Reverberation – Dereverberation
The promise of hybrid models

Gaël RICHARD*

Professor, Telecom Paris, Institut polytechnique de Paris

Dafx'25, Sept. 4th, 2025

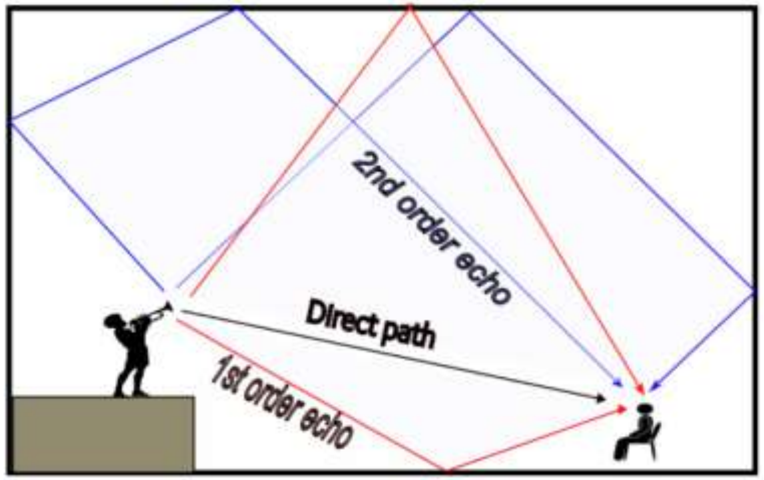
*work with collaborators and in particular H. Bai, L. Daudet, L. Bahrman, M. Fontaine

With support from *the European Union (ERC, HI-Audio - Hybrid and Interpretable Deep neural audio machines, 101052978).*
Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them*

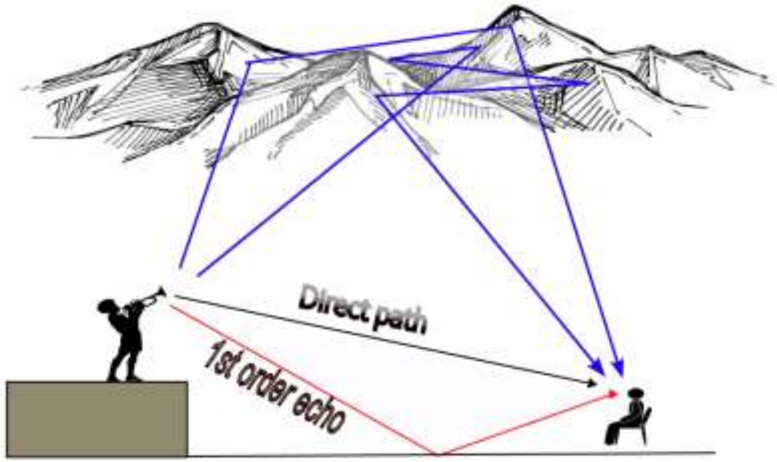
Reverberation : definition

G. Richard

- “In acoustics, **reverberation** is a persistence of sound after it is produced” [1]
- It is often created when a sound is reflected on surfaces, causing multiple reflections that build up and then decay as the sound is absorbed by the surfaces of objects in the space [2]



Reverberation in a room



Reverberation in an open space

[1] Wikipedia, from Valente, Michael; Holly Hosford-Dunn; Ross J. Roeser (2008). Audiology. Thieme. pp. 425–426. ISBN 978-1-58890-520-8.
[2] Wikipedia, from Lloyd, Llewelyn Southworth (1970). Music and Sound. Ayer Publishing. pp. 169. ISBN 978-0-8369-5188-2.

Situations with no reverberation

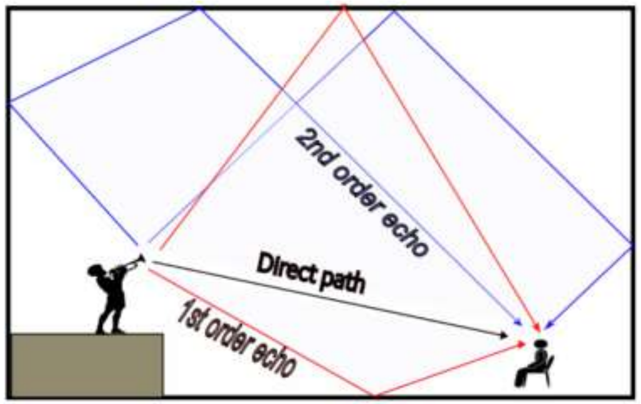
G. Richard

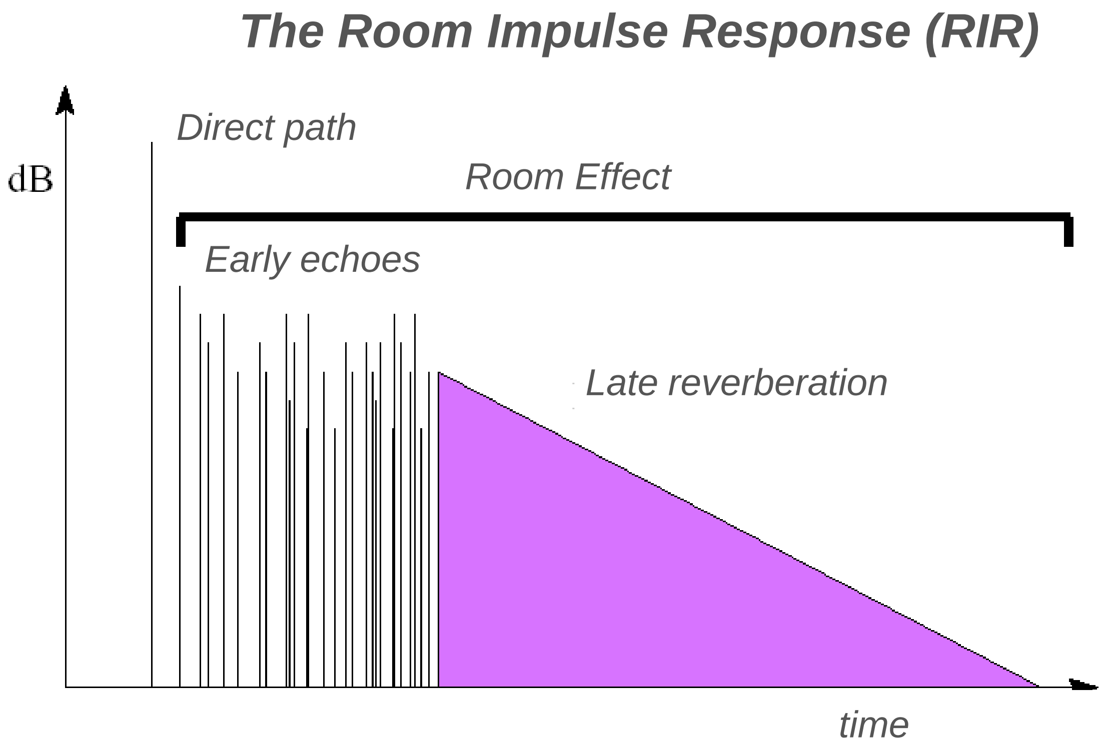
- When in an anechoic room …

- .. Or when in “free field”

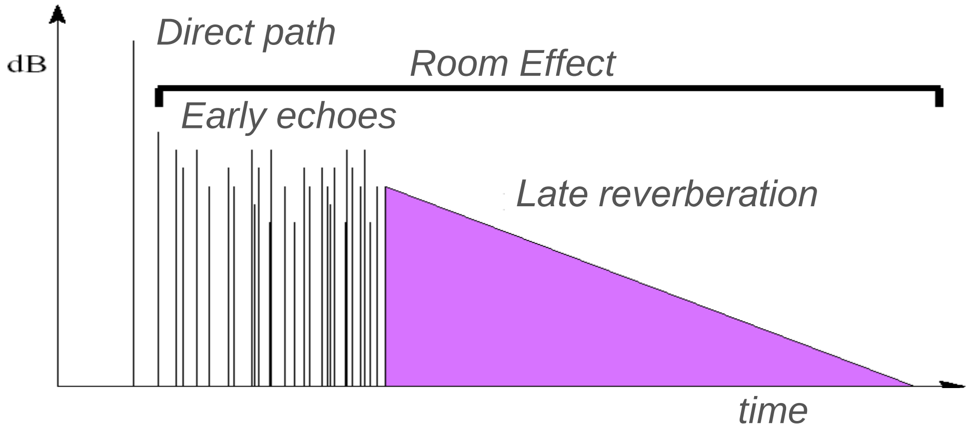
Reverberation: Room effect

- Room effect can be decomposed in:
  - A contribution due to **early echoes** or early reflexions (which depends on the room geometry and on the positions of the source and microphone)
  - A contribution due to **late reverberation** (which mainly depends on the volume and global absorption of the room)





Reverberation: Room effect



- Room effect = filtering effect

$$y(t) = \int_0^\infty x(t-u)h(u)du$$

- or

$$y(n) = \sum_{i=0}^\infty x(n-i)h(i)$$

The Room Impulse Response (RIR) (or acoustic channel)

Applications: Reverberation and Dereverberation

- **Artificial reverberation** : generating a new signal with different reverberation characteristics:

$$\hat{y}(n) = \sum_{i=0}^{\infty} x(n-i)h(i)$$

  - Applications:
    - Studio recordings and mixing
    - Live music (reverb pedals, synthesizers,…)
    - Virtual reality and movie production

- **Dereverberation:** removing the reverberation effect to retrieve the original source (or « dry » signal)

$$\text{"Recovering } \hat{x}(n) \text{ from the reverberated signal } y(n)\text{"}$$

  - Applications:
    - Speech enhancement (especially late reverberation removal to increase intelligibility)
    - Robust speech recognition
    - Acoustic transfer

Content

- What is reverberation ?
- **Artificial reverberation**
  - A long history ..
  - Towards hybrid models for late reverberation synthesis: From Radiance Transfer to Feedback Delay Networks
- **Dereverberation**
  - Short overview
  - Towards hybrid models for Weakly-Supervised (and Unsupervised) Speech Dereverberation
- **Conclusion**

Artificial reverberation: a long history …

(from [1, 2])

- **From analog devices in 1920's** …

(..transmit the sound into an empty acoustic space, and recording the response of the space via a microphone..)

  - *.. to spring resonator (late 1920's)..*
  - *..to plate reverberator, such as the EMT140 (in the 1950's)*
  - *.. to"Bucket-Brigade" Device (BBD), by Philips (in the 1960's)*

- .. ***to Digital methods ..***

  - ***…** delay networks* (as Schroeder reverberator in late 1960's) : (..the input signal is delayed, filtered and fed back along a number of paths according to parametrized reverberation characteristics)
  - … *convolutional* (typically, the input signal is convolved with a recorded or estimated impulse response of an acoustic space)
  - … *physical models* (typically the input signal drives a simulation of acoustic energy propagation in the modeled geometry).

