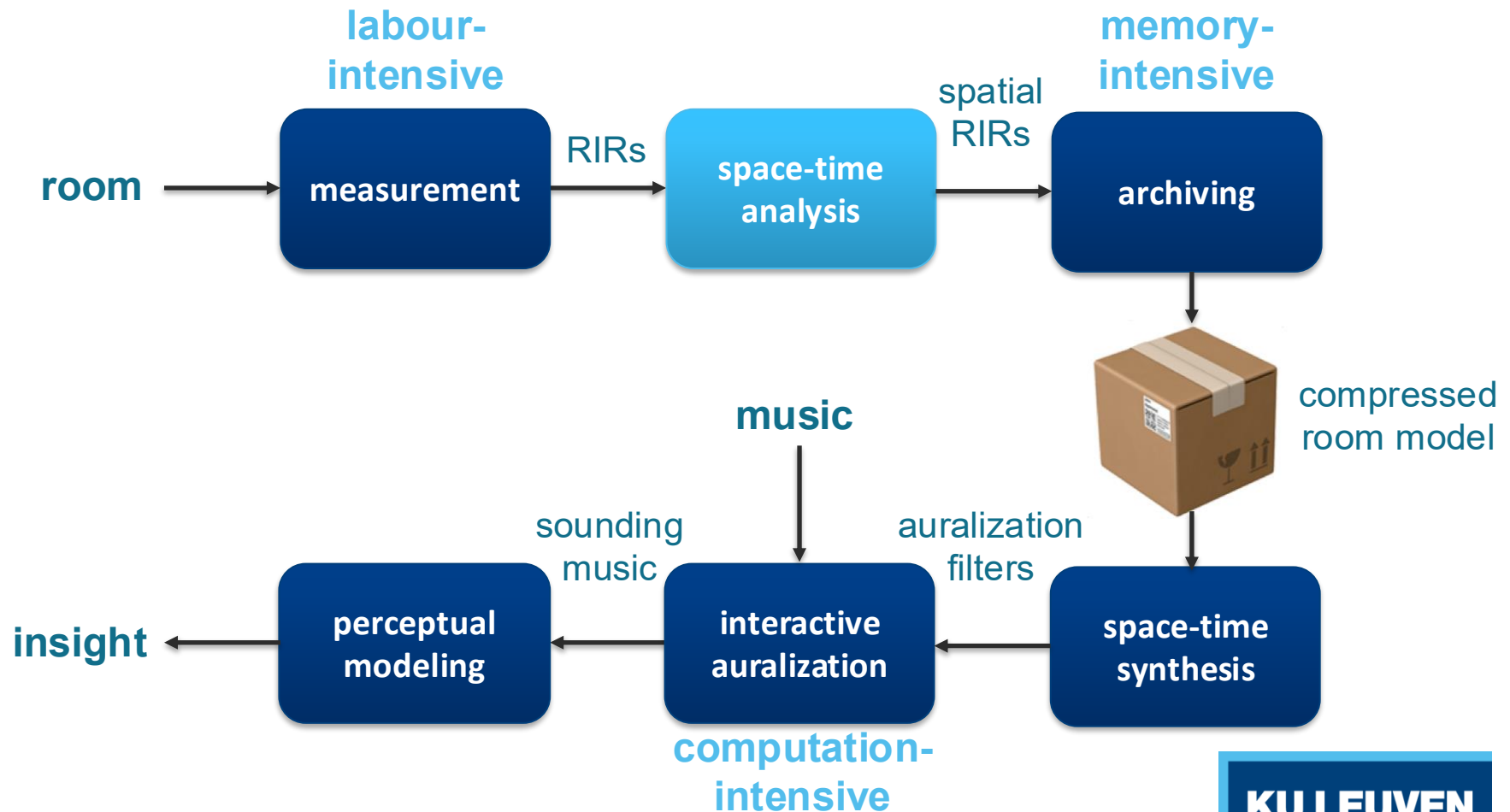
$$\min_{\mathbf{W}} \|D(\mathbf{W}) - \tilde{\mathbf{P}}\|_F^2$$



\*worst case result  
 \*\*average result

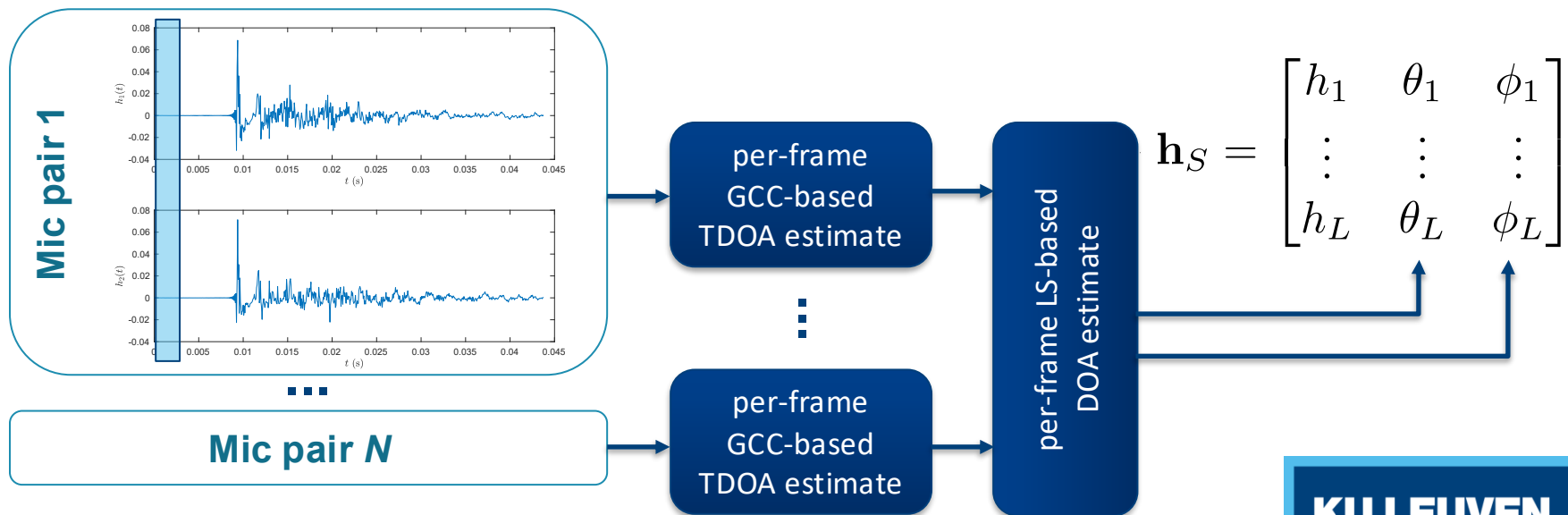
# Outline

- **System overview & challenges**



# Space-time analysis


- **Key model: Spatial room impulse response (SRIR)**
  - Include **direction of arrival information** (azimuth and elevation angles  $\theta, \phi$ ) to each RIR “peak” representing room reflection
- **Key method: Spatial decomposition method (SDM)**
  - TDOA estimation of corresponding RIR “peaks” in each mic pair
  - Least-squares DOA estimation from TDOAs of all mic pairs in array



