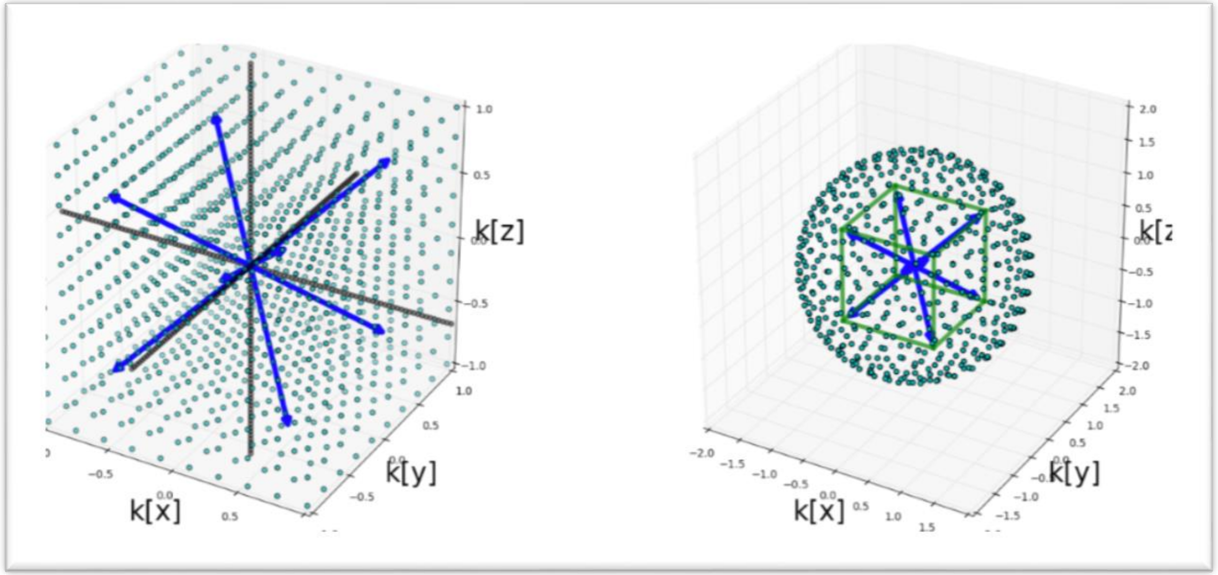
JOINT ESTIMATION OF THE ROOM GEOMETRY AND MODES WITH COMPRESSED SENSING



Helena Peić Tukuljac, Thach Pham Vu,
Hervé Lissek and Pierre Vandergheynst

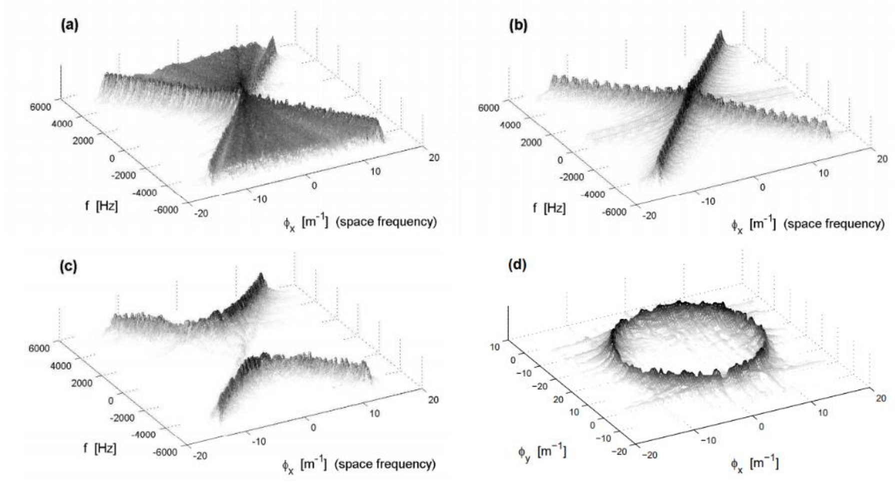
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

April 19, 2018

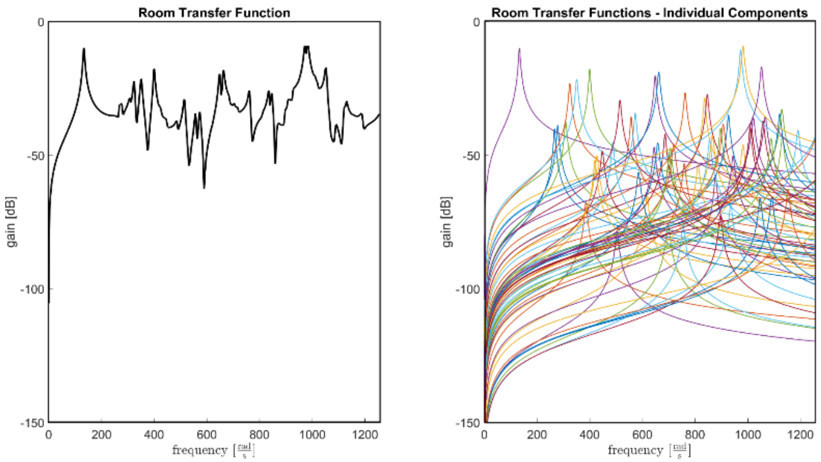
Content Atlas

Introduction

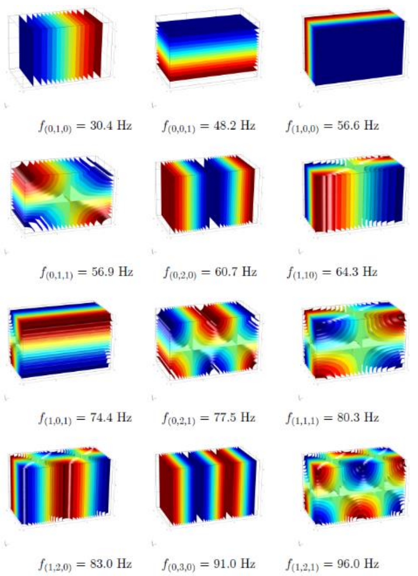
Spatio-temporal sampling relations [1]



Room transfer function [2]

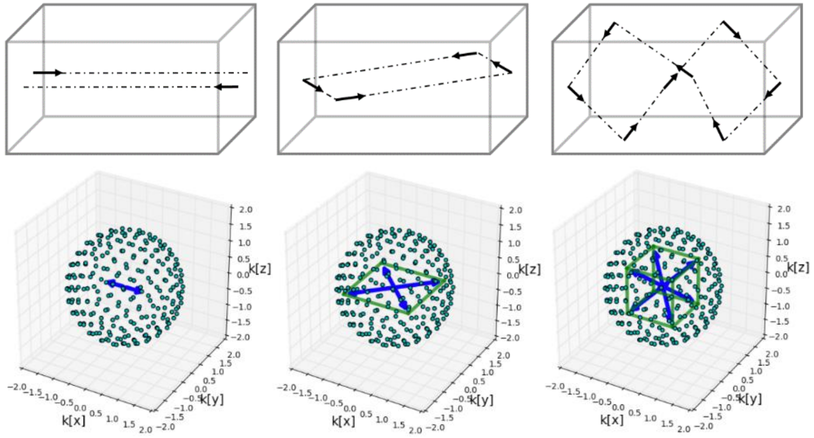


Room modes [3]

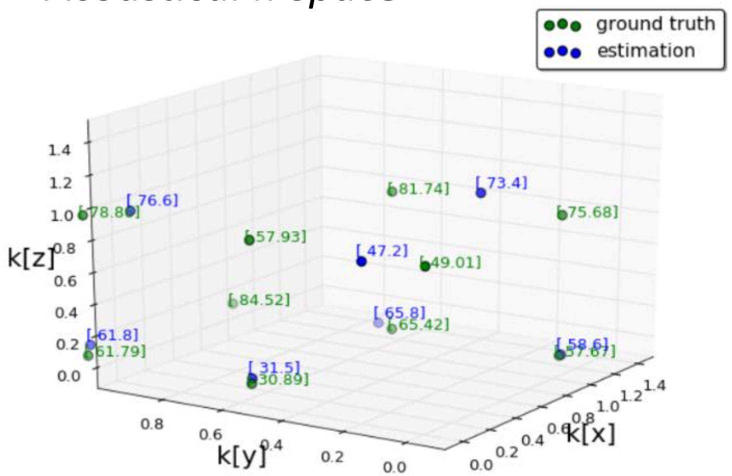


Contribution

Plane waves and wave vectors



Acoustical k-space



helena.peictukuljac@epfl.ch

[1] R. Mignot, G. Chardon and L. Daudet, "Low Frequency Interpolation of Room Impulse Responses Using Compressed Sensing," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 1, pp. 205-216, Jan. 2014.
[2] **H. Peic Tukuljac**, H. Lissek, and P. Vandergheynst, "Localization of sound sources in a room with one microphone," SPIE, Wavelets and Sparsity XVII, vol. 10394, 2017.
[3] R. Boulandet, "Tunable Electroacoustic Resonators through Active Impedance Control of Loudspeakers", PhD thesis, EPFL, Lausanne, 2012

