Multi-Microphone Speaker Localization on Manifolds

Sharon Gannot

joint work with Bracha Laufer-Goldshtein and Ronen Talmon

Bar-Ilan University, Ramat-Gan, Israel

IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC 2022), October 25, 2022

Bracha Laufer-Goldshtein, Bar-Ilan University (Currently at Massachusetts Institute of Technology – MIT)

Ronen Talmon, Technion – Israel Institute of Technology

Sharon Gannot, Bar-Ilan University

Acoustic Source Localization & Tracking

An Essential component in Speech Processing Applications

Devices & Applications

1. Hands-free devices
2. Smart homes and cars
3. Smart speakers & Personal assistants
4. Camera steering
5. Robot audition
6. Hearing aids and hearables

Prior Art

Families of localization algorithms

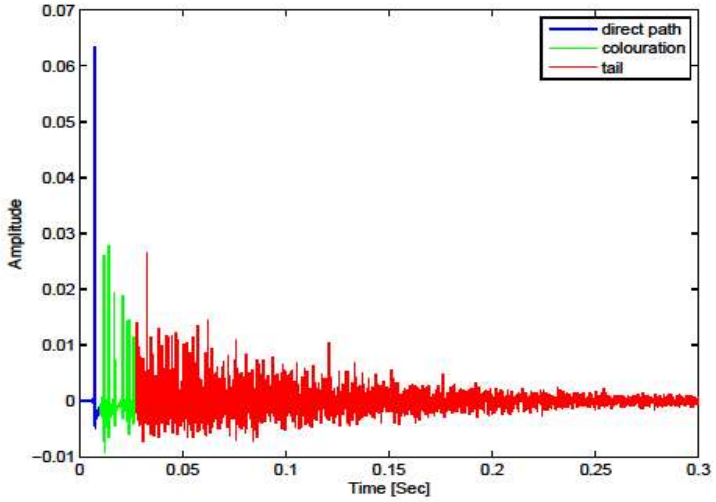
- Beamformers steered towards all potential locations (directions)
- TDOA estimation + geometric intersection
- Bayesian
- Non-Bayesian
- Learning-based methods: unsupervised (e.g. MoG-EM) and supervised (manifold-learning, deep-learning)

See a structured list of algorithms and references

Why Localization is so Difficult?

Room Acoustics Essentials

- When sound propagates in an enclosure it undergoes reflections from its surfaces
- Mathematical/statistical models of the acoustic path:
  - Virtual images beyond room walls
    [Allen and Berkley, 1979, Peterson, 1986]
  - Statistical models for late reflections
    [Polack, 1993, Schroeder, 1996, Jot et al., 1997]
  - Diffuseness of late reflections (non-directional)
    [Dal Degan and Prati, 1988, Habets and Gannot, 2007]



Describing the sound propagation is a cumbersome task

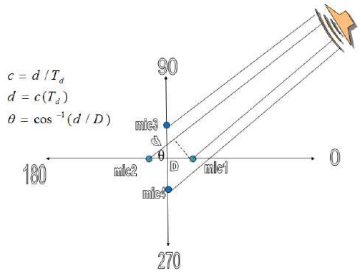
Which is the Best Model for the Problem at Hand?

How to model the acoustic path?

Classical speech processing

TDOA-only

Too simplified



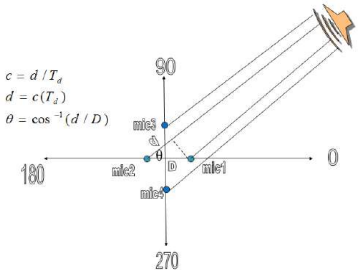
Which is the Best Model for the Problem at Hand?

How to model the acoustic path?