# Space-time analysis



- **SDM challenges**

- Find phase-matched microphones 
- Find RIR “peaks”
- Match “peaks” in different RIRs  $\rightarrow$  *generalized cross-correlation (GCC)*
- Estimate TDOAs with subsample accuracy  $\rightarrow$  *TDOA interpolation*

- **Subsample TDOA estimation**

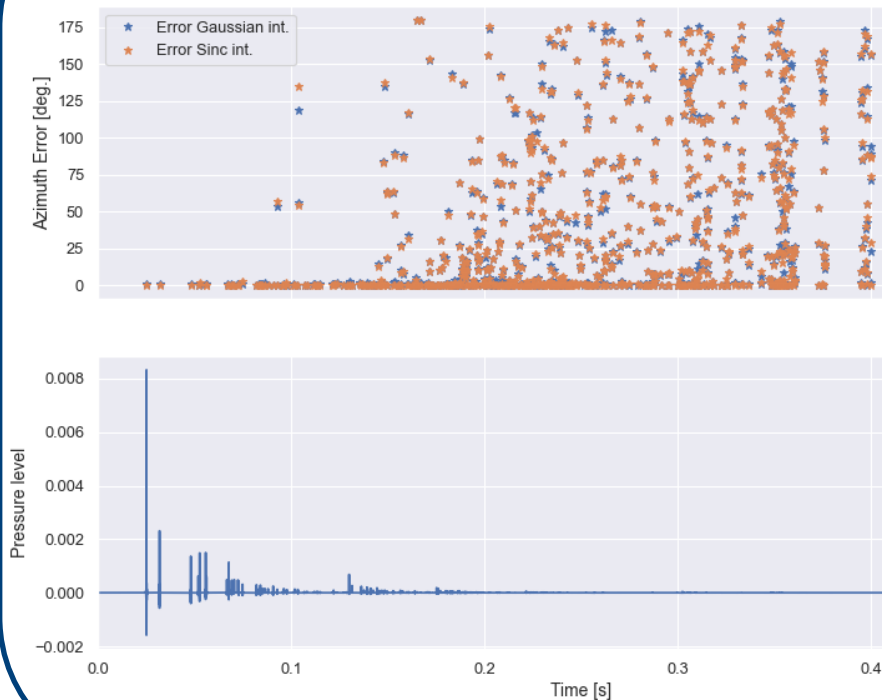
- Due to bandlimited nature of measurement system, RIR “peaks” are sinc functions rather than impulse functions
- GCC  $R_{m,n}(t)$  of two time-shifted sinc functions = sinc function
- Optimal subsample TDOA estimation method is hence based on sinc interpolation rather than parabolic/Gaussian interpolation

$$\tau_{\text{sinc}} = \arg \min_{\tau} \int_{-\infty}^{\infty} \left| \text{sinc}(\pi f_s(t - \tau)) - \frac{R_{m,n}(t)}{\max(R_{m,n}(t))} \right|^2 dt$$