Spatio-Temporal Sampling Relations

Sampling steps: Δt Δx Δy Δz

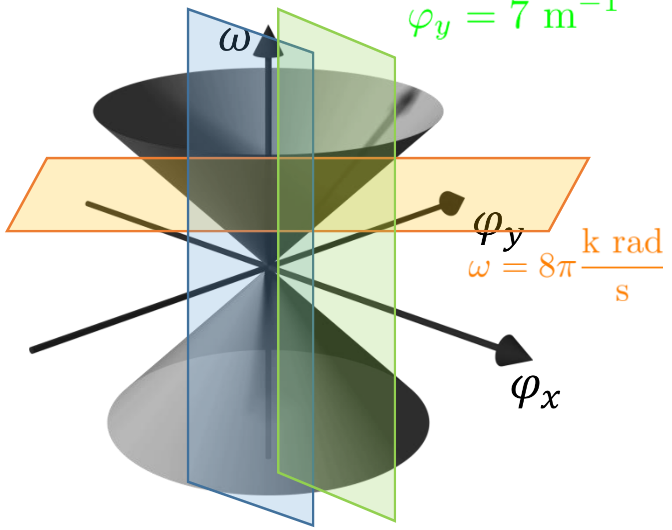
Sampling frequencies: $\omega = \frac{2\pi}{\Delta t}$ $\varphi_x = \frac{2\pi}{\Delta x}$ $\varphi_y = \frac{2\pi}{\Delta y}$ $\varphi_z = \frac{2\pi}{\Delta z}$

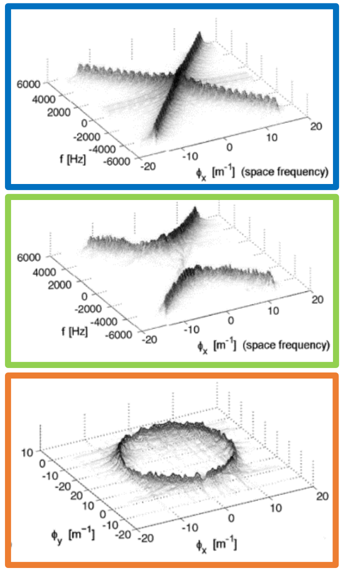
Choosing Δt is easy: [1]

$$\omega \geq 2\omega_{\text{cutoff}}$$

Two points of view on spatial sampling:

View #1: Plenacoustic function [2]





[3]

Connected plane wave sparsity and compressed sensing in 2014

View #2: Courant–Friedrichs–Lewy condition[4]

If a wave is moving across a discrete spatial grid and we want to compute its amplitude at discrete time steps of equal duration, then this duration must be less than the time for the wave to travel to *adjacent grid points*

$$\Delta t \sum_{i=1}^{n} \frac{u_{x_i}}{\Delta x_i} \leq C_{\max}$$

$$\varphi_x^2 + \varphi_y^2 \leq \frac{\omega^2}{c^2} \Leftrightarrow \frac{c\Delta t}{\Delta x_i} \leq \frac{1}{\sqrt{2}}$$

$$\varphi_x^2 + \varphi_y^2 + \varphi_z^2 \leq \frac{\omega^2}{c^2} \Leftrightarrow \frac{c\Delta t}{\Delta x_i} \leq \frac{1}{\sqrt{3}}$$

Can we introduce a parametric solution and some assumptions that will reduce the complexity of the problem?

helena.peictukuljac@epfl.ch

[1] H. Nyquist, “Certain topics in telegraph transmission theory,” Transactions of the American Institute of Electrical Engineers, vol. 47, no. 2, pp. 617–644, April 1928.
[2] T. Ajdler, L. Sbaiz, and M. Vetterli, “The plenacoustic function and its sampling,” IEEE Transactions on Signal Processing, vol. 54, no. 10, pp. 3790–3804, Oct 2006.
[3] R. Mignot, G. Chardon and L. Daudet, "Low Frequency Interpolation of Room Impulse Responses Using Compressed Sensing," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 1, pp. 205-216, Jan. 2014
[4] Lewy, H., Friedrichs, K., and Courant, R.. "Über die partiellen Differenzengleichungen der mathematischen Physik." *Mathematische Annalen* 100 (1928): 32-74.

Wave equation and its parametrized solution

$$\Delta p(t, \mathbf{X}) - \frac{1}{c^2} \frac{\partial^2}{\partial^2 t} p(t, \mathbf{X}) = 0$$

$$p(t, \mathbf{X}) = \sum_{q \in \mathbb{I}} A_q \Phi_q(\mathbf{X}) g_q(t)$$

Spatial dependency: $\Phi_q(\mathbf{X}) \approx \sum_{r=1}^{R} a_{q,r} e^{j \mathbf{k}_{q,r} \cdot \mathbf{X}}$

Temporal dependency: $g_q(t) = e^{j k_q c t}$ $\quad k_q = \frac{\omega_q - j\xi_q}{c}$

$$p(t, \mathbf{X}) = \sum_{q,r} \alpha_{q,r} e^{j(k_q c t + \mathbf{k}_{q,r} \cdot \mathbf{X})}$$

The goal: Fast and efficient **parameter learning** from a *small* number of microphone measurements

$$p(t, \mathbf{X}) = \sum_{q,r} \alpha_{q,r} e^{j((\omega_q - j\xi_q)t + \mathbf{k}_{q,r} \cdot \mathbf{X})}$$

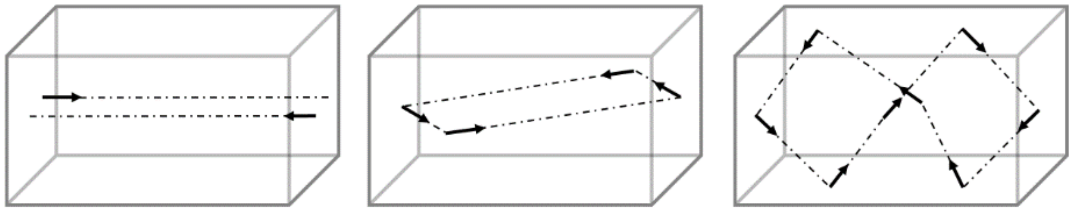
[1] R. Mignot, G. Chardon and L. Daudet, "Low Frequency Interpolation of Room Impulse Responses Using Compressed Sensing," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 1, pp. 205-216, Jan. 2014

helena.peictukuljac@epfl.ch

A set of assumptions for a well-posed problem

What natural **assumptions** can we introduce to reduce the complexity of the mathematical model at a low cost of approximation losses?

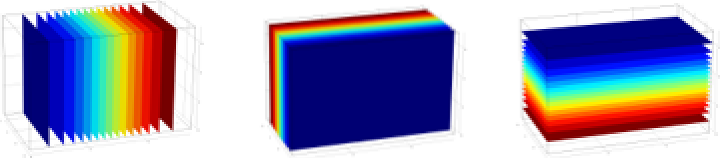
Assumption #1: simple (rectangular) room geometries $L_x \times L_y \times L_z$



x-axial mode | *xy*-tangential mode | oblique mode

Axial modes imply room shape

$$\omega_n = \pi c \sqrt{\left(\frac{n_x}{L_x}\right)^2 + \left(\frac{n_y}{L_y}\right)^2 + \left(\frac{n_z}{L_z}\right)^2}$$



$$\omega_{[1,0,0]} = \frac{\pi c}{L_x} \qquad \omega_{[0,1,0]} = \frac{\pi c}{L_y} \qquad \omega_{[0,0,1]} = \frac{\pi c}{L_z}$$

Assumption #2: lightly damped rooms

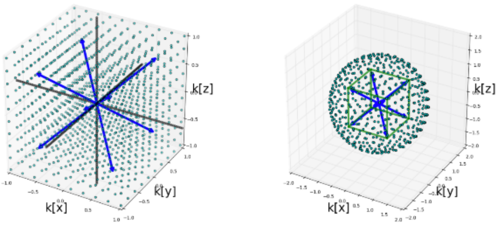
$$\mathbf{k}_q = [k_x \; k_y \; k_z]^T \qquad k_q = \frac{\omega_q - j\xi_q}{c}$$

$$\xi \ll \omega \qquad |\mathbf{k}| \approx \left|\frac{\omega}{c}\right| = \sqrt{k_x^2 + k_y^2 + k_z^2}$$

