Slides: https://tvanwate.github.io/iwaenc2022.pdf

KU LEUVEN

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

IWAENC 2022, Bamberg, Germany, Sept. 2022

Toon van Waterschoot

STADIUS Center for Dynamical Systems, Signal Processing, and Data Analytics
ESAT – Department of Electrical Engineering
KU Leuven, Belgium

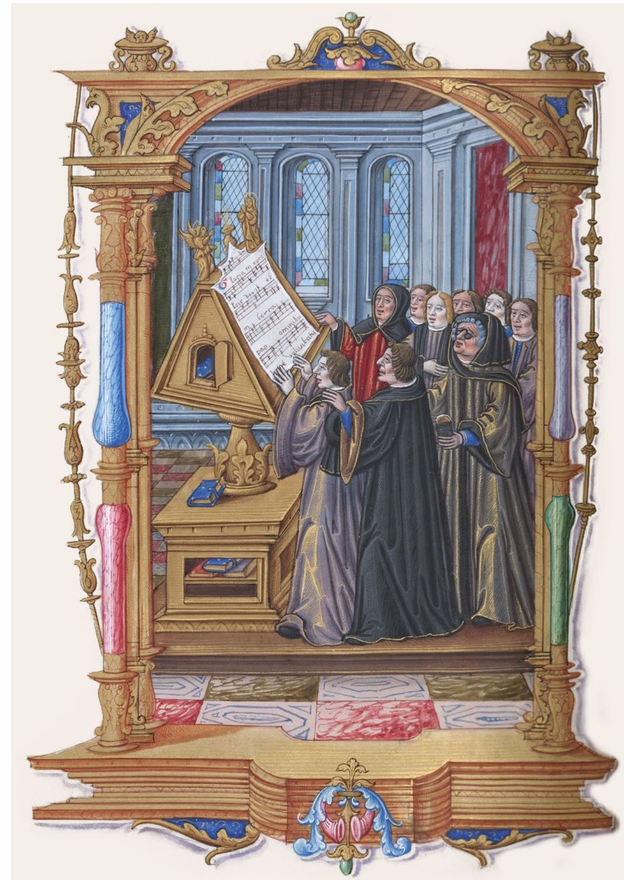
Context & Motivation

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

Flemish primitives

KU LEUVEN

Context & Motivation

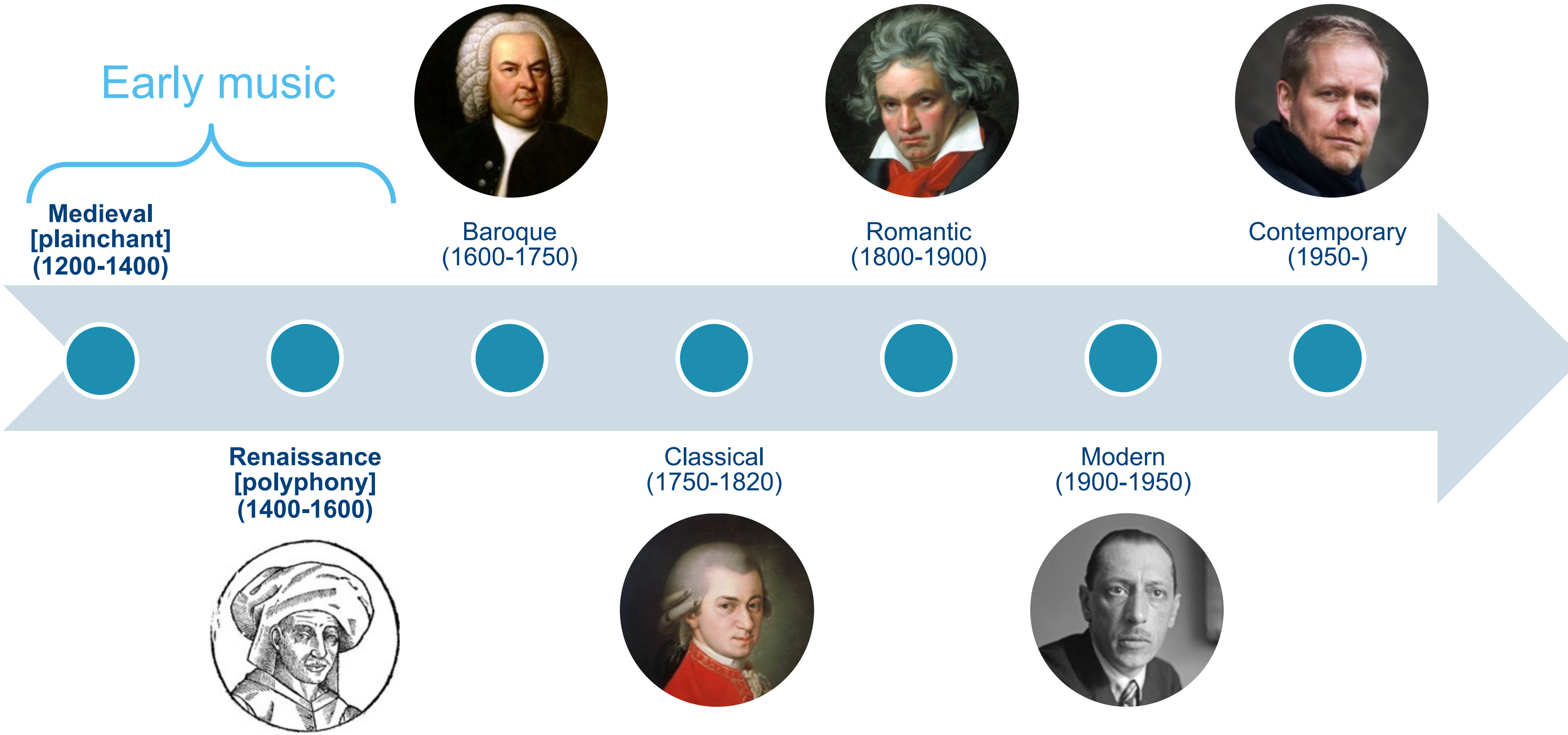


Nymphes des bois (J. des Prez, 1497, performed by Capella Pratensis)

Franco-Flemish polyphony

Image credits: Biblioteca Medicea Laurenziana, Bibliothèque nationale de France – *Audio credit:* Capella Pratensis

Context & Motivation



A concise timeline of Western music

Image credits: Wikimedia Commons, US Library of Congress, Mike Terry

Context & Motivation

Early Music Research

Source

Performer

Environment

Image credits: Alamire Foundation, Cappella Pratensis, TvW

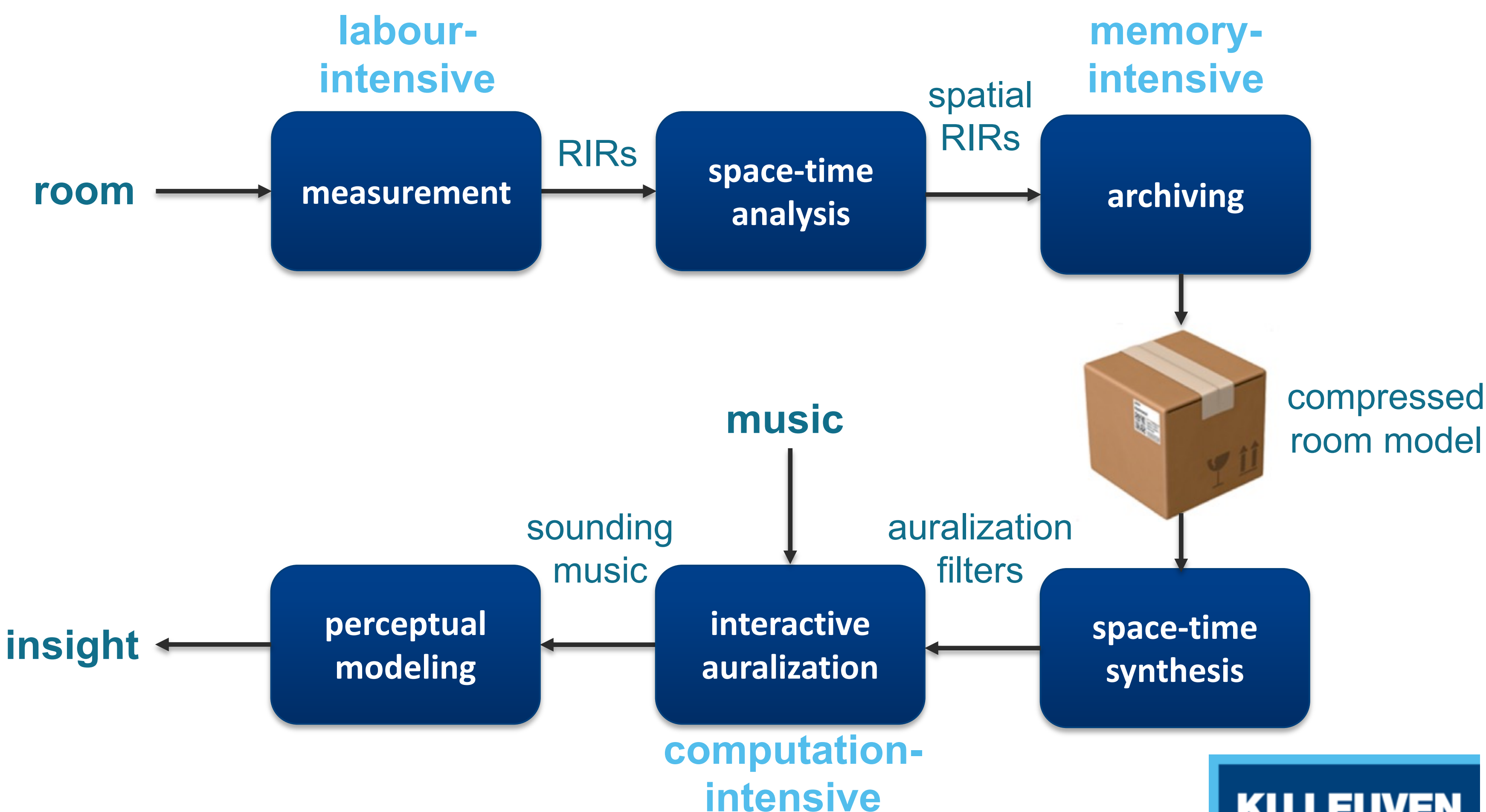
Context & Motivation

- **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?

