AUD-19.2

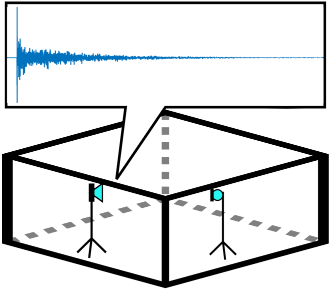
Region-to-region kernel interpolation of acoustic transfer function with directional weighting

Juliano G. C. Ribeiro, Shoichi Koyama, and Hiroshi Saruwatari

Graduate School of Information Science and Technology, The University of Tokyo

Background

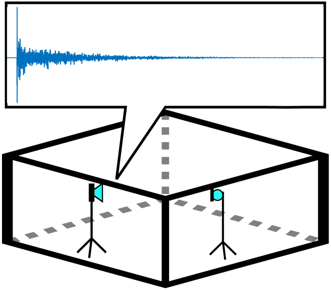
- Soundwave propagation inside an environment isn't predictable.
- The relation between source and receiver signal can be estimated with the **acoustic transfer function** (ATF) of the space.



Background

ATF interpolation with variable source and receiver

- Most ATF interpolation methods have a fixed source position.
- Our objective is to describe how the ATF changes for variable source and receiver within assigned regions. A **region-to-region** interpolation.



Background

⇒There are established region-to-region interpolation methods.

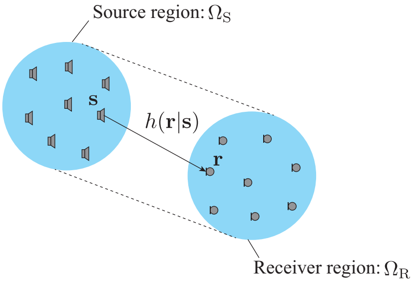
- In [Samarasinghe+, 2015], an ATF interpolation method using a spherical wavefunction expansion was proposed.
- This formulation was compared to a kernel ridge regression method with a specialized kernel in [Ribeiro+, 2020].
- The kernel method outperformed the wavefunction expansion for every frequency.

⇒The kernel method still has issues

- Neither method takes into consideration the distribution of the sources and receivers.
- The methods are rather susceptible to noise contamination, and as such are vulnerable to outliers in the data.

Problem statement

- Our objective is to accurately estimate the ATF between a variable receiver/source pair $\mathbf{r}|\mathbf{s}$ within regions in a space $\Omega \subset \mathbb{R}^3$.
- We distribute L loudspeakers in a source region $\Omega_\mathrm{S} \subset \Omega$ and M microphones in a receiver region $\Omega_\mathrm{R} \subset \Omega$.
- In order to interpolate the ATF between regions, we must derive an interpolation function from the $N = LM$ measurements.



Problem statement

Properties of the ATF

- The ATF is the superposition of two components:

$$h(\mathbf{r}|\mathbf{s},k) = h_{\mathrm{D}}(\mathbf{r}|\mathbf{s},k) + h_{\mathrm{R}}(\mathbf{r}|\mathbf{s},k).$$

- The direct component h_{D} is considered to be the equivalent of a recorded signal in the free-field, caused by a point source.

$$h_{\mathrm{D}}(\mathbf{r}|\mathbf{s},k) = G_0(\mathbf{r}|\mathbf{s},k) = \frac{e^{ik\|\mathbf{r}-\mathbf{s}\|}}{4\pi\|\mathbf{r}-\mathbf{s}\|}$$

- The reverberant component h_{R} satisfies the Helmholtz equation on both position variables:

$$(\nabla_{\mathbf{r}}^2 + k^2)h_{\mathrm{R}}(\mathbf{r}|\mathbf{s},k) = (\nabla_{\mathbf{s}}^2 + k^2)h_{\mathrm{R}}(\mathbf{r}|\mathbf{s},k) = 0$$

- Reciprocity for every pair $\mathbf{r}|\mathbf{s}$: $h(\mathbf{r}|\mathbf{s},k) = h(\mathbf{s}|\mathbf{r},k)$

Optimization objective

- Since h_{D} is considered known, we focus our efforts on h_{R}.
- The interpolation function $\hat{h}_{\mathrm{R}}$ obtained from our reverberant field measurements $\mathbf{y} = [y_1,\ y_2,\ \ldots,\ y_N]$.
- We create the cost to be minimized:

$$\mathcal{J}(f) := \sum_{n=1}^{N} |y_n - f(\mathbf{q}_n)|^2 + \lambda \|f\|_{\mathscr{H}}^2,\ f \in \mathscr{H}$$

- The vector $\mathbf{q}_n$ represents the n-th source/receiver position pair.
- The measurement vector $\mathbf{y}$ is obtained by removing the direct component from all impulse response recordings.