- … **to deep learning** (e.g. data based) methods

  - ..for instance learning the parameters of a reverberation model using deep learning (as in [3])

[1] V. Valimaki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel. Fifty years of artificial reverberation. IEEE Transactions on Audio, Speech, and Language Processing, 20(5):1421–1448, 2012.
[2] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "More than 50 years of artificial reverberation," in proceedings of AES Conf. Feb. 2016.
[3] S. Lee, H. -S. Choi and K. Lee, "Differentiable Artificial Reverberation," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 2541-2556, 2022,

G. Richard

Reverberation - Dereverberation

Artificial reverberation

the interest for Hybrid methods

- On one side **Physics-based methods**
  - Accurate sound propagation modeling
  - Relatively high complexity
    - Image source method
    - Radiance transfer method
    - Beam tracing method, · · ·
- On the other side « **Perception** »**-based methods**
  - Computational efficiency
  - Not based on room geometry
    - Schroeder reverberation model
    - Feedback delay networks, · · ·
- The interest for **Hybrid methods**
  - Can link geometry-based models and perception-based models [1]
  - Can exploit machine learning (deep learning) to learn model parameters [3]

[1] H. Bai, G. Richard, and L. Daudet. "Geometric-based reverberator using acoustic rendering networks." In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pages 1–4, New Paltz, NY, October 2015.
[2] H. Bai, G. Richard, and L. Daudet. "Late reverberation synthesis : From radiance transfer to feedback delay networks." Audio, Speech, and Language Processing, IEEE/ACM Transactions on, 23(12) :2260–2271, 2015.
[3] S. Lee, H. -S. Choi and K. Lee, "Differentiable Artificial Reverberation," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 2541-2556, 2022,

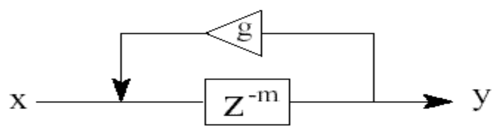
G. Richard

Artificial reverberation

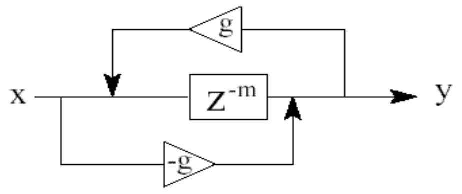
« Perception »-based methods

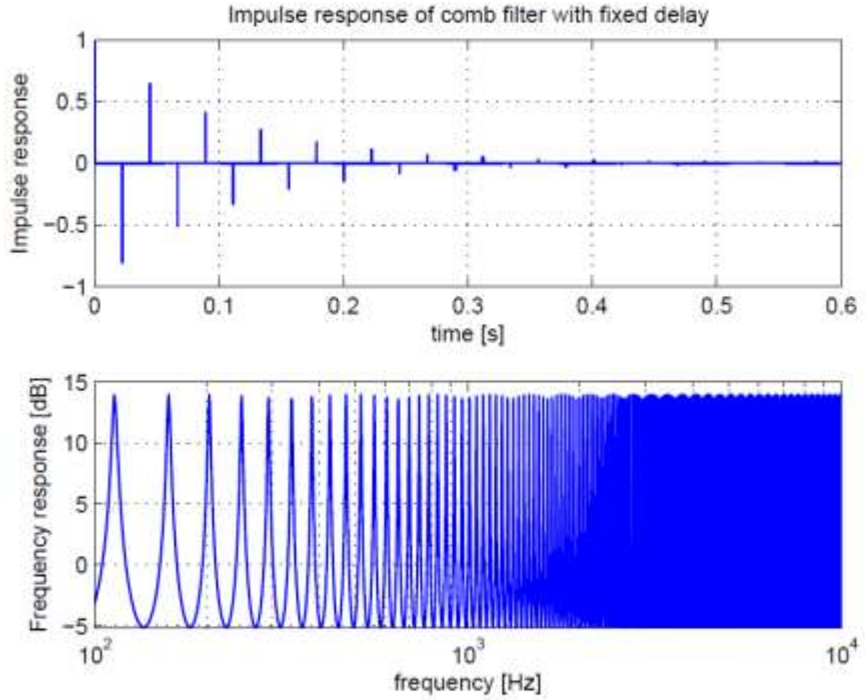
- Use of comb filters or/and all-pass filters

$$C(z) = \frac{z^{-m}}{1 - g z^{-m}}$$



$$A(z) = \frac{-g + z^{-m}}{1 - g z^{-m}}$$





- But, induce coloration… and low density of echos

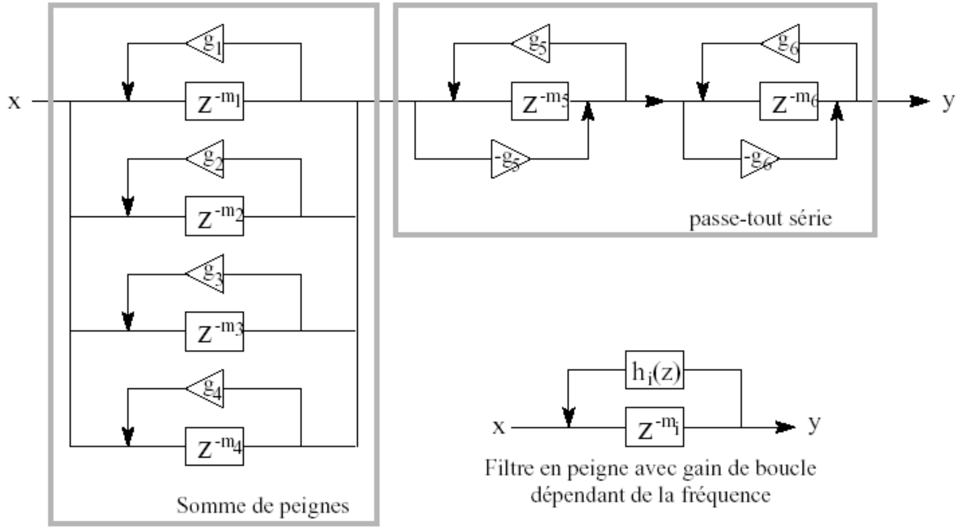
G. Richard

Reverberation - Dereverberation

Artificial reverberation

« Perception »-based methods

- The schroeder reverberation model



- Pameters can be related to reverberation time but not easily or directly to the exact room characteristics

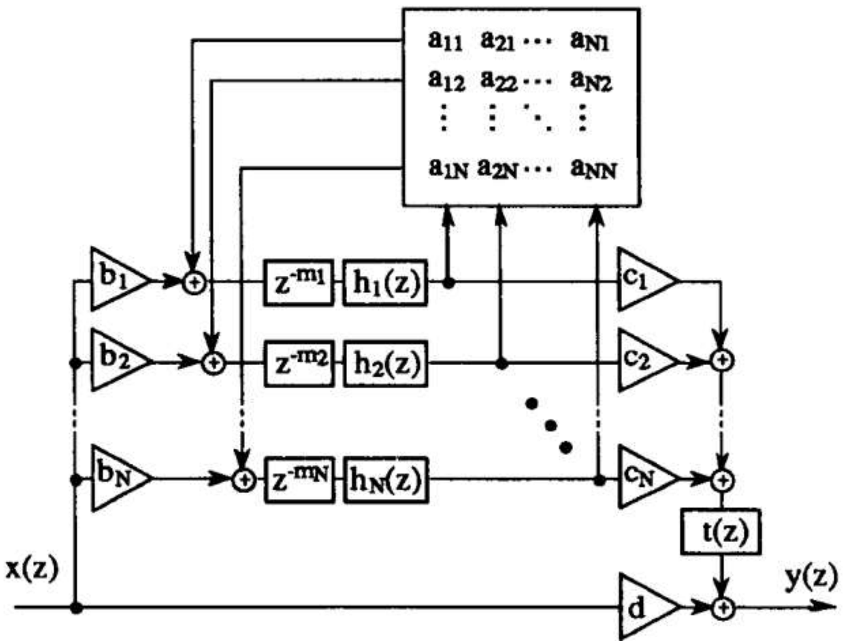
[1] M. R. Schroeder and B. F. Logan, "Colorless artificial reverberation",J. Audio Eng. Soc., vol. 9, no. 3, pp. 192–197, Jul. 1961.

G. Richard

Artificial reverberation

« Perception »-based methods

- Generalisation to feedback Delay networks



- ….still parameters not easily or directly linked to the exact room characteristics

[1] J.-M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," in Proc. 90th Conv. Audio Eng. Soc., Paris, France, Jan. 1991.

Artificial reverberation

« Perception »-based methods

- Another interesting model : The statistical polack model [1]
  - **Assumption:** the reverberant tail of a room impulse response (RIR) can be modeled as an exponentially decaying stochastic process.

$$h_r(n) = b(n)e^{-n/\tau},$$

  - With $b(n) \sim \mathcal{N}(0, \sigma^2)$ and $\tau = \frac{\mathrm{RT}_{60} f_s}{3 \ln(10)}.$

  - ….still parameters not easily or directly linked to the exact room characteristics
  - … but links with the statistical wave-field theory [2] can be made

[1] J.-D. Polack, "La transmission de l'energie sonore dans les salles," Ph.D. dissertation (in French), Université du Maine, 1988
[2] R. Badeau. Statistical wave field theory. *Journal of the Acoustical Society of America*, 2024, 156 (1), pp.573 - 599.

Artificial reverberation

« Physics-based methods »

- **Solving (or approximately) solving the wave equation**
  - Finite-difference time-domain (FDTD) method (in time domain)
  - Finite element method (FEM) and boundary element method (BEM) (in the frequency domain).
  - Statistical wave-field theory [1]
  - …

- **Geometrical acoustics**
  - Image-source method [2]
  - Ray-tracing [3]
  - Radiance transfer [4] extended to also represent specular reflexions [5]
  - …

[1] R. Badeau. Statistical wave field theory. *Journal of the Acoustical Society of America*, 2024, 156 (1), pp.573 - 599.
[2] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Amer., vol. 65, no. 4, pp. 943–950, Apr. 1979
[3] A. Krokstad, S. Strøm, and S. Sørsdal, "Calculating the acoustical room response by the use of a ray tracing technique," J. Sound Vibr., vol. 8, no. 1, pp. 118–125, Jan. 1968.
[4] T. Lewers. A combined beam tracing and radiatn exchange computer model of room acoustics. Applied Acoustics, 38(2):161–178, 1993.
[5] S. Siltanen, T. Lokki, and L. Savioja, "Frequency domain acoustic radiance transfer for real-time auralization," Acta Acustica United With Acustica, vol. 95, no. 1, pp. 106–117, Jan. 2009.

An alternative method (physical based):

Radiance Transfer Method (RTM)

- Radiance Transfer Method (RTM), a ray-based geometric method, can efficiently model the diffuse reflections of RIRs and the sound energy decay of the late reverberation.