Image credits: Alamire Foundation, Cappella Pratensis, TvW

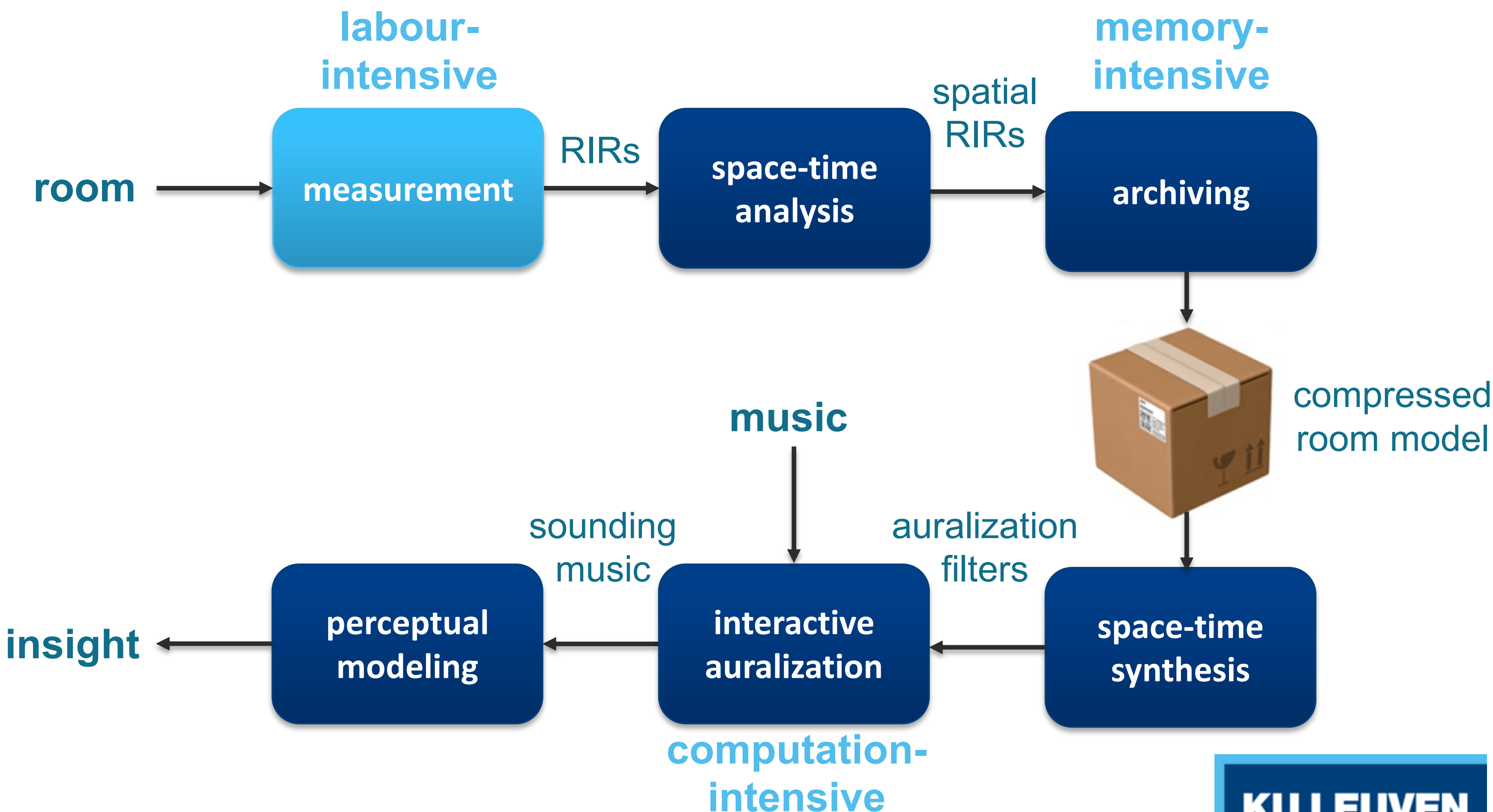
Outline

- **System overview & challenges**



Outline

- **System overview & challenges**

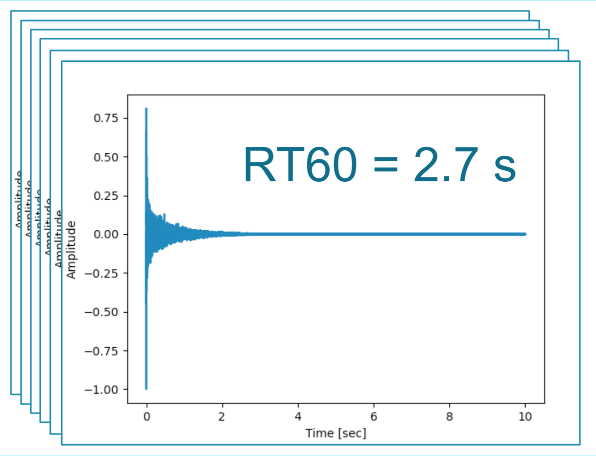


Measurement

- **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



KU LEUVEN

Image credits: Hannes Rosseel, TvW – *Reference:* A. Novak, P. Lotton, and L. Simon, “Synchronized Swept-Sine: Theory, Application, and Implementation,” *J. Audio Eng. Soc.*, vol. 63, no. 10, pp. 786–798, Nov. 2015.

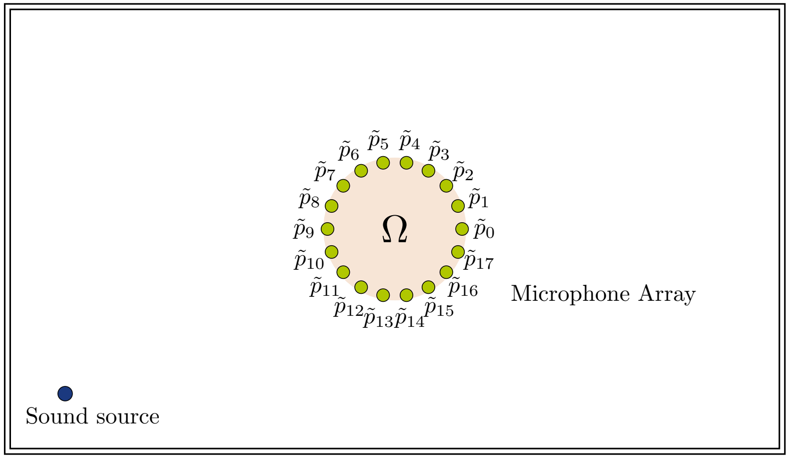
Measurement

- **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

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

find RIR for same source at any mic position **r** in Ω



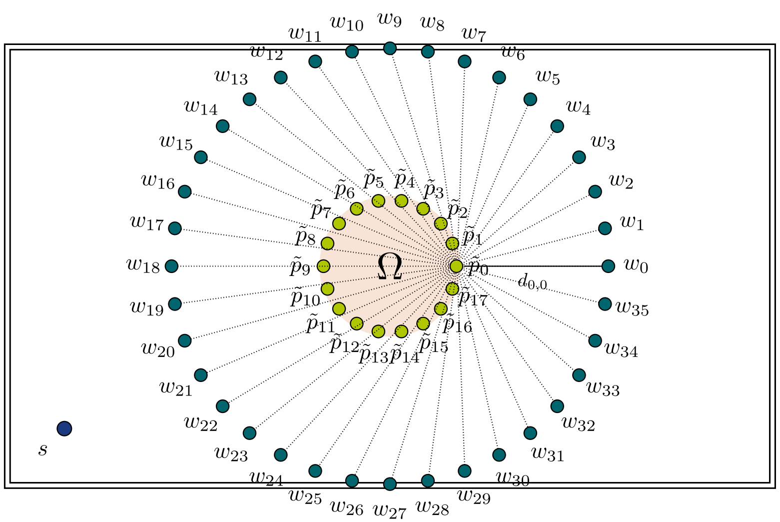
KU LEUVEN

Image credit: Niccolò Antonello – *Reference:* N. Antonello, E. De Sena, M. Moonen, P. A. Naylor, and T. van Waterschoot, "Room impulse response interpolation using a sparse spatio-temporal representation of the sound field," *IEEE/ACM TASLP*, 25(10): 1929-1941, 2017.