The functional space $\mathscr{H}$ must be defined to be representative of the data

Kernel ridge regression

- When $\mathscr{H}$ is a reproducing kernel Hilbert space (RKHS) of reproducing kernel κ, the minimizer of $\mathcal{J}$ has a closed form:

$$\hat{h}_{\mathrm{R}}(\mathbf{r}|\mathbf{s}) = \boldsymbol{\kappa}(\mathbf{r}|\mathbf{s})(\mathbf{K} + \lambda \mathbf{I})^{-1}\mathbf{y},$$

  where:

$$\boldsymbol{\kappa}(\mathbf{r}|\mathbf{s}) = [\kappa(\mathbf{r}|\mathbf{s}, \mathbf{q}_1), \ldots, \kappa(\mathbf{r}|\mathbf{s}, \mathbf{q}_N)],$$

$$\mathbf{K} = \begin{bmatrix} \kappa(\mathbf{q}_1, \mathbf{q}_1) & \kappa(\mathbf{q}_1, \mathbf{q}_2) & \ldots & \kappa(\mathbf{q}_1, \mathbf{q}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \kappa(\mathbf{q}_N, \mathbf{q}_1) & \kappa(\mathbf{q}_N, \mathbf{q}_2) & \ldots & \kappa(\mathbf{q}_N, \mathbf{q}_N) \end{bmatrix},$$

  and $\lambda > 0$.

- In [Ribeiro+, 2020] we have shown that if the RKHS is defined with the properties of the ATF in mind, we can achieve accurate interpolations.

Definition of the generalized Hilbert space

- We express h_{R} using the Herglotz wavefunction $h_{\mathrm{R}}(\mathbf{r}|\mathbf{s}) = \mathcal{I}\left(\tilde{h}_{\mathrm{R}};\mathbf{r}|\mathbf{s}\right)$, where:

$$\mathcal{I}\left(f;\mathbf{r}|\mathbf{s}\right) := \int_{\mathbb{S}^2\times\mathbb{S}^2} e^{\mathrm{i}k(\hat{\mathbf{r}}\cdot\mathbf{r}+\hat{\mathbf{s}}\cdot\mathbf{s})} f(\hat{\mathbf{r}},\hat{\mathbf{s}})\mathrm{d}\hat{\mathbf{r}}\mathrm{d}\hat{\mathbf{s}}.$$

- We can thus define the Hilbert space $(\mathscr{H}, \langle\cdot,\cdot\rangle_{\mathscr{H}})$ as:

$$\mathscr{H} = \Big\{h_{\mathrm{R}} = \mathcal{I}\left(\tilde{h}_{\mathrm{R}};\mathbf{r}|\mathbf{s}\right) : \tilde{h}_{\mathrm{R}} \in L^2(W,\mathbb{S}^2\times\mathbb{S}^2), \quad \tilde{h}_{\mathrm{R}}(\hat{\mathbf{r}},\hat{\mathbf{s}}) = \tilde{h}_{\mathrm{R}}(\hat{\mathbf{s}},\hat{\mathbf{r}})\ \forall\hat{\mathbf{r}},\hat{\mathbf{s}}\in\mathbb{S}^2\Big\}$$

$$\langle f,g\rangle_{\mathscr{H}} = \int_{\mathbb{S}^2\times\mathbb{S}^2} \frac{\overline{\tilde{f}(\hat{\mathbf{r}},\hat{\mathbf{s}})}\tilde{g}(\hat{\mathbf{r}},\hat{\mathbf{s}})}{W(\hat{\mathbf{r}},\hat{\mathbf{s}})}\mathrm{d}\hat{\mathbf{r}}\mathrm{d}\hat{\mathbf{s}},\ \forall f,g\in\mathscr{H}$$

Directional kernel formulation

- The weight function W gives the reproducing kernel κ as

$$\kappa(\mathbf{r}|\mathbf{s},\mathbf{r}'|\mathbf{s}') = \mathcal{I}\left(W(\hat{\mathbf{r}},\hat{\mathbf{s}})\left(\frac{e^{-\mathrm{i}k(\hat{\mathbf{r}}\cdot\mathbf{r}'+\hat{\mathbf{s}}\cdot\mathbf{s}')}+e^{-\mathrm{i}k(\hat{\mathbf{r}}\cdot\mathbf{s}'+\hat{\mathbf{s}}\cdot\mathbf{r}')}}{2}\right);\mathbf{r}|\mathbf{s}\right)$$

- In [Ribeiro+, 2020], the relative position of the regions within Ω was not taken into account.
- The introduction of a weight function enables the choice of a kernel function better adapted for the region configurations.
- The weight will be separable, $W(\hat{\mathbf{r}},\hat{\mathbf{s}}) = w(\hat{\mathbf{r}})w(\hat{\mathbf{s}})$ in order to simplify calculations and guarantee reciprocity.
- As the direct component is removed, plane wave components in the direct path are expected to be less significant.