Spherical Search space

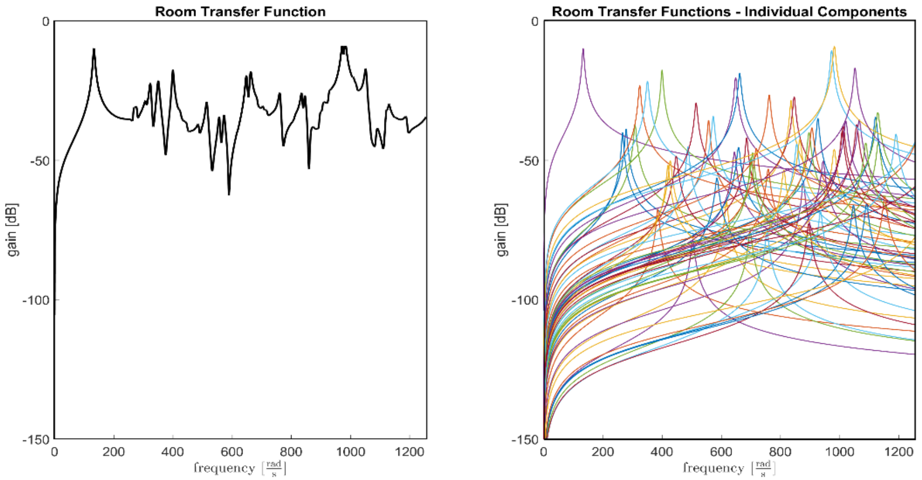
Separability of spatial and temporal parameter estimation:

1. *temporal:* estimate ω and build a ball $r = \frac{\omega}{c}$
2. *spatial*: estimate the direction of $\mathbf{k}$



Assumption #3: low frequency domain analysis

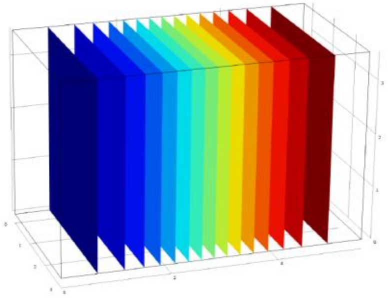
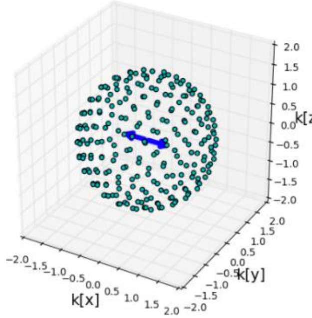
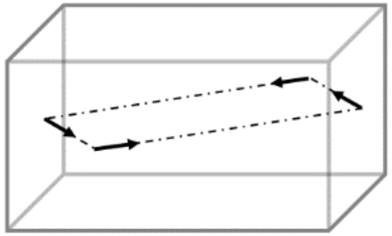
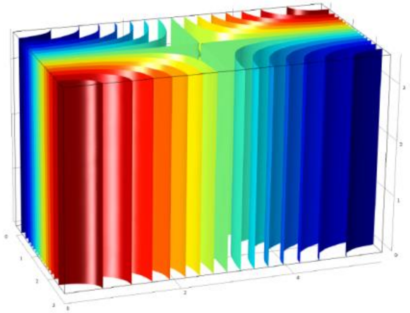
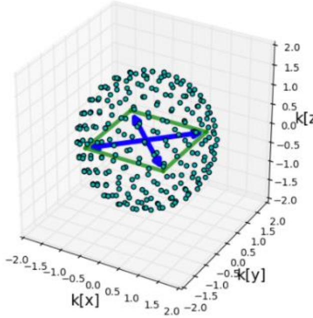
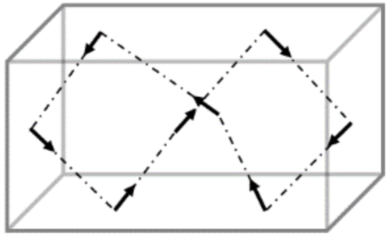
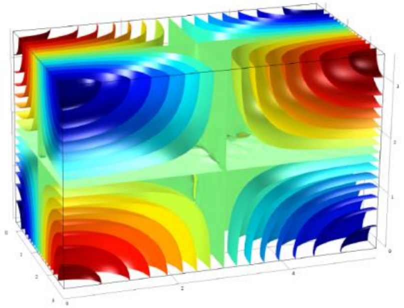
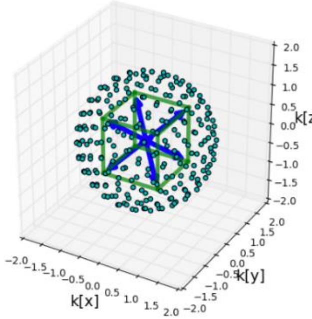
Room mode sparsity



helena.peictukuljac@epfl.ch

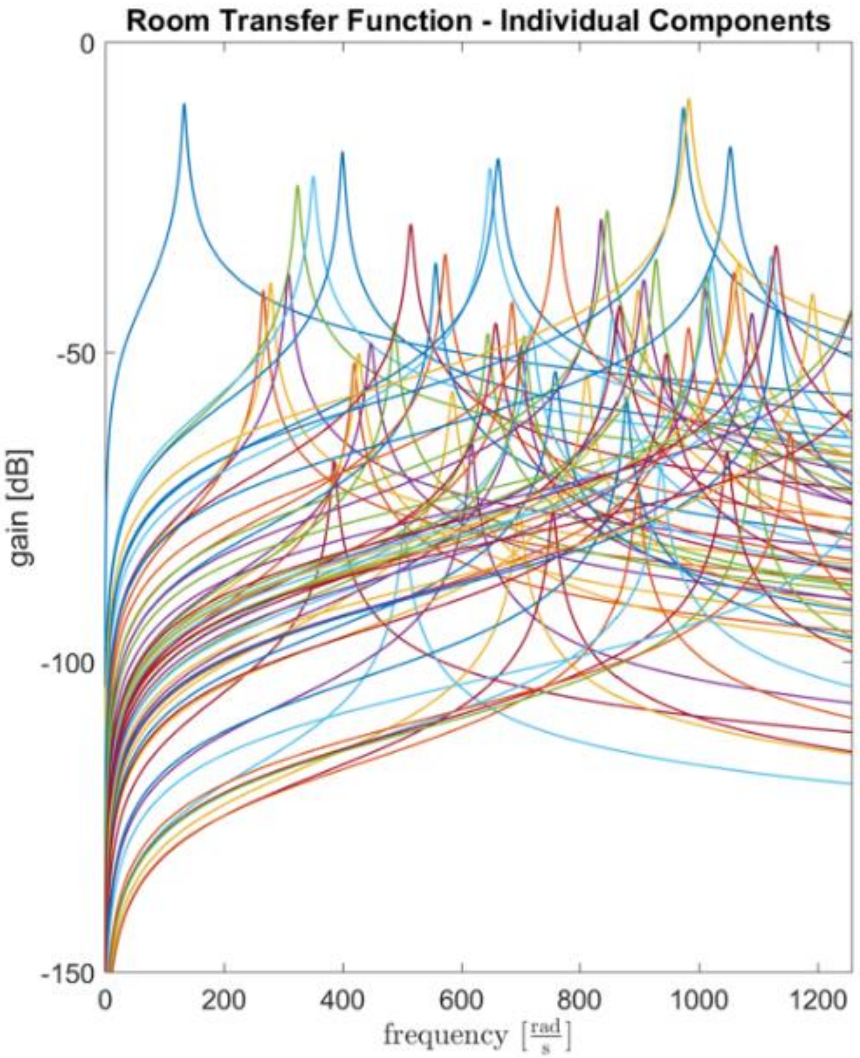
[1] Kuttruff, H. and Mommertz, E., "Room Acoustics", Springer Berlin Heidelberg, Berlin, Heidelberg(2013).
[2] R. Mignot, G. Chardon and L. Daudet, "Low Frequency Interpolation of Room Impulse Responses Using Compressed Sensing," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 1, pp. 205-216, Jan. 2014
[3] **H. Peic Tukuljac**, H. Lissek, and P. Vandergheynst, "Localization of sound sources in a room with one microphone," SPIE, Wavelets and Sparsity XVII, vol. 10394, 2017.