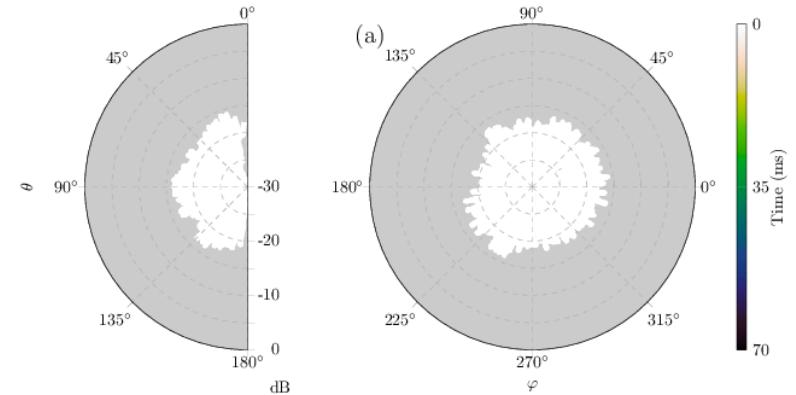
# Space-time analysis

- Resulting space-time model

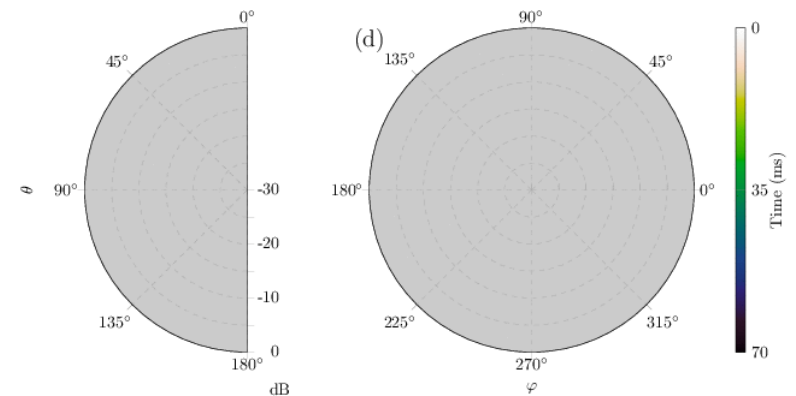
## TDOA estimation errors



## SRIR visualization of interpolated RIRs



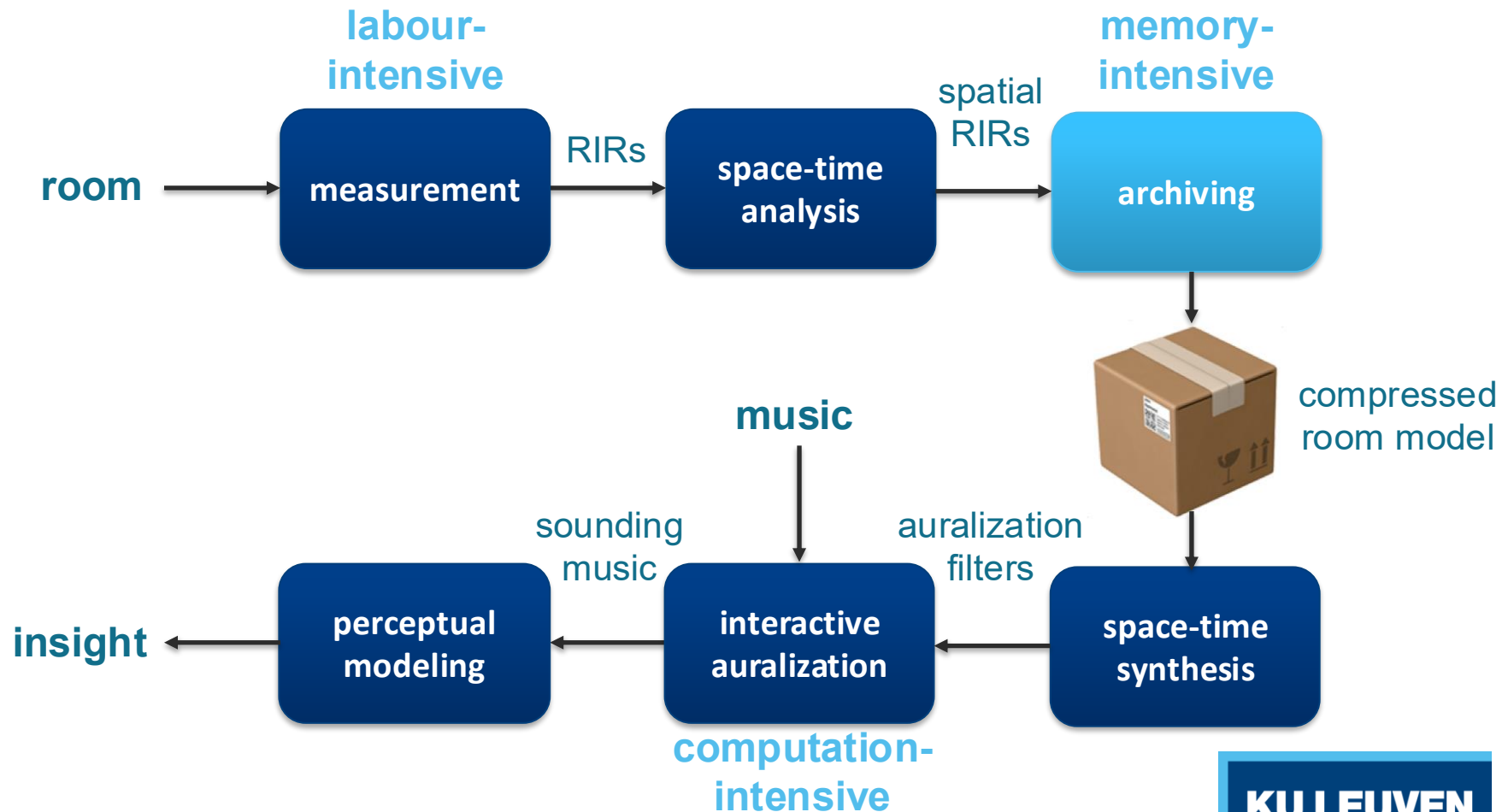
*without sparse regularization*



*with sparse regularization*

# Outline

- **System overview & challenges**



# Archiving

- **Why is archiving of church acoustics a challenge?**
  - Impulse response contains  $L \sim 10^4$  samples  $\sim 10^5$  bits
  - 1 spatial impulse response = 6 impulse responses  $\sim 10^6$  bits
  - Plenacoustic sampling theory: accurate sound field reconstruction requires spatial resolution of  $\sim 10$  cm
    - $\sim (100)^3$  source positions x  $(100)^3$  observer positions
    - $\sim 10^{12}$  spatial impulse responses  $\sim 10^{18}$  bits  $\sim$  **100 petabyte**



# Archiving

- How to “compress” a room impulse response?
  - Truncation
  - Hard thresholding
  - Sparse approximation
  - Low-rank approximation

## Physical justification

*non-homogenous Helmholtz equation*

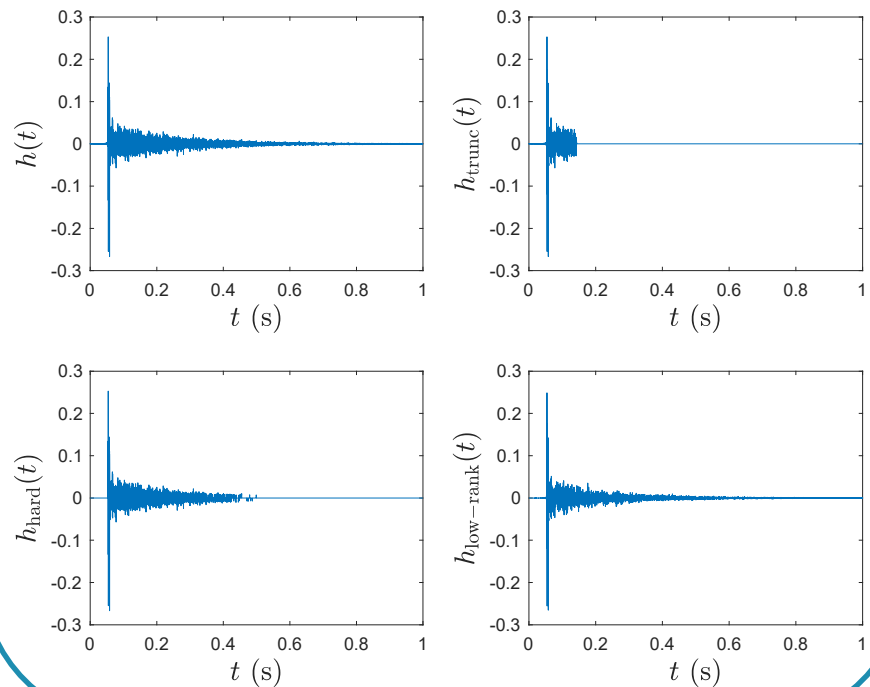
$$(\nabla^2 + k^2)H(\mathbf{r}, \mathbf{r}_s, \omega) = -j\omega\rho_0 Q \delta(\mathbf{r} - \mathbf{r}_s)$$



**RIR  $\approx$  sum of damped sinusoids**

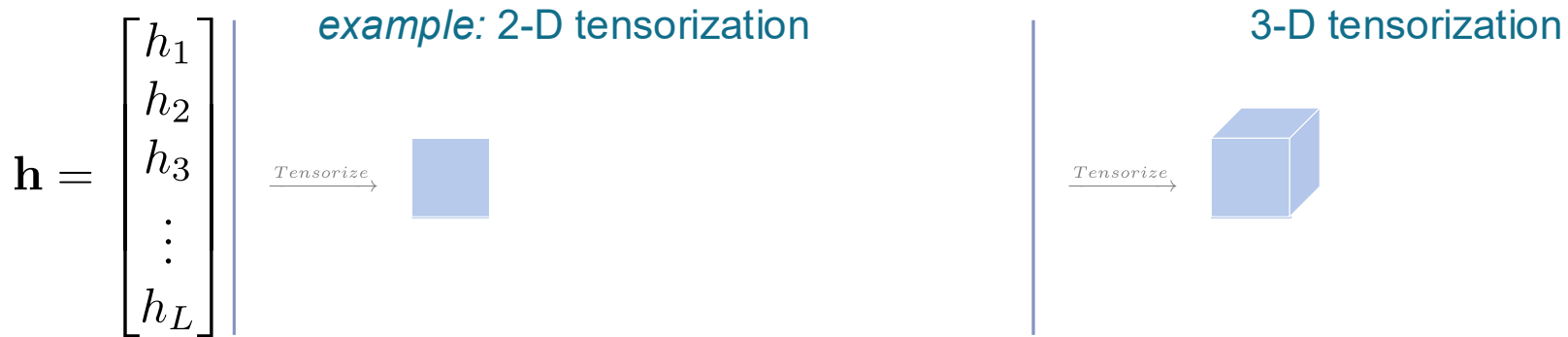
$$h(t) \approx \sum_{m=1}^M \mu_m e^{-\beta_m t} \cos(\omega_m t + \phi_m), \quad t \geq 0$$