Classical speech processing

TDOA-only

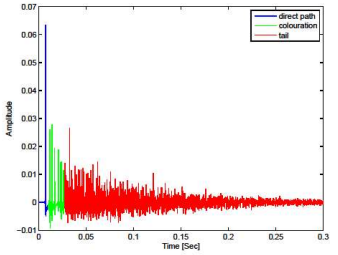
Too simplified



To cope with reverberations

Entire Impulse Response

Perfect but Complicated: No simple Model



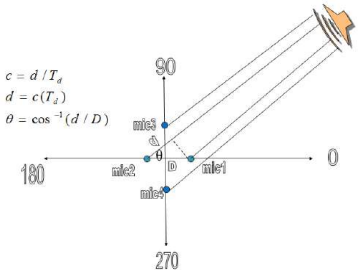
Which is the Best Model for the Problem at Hand?

How to model the acoustic path?

Classical speech processing

TDOA-only

Too simplified



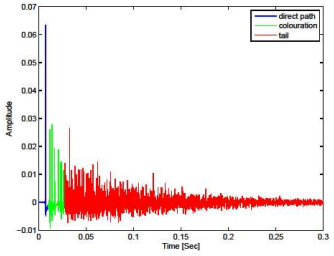
Manifold Learning

Simple to describe and to implement

To cope with reverberations

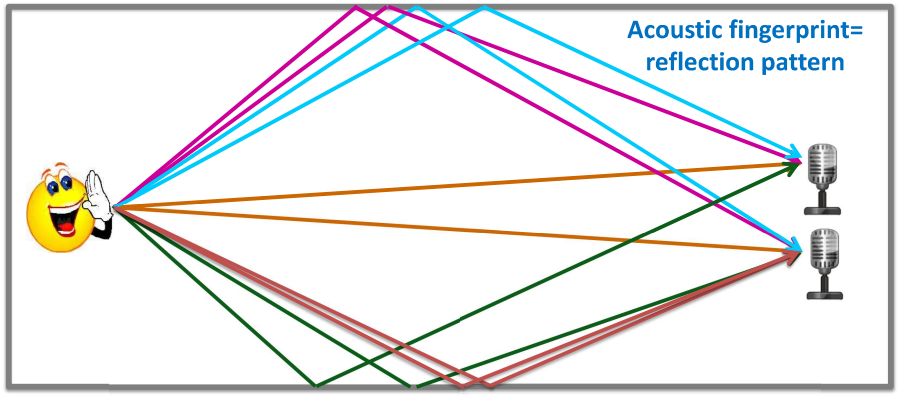
Entire Impulse Response

Perfect but Complicated: No simple Model



Our Proposed Localization & Tracking Methodology

Most classical localization methods are ignoring the richness of the acoustic propagation path



Environment-aware & data-driven acoustic source localization scheme

- Take advantage of the intricate reflection pattern and define an acoustic fingerprint characterizing source position

  [Gannot et al., 2001];[Markovich et al., 2009]

- Data-driven paradigms are harnessed to:
  - Show that the collection of these acoustic fingerprints pertain to a low-dimensional acoustic manifold
  - Extract the geometrical structure of the acoustic manifold and reveal its intrinsic degrees of freedom (DoFs) associated with the location
  - Infer state-space models from the manifold structure moving speaker tracking scenarios