Measurement

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

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



independent of **r** !

$$\mathbf{W} = \left[\mathbf{w}_0 \ldots \mathbf{w}_{N_w-1}\right]$$

$$\mathcal{D}(\mathbf{r})[\cdot] = \text{linear operator}$$

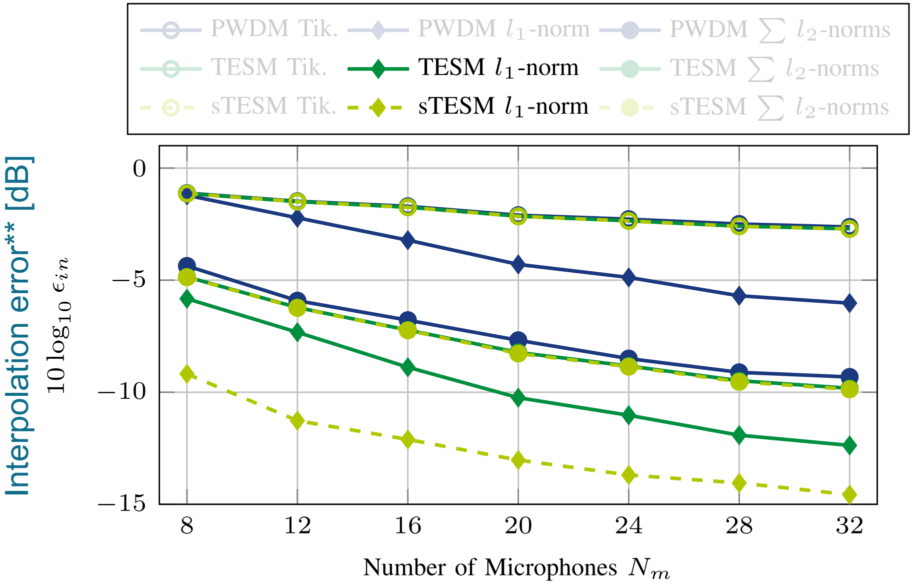
$$(\text{delays} + \text{attenuations})$$

Image credit: Niccolò Antonello – *Reference:* N. Antonello, E. De Sena, M. Moonen, P. A. Naylor, and T. van Waterschoot, "Room impulse response interpolation using a sparse spatio-temporal representation of the sound field," *IEEE/ACM TASLP*, 25(10): 1929-1941, 2017.

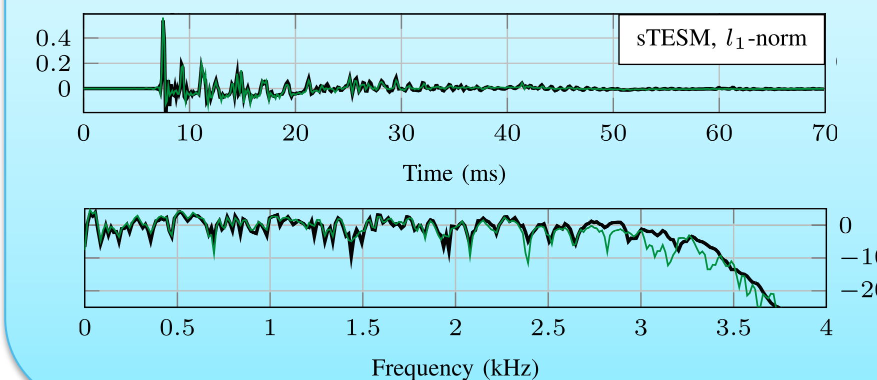
Measurement

- **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**

$$\min_{\mathbf{W}} \|D(\mathbf{W}) - \tilde{\mathbf{P}}\|_F^2 + \lambda \|\mathrm{vec}(\mathbf{W})\|_1$$



Interpolated RIR + magnitude response*



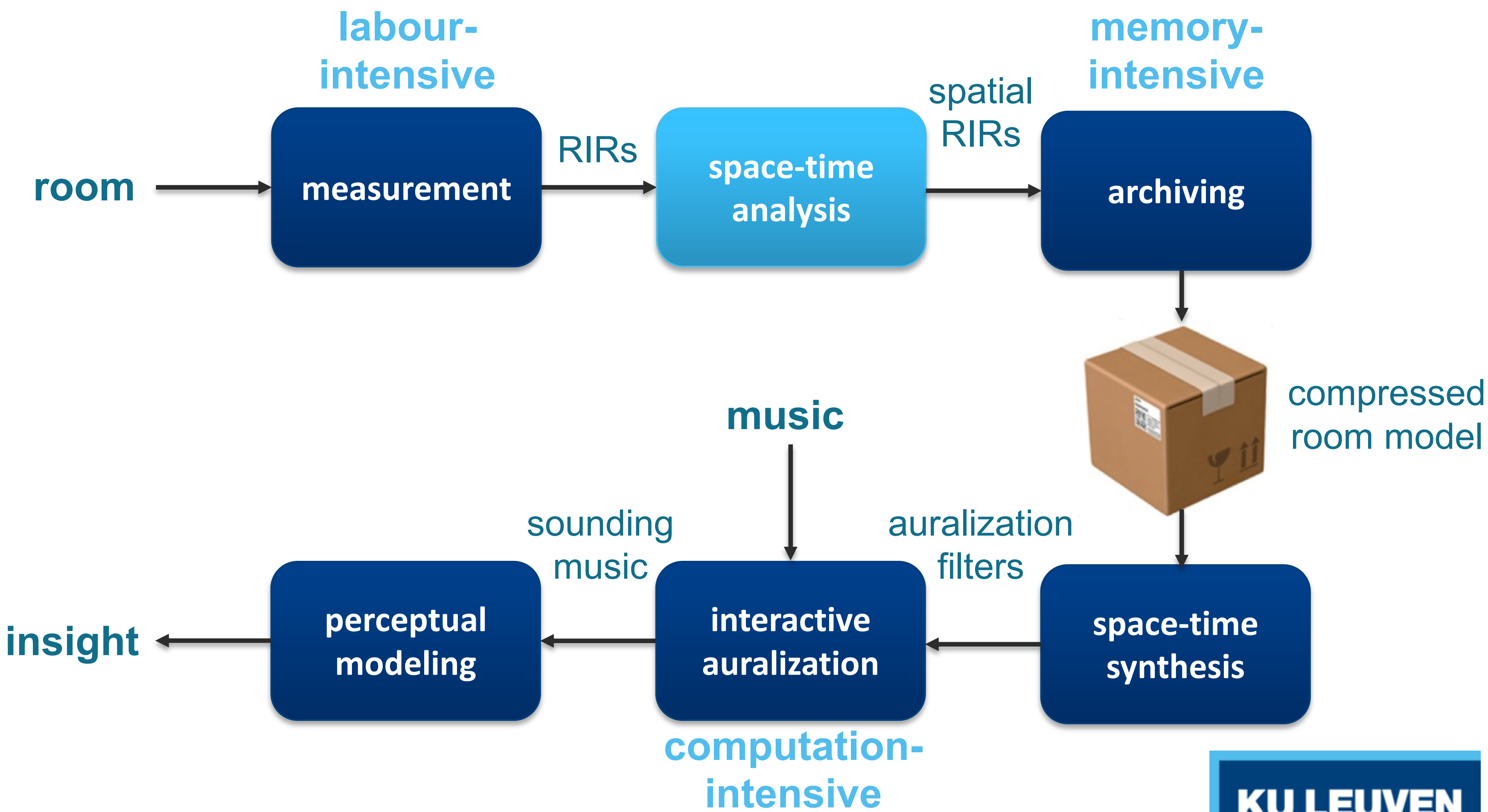
*worst case result
**average result

KU LEUVEN

Image credit: Niccolò Antonello – *Reference:* N. Antonello, E. De Sena, M. Moonen, P. A. Naylor, and T. van Waterschoot, "Room impulse response interpolation using a sparse spatio-temporal representation of the sound field," *IEEE/ACM TASLP*, 25(10): 1929-1941, 2017.

Outline

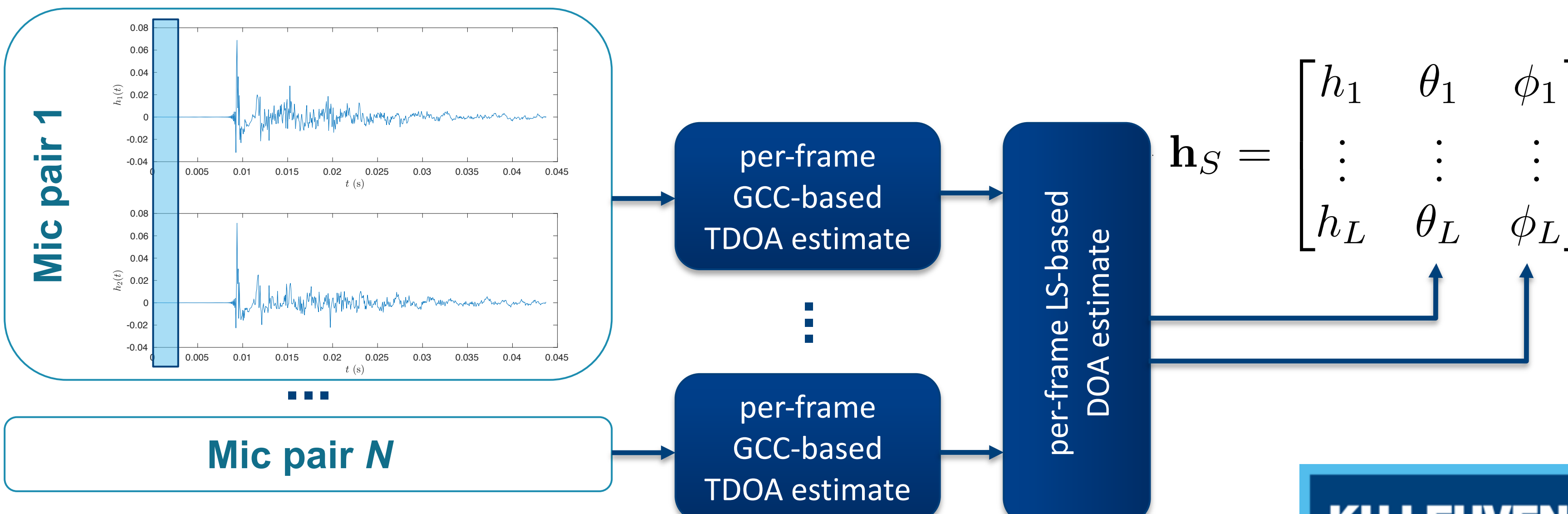
- **System overview & challenges**



KU LEUVEN

Space-time analysis

- **Key model: Spatial room impulse response (SRIR)**
  - Include **direction of arrival information** (azimuth and elevation angles θ,ϕ) 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



KU LEUVEN

Image credit: TvW – *References:* J. Merimaa and V. Pulkki, “Spatial impulse response rendering I: Analysis and synthesis,” *JAES*, 53(12): 1115-1127, 2005. – S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, “Spatial decomposition method for room impulse responses,” *JAES*, 61(1/2): 17–28, 2013.