RIR with three compression methods  
(compression rate = 85 %)



# Archiving

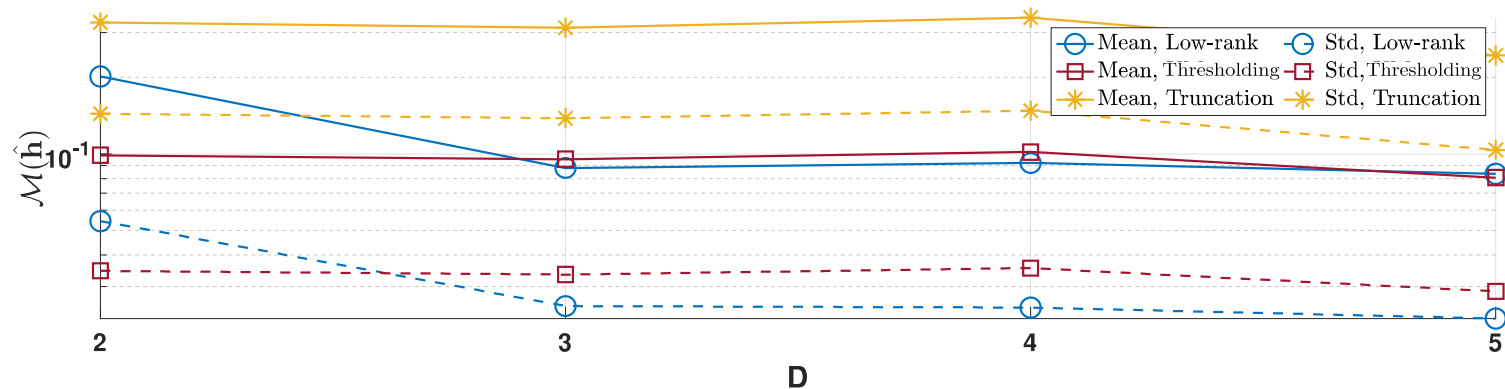
- **Key idea: Sum of  $M$  (damped) sinusoids admits matricization/tensorization of rank  $2M$** 
  - Impulse response vector  $\mathbf{h}$  (= 1-D array) of length  $L$  can be **reshaped** into  $N$ -D array of dimensions  $\sqrt[N]{L} \times \sqrt[N]{L} \times \dots \times \sqrt[N]{L}$
  - $N$ -D array can be **approximated** as sum of  $2M$  rank-1 terms (SVD or canonical polyadic decomposition)



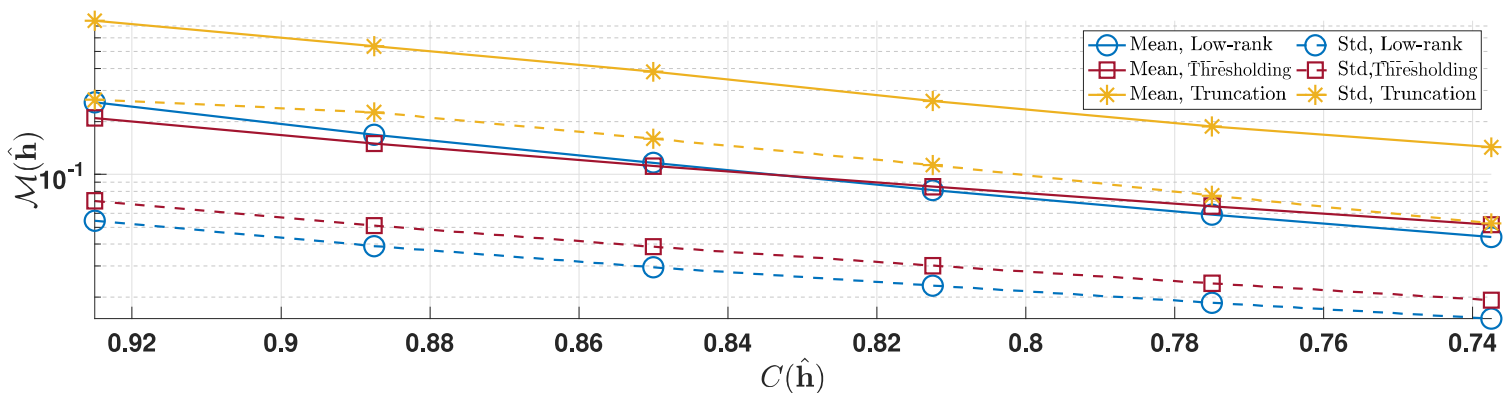
$$\mathbf{h} = \text{vec} \left( \sum_{m=1}^{2M} \mathbf{u}_m^{(1)} \circ \mathbf{u}_m^{(2)} \circ \dots \circ \mathbf{u}_m^{(N)} \right)$$

# Archiving

- **Misadjustment results after RIR compression**
  - Methods: truncation, hard thresholding, low-rank approximation
  - Misadjustment vs. tensorization dimension  $D$  & compression rate  $C$