Proposed directional kernel weight function

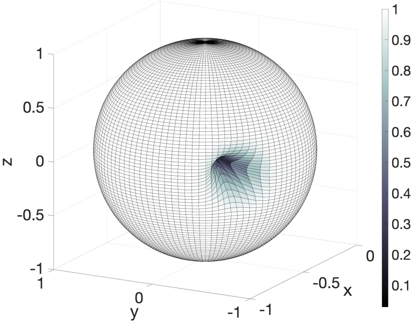
- The weight we propose for the interpolation method is:

$$w(\hat{\mathbf{v}}) = \frac{1}{4\pi}\left(1 + \gamma^2 - \frac{\cosh(\beta \hat{\mathbf{v}} \cdot \hat{\mathbf{v}}_0)}{\cosh(\beta)}\right), \ \hat{\mathbf{v}} \in \mathbb{S}^2$$

- The direction $\hat{\mathbf{v}}_0$ is the direction connecting the centers of Ω_{S} and Ω_{R}.
- The hyperparameter β adjusts the selectivity around the direct component.
- The hyperparameter γ adjusts the minimum gain baseline of the weight.

Proposed directional kernel weight function

- Below we have an example of the weight function, represented in a gain plot.
- For a direction $\hat{\mathbf{v}}$, the distance from the center to the surface is $w(\hat{\mathbf{v}})$.



$\Rightarrow$ But how do we choose β and γ?

Leave-one-out cross-validation

- Optimizing the loss function to our hyperparameters might over-condition the method to the measured data, which has noise.
- We opted instead to minimize the leave-one-out cross-validation error (LOO).

$$\mathrm{LOO}(\mathbf{y}, \ell) = \frac{1}{N} \sum_{n=1}^{N} \ell \left(\hat{f}_n(\mathbf{q}_n) - y_n \right)$$

- The loss ℓ was either square error (SQE) or Tukey loss.

$$\mathrm{SQE}(z) = |z|^2$$

$$\mathrm{Tukey}(z) = \begin{cases} \frac{\sigma^2}{6} \left(1 - \left(1 - \frac{|z|^2}{\sigma^2} \right)^3 \right), & |z| \leqslant \sigma \\ \frac{\sigma^2}{6}, & |z| > \sigma \end{cases}$$

The uniform weight kernel

- For $\gamma = 1$, $\beta = 0$, we have a uniform weight $w = 1/4\pi$, κ is known to be:

$$\kappa(\mathbf{r}|\mathbf{s}, \mathbf{r}'|\mathbf{s}') =$$
$$\frac{1}{2}\left(j_0(k\|\mathbf{r} - \mathbf{r}'\|)j_0(k\|\mathbf{s} - \mathbf{s}'\|) + j_0(k\|\mathbf{s} - \mathbf{r}'\|)j_0(k\|\mathbf{r} - \mathbf{s}'\|)\right)$$

- This kernel function coincides with the one used in [Ribeiro+, 2020], making this estimation identical.
- The weighted kernel is an extension of this kernel function.
- This method will be the standard of comparison.

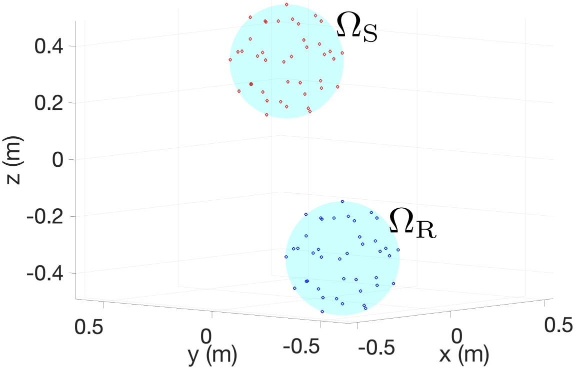
Experimental simulations

- We conducted numerical simulations with the image source method to compare both interpolation functions
- The arrays in both source and receiver regions had $L = M = 41$ points.
- Simulator conditions:

Room dimensions	$[3.2, 4.0, 2.7]$ m
Reverberation time T_{60}	0.45 s
Reflection coefficients	$[0.802, 0.866, 0.945]$
Inner radius of the array	0.19m
Outer radius of the array	0.20m
Signal-to-noise ratio	20 dB

Experimental simulations

- The center of the cartesian system is at the geometric center of the room.
- The centers of the source and receiver region were $\mathbf{s}_0 = [0.35, 0.43, 0.29]^\mathsf{T}$ m and $\mathbf{r}_0 = [-0.35, -0.43, -0.29]^\mathsf{T}$ m.