- Analytical acoustic radiance transfer model

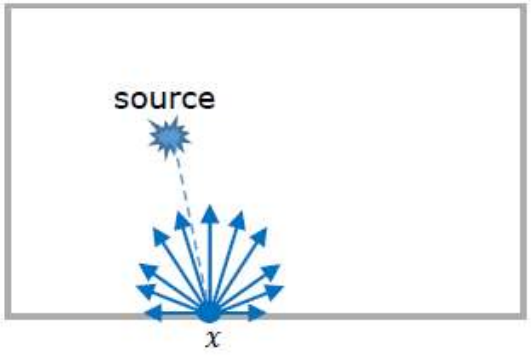
$$I(x,t) = I_0(x,t) + \int_S R(x,x',t) I(x', t - \frac{|x - x'|}{c}) dx'$$

G. Richard

Une approche alternative (physique):
Radiance Transfer Method (RTM)

- Analytical acoustic radiance transfer model

$$I(x,t) = \boxed{I_0(x,t)} + \int_S R(x,x',t)I(x',t-\frac{|x-x'|}{c})dx'$$



(a) Direct contribution

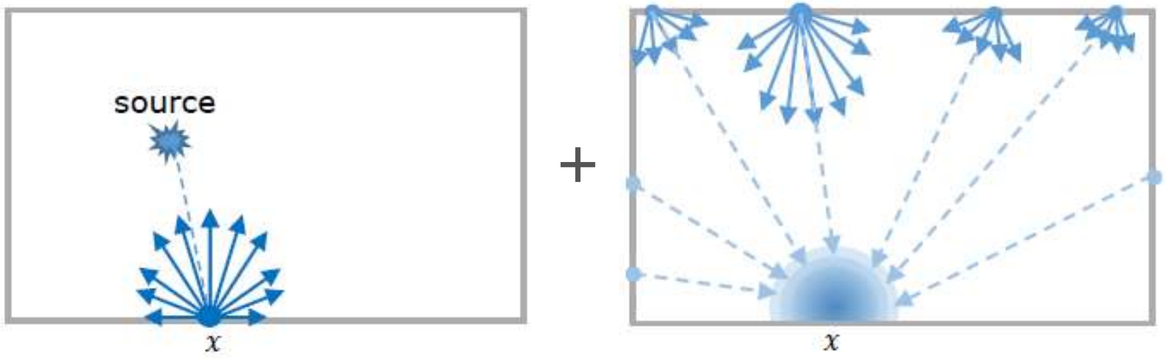
G. Richard

Une approche alternative (physique):

Radiance Transfer Method (RTM)

- Analytical acoustic radiance transfer model

$$I(x,t) = I_0(x,t) + \int_S R(x,x',t)\, I(x',t - \frac{|x - x'|}{c})\, dx'$$



(a) Direct contribution (b) Indirect contribution

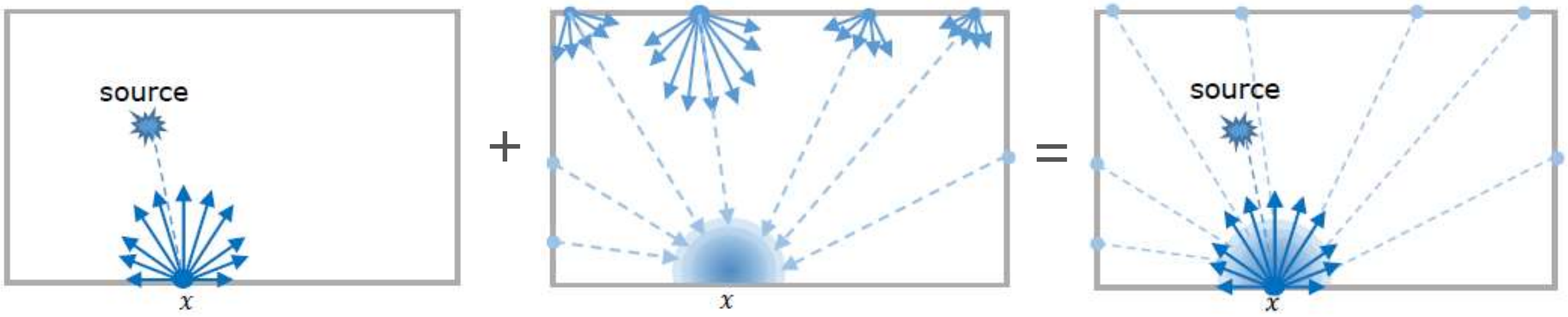
G. Richard

Une approche alternative (physique):

Radiance Transfer Method (RTM)

- Analytical acoustic radiance transfer model

$$I(x,t) = I_0(x,t) + \int_S R(x,x',t) I(x',t - \frac{|x-x'|}{c}) dx'$$



(a) Direct contribution (b) Indirect contribution (c) Total radiance

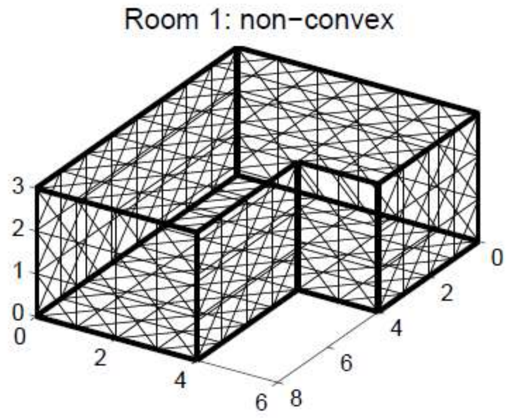
G. Richard

Reverberation - Dereverberation

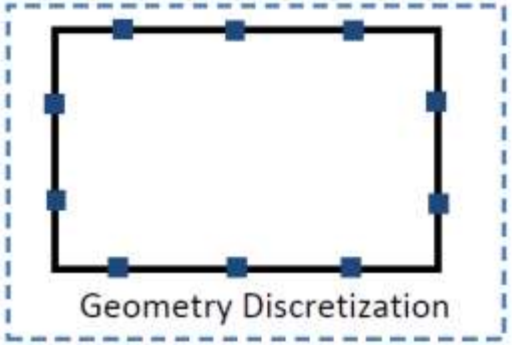
Radiance Transfer Method:

Digital simulation

- Room discretization
  - Room is divised in patches
  - Iterative expression



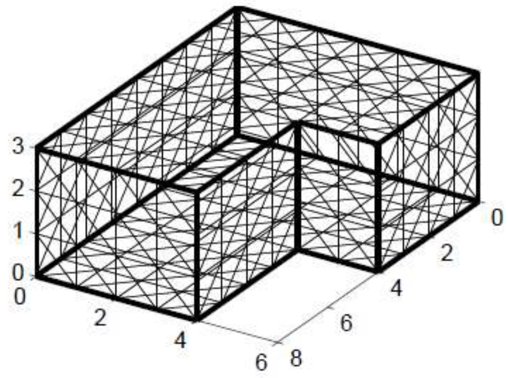
$$I_i^{(n)}(t) = I_i^{(n-1)}(t) + \sum_{j=1, j\neq i}^{M} F_{i,j}^{(1)} I_j^{(n-1)}(t - \frac{r_{i,j}}{c}))$$



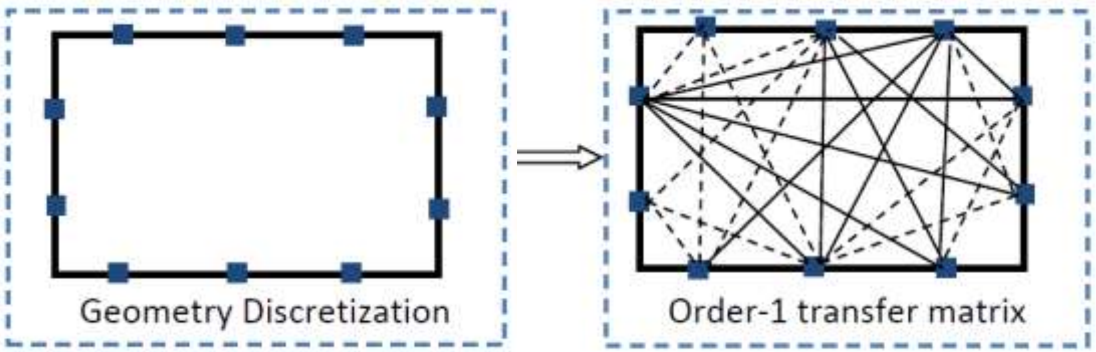
Radiance Transfer Method:

Digital simulation

- Room discretization
  - Room is divised in patches
  - Iterative expression



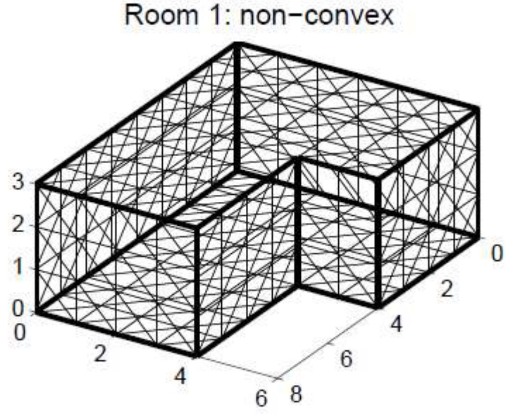
$$I_i^{(n)}(t) = I_i^{(n-1)}(t) + \sum_{j=1, j \neq i}^{M} F_{i,j}^{(1)} I_j^{(n-1)}\left(t - \frac{r_{i,j}}{c}\right)$$



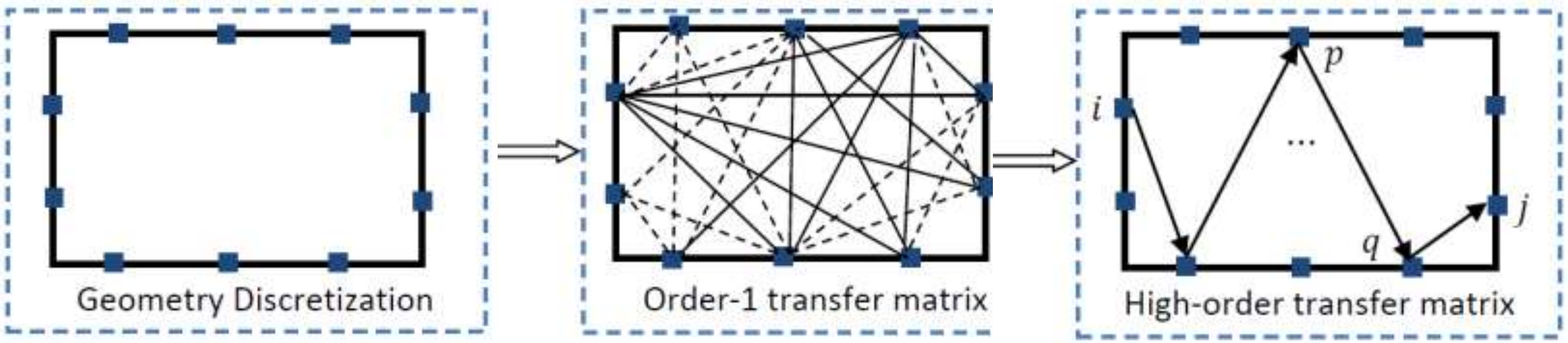
Radiance Transfer Method:

Digital simulation

- Room discretization
  - Room is divised in patches
  - Iterative expression



$$I_i^{(n)}(t) = I_i^{(0)}(t) + \sum_{j=1, j\neq i}^{M} F_{i,j}^{(n)} I_j^{(0)}(t - \frac{r_{i,j}}{c}))$$