Space-time analysis

- **SDM challenges**
  - Find phase-matched microphones 💸💸💸
  - Find RIR "peaks"
  - Match "peaks" in different RIRs ⇢ *generalized cross-correlation (GCC)*
  - Estimate TDOAs with subsample accuracy ⇢ *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

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

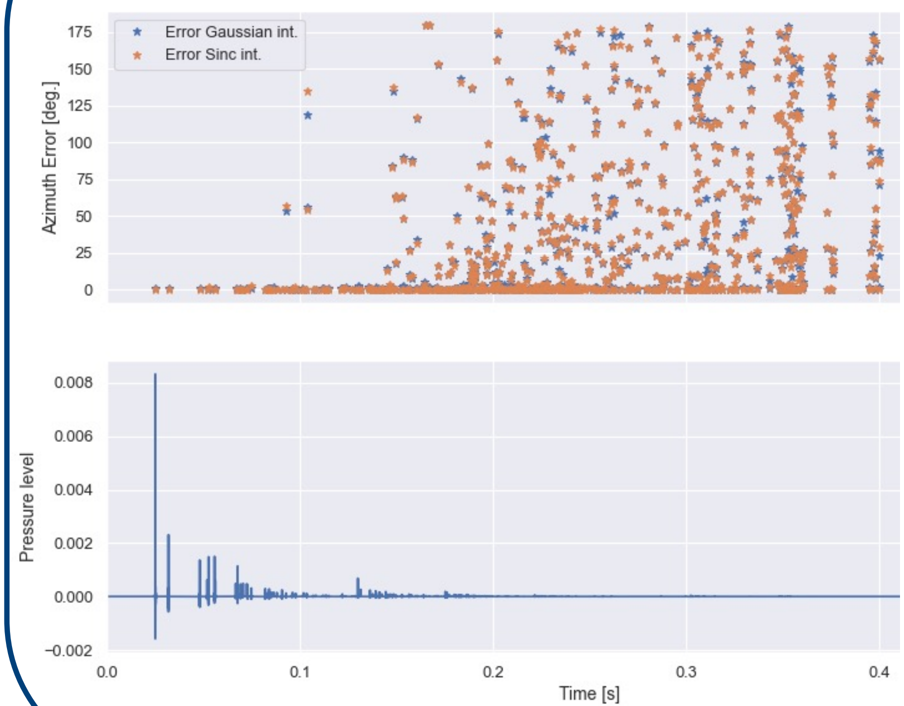
KU LEUVEN

Image credit: G.R.A.S. – *Reference:* H. Rosseel and T. van Waterschoot, "Improved acoustic source localization by time delay estimation with subsample accuracy," in *Proc. Int. Conf. Immersive & 3D Audio (I3DA '21)*, Bologna, Sep. 2021.

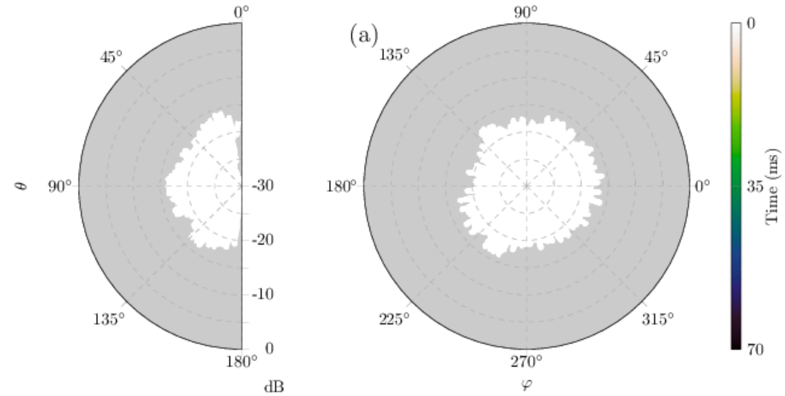
Space-time analysis

- **Resulting space-time model**

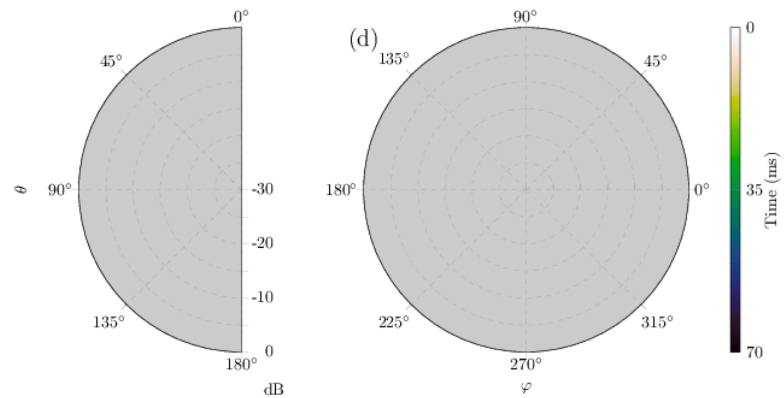
TDOA estimation errors



SRIR visualization of **interpolated** RIRs



without sparse regularization



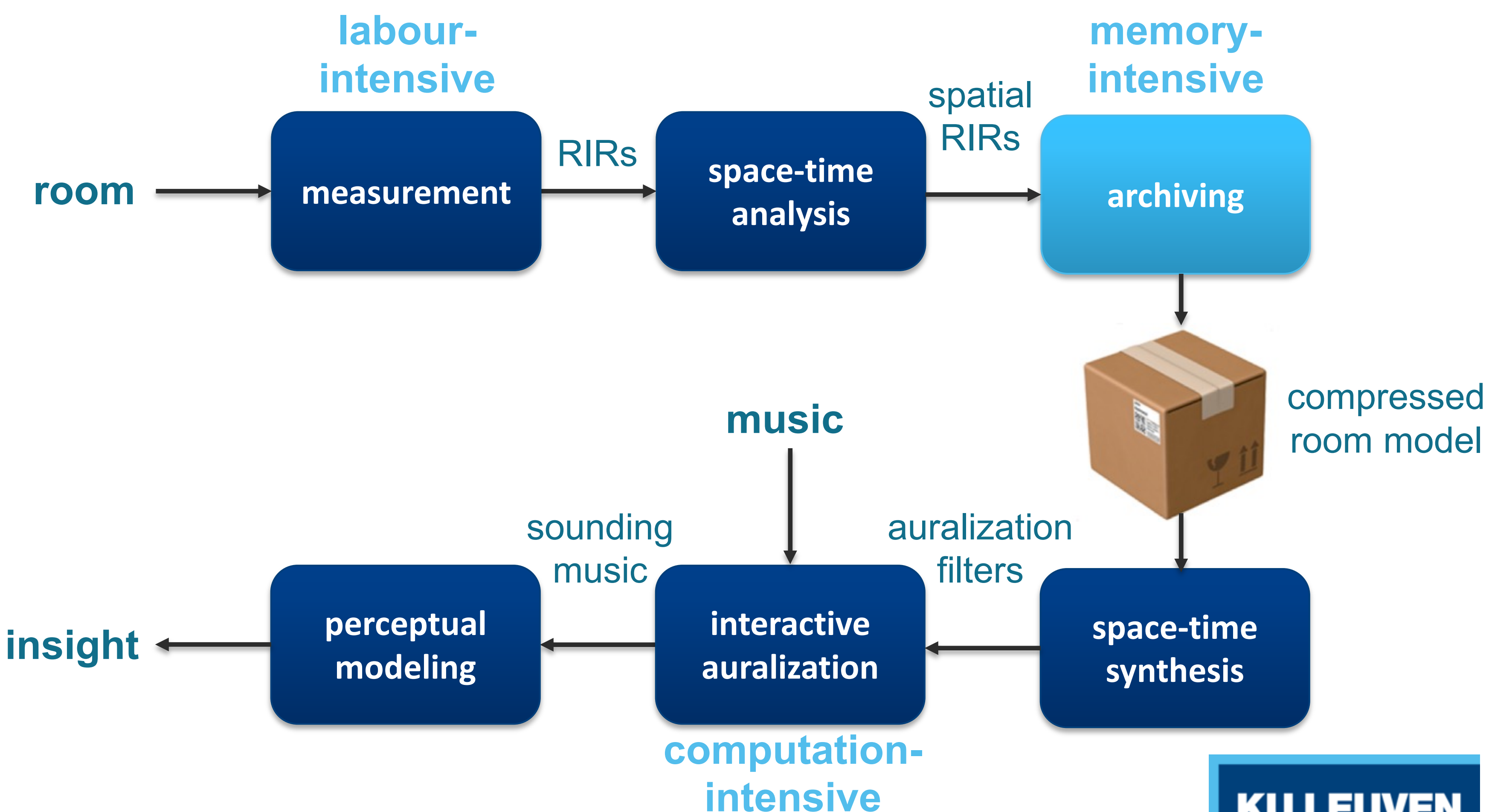
with sparse regularization

KU LEUVEN

Image credits: H. Rosseel, N. Antonello – *Reference:* N. Antonello, E. De Sena, M. Moonen, P. A. Naylor, and T. van Waterschoot, ”Room impulse response interpolation using a sparse spatio-temporal representation of the sound field,” *IEEE/ACM TASLP*, 25(10): 1929-1941, 2017.