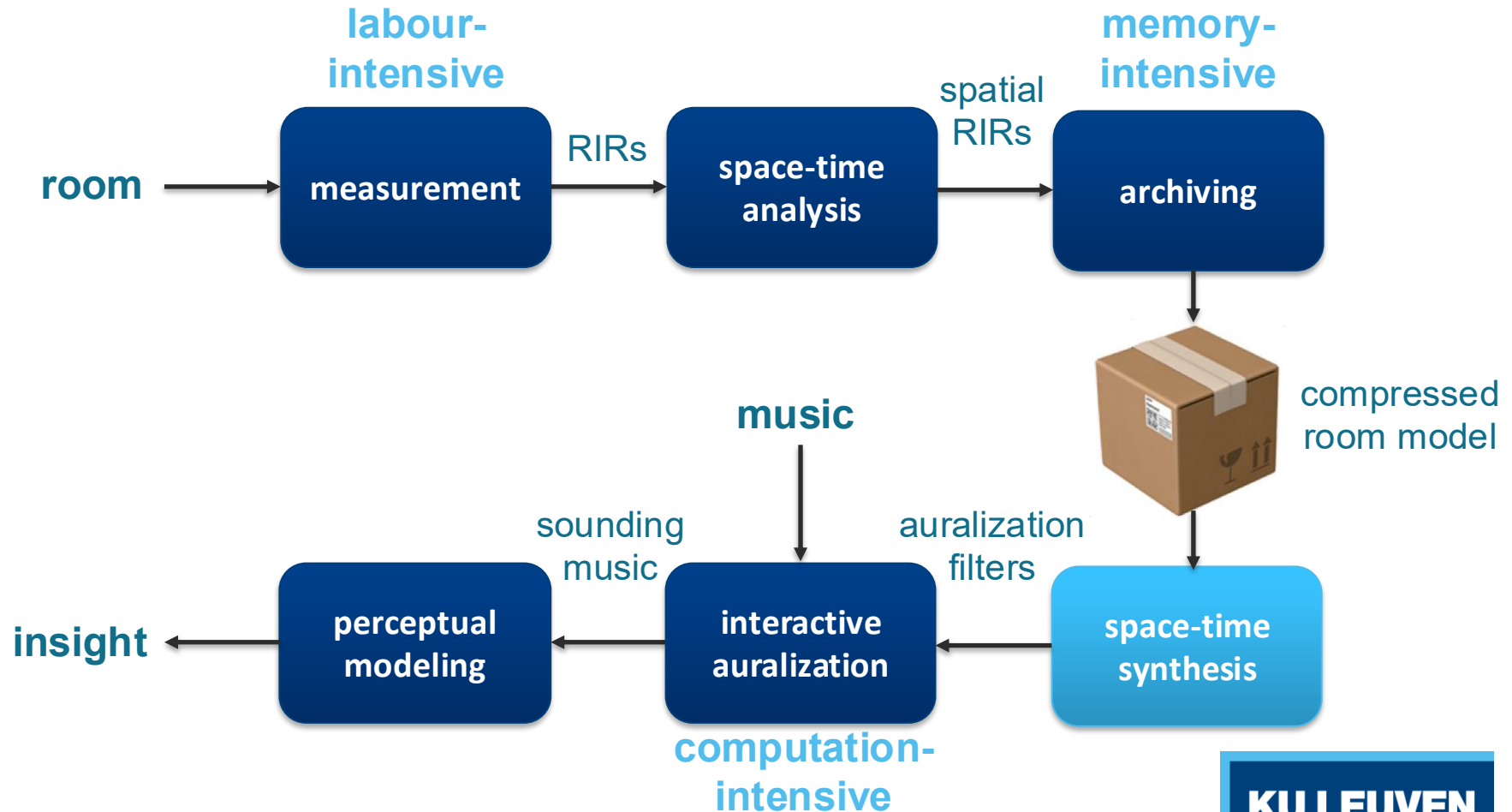
$C = 80\%$



$D = 3$

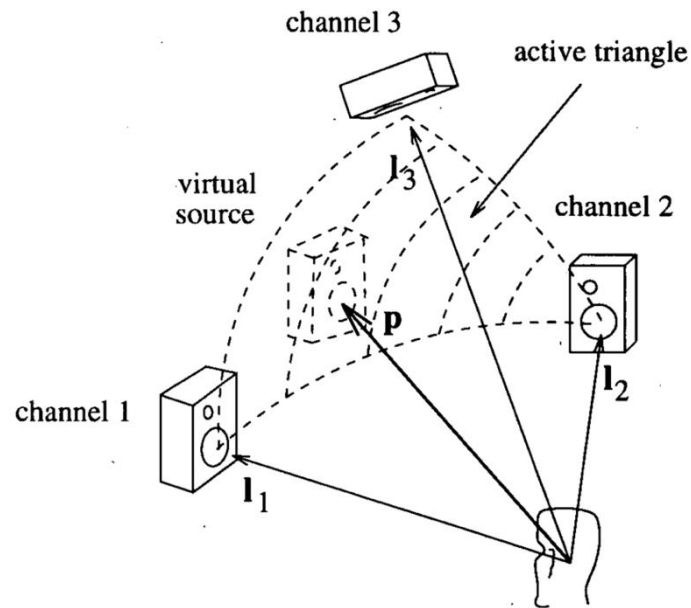
# Outline

- **System overview & challenges**



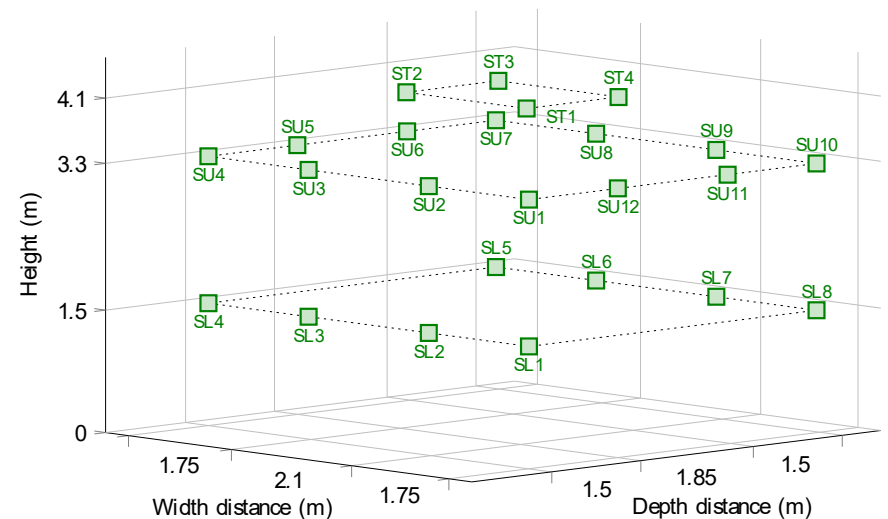
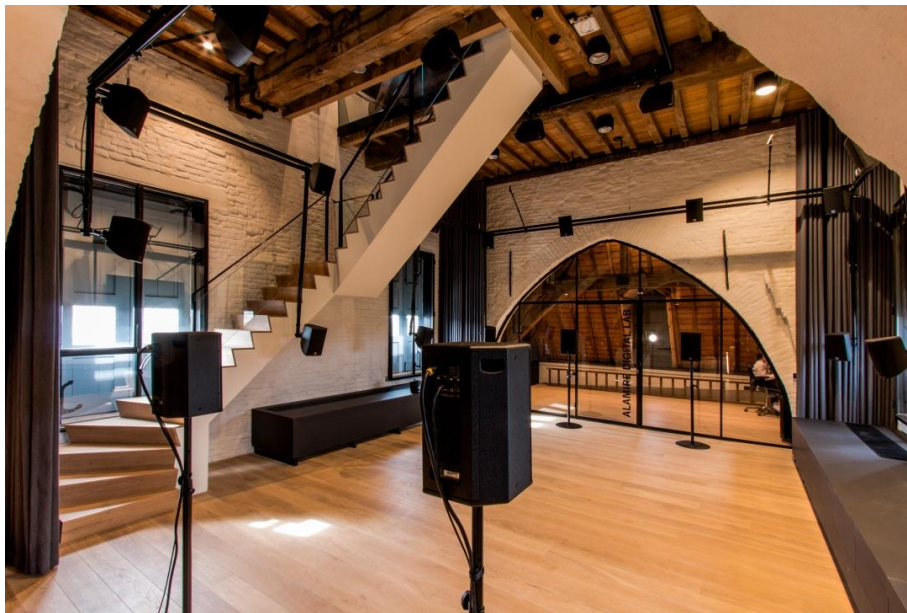
# Space-time synthesis

- **Key idea: Mapping of room reflections to loudspeakers**
  - Conversion of SRIR to set of per-loudspeaker impulse responses
    - Vector base amplitude panning (VBAP)
    - Nearest-loudspeaker mapping