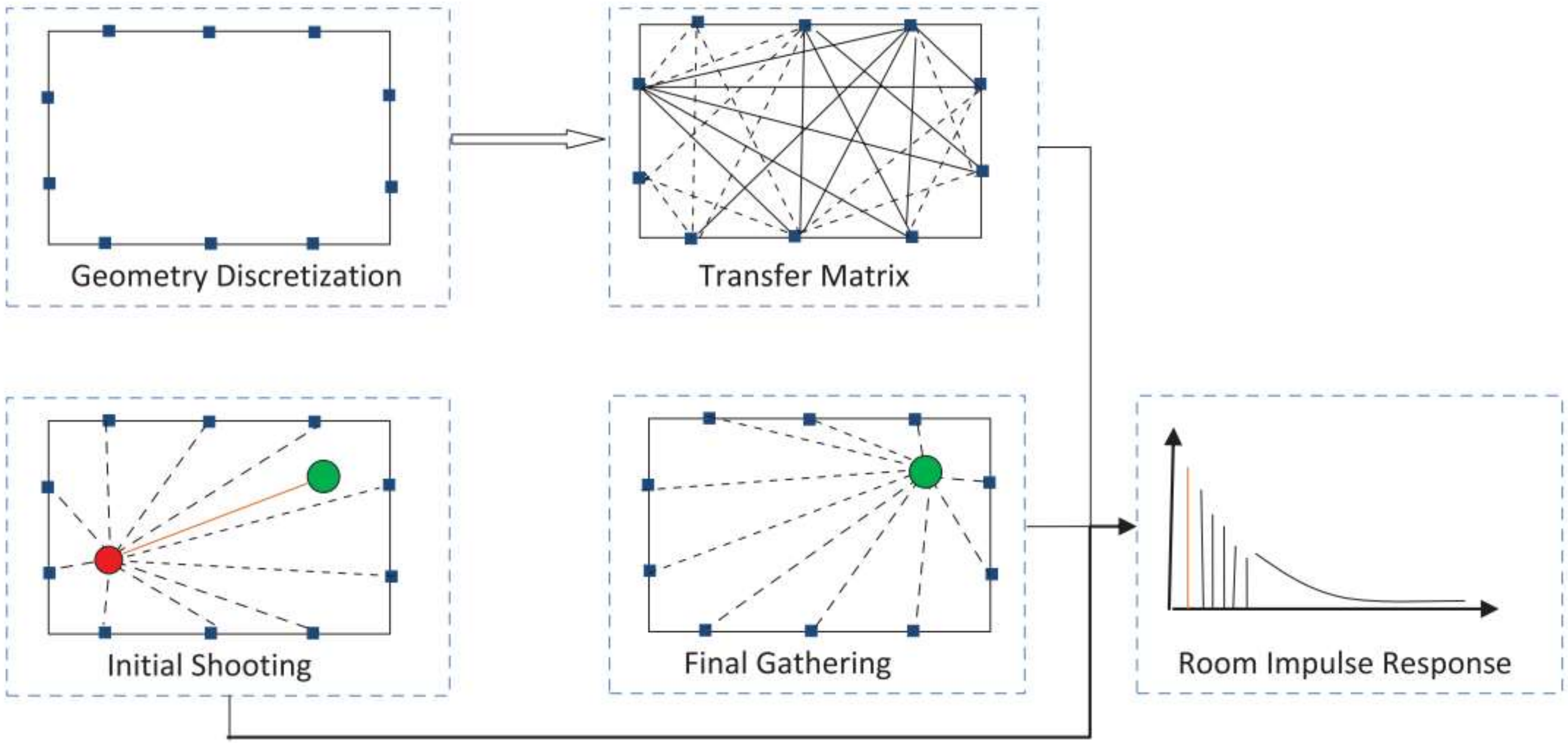


G. Richard

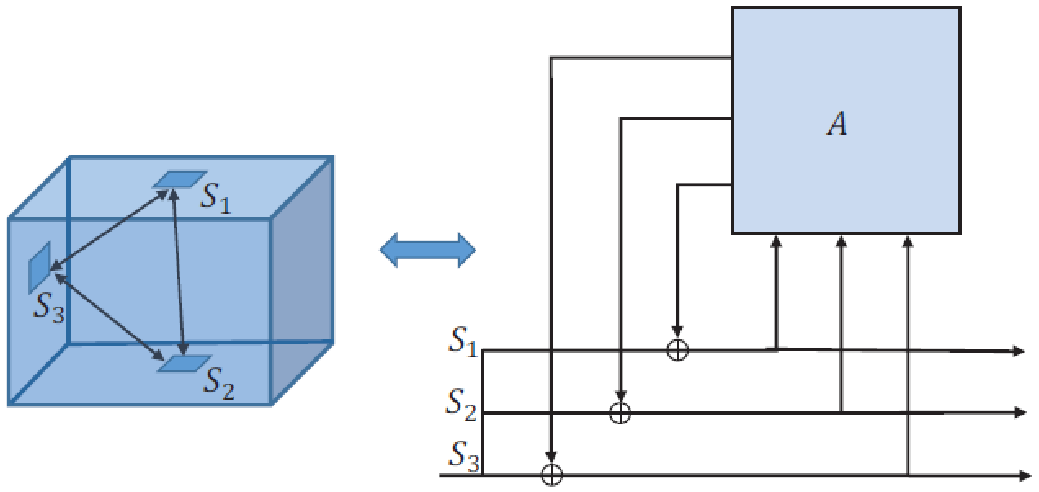
An alternative approach:

Radiance Transfer Method (RTM)

- In summary



Links between RTM and linear systems with reverberant filters



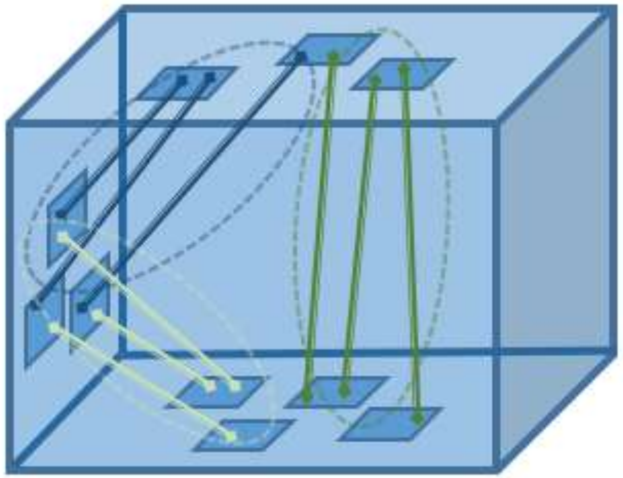
- The exchange of energy between patches of RTM can be linked to the recursive structure of the filter networks
- The exchange of energy of high order is equivalent to the infinite feedback loops of filter networks
- Brings efficient implementation of the RTM methods

[1] H. Bai, G. Richard, and L. Daudet. "Geometric-based reverberator using acoustic rendering networks." In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pages 1–4, New Paltz, NY, October 2015.
[2] H. Bai, G. Richard, and L. Daudet. "Late reverberation synthesis : From radiance transfer to feedback delay networks." Audio, Speech, and Language Processing, IEEE/ACM Transactions on, 23(12) :2260–2271, 2015.

The hybrid RTM-FDN method

(exploiting Radiance Transfer and Feedback Delay Networks)

- High number of patches leads to a very high number of delay lines ….
- For late reverberation, it is possible to group patch-to-patch interactions



- **Mapping:**
  - Each group ⇔ delay line of feedback delay networks
  - Parameters are the statistical average of each group

[1] H. Bai, G. Richard, and L. Daudet. "Geometric-based reverberator using acoustic rendering networks." In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pages 1–4, New Paltz, NY, October 2015.
[2] H. Bai, G. Richard, and L. Daudet. "Late reverberation synthesis : From radiance transfer to feedback delay networks." Audio, Speech, and Language Processing, IEEE/ACM Transactions on, 23(12) :2260–2271, 2015.

The hybrid RTM-FDN method

(exploiting Radiance Transfer and Feedback Delay Networks)

- Feedback coefficients:
  - a_{mn} describes the proportion of energy transported by the patch-to-patch interactions within group m, that will be diffusely reflected and go to some other patch-to-patch interactions in group n (order 1 diffuse reflection).
  - $\ell_m = \sum_{i \to j \in \wp_m} F_{i,j},$: the total energy transported by the patch-to-patch energy exchange in group *m*
  - $\ell_{m,n} = \sum_{i \to j \in \wp_m} \sum_{j \to k \in \wp_n} F_{i,j} F_{j,k}$: the total energy received by group *n* from the diffuse reflections of group *m*

- We have:

$$a_{mn} = \frac{\ell_{m,n}}{\ell_m},$$

$$\textit{subject to} \sum_{n=1}^{N} a_{mn} = 1$$

and the feedback Matrix:

$$A = \begin{pmatrix} \sqrt{a_{11}} & \cdots & \sqrt{a_{1N}} \\ \vdots & \ddots & \vdots \\ \sqrt{a_{N1}} & \cdots & \sqrt{a_{NN}} \end{pmatrix}$$

- Other parameters can also be easily obtained

[1] H. Bai, G. Richard, and L. Daudet. "Geometric-based reverberator using acoustic rendering networks." In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pages 1–4, New Paltz, NY, October 2015.
[2] H. Bai, G. Richard, and L. Daudet. "Late reverberation synthesis : From radiance transfer to feedback delay networks." Audio, Speech, and Language Processing, IEEE/ACM Transactions on, 23(12) :2260–2271, 2015.

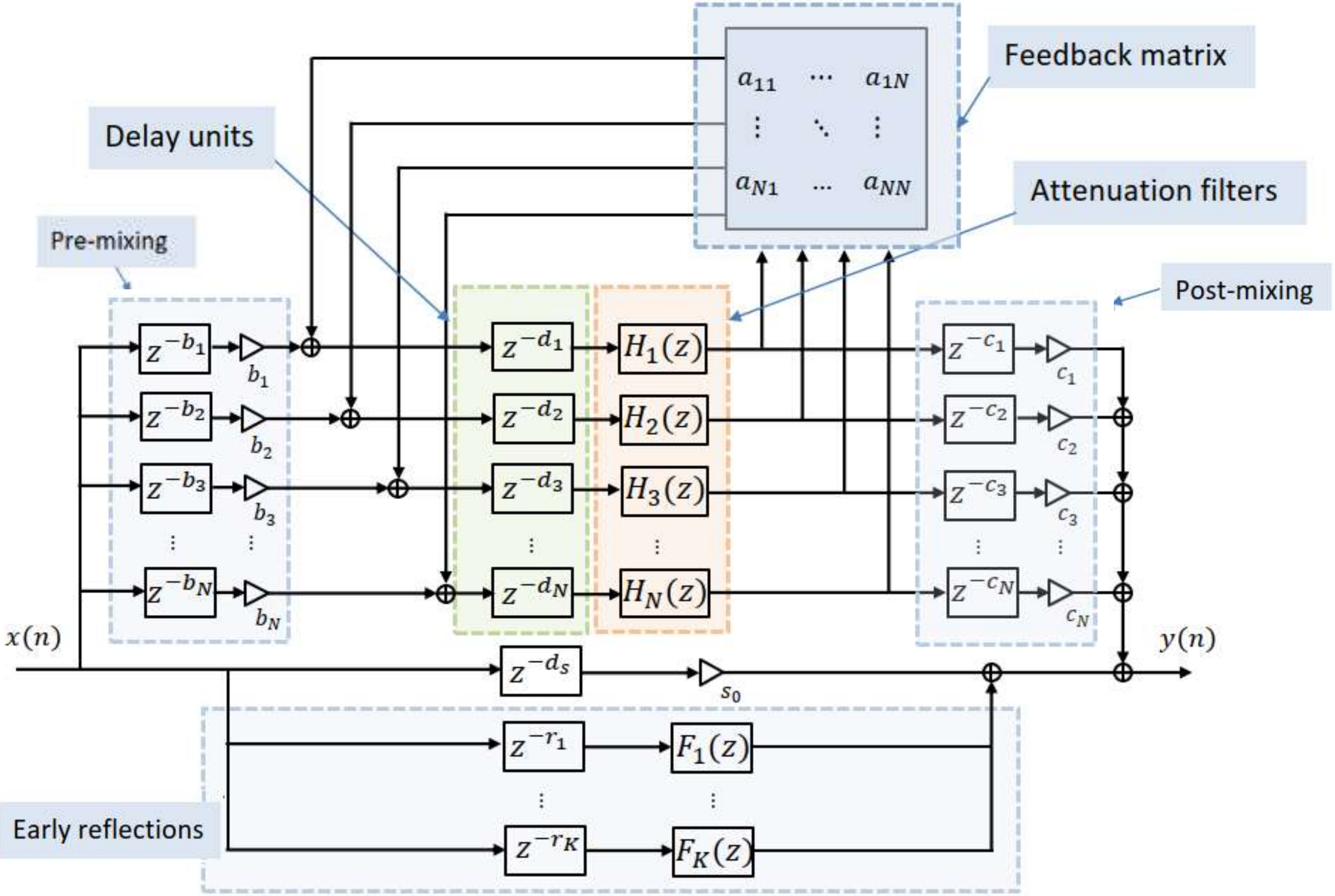
G. Richard

Reverberation - Dereverberation

The hybrid RTM-FDN method

(exploiting Radiance Transfer and Feedback Delay Networks)

- The structure



[1] H. Bai, G. Richard, and L. Daudet. "Geometric-based reverberator using acoustic rendering networks." In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pages 1–4, New Paltz, NY, October 2015.
[2] H. Bai, G. Richard, and L. Daudet. "Late reverberation synthesis : From radiance transfer to feedback delay networks." Audio, Speech, and Language Processing, IEEE/ACM Transactions on, 23(12) :2260–2271, 2015.