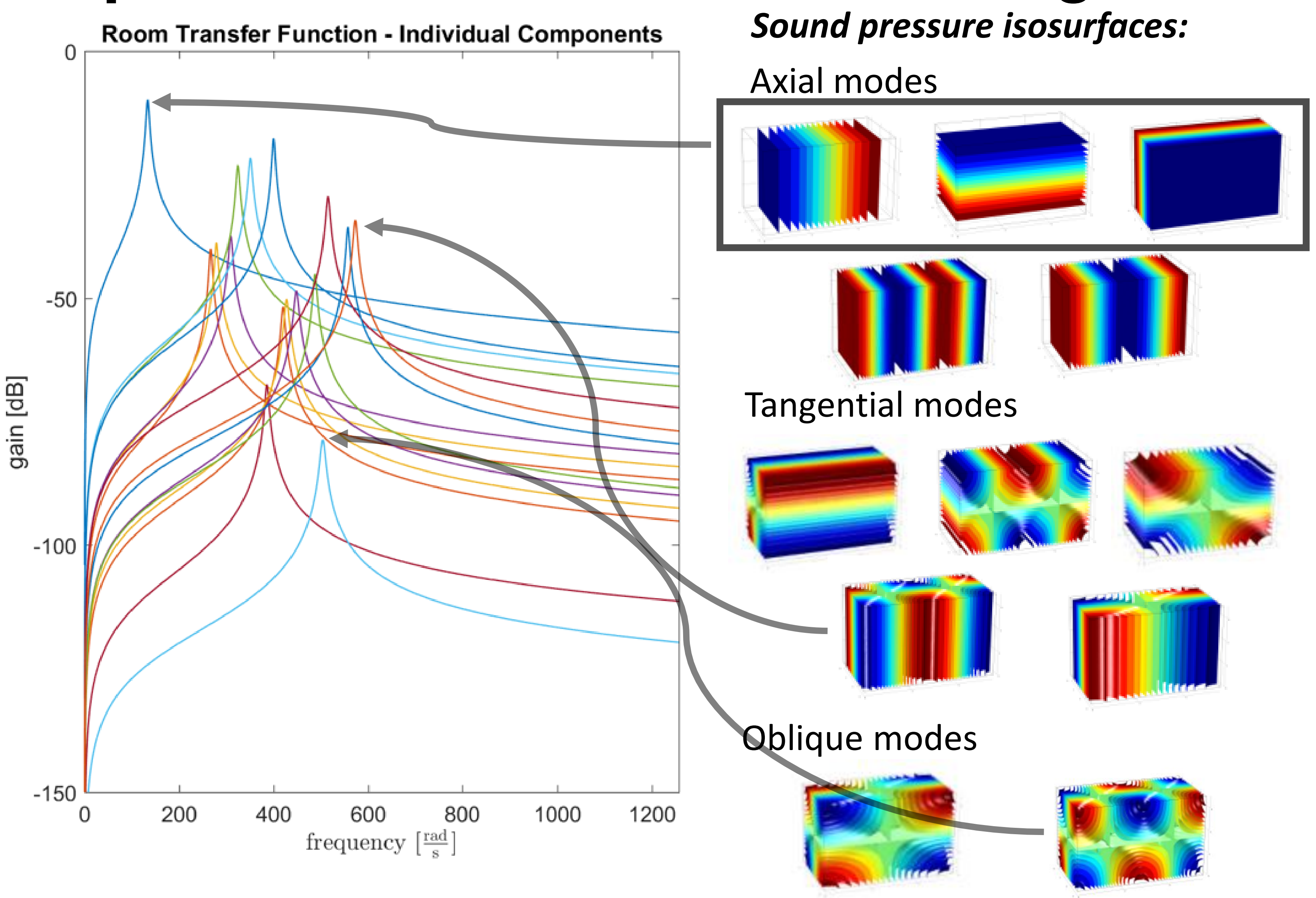
Three points of view on room modes

	Plane waves	Pressure isosurfaces	Wave vectors in search space $r = \frac{\omega}{c}$
***x*-axial mode**			
***xy*-tangential mode**			
oblique mode			

helena.peictukuljac@epfl.ch

Focusing on the low frequencies and room mode assignment





helena.peictukuljac@epfl.ch

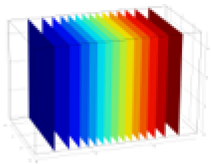
[1] **H. Peic Tukuljac**, H. Lissek, and P. Vandergheynst, "Localization of sound sources in a room with one microphone," SPIE, Wavelets and Sparsity XVII, vol. 10394, 2017.
[2] R. Boulandet, "Tunable Electroacoustic Resonators through Active Impedance Control of Loudspeakers", PhD thesis, EPFL, Lausanne, 2012

ReSEMblE algorithm

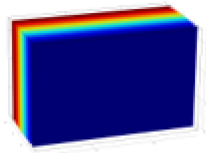
How to choose the partitioning frequency f_p (ω_p) ?

Based on the approximate room size:
Such that we capture all the basic axial modes

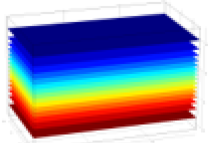
$$\omega_c = \frac{\pi c}{\tilde{L}_{\min}}$$



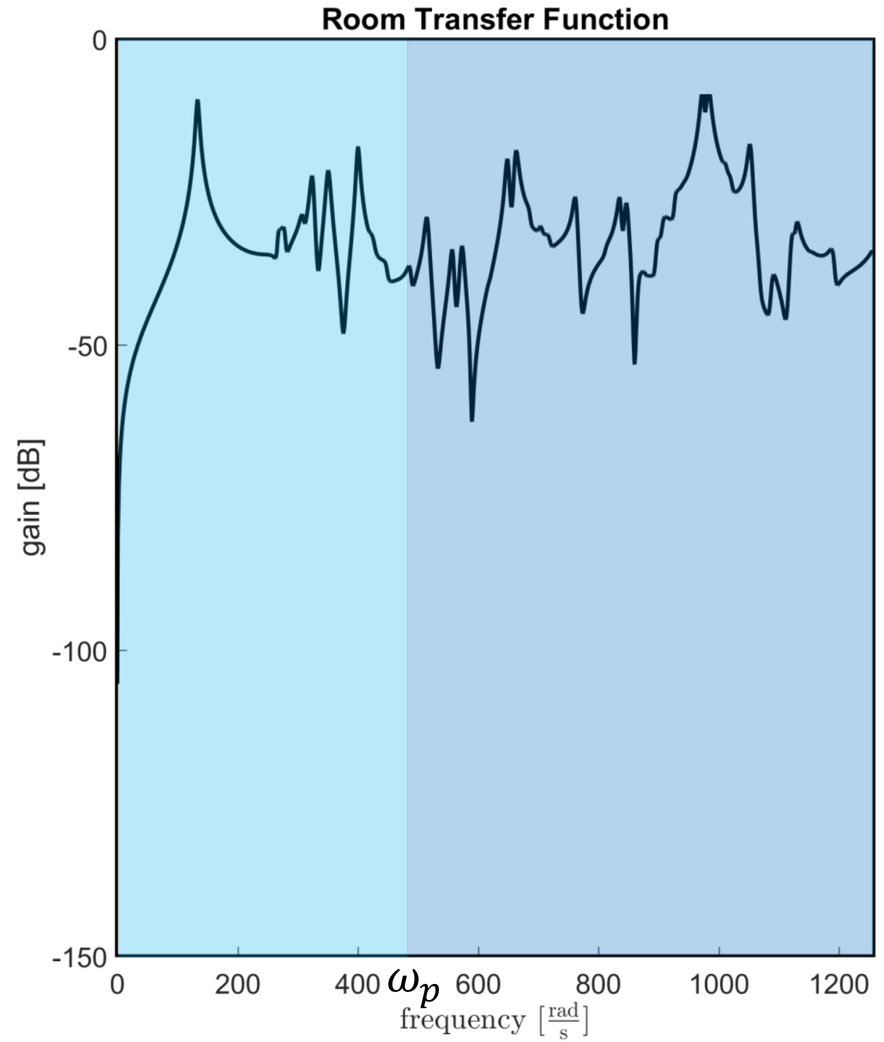
$\omega_{[1,0,0]} = \frac{\pi c}{L_x}$



$\omega_{[0,1,0]} = \frac{\pi c}{L_y}$



$\omega_{[0,0,1]} = \frac{\pi c}{L_z}$



procedure ReSEMBLE(**R**, **X**)
  Separate the measurements with f_p: $\mathbf{R} = \mathbf{R}^l + \mathbf{R}^h$.
  for $i_l \in \{1, ..., N_l\}$ **do**
    step 1: estimate $(\omega_{i_l}, \xi_{i_l})$ from $\mathbf{R}^l_{i_l}$
    step 2: estimate $\mathbf{k}_{i_l}$ from $\mathbf{r}^l_{i_l}$
    step 3: compute new residual $\mathbf{R}^l_{i_l+1}$
  end for