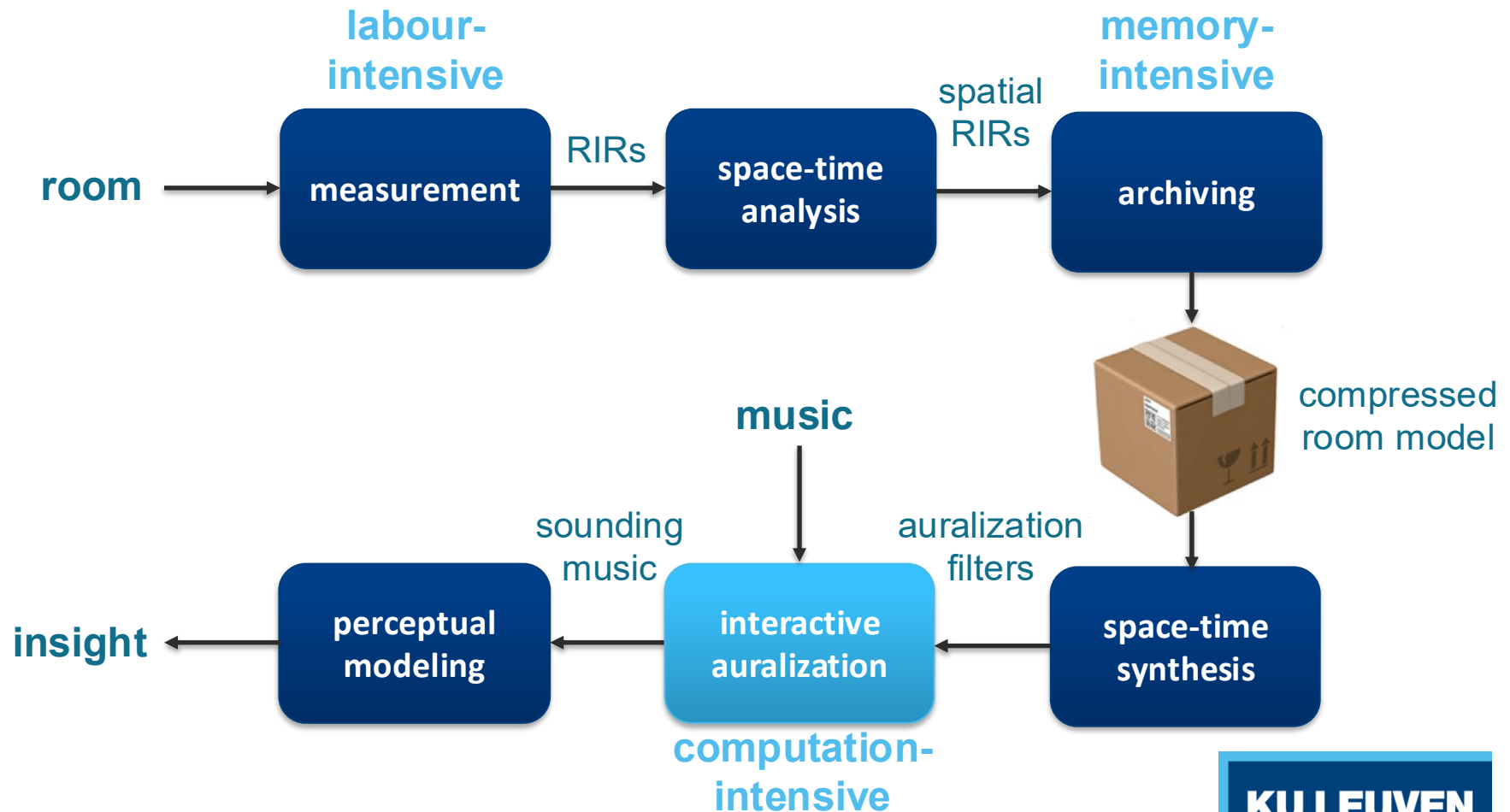
# Space-time synthesis

- **Loudspeaker set-up: Library of Voices (Leuven)**
  - 3-D array of 20–24 Martin Audio 6,5” CDD loudspeakers
  - Reproduction room: RT60 = 0.5 s (with curtains closed)



# Outline

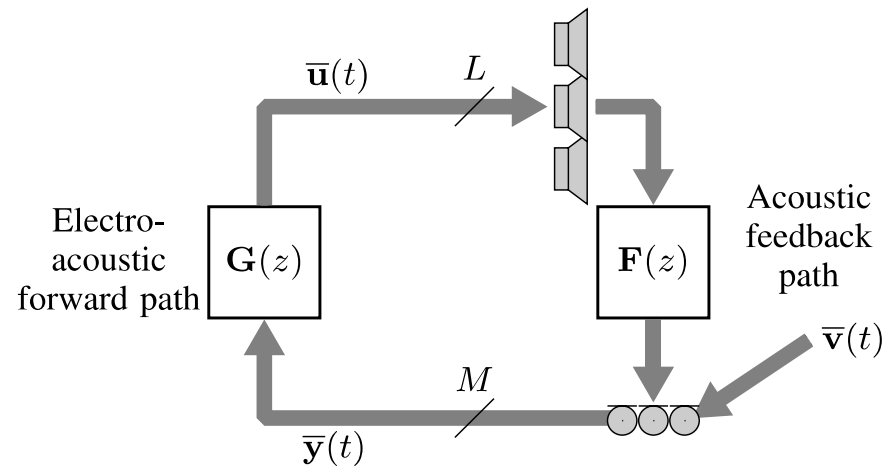
- **System overview & challenges**



# Interactive auralization

- **Challenges:**

- Multi-channel acoustic feedback control
- Low-latency real-time multi-channel convolution



# Interactive auralization

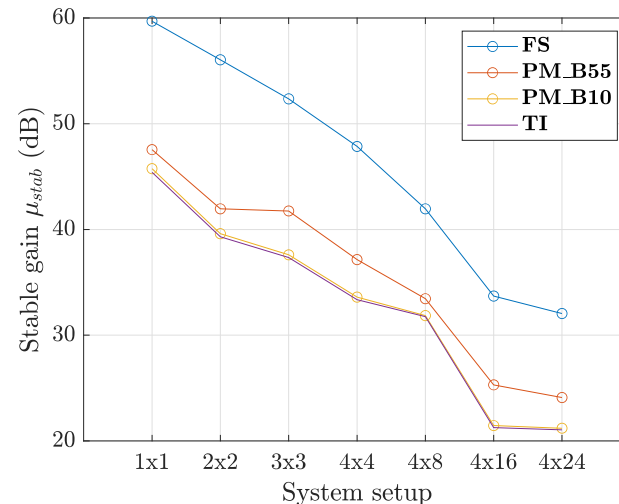
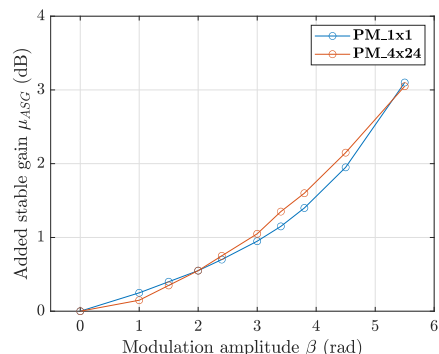
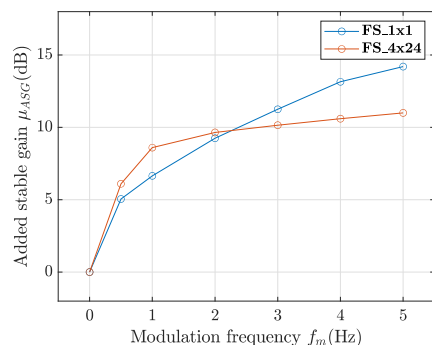
- **Multi-channel acoustic feedback control**