The hybrid RTM-FDN method

(exploiting Radiance Transfer and Feedback Delay Networks)

- Sound examples : african drum
  - FND-RTM 4 delay lines
  - FND-RTM 8 delay lines
  - FND-RTM 16 delay lines
  - Radiance transfer method

- **Sound quality:**
  - Sound quality increases with the number of delaylines
  - Sound quality approaches RTM using 16 delaylines

[1] H. Bai, G. Richard, and L. Daudet. "Geometric-based reverberator using acoustic rendering networks." In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pages 1–4, New Paltz, NY, October 2015.
[2] H. Bai, G. Richard, and L. Daudet. "Late reverberation synthesis : From radiance transfer to feedback delay networks." Audio, Speech, and Language Processing, IEEE/ACM Transactions on, 23(12) :2260–2271, 2015.

Preliminary conclusion on reverberation

- Many solutions for artificial reverberation do exist but:

  - To exactly model a reverberant space calls for complex methods even with simple systems such as with unitary feedback matrix.
  - There is a clear interest for “hybrid” methods (perceptual / physical) (see [1] for instance)
  - … and obviously also towards hybrid deep methods exploiting data …

(but only discussed in the framework of dereverberation in the next part)

M. Scerbo, L. Savioja, E. De Sena, Room Acoustic Rendering Networks with Control of Scattering and Early Reflections, IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 32, Aug. 2024

Dereverberation

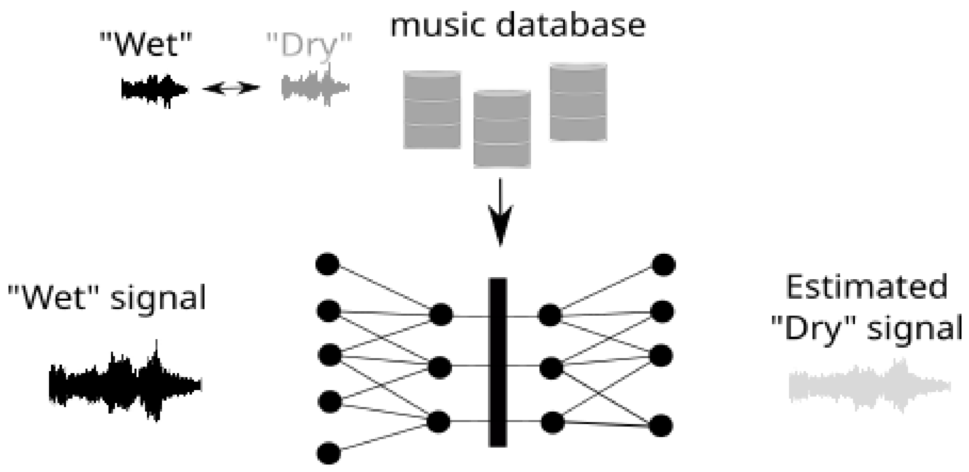
- …also a long history of methods… but, now, many methods exploit machine learning (and in particular deep learning)

- Four categories of methods :
  - **Fully data-driven model with supervised learning using both "wet" and "dry"** data with no explicit reverberation model (for instance [1-3])
  - **Autoregressive models combined with speech priors** such as in the Weighted Prediction Error (WPE) [4] and its Deep learning extensions
  - **Supervised discriminative models,** trained using pairs of dry and wet signals, and accurate AR models : these approaches are used to inform a dereverberation model using acoustic information (for instance [5, 6])
  - **Generative approaches**: use a prior that is pre-trained on dry data only. For instance with using the RIR itself with a diffusion-based prior [7], or leverage autoregressive models of reverberation combined with Recurrent VAE [8], or diffusion [9] priors.

[1] F. Weninger & al. "Speech Enhancement with LSTM Recurrent Neural Networks and its Application to Noise-Robust ASR," in Latent Variable Analysis and Signal Separation, E. Vincent, A. Yeredor, Z. Koldovsk ́and P. Tichavsk ́y, Eds. Cham:Springer International Publishing, 2015, pp. 91–99.
[2] X. Hao, X. Su, R. Horaud, and X. Li, "Fullsubnet: A Full-Band and Sub-Band Fusion Model for Real-Time Single-Channel Speech Enhancement," in Proc. ICASSP, Jun. 2021,
[3] K. Saijo, G. Wichern, F. G. Germain, Z. Pan, and J. L. Roux, "TF-Locoformer: Transformer with Local Modeling by Convolution for Speech Separation and Enhancement," in Proc. IWAENC. Sep. 2024.
[4] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B.-H. Juang, "Speech Dereverberation Based on Variance-Normalized Delayed Linear Prediction," IEEE Trans. ASLP, vol. 18, no. 7, Sep. 2010.
[5] B. Wu, K. Li, M. Yang, and C.-H. Lee, "A Reverberation-Time-Aware Approach to Speech Dereverberation Based on Deep Neural Networks," IEEE/ACM Trans. ASLP, vol. 25, no. 1, Jan. 2017.
[6] N. K. S. Rao, S. R. Chetupalli, S. S. Shetu, E. A. P. Habets, and O. Thiergart, "Low-Complexity Neural Speech Dereverberation With Adaptive Target Control," in Proc. ICASSP, Apr. 2025,
[7] M. Lemercier, S. Welker, and T. Gerkmann, "Diffusion posterior sampling for informed single-channel dereverberation," in Proc. WASPAA, 2023
[8] P. Wang and X. Li, "RVAE-EM: Generative Speech Dereverberation Based On Recurrent Variational AutoEncoder And Convolutive Transfer Function," in Proc. ICASSP, Apr. 2024,
[9] J.-M. Lemercier, E. Moliner, S. Welker, V. Valimaki, and T. Gerkmann, "Unsupervised Blind Joint Dereverberation and Room Acoustics Estimation With Diffusion Models," IEEE Trans. ASLP, vol 33., 2025

Towards model-based deep learning approaches

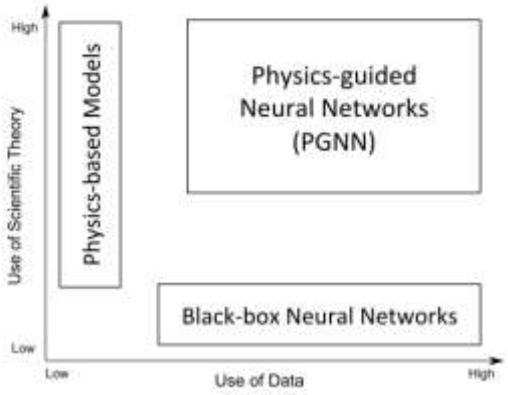
- Machine learning: a growing trend towards pure “Data-driven” deep learning approaches



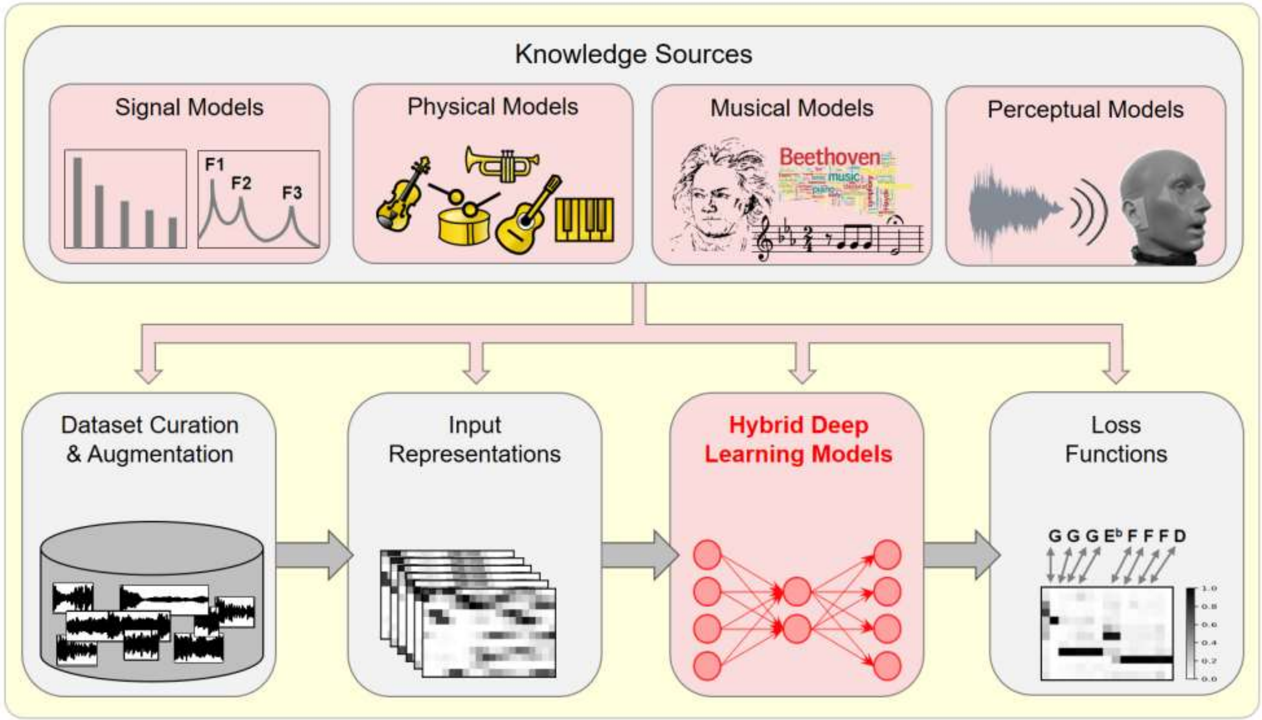
- High performances but some main limitations:
  - *“Knowledge” is learned (only) from data*
  - *Complexity: overparametrized models (>> 100 millions parameters)*
  - Overconsumption regime
  - Non-interpretable/non-controllable

Towards model-based deep learning approaches

- Coupling model-based and deep learning:



Example with Hybrid deep model for Music signals

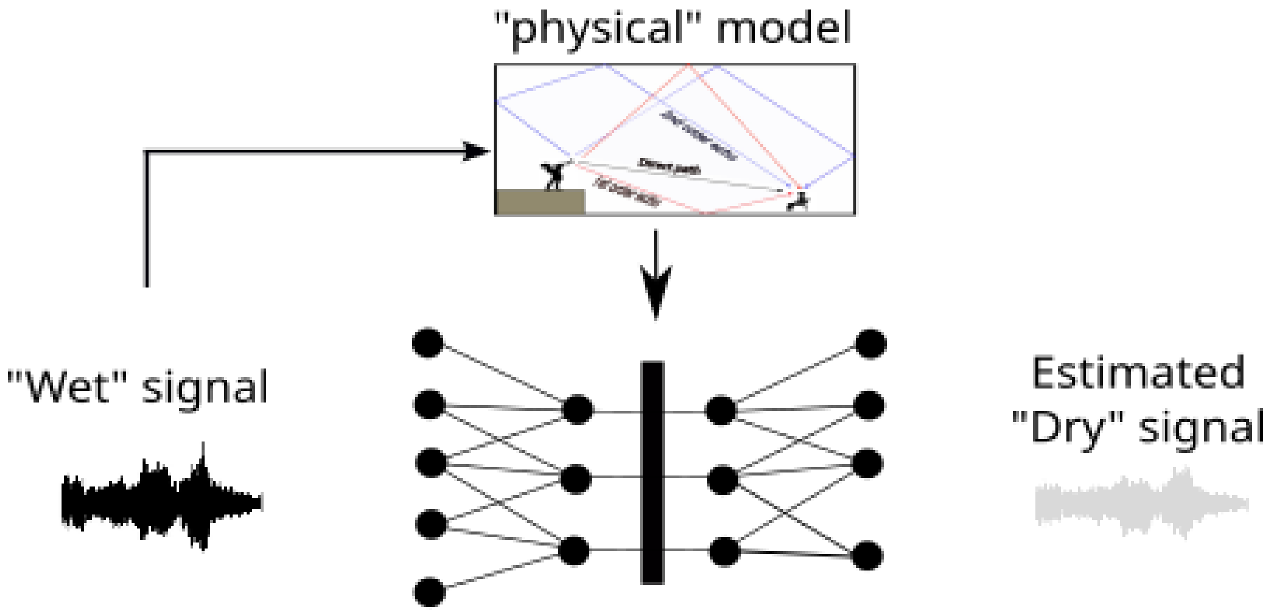


G. Richard, V. Lostanlen, Y.-H. Yang, M. Müller, "Hybrid Deep Learning for Music Information Research", IEEE Signal Processing Magazine - Special Issue on Model-based and Data-Driven Audio Signal Processing, 2025

Hi-Audio, Hybrid and Interpretable Deep neural audio machines, European Research Council "Advanced Grant" (AdG) project - https://hi-audio.imt.fr/

Towards model-based deep dereverberation

- Exploiting a physical model of reverberation



Louis Bahrman, Mathieu Fontaine, Gaël Richard, U-DREAM: Unsupervised Dereverberation guided by a Reverberation Model, 2025, preprint https://hal.science/hal-05158698v1

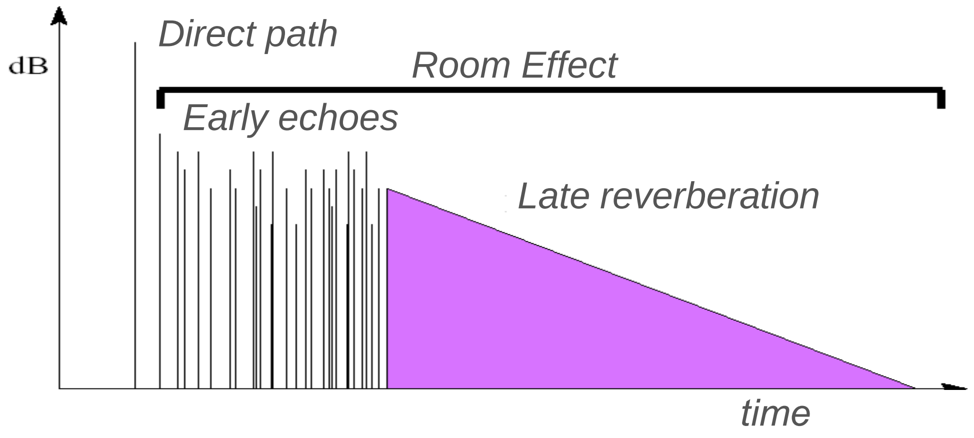
Towards model-based deep dereverberation

Exploiting a room impulse response model

- The reverberant signal : $y(n) = (s \star h)(n) + \epsilon(n),$

- The room impulse model

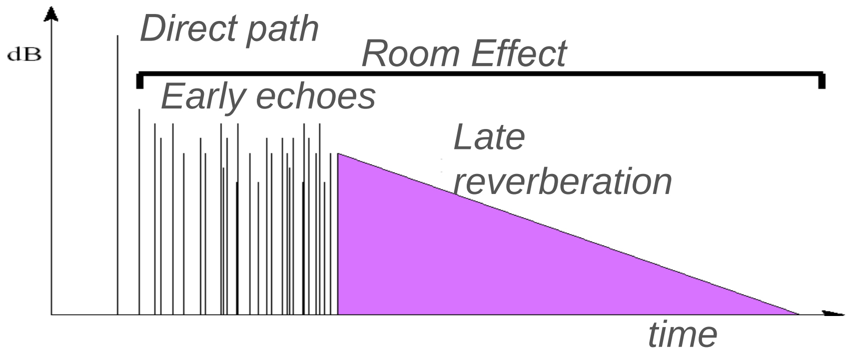
$$h(n) = h_e + h_r$$



Louis Bahrman, Mathieu Fontaine, Gaël Richard, U-DREAM: Unsupervised Dereverberation guided by a Reverberation Model, 2025, preprint https://hal.science/hal-05158698v1

Towards model-based deep dereverberation

Exploiting a room impulse response model



- The RIR model: important parameters:
  - **Direct-to-Reverberant ratio (DRR):** quantifies the energy balance between the direct path and the reverberant tail

$$\mathrm{DRR}_{dB} = 10\log_{10}\left(\frac{\sum_{n=0}^{n_d} h^2(n)}{\sum_{n=n_d+1}^{\infty} h^2(n)}\right)$$

  - **Reverberation time** RT_{60} **:** can be estimated (Under idealized conditions) from the slope of the energy decay curve (EDC)

$$\mathrm{EDC}_h(t) = \int_t^{+\infty} h(u)du,$$

P. A. Naylor and N. D. Gaubitch, Eds., Speech Dereverberation, ser. Signals and Communication Technology. London: Springer London, 2010.

The statistical Polack model

- DRR and RT60 are sufficient to characterize the Polack (late) reverberation model [1]

$$h_r(n) = b(n)e^{-n/\tau},$$

- With $b(n) \sim \mathcal{N}(0, \sigma^2)$ and $\tau = \frac{\mathrm{RT}_{60} f_s}{3\ln(10)}.$

- For reverberation, the polack model is valid after the « mixing time » $n_m = (4V f_s)/(cA)$, where V , f_s, c, A are respectively the room volume, the sampling frequency, the speed of sound and the area of the walls.

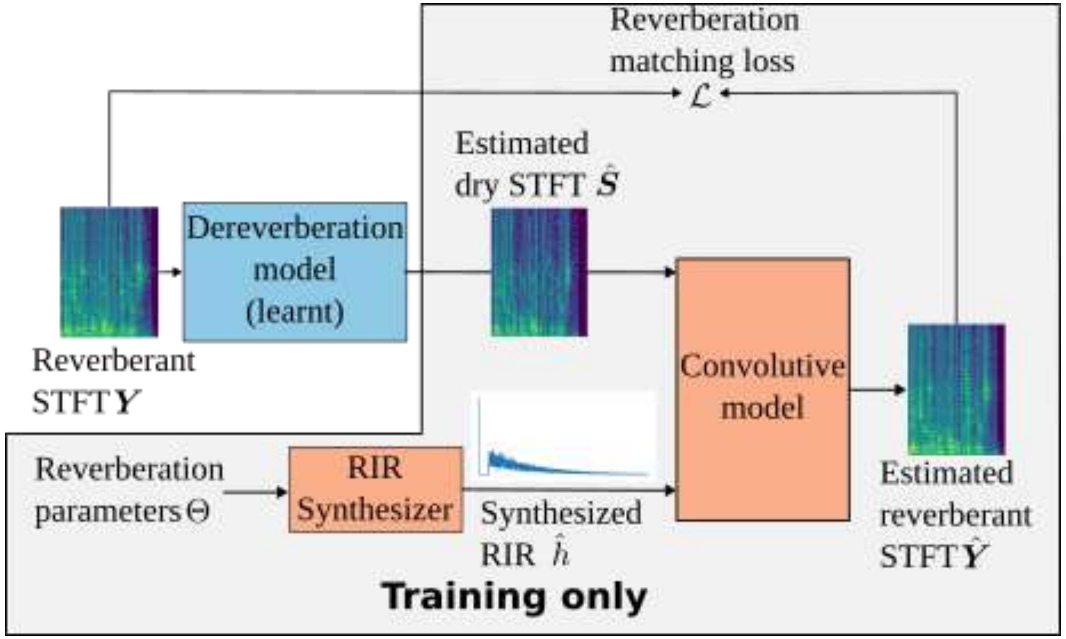
[1] J.-D. Polack, “La transmission de l’energie sonore dans les salles,” Ph.D. dissertation (in French), Université du Maine, 1988

Towards model-based deep dereverberation

Exploiting a room impulse response model

G. Richard

Reverberation - Dereverberation



Regularization term (for low amplitudes)

- Reverberation Loss used: $\mathcal{L} = \sum_{f,t} \left[|\hat{Y}_{f,t} - Y_{f,t}|^2 + \lambda \left| \log \left(\frac{1+\gamma|\hat{Y}_{f,t}|}{1+\gamma|Y_{f,t}|} \right) \right|^2 \right]$

L. Bahrman, M. Fontaine, and G. Richard, "A Hybrid Model for Weakly Supervised Speech Dereverberation," in ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing 873 (ICASSP), Apr. 2025,
S. Schwär and M. Müller, "Multi-Scale Spectral Loss Revisited," IEEE Signal Process. Lett., vol. 30, pp. 1712–1716, 2023.