  Recover the room size $\tilde{L}_x, \tilde{L}_y, \tilde{L}_z$ from basic axial room modes and form the regular wave vector grid.
  for $i_h \in \{N_l + 1, ..., N\}$ **do**
    step 1: get ω_{i_h} and $\mathbf{k}_{i_h}$ from the wave vector grid
    step 2: estimate ξ_{i_h} from $\mathbf{R}^h_{i_h}$
    step 3: compute new residual $\mathbf{R}^h_{i_h+1}$
  end for
  Estimate the expansion coefficients $\{\alpha\}_{n=1,v=1}^{N,V}$ using least square approach.
end procedure

ReSEMblE algorithm

Temporal dictionary: $\theta[i] = e^{\xi_n[i]t} e^{j\omega_n[i]t}$

Spatio-temporal dictionary: $\Sigma[:, i] = e^{\xi_{i_l} t} e^{j\omega_{i_l} t} e^{\mathbf{X}\cdot\mathbf{k}[i]}$

procedure RESEMBLE($\mathbf{R}$, $\mathbf{X}$)

Separate the measurements with f_p: $\mathbf{R} = \mathbf{R}^l + \mathbf{R}^h$.

for $i_l \in \{1, ..., N_l\}$ **do**

step 1: estimate $(\omega_{i_l}, \xi_{i_l})$ from $\mathbf{R}^l_{i_l}$

step 2: estimate $\mathbf{k}_{i_l}$ from $\mathbf{r}^l_{i_l}$

step 3: compute new residual $\mathbf{R}^l_{i_l+1}$

end for

Recover the room size $\tilde{L}_x, \tilde{L}_y, \tilde{L}_z$ from basic axial room modes and form the regular wave vector grid.

for $i_h \in \{N_l + 1, ..., N\}$ **do**

step 1: get ω_{i_h} and $\mathbf{k}_{i_h}$ from the wave vector grid

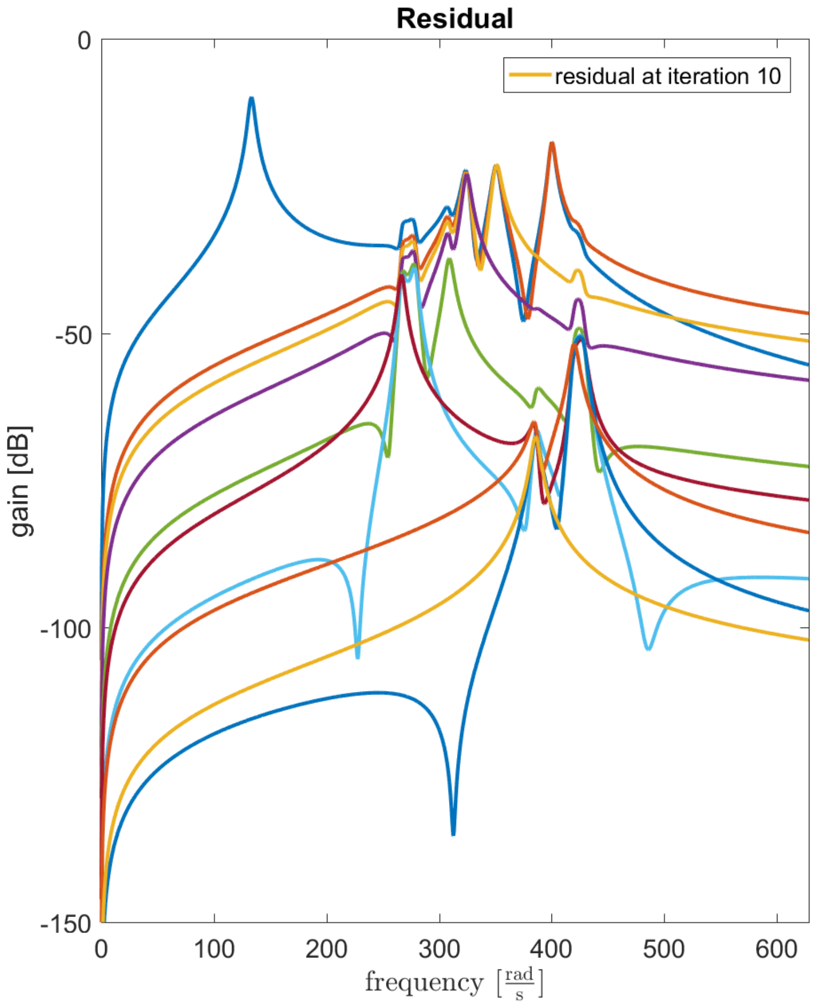
step 2: estimate ξ_{i_h} from $\mathbf{R}^h_{i_h}$

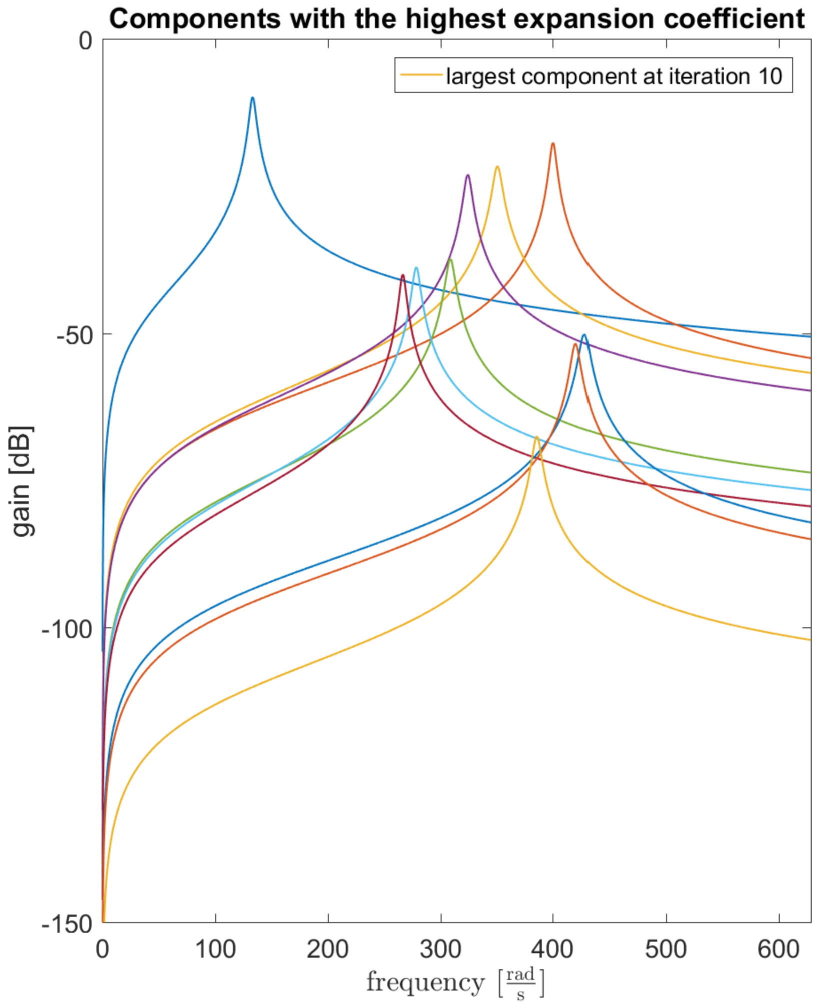
step 3: compute new residual $\mathbf{R}^h_{i_h+1}$

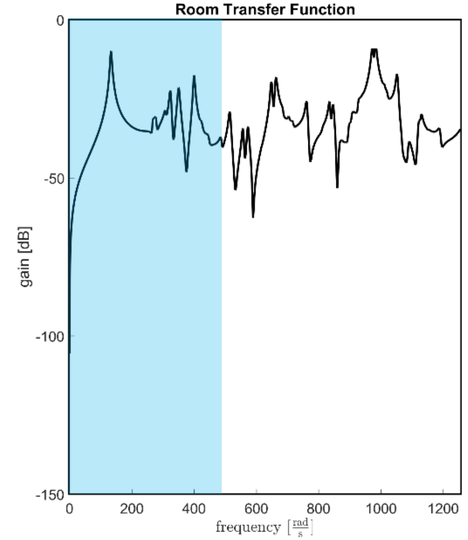
end for

Estimate the expansion coefficients $\{\alpha\}_{n=1,v=1}^{N,V}$ using least square approach.