Outline

- **System overview & challenges**



KU LEUVEN

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 ~ 10 cm

    ~ $(100)^3$ source positions x $(100)^3$ observer positions

    ~ 10^{12} spatial impulse responses ~ 10^{18} bits ~ **100 petabyte**

Image credit: Elbowroom – *Reference:* T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3790 –3804, Oct. 2006.

Archiving

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

Physical justification

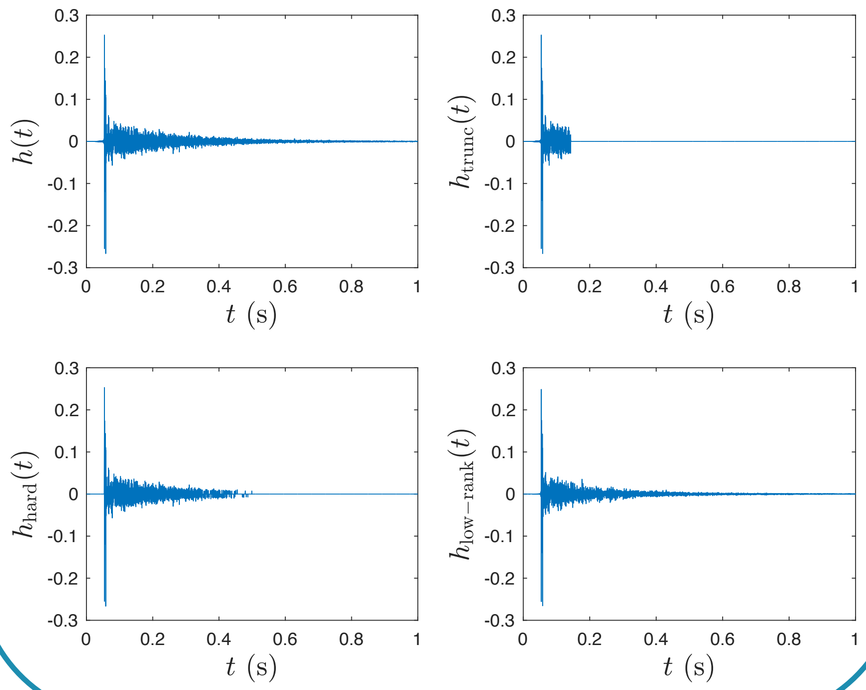
non-homogenous Helmholtz equation

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

RIR ≈ sum of damped sinusoids

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

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

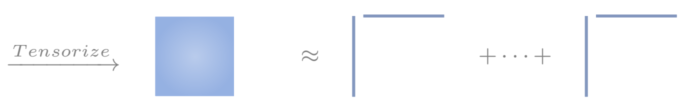


Archiving

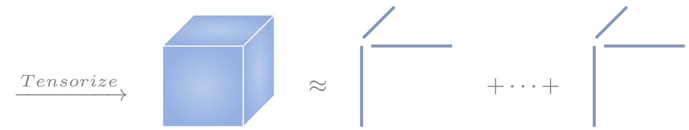
- **Key idea:** **Sum of *M* (damped) sinusoids admits matricization/tensorization of rank 2*M***
  - Impulse response vector **h** (= 1-D array) of length *L* can be **reshaped** into *N*-D array of dimensions $\sqrt[N]{L} \times \sqrt[N]{L} \times \ldots \times \sqrt[N]{L}$
  - *N*-D array can be **approximated**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

$$\mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ \vdots \\ h_L \end{bmatrix}$$

example: 2-D tensorization



3-D tensorization

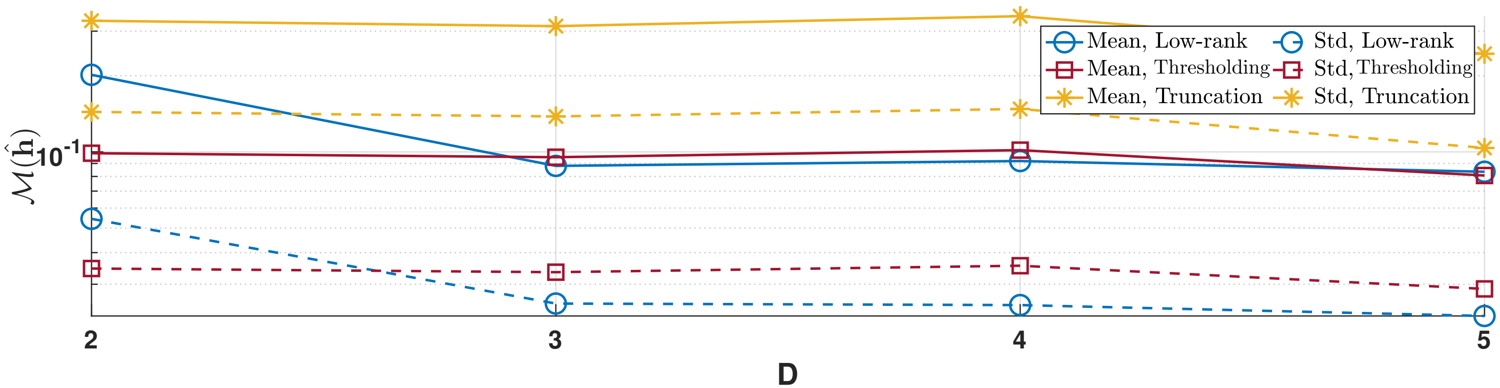


$$\mathbf{h} = \mathrm{vec}\left(\sum_{m=1}^{2M} \mathbf{u}_m^{(1)} \circ \mathbf{u}_m^{(2)} \circ \ldots \circ \mathbf{u}_m^{(N)} \right)$$

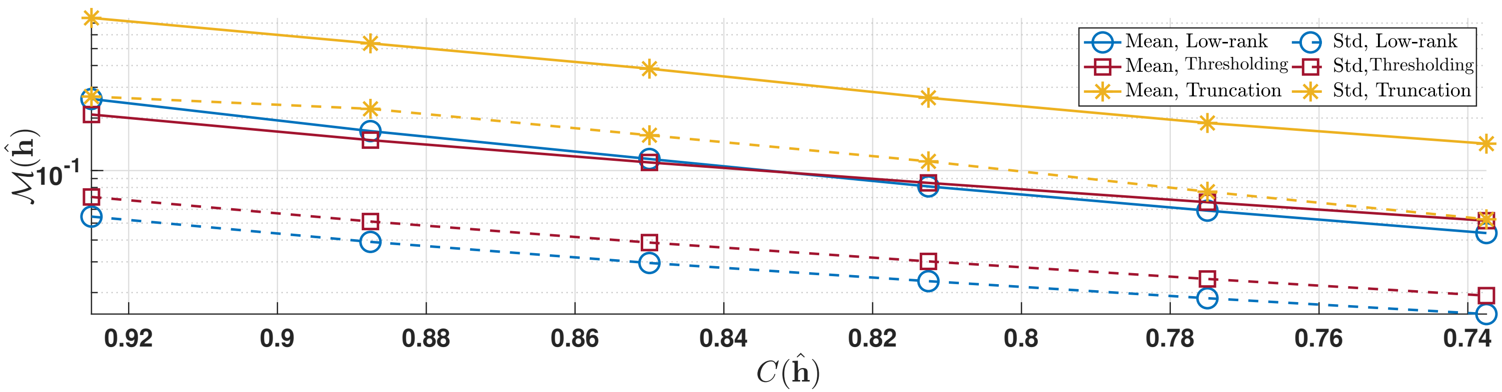
Image credit: Martijn Boussé – *Reference:* M. Jälmby, F. Elvander, and T. van Waterschoot, “Low-rank tensor modeling of room impulse responses,” in *Proc. EUSIPCO 2021*, pp. 111–115, 2021.