- Gain before instability decreases with increasing # channels
- Feedback-related artefacts: howling, ringing, coloration

- **Acoustic feedback control methods:**

- Phase modulation methods → *Reverberation enhancement systems*

- Gain reduction methods
- Spatial filtering methods
- Room modeling methods



FS = frequency shifting (5 Hz)

PM = phase modulation with modulation index  $\beta=1.0$  or 5.5

# Interactive auralization

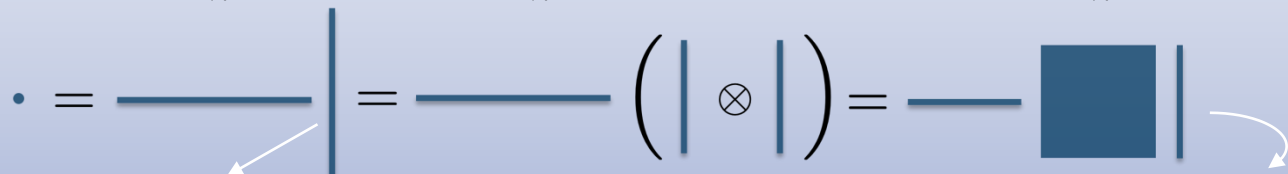
- **Low-latency real-time multi-channel convolution**

- **Key idea:** Rewrite convolution operation using compressed RIR

- **Low latency:** no buffering required by staying in time domain
- **Real-time:** reduced # input buffer reads and # FLOPS
  - Truncation  $\rightarrow$  shorter FIR filter
  - Hard thresholding  $\rightarrow$  FIR filter with fewer non-zero coefficients
  - Sparse approximation  $\rightarrow$  short “warped” FIR filter (mixed-Kautz model)
  - Low-rank approximation  $\rightarrow$  “low-rank convolution”

**Low-rank convolution:** illustration for rank-1 model  $\mathbf{h}_{L \times 1} = \mathbf{u}_{P \times 1} \otimes \mathbf{v}_{Q \times 1}$

$$y_k = \mathbf{x}_k^T \mathbf{h} = \mathbf{x}_k^T (\mathbf{u} \otimes \mathbf{v}) = \mathbf{v}^T \mathbf{X}_k^T \mathbf{u}$$