Error criteria

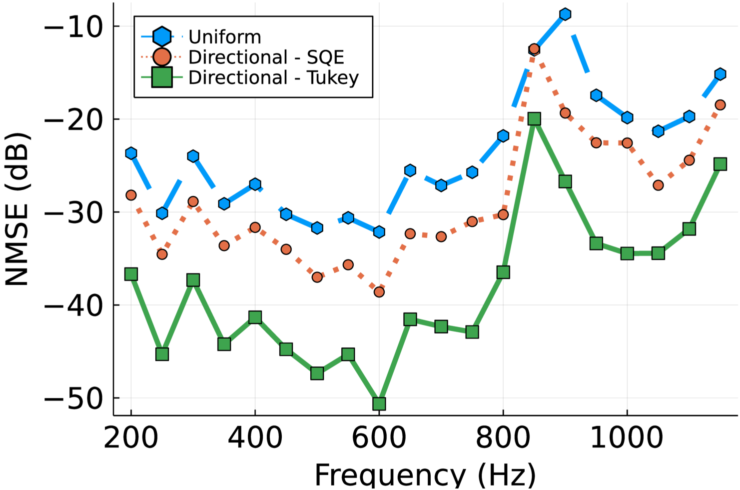
- The first criterion was the normalized mean square error (NMSE)

$$\mathrm{NMSE} = 10\log_{10}\left(\frac{\sum_n \left|\hat{h}(\mathbf{q}'_n) - h(\mathbf{q}'_n)\right|^2}{\sum_n |h(\mathbf{q}'_n)|^2}\right)$$

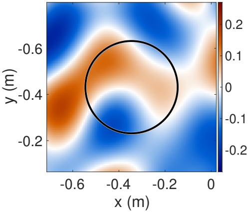
- We analyzed the NMSE for 9025 possible densely-distributed source-receiver evaluation pairs given as $\{\mathbf{q}'_n\}_{n=1}^{9025}$.
- The second criterion was the normalized square error (NSE) distribution in the regions.
- The frequency of analysis for the NSE was $950\ \mathrm{Hz}$

Normalized mean square error

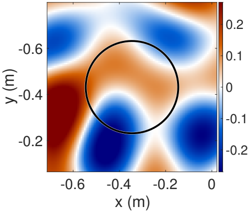
Normalized mean square error comparison:



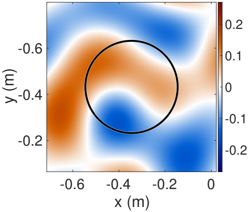
Colormaps of the real part, comparing the reconstruction of the signal generated by a single source in the center of Ω_{S}:



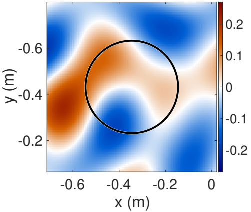
(a) Original



(b) Uniform

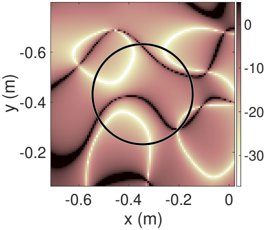


(c) SQE

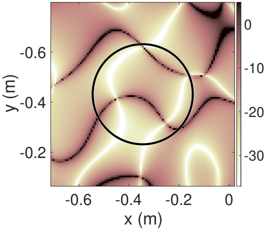


(d) Tukey

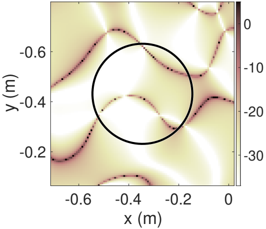
Normalized square error distributions:



(a) Uniform



(b) SQE



(c) Tukey

Conclusion

In summary:

- We defined a function space $\mathscr{H}$ using the physical properties of the ATF, which allowed us to interpolate its value for variable source and receiver positions.
- We generalized a previously established kernel formulation by adding directionality based on the expected profile of the ATF.
- This proposed formulation can be optimized using the same data points used to derive the model.
- The directional kernel estimations outperformed the uniform kernel in both mean error by frequency and in reconstructing the ATF spatially.
- Additionally, the use of a robust loss criterion also gave us better results than the standard square error.