Archiving

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



C = 80 %



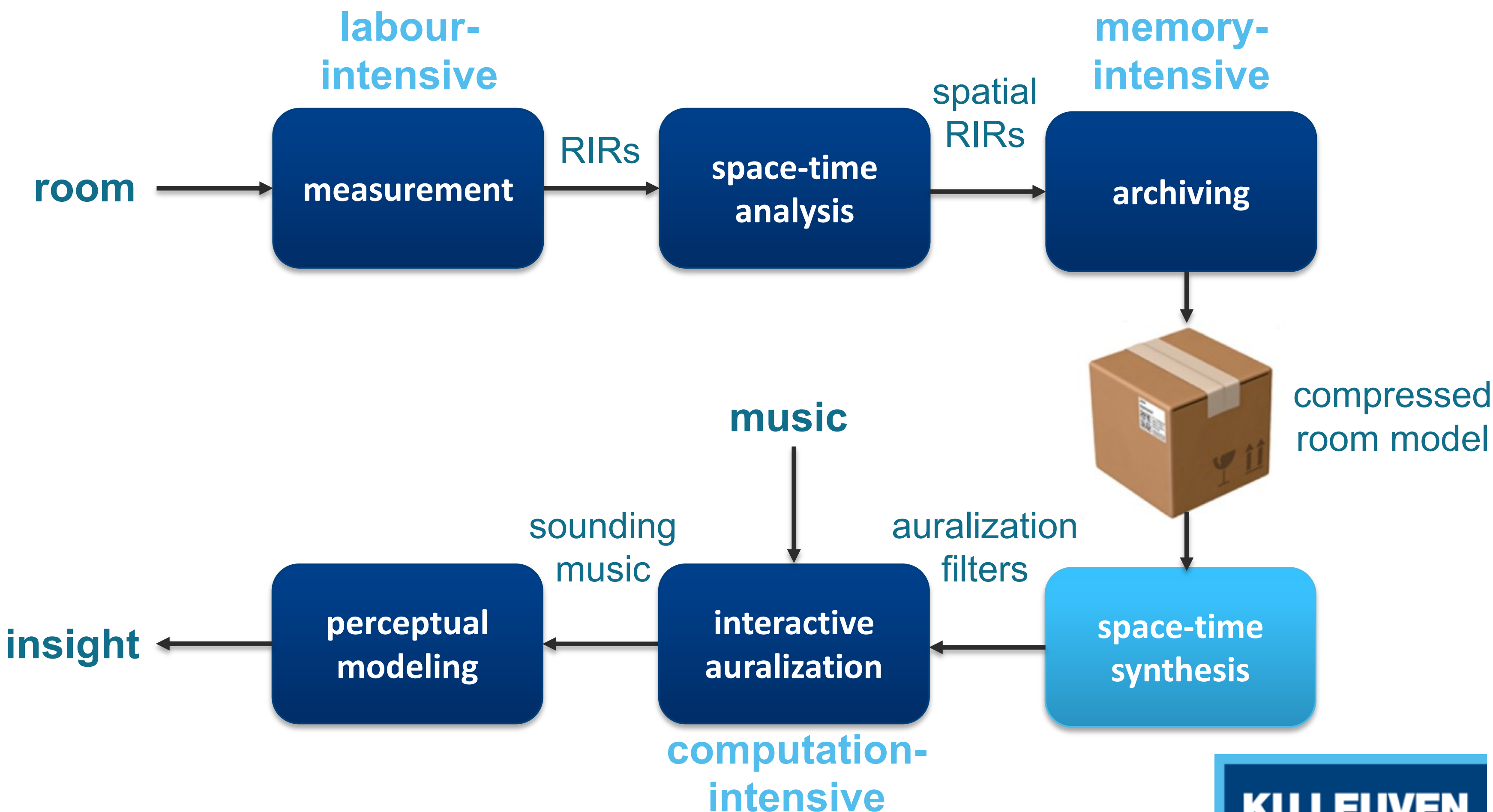
D = 3

KU LEUVEN

Image credit: Martin Jälmby – *Reference:* M. Jälmby, F. Elvander, and T. van Waterschoot, “Low-rank tensor modeling of room impulse responses,” in *Proc. EUSIPCO 2021*, pp. 111–115, 2021.

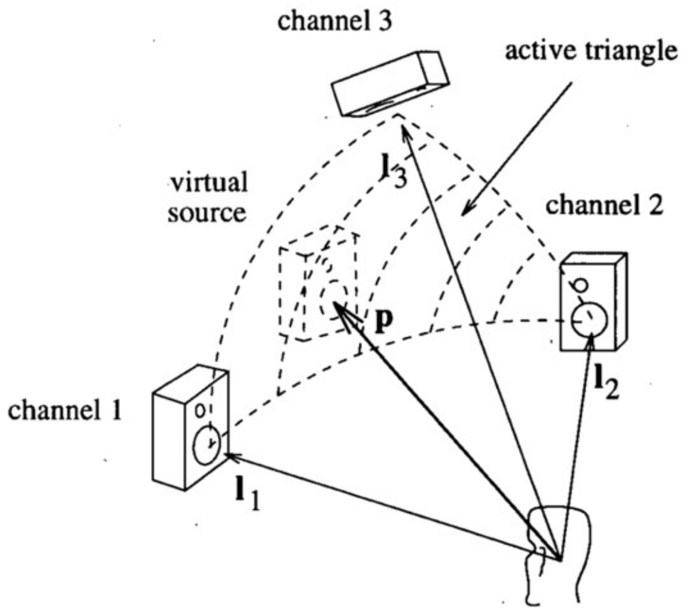
Outline

- **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



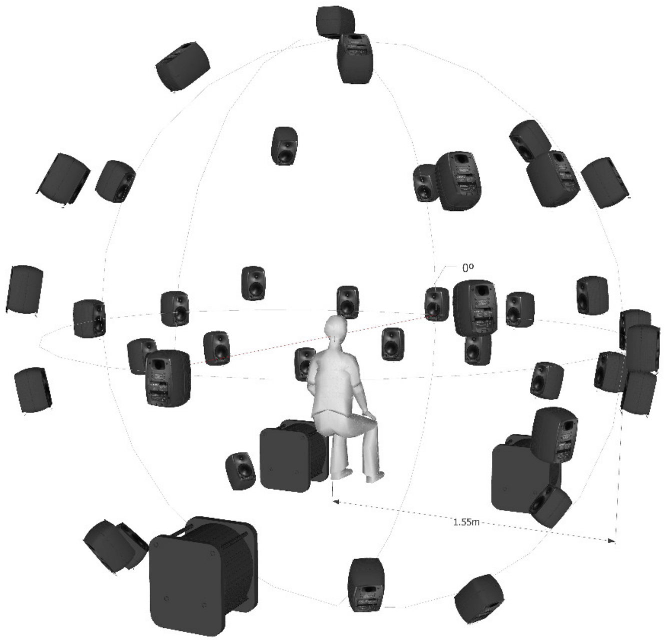
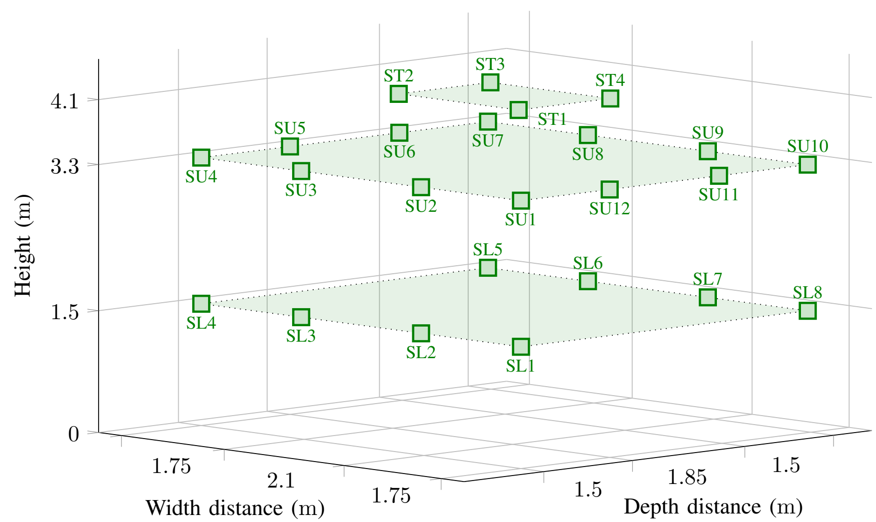


Image credits: Ville Pulkki, Neo Kaplanis – *Reference:* V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466, 1997.

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)