Outline

Outline

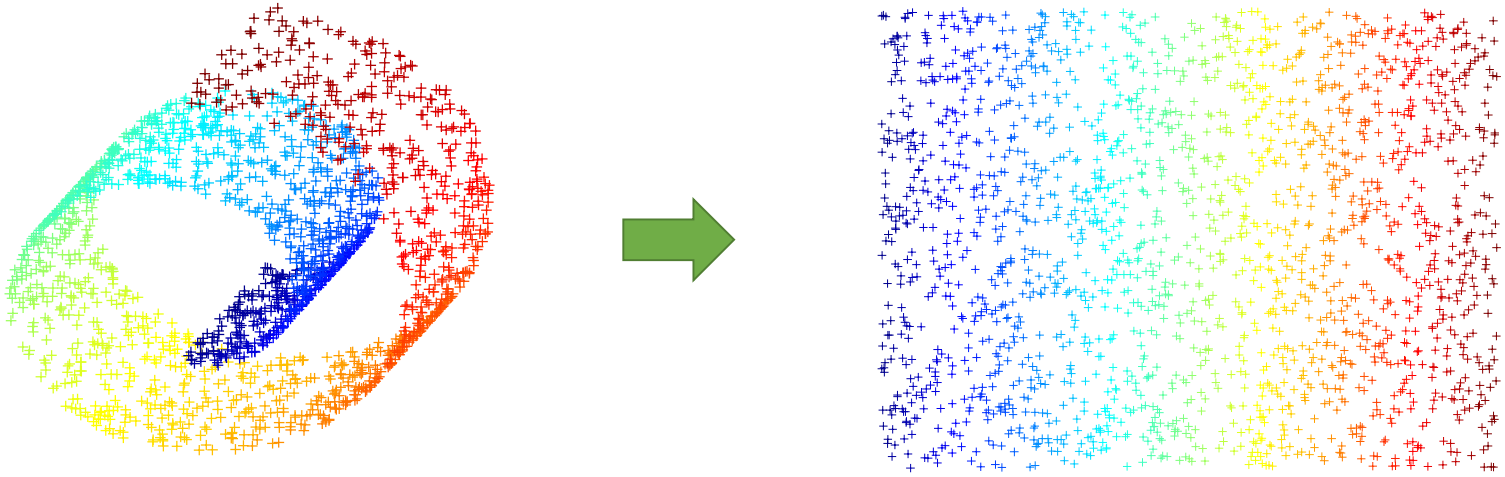
Data Representation & Manifolds

Measured data often:

- Exhibit highly redundant representations
- Controlled by a small set of parameters
- Lie on a low-dimensional manifold

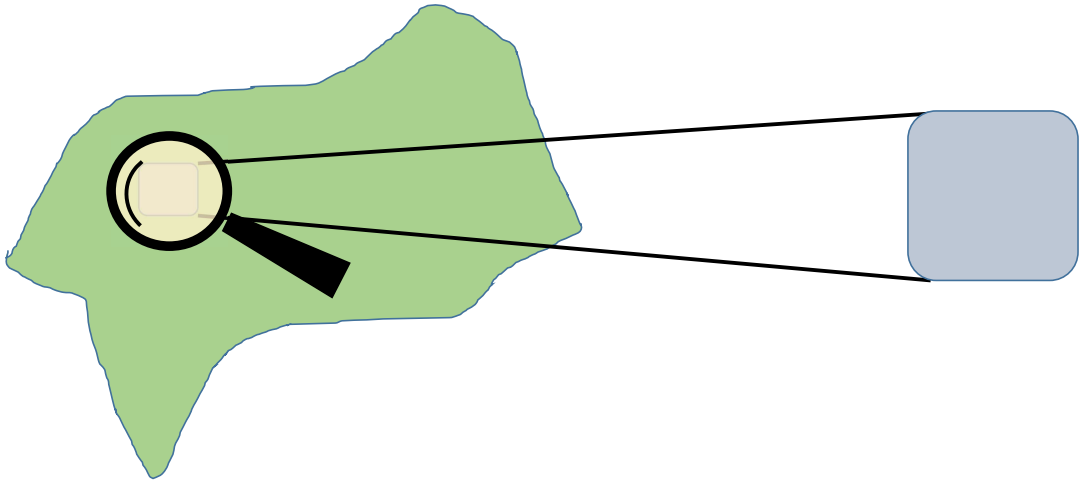
Dimensionality reduction

- Consider n high-dimensional features $\mathbf{h}_i \in \mathbb{R}^D$ extracted from the data
- Construct a low-dimensional representation $\mathbf{y}_i \in \mathbb{R}^d$ of $\mathbf{h}_i$, $d < D$ that respects the manifold geometric structure