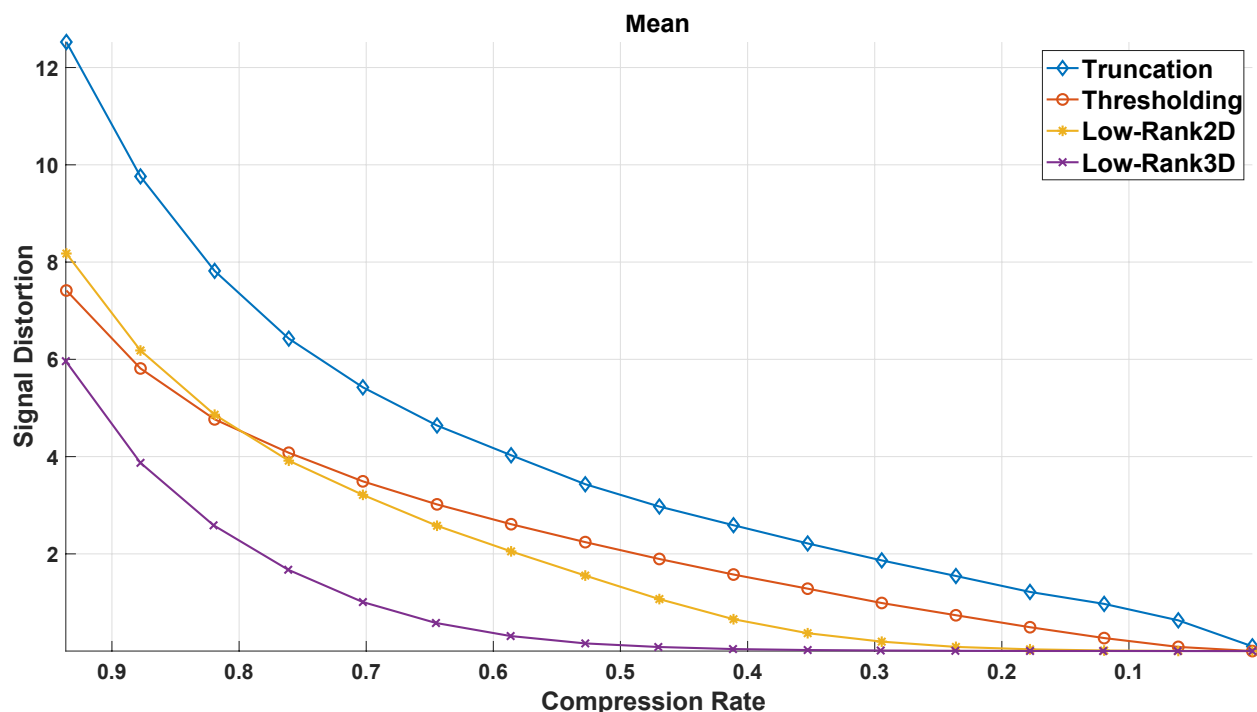


$L = PQ$  multiplications/sample

$L = P+Q$  multiplications/sample

# Interactive auralization

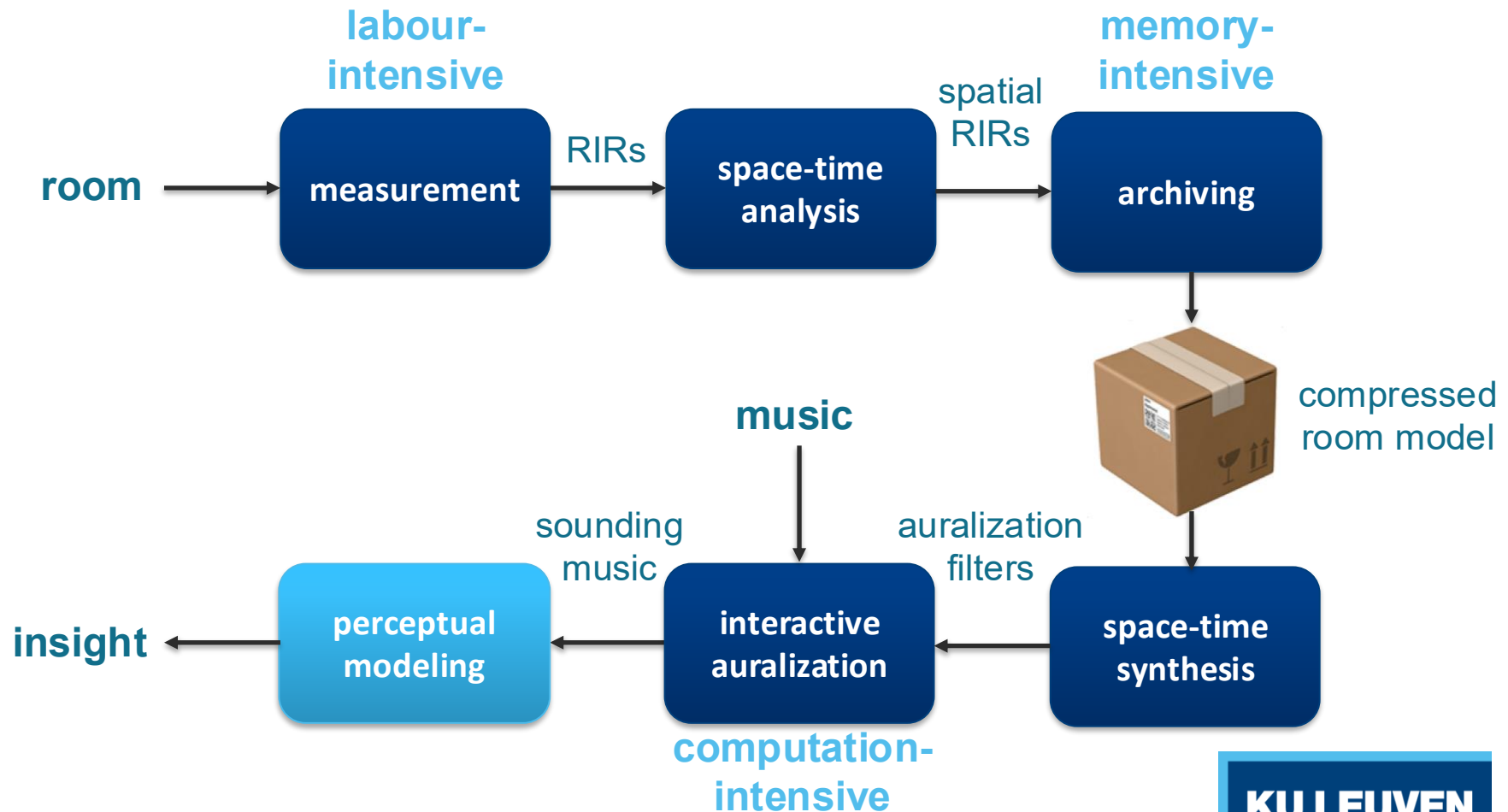
- **Signal distortion after compressed RIR convolution**
  - Methods: truncation, hard thresholding, low-rank approximation
  - Mean ERB-weighted signal distortion vs. RIR compression rate  $C$



input = dry vocal music  
RIR = Nassau Chapel RIR,  
RT60 = 2.7 s,  $L = 117649$

# Outline

- **System overview & challenges**



# Perceptual modeling

- **Aim:** understand relation between acoustic characteristics of church and their perceived quality (broadly defined)
- **Key methodology: Flash Profile rapid sensory analysis**
  - Originally developed in frame of **food tasting** experiments
  - Validated for perceptual modeling of virtual acoustics **by listening** (e.g. auralization of concert halls, car cabins, domestic rooms)
  - Evaluated in preliminary experiments for perceptual modeling of virtual church acoustics **by listening while singing**



## Materials

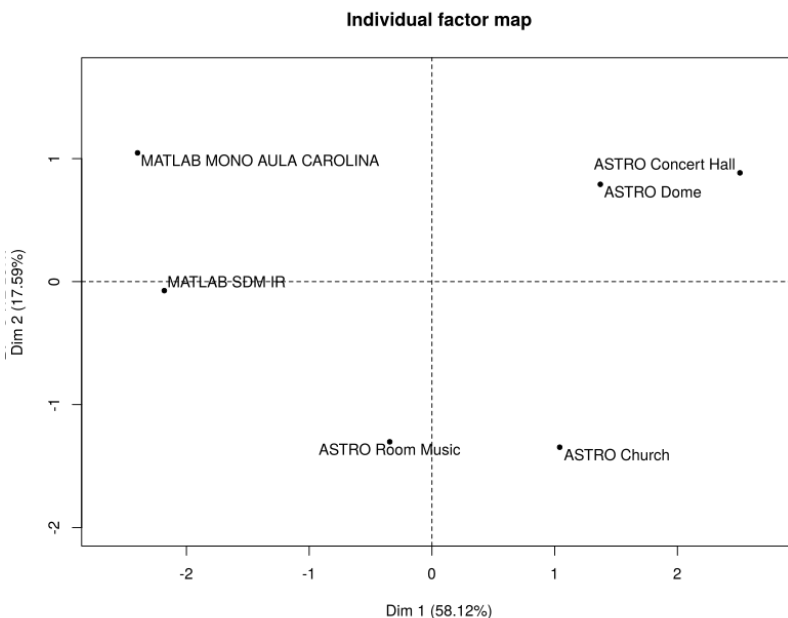
- 6 virtual acoustic spaces
- 4 expert listeners: male singers from Cappella Pratensis
- 2 pieces: plainchant + polyphonic





# Perceptual modeling

- **Flash Profile Methodology:** statistical data analysis
  - Multi-factor principal component analysis + clustering



- 75% of perceived variance among spaces is modeled by two largest principal components
- two largest principal components correlate to perceptual attributes of **spaciousness/reverberance** (Dim 1) and **spectral content** (Dim 2)
- six virtual spaces can then be mapped into this 2-D principal component subspace

# Conclusion

- **Some take-home messages...**
  - **Acoustic signal processing in digital humanities:** case studies in humanities/musicology/cultural heritage/... may carry remarkably many research challenges relevant to our community
  - **Fundamental signal processing research** (sampling & interpolation, sparse & low-rank modeling, ...) can be highly relevant to address application-specific bottlenecks relating to labour, memory, and computational resources
  - **Interdisciplinary research** is much more than putting together diverse team – it requires to learn common language and mutual scientific understanding, attract funding, develop proper research methodology, ... which is not easy but very enriching
  - Looking for challenging research topic in acoustic signal enhancement? Consider **multi-channel acoustic feedback control**

# Thank you...

- **This presentation is the result of joint work with:**  
*Niccolò Antonello, Stratton Bull, David Burn, Bart De Moor, Enzo De Sena, Bart Demuyt, Efren Fernandez-Grande, Ann Kelders, Martin Jälmy, Rudi Knoops, Hannes Rosseel, Wannes Van Ransbeeck ... and many more*
- **The work presented here has been carried out and supported by:**



DEPARTEMENT  
ECONOMIE  
WETENSCHAP &  
INNOVATIE



- **Contact**

[toon.vanwaterschoot@esat.kuleuven.be](mailto:toon.vanwaterschoot@esat.kuleuven.be)

<https://tvanwate.github.io>