end procedure







helena.peictukuljac@epfl.ch

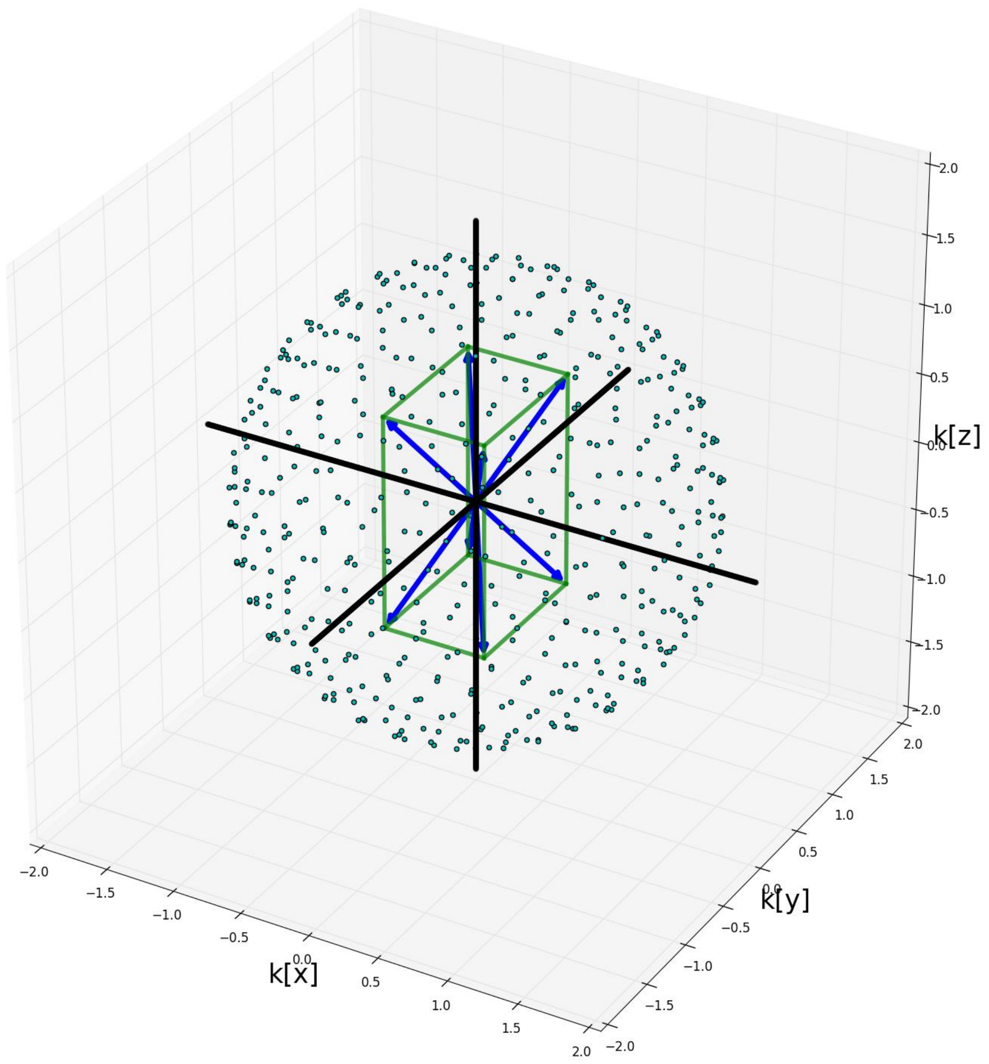
[1] J. A. Tropp, A. C. Gilbert and M. J. Strauss, "Simultaneous sparse approximation via greedy pursuit," *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, 2005, pp. v/721-v/724 Vol. 5. doi: 10.1109/ICASSP.2005.1416405

Structured group sparsity

Separability of spatial and temporal parameter estimation:

1. temporal: estimate ω and build a ball $r = \frac{\omega}{c}$

2. *spatial*: estimate the direction of $\mathbf{k}$

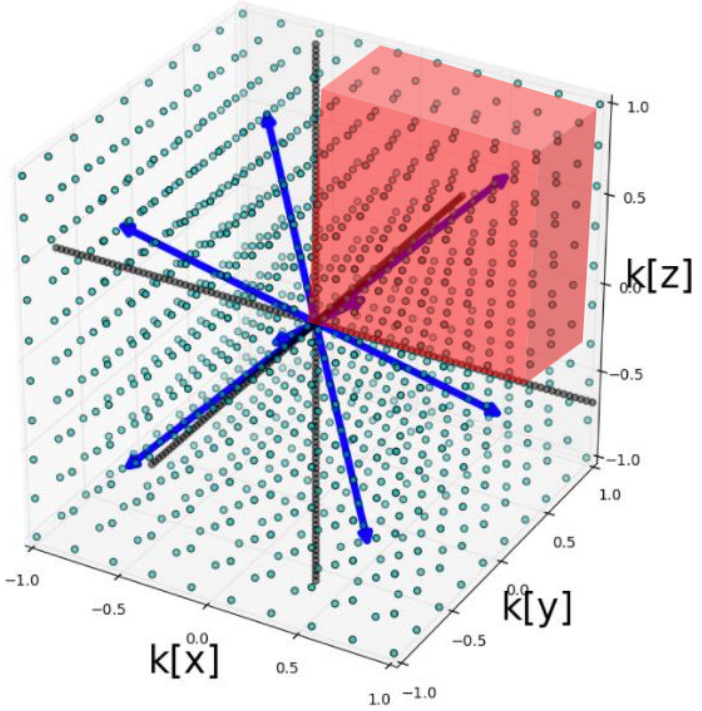


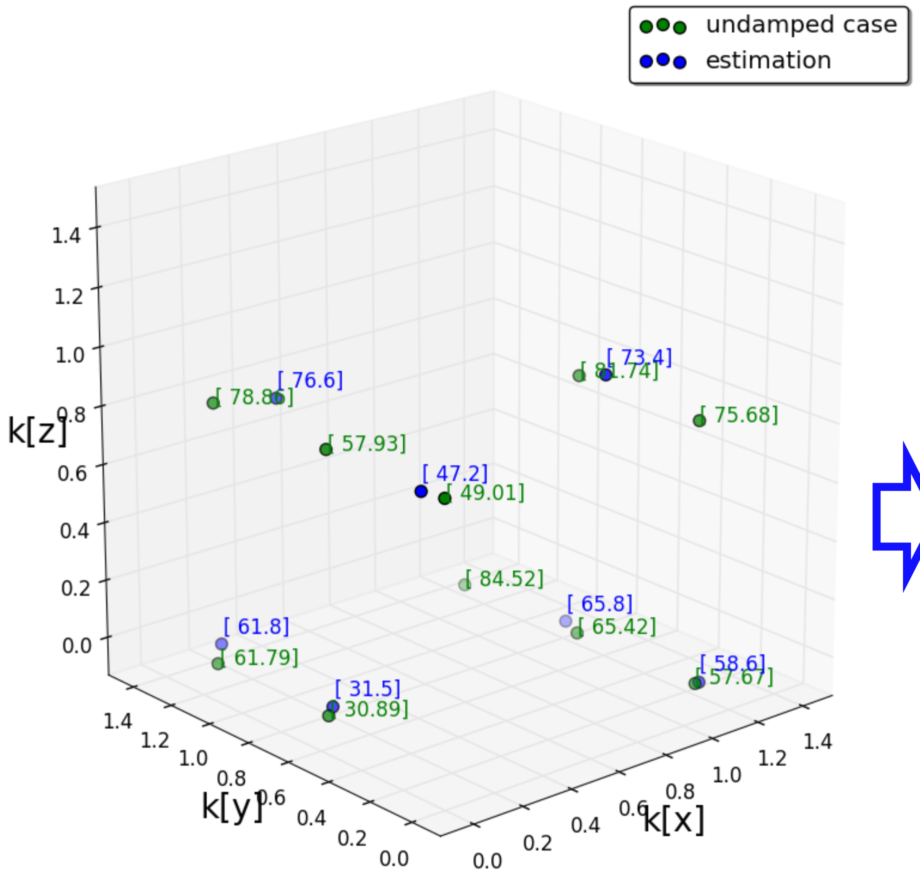
k-space estimation results

$$|k| \approx \left|\frac{\omega}{c}\right|$$

Why the position of points deviates the most over *z*-axis?
This is why the *x*-, *y*- and *xy*-modes are precise and *z*-, *xz*-, *xy*- and *xyz*-modes are far off from the ground truth