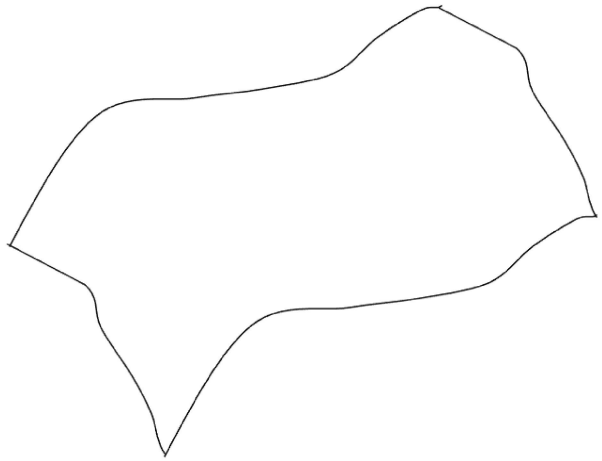
Laplacian

- Roughly, a manifold is a space that is locally Euclidean
- The Laplacian Δ is an operator defined by the divergence of the gradient of a function in a Euclidean space: $\Delta = \nabla \cdot \nabla$
- The Laplace–Beltrami operator $\mathcal{L}$: Extension to Riemannian manifolds
- The Laplacian contains all the information about the manifold geometry and induces a local coordinate system
- The Laplacian describes the time-evolution of a diffusion process (heat equation)



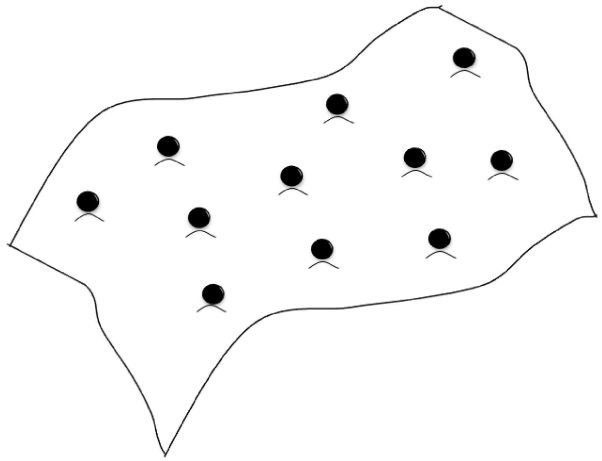
Discretization of the Manifold

- The Laplacian is an infinite-dimension operator defined on continuous spaces



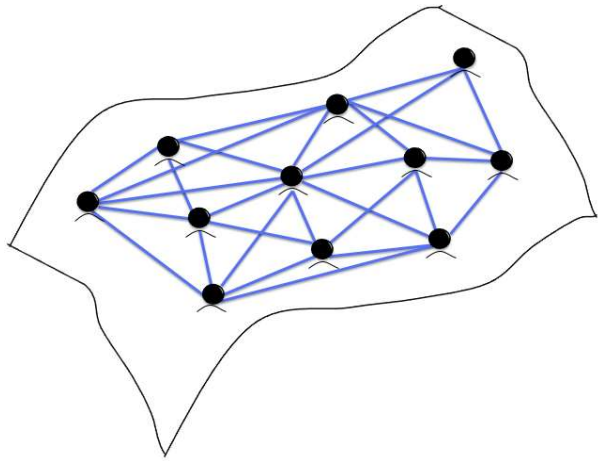
Discretization of the Manifold

- The Laplacian is an infinite-dimension operator defined on continuous spaces
  - We are typically given a finite set of observations in discrete spaces
  - What is the finite-dimension counterpart of the Laplacian?



Discretization of the Manifold

- The Laplacian is an infinite-dimension operator defined on continuous spaces
  - We are typically given a finite set of observations in discrete spaces
  - What is the finite-dimension counterpart of the Laplacian?
- The manifold can be empirically represented by a graph
  - The observations are the graph nodes
  - Define a finite operator (matrix) – the graph Laplacian (will be explicitly defined later)