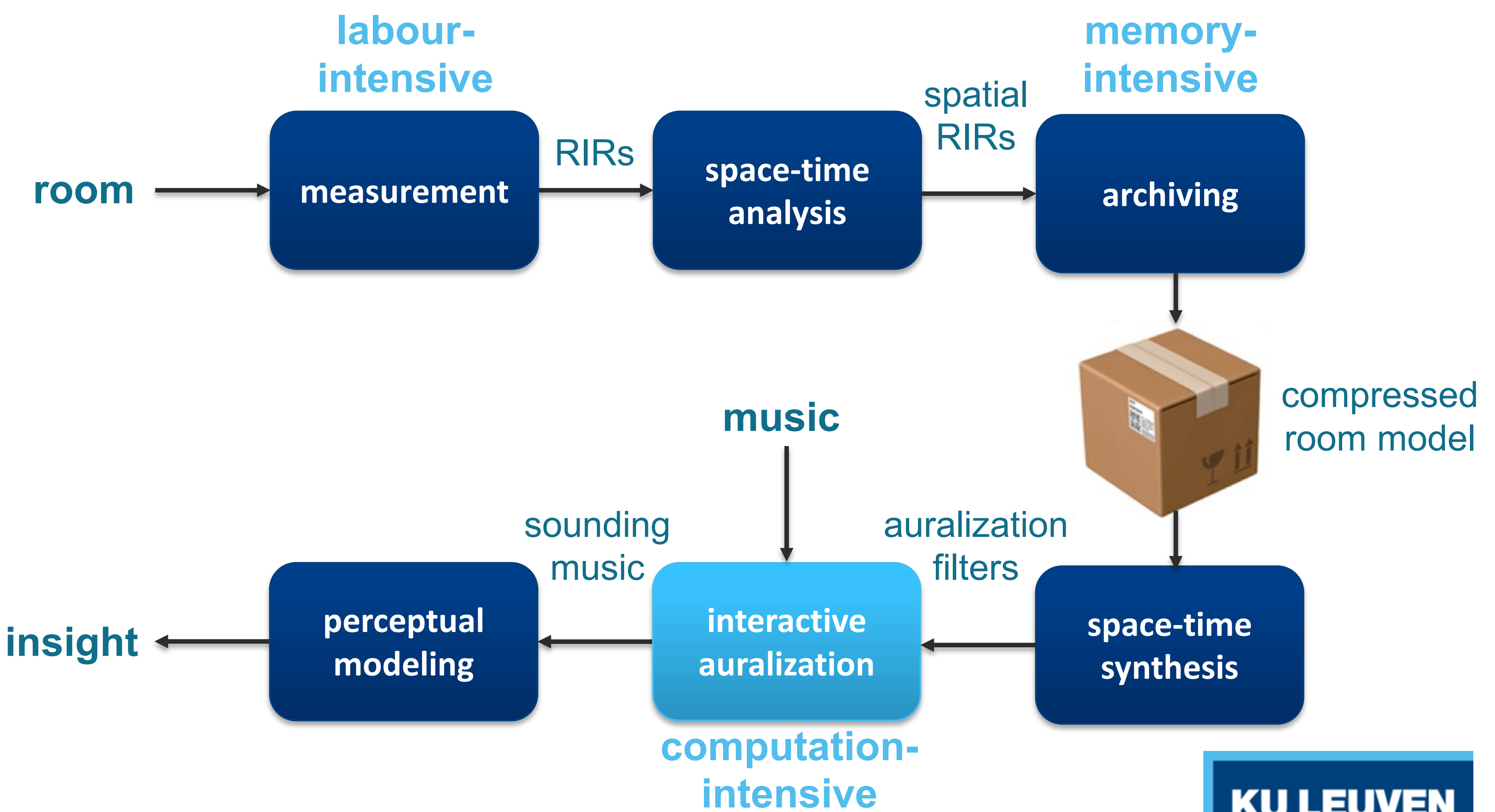


KU LEUVEN

Image credits: FACE, Thomas Dietzen – *Acknowledgement:* FWO Large-scale research infrastructure "The Library of Voices - Unlocking the Alamire Foundation’s Music Heritage Resources Collection through Visual and Sound Technology" (I013218N)

Outline

- **System overview & challenges**



Interactive auralization

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

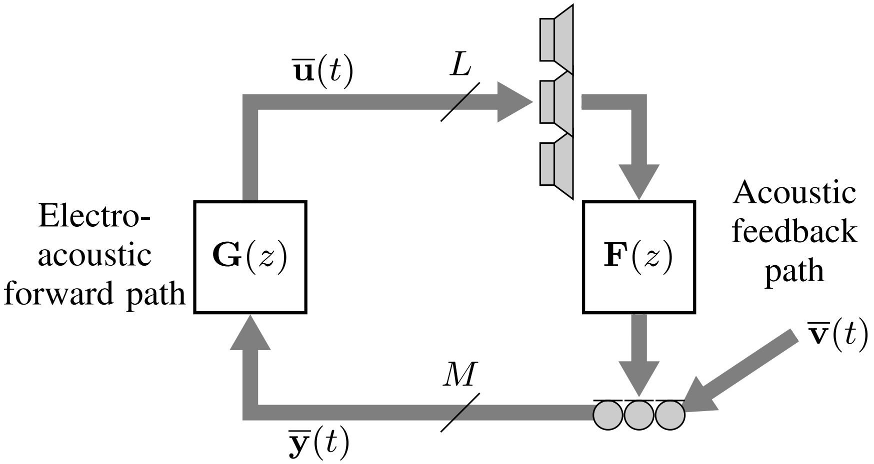
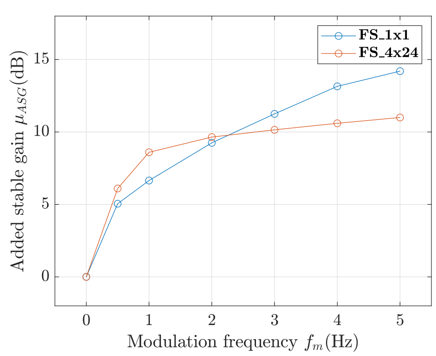
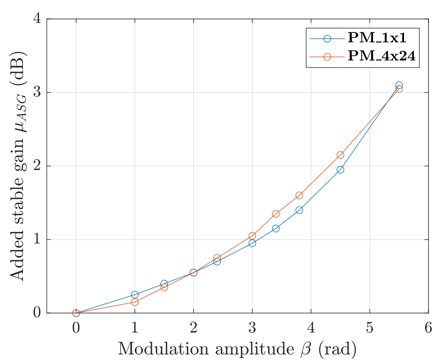


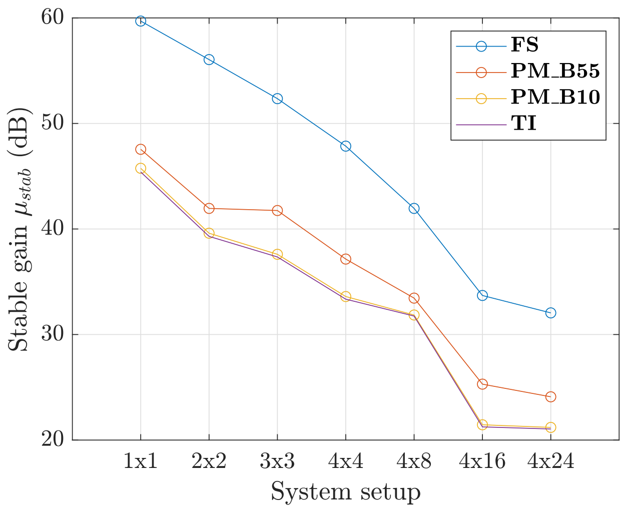
Image credits: TvW, David Pelegrin Garcia – *Audio credit:* TvW, Capella Pratensis

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 β=1.0 or 5.5

KU LEUVEN

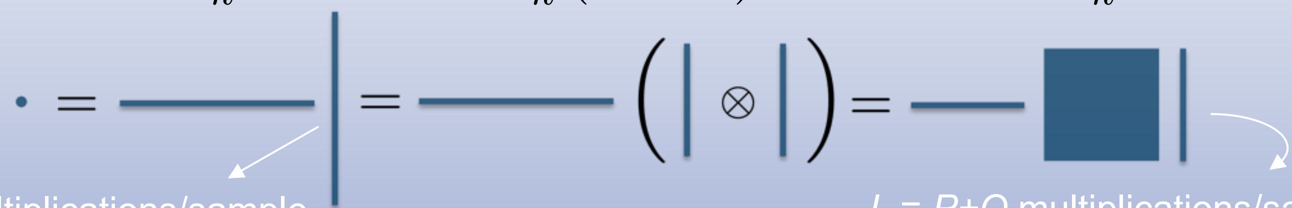
Image credit: Wannes Van Ransbeeck – *References:* M. A. Poletti, "The stability of multichannel sound systems with frequency shifting," *JASA*, 116(2): 853–871, Aug. 2004. – P. Svensson, "On reverberation enhancement in auditoria," Ph.D. dissertation, *Chalmers Univ. Technol.*, 1994. – T. van Waterschoot and M. Moonen, Fifty years of acoustic feedback control: state of the art and future challenges, *PIEEE*, 99(2): 288-327, 2011.

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 ⇢ shorter FIR filter
      - Hard thresholding ⇢ FIR filter with fewer non-zero coefficients
      - Sparse approximation ⇢ short "warped" FIR filter (mixed-Kautz model)
      - Low-rank approximation ⇢ "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

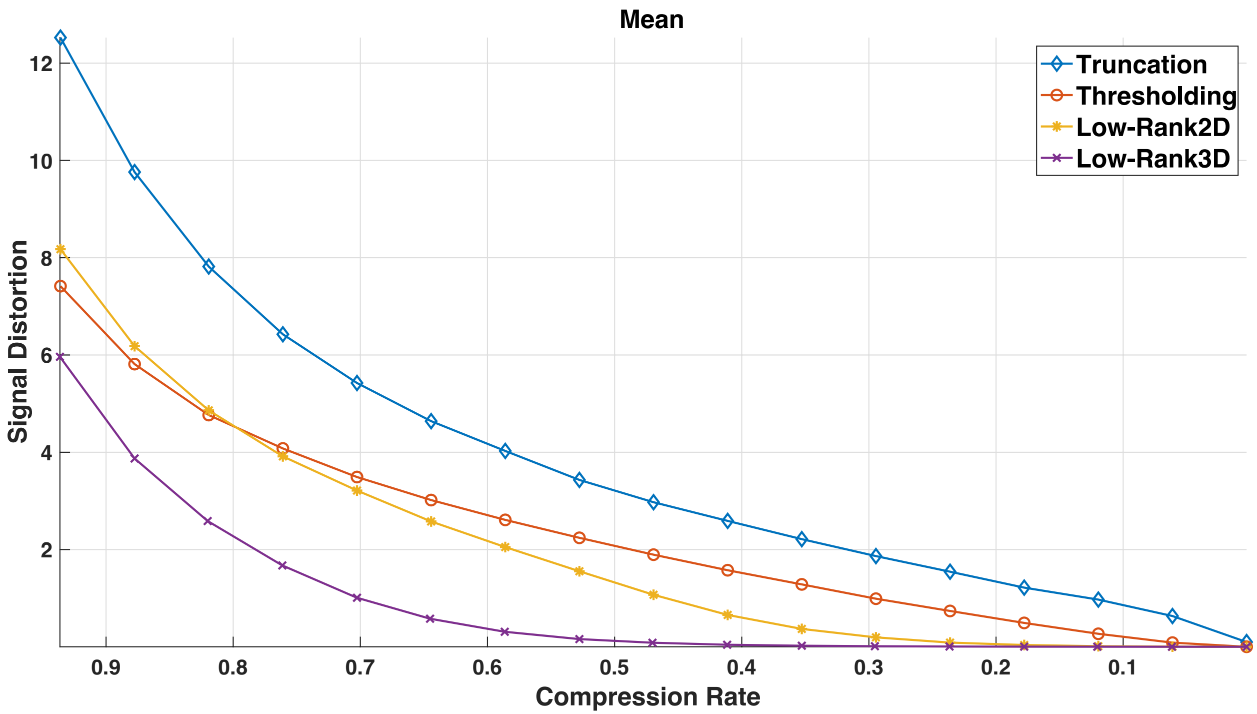
L = *P*+*Q* multiplications/sample

KU LEUVEN

Image credit: TvW – *References:* 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 TASLP*, 25(7): 1547-1561, 2017. – M. Jälmby and T. van Waterschoot, "Low-rank convolution," *Tech. Rep.*, KU Leuven, 2022.