Towards model-based deep dereverberation

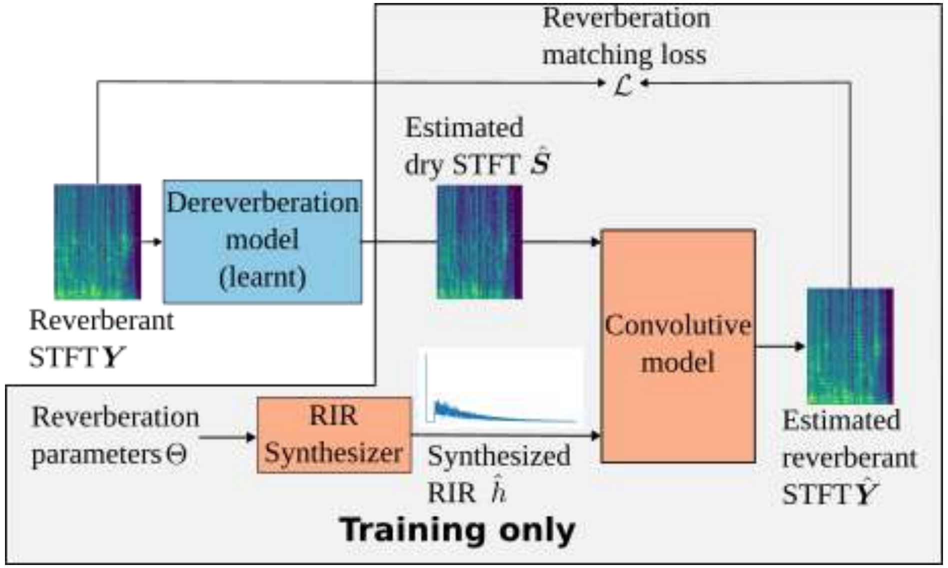
Exploiting a room impulse response model

- Main advantages of the model

  - Can be trained in an unsupervised way (no needs of pairs Wet- dry of signals)

  - The dereverberation model is more interpretable and controllable (e.g. use « physical » constraints)

  - Smaller network may be sufficient to obtain similar performances than bigger networks trained in a supervised way



L. Bahrman, M. Fontaine, and G. Richard, "A Hybrid Model for Weakly Supervised Speech Dereverberation," in ICASSP 2025 - 2025 IEEE 872 International Conference on Acoustics, Speech and Signal Processing 873 (ICASSP), Apr. 2025,

G. Richard

Towards model-based deep dereverberation

Exploiting a room impulse response model

- Different levels of supervision:

Weak supervision variants include using Polack's model with either :

- $\Theta \triangleq \{\mathrm{RT}_{60}, \sigma, V, A\}$: all the parameters, including those used to estimate the mixing time.
- $\{\mathrm{RT}_{60}, \sigma\}$: a fixed mixing time set as 20ms after the peak, corresponding to the mean of all mixing times in the training dataset.
- $\{\mathrm{RT}_{60}\}$: a fixed mixing time at 20ms and a median value of Polack's variance over the training dataset of $\sigma = 0.02$

Strong supervision

- the exact RIR *h* as an oracle RIR synthesis model.
- ..or each model's original paired training loss as supervision.

- Note that we have (*Assuming the direct-path energy is normalized to 1*):

$$\sigma = \sqrt{\frac{2e^{2n_D/\tau}}{\tau \mathrm{DRR}}}$$

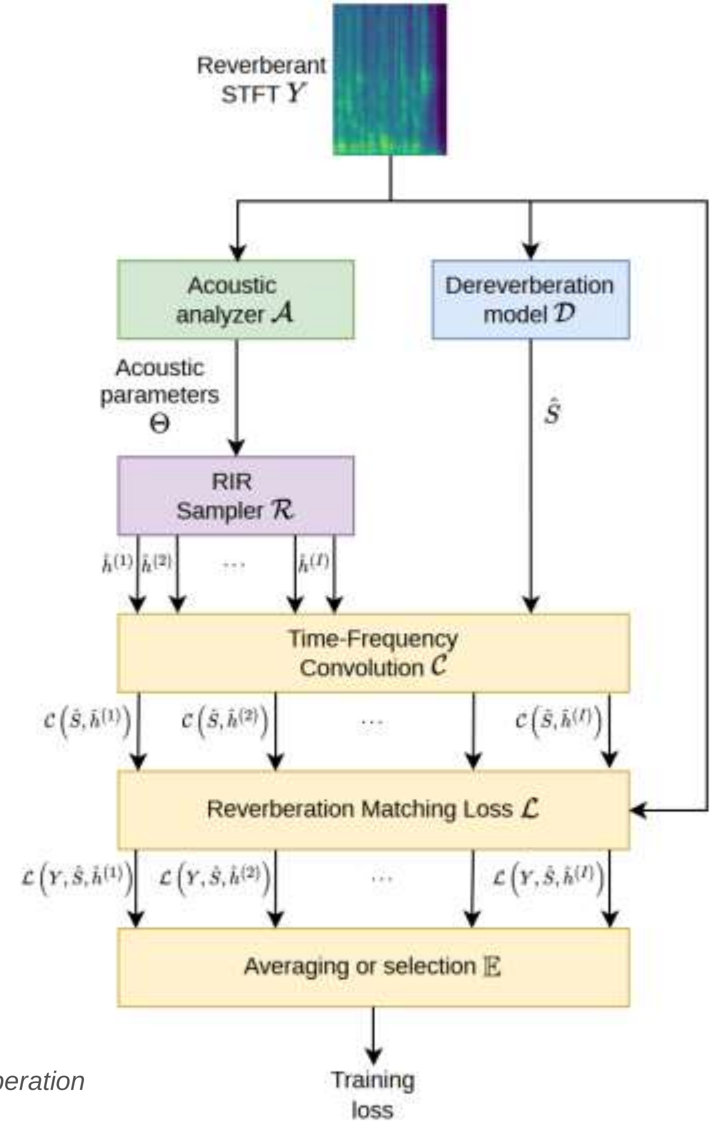
U-DREAM: the extension to "Unsupervised Dereverberation" guided by a Reverberation Model

- The optimization problem

$$\hat{\boldsymbol{S}}, \hat{\Theta} = \underset{\boldsymbol{S},\Theta}{\operatorname{argmin}} \, \mathbb{E}_{p(h|\Theta)} \left[\|\boldsymbol{Y} - \mathcal{C}(\boldsymbol{S}, h)\|_F^2 \right]$$

- An **Acoustic Analyzer** to estimate acoustic parameters for sampling candidate Room Impulse Responses

- **RIR sampler**, using Polack's model as previously, but several draws possible



Louis Bahrman, Mathieu Fontaine, Gaël Richard, U-DREAM: Unsupervised Dereverberation guided by a Reverberation Model, 2025, preprint https://hal.science/hal-05158698v1

Towards model-based deep dereverberation

Exploiting a room impulse response model

Some results

- **Dataset used:** EARS-ISM (synthetic RIR) - EARS-Reverb (Real RIRs)
- **Dereverberation model used**: BiLSTM *(2-layer 599 bidirectional LSTM model followed by a linear layer, performing subband processing of the STFT magnitudes).*
- **Pre-trained Acoustic Analyzer:** Parameter MSE loss, trained with 100 samples of couple (y, $\Theta = \{\mathrm{DRR}, \mathrm{RT}_{60}\}$)
- **Evaluation (objective) metrics**
  - SI-SDR (« signal distorsion »),
  - PESQ (« perceptual quality »
  - STOI (« intelligibility »),
  - SRMR (« reverberation »)

L. Bahrman, M. Fontaine, and G. Richard, “A Hybrid Model for Weakly Supervised Speech Dereverberation,” in ICASSP 2025,Apr. 2025,
L. Bahrman, M. Fontaine, G. Richard, U-DREAM: Unsupervised Dereverberation guided by a Reverberation Model, 2025, preprint https://hal.science/hal-05158698v1
(EARS): J. Richter, Y.-C. Wu, S. Krenn, S. Welker, B. Lay, S. Watanabe, A. Richard, and T. Gerkmann, “EARS: An Anechoic Fullband Speech 1001 Dataset Benchmarked for Speech Enhancement and Dereverberation,” 1002 in *Interspeech 2024*.
(BiLSTM): F. Weninger & al. “Speech Enhancement with LSTM Recurrent Neural Networks and its Application to Noise-Robust ASR,” in Latent Variable Analysis and Signal Separation, E. Vincent, A. Yeredor, Z. Koldovsk ́and P. Tichavsk ́y, Eds. Cham:Springer International Publishing, 2015, pp. 91–99.
(WPE) T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B.-H. Juang, “Speech Dereverberation Based on Variance-Normalized Delayed Linear Prediction,” IEEE Trans. ASLP, vol. 18, no. 7, Sep. 2010.

Towards model-based deep dereverberation

Exploiting a room impulse response model

Some results

		Synthetic RIRs				Real RIRs			
Supervision type	Supervision	↑ SISDR	ESTOI	WB-PESQ	SRMR	↑ SISDR	ESTOI	WB-PESQ	SRMR
strong	Dry speech	-2.0 ± 6.1	0.75 ± 0.12	2.15 ± 0.64	7.7 ± 3.6	-14.5 ± 9.2	0.61 ± 0.13	1.73 ± 0.41	6.5 ± 2.9
	Exact RIR	-2.3 ± 5.8	0.72 ± 0.13	1.99 ± 0.66	8.5 ± 3.6	-15.6 ± 10.6	0.61 ± 0.14	1.75 ± 0.46	6.5 ± 2.8
weak	Oracle parameters	-1.7 ± 5.4	0.67 ± 0.15	1.74 ± 0.62	6.4 ± 3.0	-14.5 ± 8.1	0.58 ± 0.13	1.64 ± 0.39	5.4 ± 2.6
unsupervised	Pretrained Acoustic Analyzer	-3.6 ± 5.1	0.64 ± 0.12	1.62 ± 0.43	8.0 ± 3.4	-14.5 ± 8.7	0.57 ± 0.12	1.58 ± 0.31	6.2 ± 2.9
	WPE	-2.1 ± 5.0	0.72 ± 0.14	1.94 ± 0.76	6.9 ± 3.4	-15.8 ± 9.1	0.54 ± 0.17	1.54 ± 0.43	5.2 ± 3.2
	Reverberant	-6.7 ± 6.4	0.67 ± 0.15	1.79 ± 0.64	8.2 ± 5.9	-16.1 ± 9.3	0.52 ± 0.17	1.48 ± 0.36	4.8 ± 2.9

L. Bahrman, M. Fontaine, and G. Richard, "A Hybrid Model for Weakly Supervised Speech Dereverberation," in ICASSP 2025,Apr. 2025,
L. Bahrman, M. Fontaine, G. Richard, U-DREAM: Unsupervised Dereverberation guided by a Reverberation Model, 2025, preprint https://hal.science/hal-05158698v1
(EARS): J. Richter, Y.-C. Wu, S. Krenn, S. Welker, B. Lay, S. Watanabe, A. Richard, and T. Gerkmann, "EARS: An Anechoic Fullband Speech 1001 Dataset Benchmarked for Speech Enhancement and Dereverberation," 1002 in *Interspeech 2024*.
(BiLSTM): F. Weninger & al. "Speech Enhancement with LSTM Recurrent Neural Networks and its Application to Noise-Robust ASR," in Latent Variable Analysis and Signal Separation, E. Vincent, A. Yeredor, Z. Koldovsk´and P. Tichavsk´y, Eds. Cham:Springer International Publishing, 2015, pp. 91–99.
(WPE) T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B.-H. Juang, "Speech Dereverberation Based on Variance-Normalized Delayed Linear Prediction," IEEE Trans. ASLP, vol. 18, no. 7, Sep. 2010.