$$k_q = \frac{\omega_q - j\xi_q}{c} \qquad \xi \ll \omega$$





procedure RESEMBLE($\mathbf{R}$, $\mathbf{X}$)
  Separate the measurements with f_p: $\mathbf{R} = \mathbf{R}^l + \mathbf{R}^h$.
  for $i_l \in \{1, ..., N_l\}$ **do**
    step 1: estimate $(\omega_{i_l}, \xi_{i_l})$ from $\mathbf{R}^l_{i_l}$
    step 2: estimate $\mathbf{k}_{i_l}$ from $\mathbf{r}^l_{i_l}$
    step 3: compute new residual $\mathbf{R}^l_{i_l+1}$
  end for
  Recover the room size $\tilde{L}_x, \tilde{L}_y, \tilde{L}_z$ from basic axial room modes and form the regular wave vector grid.
  for $i_h \in \{N_l + 1, ..., N\}$ **do**
    step 1: get ω_{i_h} and $\mathbf{k}_{i_h}$ from the wave vector grid
    step 2: estimate ξ_{i_h} from $\mathbf{R}^h_{i_h}$
    step 3: compute new residual $\mathbf{R}^h_{i_h+1}$
  end for
  Estimate the expansion coefficients $\{\alpha\}^{N,V}_{n=1,v=1}$ using least square approach.
end procedure

$$\tilde{L}_x, \tilde{L}_y, \tilde{L}_z$$

helena.peictukuljac@epfl.ch

[1] J. A. Tropp, A. C. Gilbert and M. J. Strauss, "Simultaneous sparse approximation via greedy pursuit," *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, 2005, pp. v/721-v/724 Vol. 5. doi: 10.1109/ICASSP.2005.1416405

ReSEMblE algorithm

Imply the wave vectors from the reconstructed room shape:

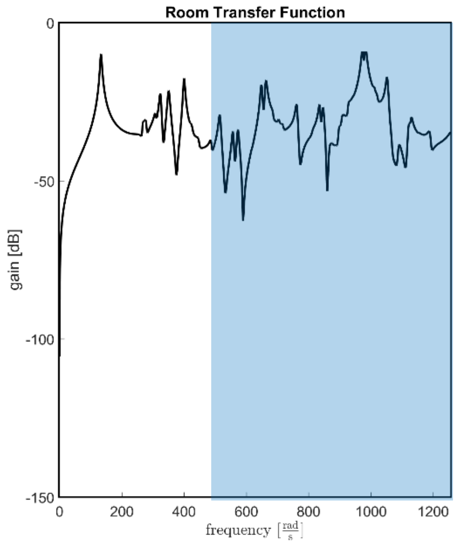
$$(n_x, n_y, n_z) \in \mathcal{N}_0^3 \backslash (0,0,0)$$

$$k_x = n_x \frac{\pi}{\tilde{L}_x} \qquad k_y = n_y \frac{\pi}{\tilde{L}_y} \qquad k_z = n_z \frac{\pi}{\tilde{L}_z}$$

After applying the *high* part of the algorithm, the Pearson correlation coefficient that the approximation is good (e.g. 82% for only 19-microphone setting and $f_c = 200\text{Hz}$), but it should be further improved once the deviation of the wave vectors is efficiently characterized

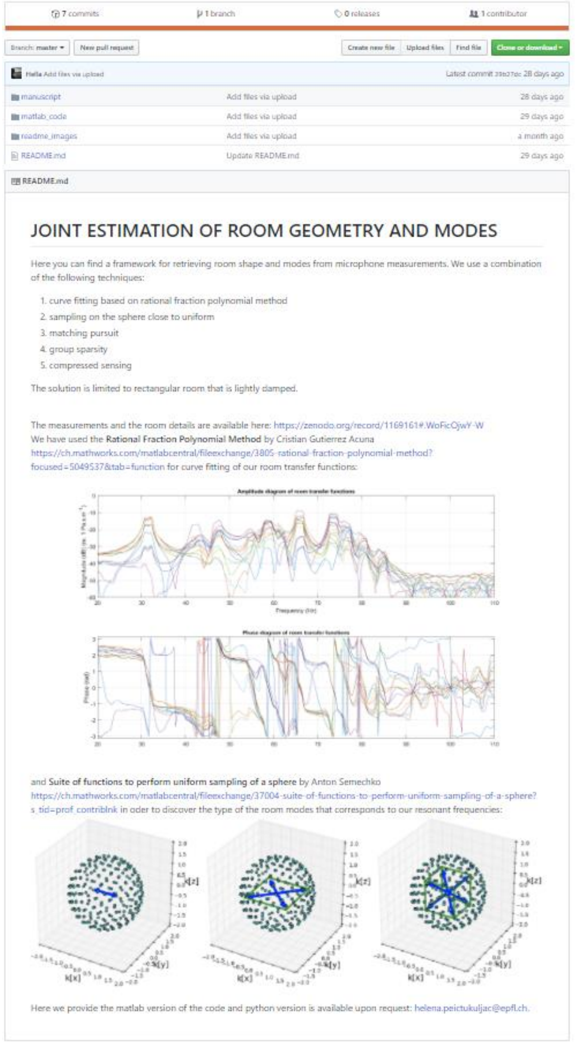
procedure RESEMBLE($\mathbf{R}$, $\mathbf{X}$)
  Separate the measurements with f_p: $\mathbf{R} = \mathbf{R}^l + \mathbf{R}^h$.
  for $i_l \in \{1, ..., N_l\}$ **do**
    step 1: estimate $(\omega_{i_l}, \xi_{i_l})$ from $\mathbf{R}^l_{i_l}$
    step 2: estimate $\mathbf{k}_{i_l}$ from $\mathbf{r}^l_{i_l}$
    step 3: compute new residual $\mathbf{R}^l_{i_l+1}$
  end for

  Recover the room size $\tilde{L}_x, \tilde{L}_y, \tilde{L}_z$ from basic axial room modes and form the regular wave vector grid.
  for $i_h \in \{N_l + 1, ..., N\}$ **do**
    step 1: get ω_{i_h} and $\mathbf{k}_{i_h}$ from the wave vector grid
    step 2: estimate ξ_{i_h} from $\mathbf{R}^h_{i_h}$
    step 3: compute new residual $\mathbf{R}^h_{i_h+1}$
  end for
  Estimate the expansion coefficients $\{\alpha\}^{N,V}_{n=1,v=1}$ using least square approach.
end procedure