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*

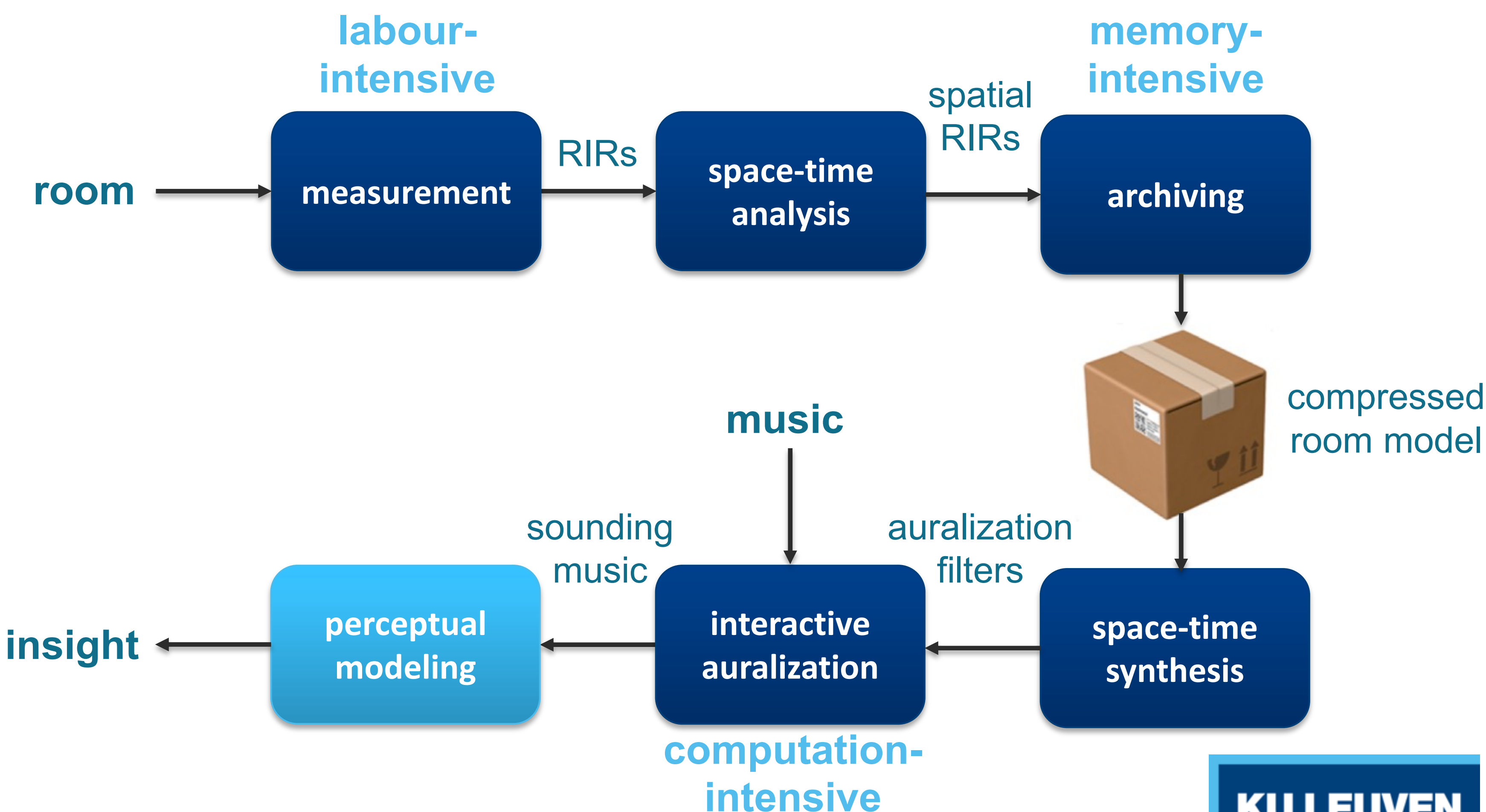


input = dry vocal music
RIR = Nassau Chapel RIR,
RT60 = 2.7 s, L = 117649

Image credit: Martin Jälmby – *Reference:* M. Jälmby and T. van Waterschoot, “Low-rank convolution,” *Tech. Rep.*, KU Leuven, 2022.

Outline

- **System overview & challenges**



KU LEUVEN

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

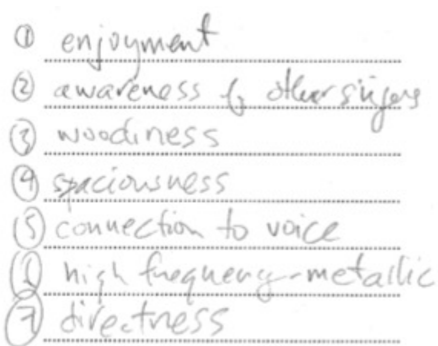
Image credit: TvW – *References:* V. Dairou and J. M. Sieffermann, “A comparison of 14 jams characterized by conventional profile and a quick original method, the flash profile”, *J. Food Sci*. 67, 2002. – N. Kaplanis, S. Bech, T. Lokki, T. van Waterschoot, and S. H. Jensen, Perception and preference of reverberation in small listening rooms for multi-loudspeaker reproduction, *JASA*, 146(5): 3562-3576, 2019.

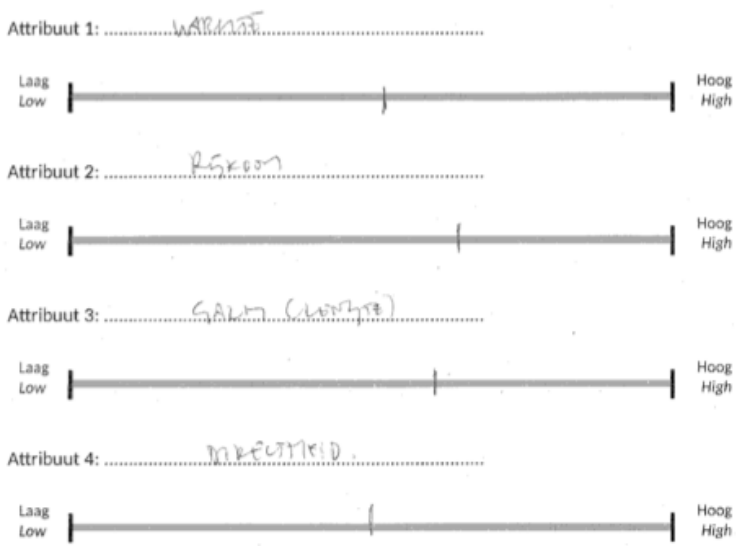
Perceptual modeling

- **Flash Profile Methodology:** experiment design
  - **1) Elicitation phase:** individual semantic definition of perceptual attributes to characterize differences between virtual spaces
  - **2) Ranking phase:** continuous-scale (low-high) quantification of each perceptual attribute for each virtual space

Expert 1

Expert 2

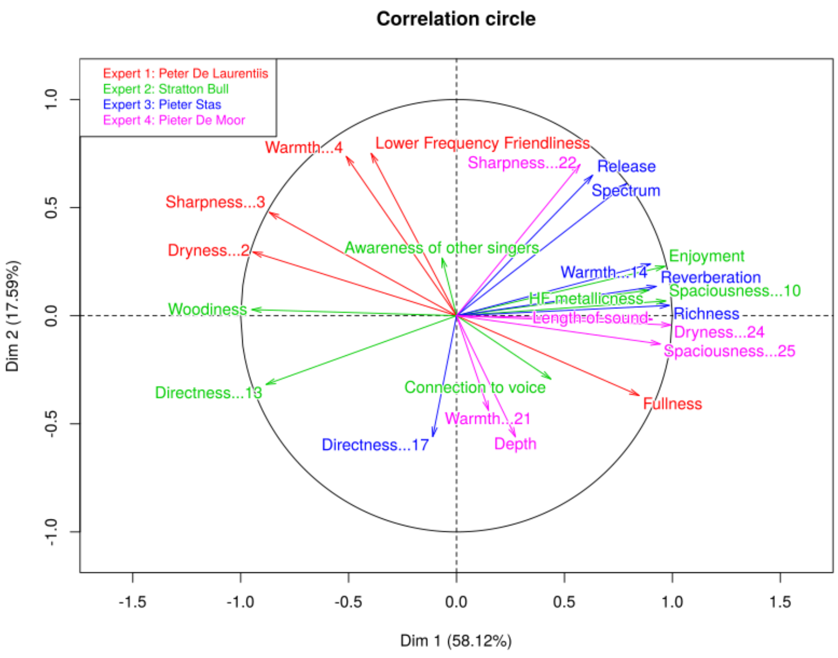




KU LEUVEN

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 prinicipal 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

KU LEUVEN

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**

KU LEUVEN

Slides: https://tvanwate.github.io/iwaenc2022.pdf

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 Jälmby, Rudi Knoops, Hannes Rosseel, Wannes Van Ransbeeck ... and many more
- **The work presented here has been carried out and supported by:**

KU LEUVEN

ALAMIRE FOUNDATION

fwo

DEPARTEMENT ECONOMIE WETENSCHAP & INNOVATIE

erc

- **Contact**

toon.vanwaterschoot@esat.kuleuven.be

https://tvanwate.github.io