Towards model-based deep dereverberation

Exploiting a room impulse response model

Some results

Supervision type	Supervision	Synthetic RIRs ↑ SISDR	Synthetic RIRs ESTOI	Synthetic RIRs WB-PESQ	Synthetic RIRs SRMR	Real RIRs ↑ SISDR	Real RIRs ESTOI	Real RIRs WB-PESQ	Real RIRs SRMR
strong	Dry speech	-2.0 ± 6.1	0.75 ± 0.12	2.15 ± 0.64	7.7 ± 3.6	-14.5 ± 9.2	0.61 ± 0.13	1.73 ± 0.41	6.5 ± 2.9
	Exact RIR	-2.3 ± 5.8	0.72 ± 0.13	1.99 ± 0.66	8.5 ± 3.6	-15.6 ± 10.6	0.61 ± 0.14	1.75 ± 0.46	6.5 ± 2.8
weak	Oracle parameters	-1.7 ± 5.4	0.67 ± 0.15	1.74 ± 0.62	6.4 ± 3.0	-14.5 ± 8.1	0.58 ± 0.13	1.64 ± 0.39	5.4 ± 2.6
unsupervised	Pretrained Acoustic Analyzer	-3.6 ± 5.1	0.64 ± 0.12	1.62 ± 0.43	8.0 ± 3.4	-14.5 ± 8.7	0.57 ± 0.12	1.58 ± 0.31	6.2 ± 2.9
	WPE	-2.1 ± 5.0	0.72 ± 0.14	1.94 ± 0.76	6.9 ± 3.4	-15.8 ± 9.1	0.54 ± 0.17	1.54 ± 0.43	5.2 ± 3.2
	Reverberant	-6.7 ± 6.4	0.67 ± 0.15	1.79 ± 0.64	8.2 ± 5.9	-16.1 ± 9.3	0.52 ± 0.17	1.48 ± 0.36	4.8 ± 2.9

- **All methods perform some level of dereverberation**

L. Bahrman, M. Fontaine, and G. Richard, "A Hybrid Model for Weakly Supervised Speech Dereverberation," in ICASSP 2025,Apr. 2025,
L. Bahrman, M. Fontaine, G. Richard, U-DREAM: Unsupervised Dereverberation guided by a Reverberation Model, 2025, preprint https://hal.science/hal-05158698v1
(EARS): J. Richter, Y.-C. Wu, S. Krenn, S. Welker, B. Lay, S. Watanabe, A. Richard, and T. Gerkmann, "EARS: An Anechoic Fullband Speech 1001 Dataset Benchmarked for Speech Enhancement and Dereverberation," 1002 in *Interspeech 2024*.
(BiLSTM): F. Weninger & al. "Speech Enhancement with LSTM Recurrent Neural Networks and its Application to Noise-Robust ASR," in Latent Variable Analysis and Signal Separation, E. Vincent, A. Yeredor, Z. Koldovsk´and P. Tichavsk´y, Eds. Cham:Springer International Publishing, 2015, pp. 91–99.
(WPE) T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B.-H. Juang, "Speech Dereverberation Based on Variance-Normalized Delayed Linear Prediction," IEEE Trans. ASLP, vol. 18, no. 7, Sep. 2010.

G. Richard

Reverberation - Dereverberation

Towards model-based deep dereverberation

Exploiting a room impulse response model

Some results

		Synthetic RIRs				Real RIRs			
Supervision type	Supervision	↑ SISDR	ESTOI	WB-PESQ	SRMR	↑ SISDR	ESTOI	WB-PESQ	SRMR
strong	Dry speech	-2.0 ± 6.1	0.75 ± 0.12	2.15 ± 0.64	7.7 ± 3.6	-14.5 ± 9.2	0.61 ± 0.13	1.73 ± 0.41	6.5 ± 2.9
	Exact RIR	-2.3 ± 5.8	0.72 ± 0.13	1.99 ± 0.66	8.5 ± 3.6	-15.6 ± 10.6	0.61 ± 0.14	1.75 ± 0.46	6.5 ± 2.8
weak	Oracle parameters	-1.7 ± 5.4	0.67 ± 0.15	1.74 ± 0.62	6.4 ± 3.0	-14.5 ± 8.1	0.58 ± 0.13	1.64 ± 0.39	5.4 ± 2.6
unsupervised	Pretrained Acoustic Analyzer	-3.6 ± 5.1	0.64 ± 0.12	1.62 ± 0.43	8.0 ± 3.4	-14.5 ± 8.7	0.57 ± 0.12	1.58 ± 0.31	6.2 ± 2.9
	WPE	-2.1 ± 5.0	0.72 ± 0.14	1.94 ± 0.76	6.9 ± 3.4	-15.8 ± 9.1	0.54 ± 0.17	1.54 ± 0.43	5.2 ± 3.2
	Reverberant	-6.7 ± 6.4	0.67 ± 0.15	1.79 ± 0.64	8.2 ± 5.9	-16.1 ± 9.3	0.52 ± 0.17	1.48 ± 0.36	4.8 ± 2.9

- **Weakly-supervised method outperforms the baseline WPE on most metrics (especially on real RIRs)**

L. Bahrman, M. Fontaine, and G. Richard, "A Hybrid Model for Weakly Supervised Speech Dereverberation," in ICASSP 2025,Apr. 2025,

L. Bahrman, M. Fontaine, G. Richard, U-DREAM: Unsupervised Dereverberation guided by a Reverberation Model, 2025, preprint https://hal.science/hal-05158698v1

(EARS): J. Richter, Y.-C. Wu, S. Krenn, S. Welker, B. Lay, S. Watanabe, A. Richard, and T. Gerkmann, "EARS: An Anechoic Fullband Speech 1001 Dataset Benchmarked for Speech Enhancement and Dereverberation," 1002 in *Interspeech 2024*.

(BiLSTM): F. Weninger & al. "Speech Enhancement with LSTM Recurrent Neural Networks and its Application to Noise-Robust ASR," in Latent Variable Analysis and Signal Separation, E. Vincent, A. Yeredor, Z. Koldovsk´and P. Tichavsk´y, Eds. Cham:Springer International Publishing, 2015, pp. 91–99.

(WPE) T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B.-H. Juang, "Speech Dereverberation Based on Variance-Normalized Delayed Linear Prediction," IEEE Trans. ASLP, vol. 18, no. 7, Sep. 2010.

Towards model-based deep dereverberation

Exploiting a room impulse response model

Some results

		Synthetic RIRs				Real RIRs			
Supervision type	Supervision	↑ SISDR	ESTOI	WB-PESQ	SRMR	↑ SISDR	ESTOI	WB-PESQ	SRMR
strong	Dry speech	-2.0 ± 6.1	0.75 ± 0.12	2.15 ± 0.64	7.7 ± 3.6	-14.5 ± 9.2	0.61 ± 0.13	1.73 ± 0.41	6.5 ± 2.9
	Exact RIR	-2.3 ± 5.8	0.72 ± 0.13	1.99 ± 0.66	8.5 ± 3.6	-15.6 ± 10.6	0.61 ± 0.14	1.75 ± 0.46	6.5 ± 2.8
weak	Oracle parameters	-1.7 ± 5.4	0.67 ± 0.15	1.74 ± 0.62	6.4 ± 3.0	-14.5 ± 8.1	0.58 ± 0.13	1.64 ± 0.39	5.4 ± 2.6
unsupervised	Pretrained Acoustic Analyzer	-3.6 ± 5.1	0.64 ± 0.12	1.62 ± 0.43	8.0 ± 3.4	-14.5 ± 8.7	0.57 ± 0.12	1.58 ± 0.31	6.2 ± 2.9
	WPE	-2.1 ± 5.0	0.72 ± 0.14	1.94 ± 0.76	6.9 ± 3.4	-15.8 ± 9.1	0.54 ± 0.17	1.54 ± 0.43	5.2 ± 3.2
	Reverberant	-6.7 ± 6.4	0.67 ± 0.15	1.79 ± 0.64	8.2 ± 5.9	-16.1 ± 9.3	0.52 ± 0.17	1.48 ± 0.36	4.8 ± 2.9

- **Unsupervised method is efficient, in particular on Real RIRs**

L. Bahrman, M. Fontaine, and G. Richard, “A Hybrid Model for Weakly Supervised Speech Dereverberation,” in ICASSP 2025,Apr. 2025,
L. Bahrman, M. Fontaine, G. Richard, U-DREAM: Unsupervised Dereverberation guided by a Reverberation Model, 2025, preprint https://hal.science/hal-05158698v1
(EARS): J. Richter, Y.-C. Wu, S. Krenn, S. Welker, B. Lay, S. Watanabe, A. Richard, and T. Gerkmann, “EARS: An Anechoic Fullband Speech 1001 Dataset Benchmarked for Speech Enhancement and Dereverberation,” 1002 in *Interspeech 2024*.
(BiLSTM): F. Weninger & al. “Speech Enhancement with LSTM Recurrent Neural Networks and its Application to Noise-Robust ASR,” in Latent Variable Analysis and Signal Separation, E. Vincent, A. Yeredor, Z. Koldovsk´and P. Tichavsk´y, Eds. Cham:Springer International Publishing, 2015, pp. 91–99.
(WPE) T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B.-H. Juang, “Speech Dereverberation Based on Variance-Normalized Delayed Linear Prediction,” IEEE Trans. ASLP, vol. 18, no. 7, Sep. 2010.

G. Richard

Reverberation - Dereverberation

Towards model-based deep dereverberation

Exploiting a room impulse response model

- Some sounds (weak-supervision results)

	Wet input	Ground truth	FSN (proposed)	FSN	BiLSTM (proposed)	BiLSTM	Baseline
WS			✓	✗	✓	✗	✓
RT60=0.6							

- More audio demo at https://louis-bahrman.github.io/Hybrid-WSSD/

L. Bahrman, M. Fontaine, and G. Richard, "A Hybrid Model for Weakly Supervised Speech Dereverberation," in ICASSP 2025 - 2025 IEEE 872 International Conference on Acoustics, Speech and Signal Processing 873 (ICASSP), Apr. 2025,
(WPE) T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B.-H. Juang, "Speech Dereverberation Based on Variance-Normalized Delayed Linear Prediction," IEEE Trans. ASLP, vol. 18, no. 7, Sep. 2010.

G. Richard

Reverberation - Dereverberation

To conclude

- Reverberation is a fascinating field ….
- Many methods and approaches exist for reverberation synthesis and dereverberation…
- As in many domains, the prominence of deep learning solutions is progressing …
- … but I believe in hybrid methods, hybrid deep learning … which bring
    - **Interpretability, Controllability, Explainability**
        - Hybrid model becomes controllable by human-understandable parameters
        - Hybrid model can lead to unsupervised methods
    - **Frugality: gain of several orders of magnitude** in the need of data and model complexity
    - **Can be applied to many audio processing problems**
        - Exploiting room acoustics for Audio dereverberation [1],
        - Exploiting physical/signal models for music synthesis [2],
        - Exploiting “audio class specific” codebooks for audio compression and separation [3]
        - Exploiting key speech attributes for controlled speech synthesis and transformation [4]
        - …

[1] Louis Bahrman, Mathieu Fontaine, Gael Richard. A Hybrid Model for Weakly-Supervised Speech Dereverberation. *IEEE ICASSP 2025*, ⟨hal-04931672⟩
[2] Lenny Renault, Rémi Mignot, Axel Roebel. Differentiable Piano Model for MIDI-to-Audio Performance Synthesis. Int. Conf.on Digital Audio Effects (DAFx20in22), Sep 2022, Vienna,
[3] Xiaoyu Bie, Xubo Liu, Gaël Richard. Learning Source Disentanglement in Neural Audio Codec. *IEEE ICASSP 2025* , ⟨hal-04902131⟩
[4] Samir Sadok, Simon Leglaive, Laurent Girin, Gaël Richard, Xavier Alameda-Pineda. AnCoGen: Analysis, Control and Generation of Speech with a Masked Autoencoder. *IEEE ICASSP 2025* , ⟨hal-04891286⟩