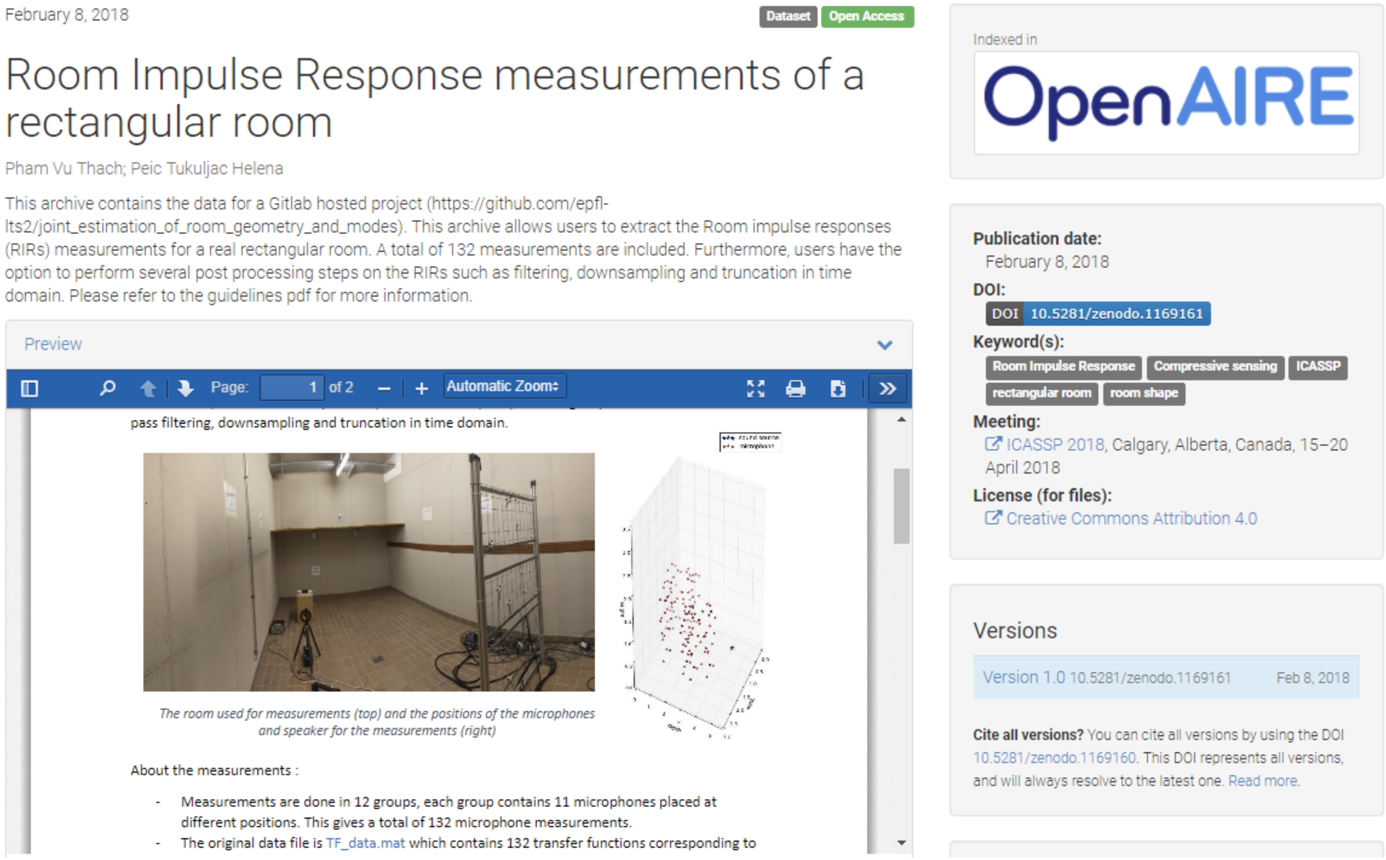


helena.peictukuljac@epfl.ch

In the spirit of *open research* and *acoustic data augmentation*



[1]

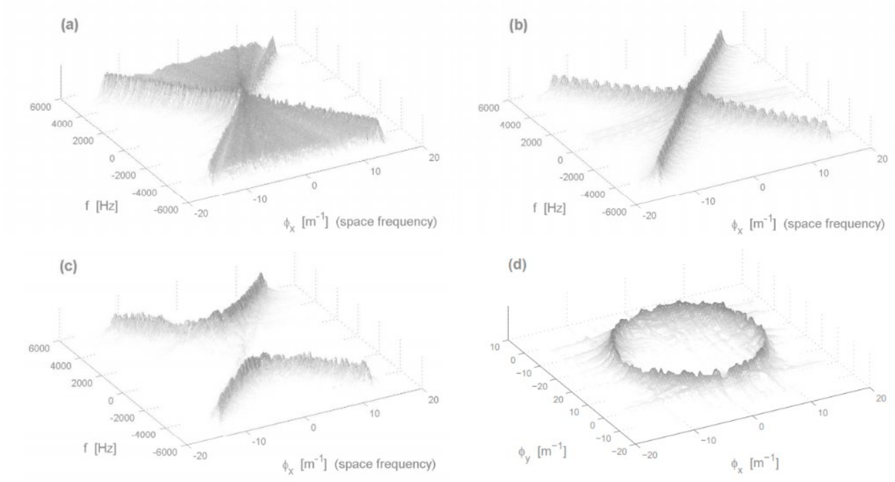


[2]

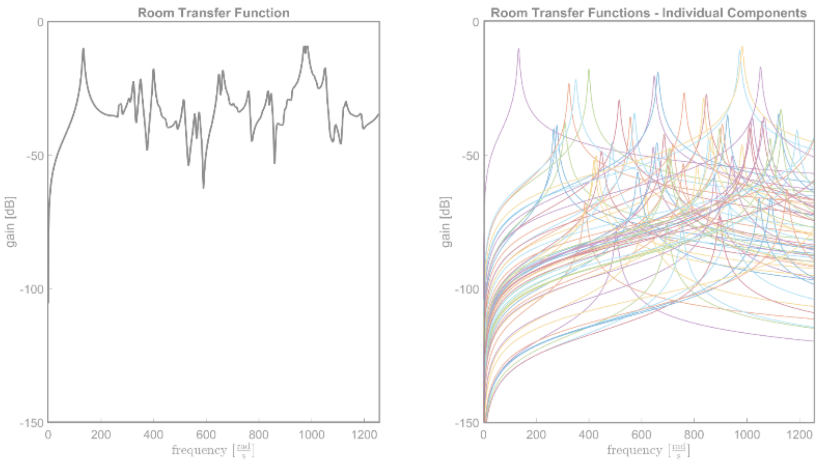
Content Atlas

Introduction

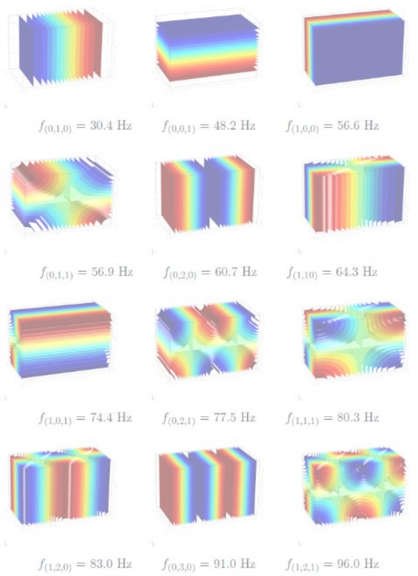
Spatio-temporal sampling relations [1]



Room transfer function [2]

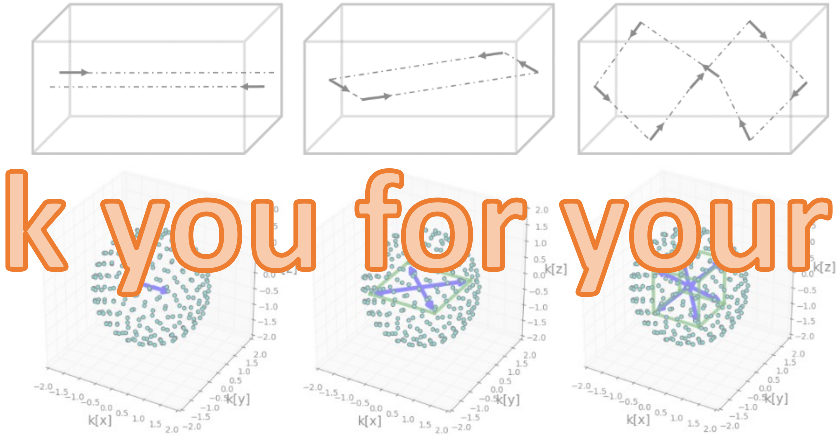


Room modes [3]

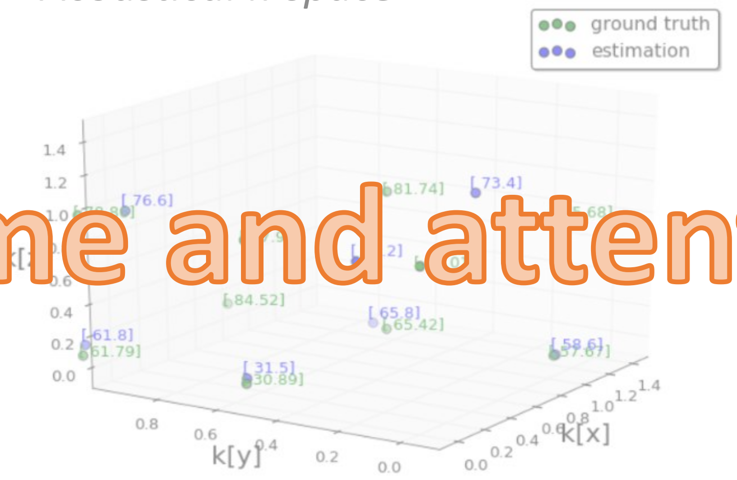


Contribution

Plane waves and wave vectors



Acoustical k-space



Thank you for your time and attention!

helena.peictukuljac@epfl.ch

[1] R. Mignot, G. Chardon and L. Daudet, "Low Frequency Interpolation of Room Impulse Responses Using Compressed Sensing," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 1, pp. 205-216, Jan. 2014.
[2] **H. Peic Tukuljac**, H. Lissek, and P. Vandergheynst, “Localization of sound sources in a room with one microphone,” SPIE, Wavelets and Sparsity XVII, vol. 10394, 2017.
[3] R. Boulandet, “Tunable Electroacoustic Resonators through Active Impedance Control of Loudspeakers”, PhD thesis, EPFL, Lausanne, 2012