Manifold Learning Paradigms

The goal of manifold learning

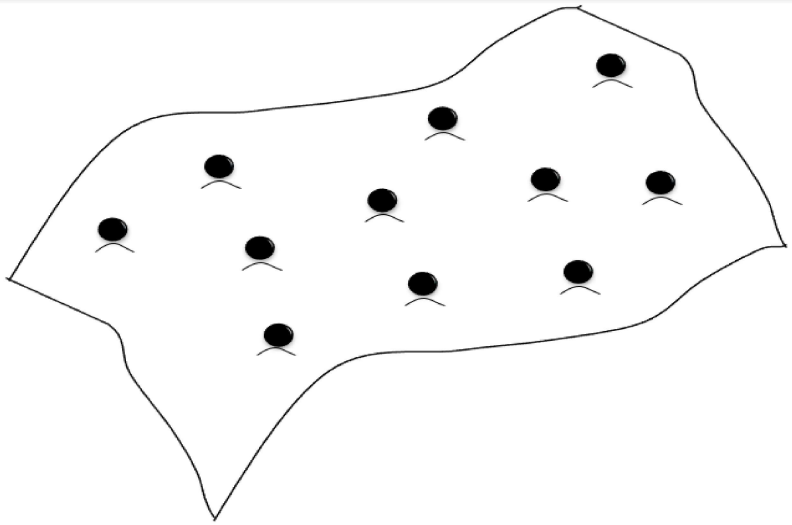
Given high-dimensional points without any prior data modelling, the goal is to recover the manifold from the data

Classical methods

- The foundations of manifold learning were laid in Science, December 2000 issue:
  - Locally linear embedding (LLE) [Roweis and Saul, 2000]
  - Isometric feature mapping (ISOMAP) [Tenenbaum et al., 2000]
- We will focus on diffusion maps due to the notion of diffusion distance [Coifman and Lafon, 2006]

Diffusion Maps [Coifman and Lafon, 2006]

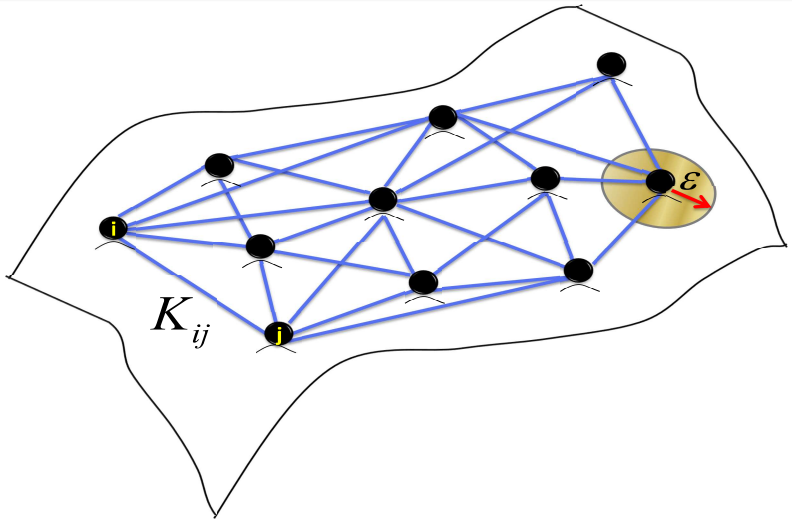
- Samples are the graph nodes



Diffusion Maps [Coifman and Lafon, 2006]

- Samples are the graph nodes
- The weights of the edges are defined using a kernel function:

$$K_{ij} = k(\mathbf{h}_i, \mathbf{h}_j) = \exp\left\{-\frac{\|\mathbf{h}_i - \mathbf{h}_j\|^2}{\varepsilon}\right\}$$



Diffusion Maps [Coifman and Lafon, 2006]

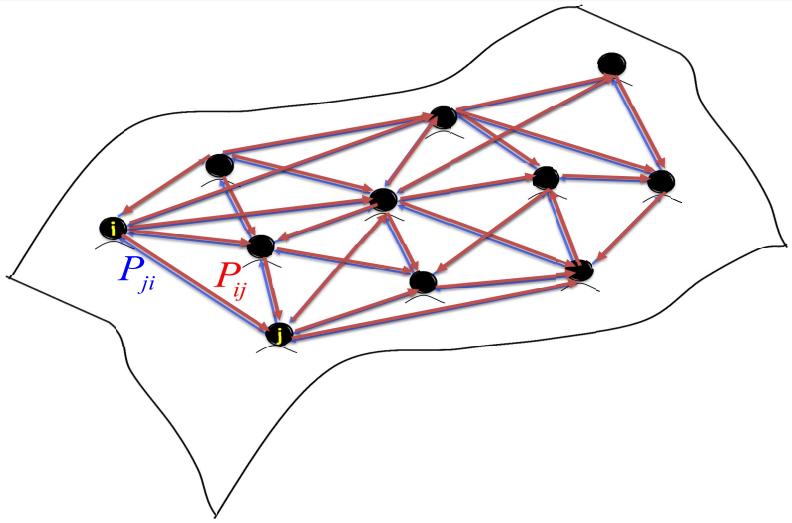
- Samples are the graph nodes
- The weights of the edges are defined using a kernel function:

$$K_{ij} = k(\mathbf{h}_i, \mathbf{h}_j) = \exp\left\{-\frac{\|\mathbf{h}_i - \mathbf{h}_j\|^2}{\varepsilon}\right\}$$

- Define a Markov process on the graph by the transition matrix:

$$P_{ij} = p(\mathbf{h}_i, \mathbf{h}_j) = K_{ij} / \sum_{r=1}^{N} K_{ir}$$

which is a discretization of a diffusion process on the manifold



Diffusion Maps [Coifman and Lafon, 2006]

- Samples are the graph nodes
- The weights of the edges are defined using a kernel function:

$$K_{ij} = k(\mathbf{h}_i, \mathbf{h}_j) = \exp\left\{-\frac{\|\mathbf{h}_i - \mathbf{h}_j\|^2}{\varepsilon}\right\}$$

- Define a Markov process on the graph by the transition matrix:

$$P_{ij} = p(\mathbf{h}_i, \mathbf{h}_j) = K_{ij} / \sum_{r=1}^{N} K_{ir}$$

which is a discretization of a diffusion process on the manifold

$\Rightarrow$ $\mathbf{P} = \mathbf{D}^{-1}\mathbf{K} \in \mathbb{R}^{n \times n}$, $\mathbf{D}$ is diagonal with $D_{ii} = \sum_{r=1}^{n} K_{ir}$

- $\mathbf{P}$ is similar to a symmetric matrix and hence has a real spectrum
- The (normalized) graph Laplacian, defined as $\mathbf{N} = \mathbf{I} - \mathbf{P}$, asymptotically ($\varepsilon \to 0$ $n \to \infty$) converges to the Laplacian $\mathcal{L}$

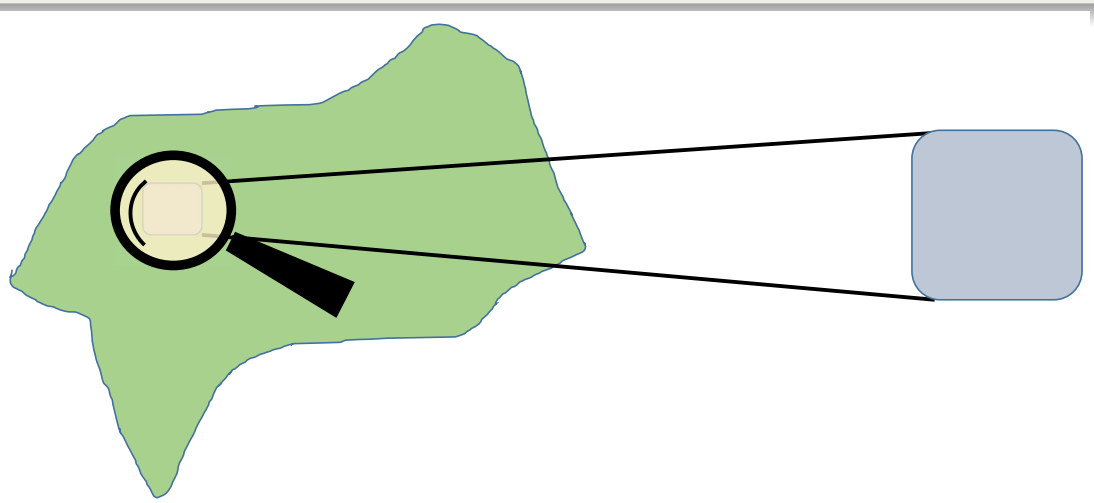
$\Rightarrow$ The normalized graph Laplacian $\mathbf{N}$ (and $\mathbf{P}$) contains the information about the manifold geometry

Diffusion Maps [Coifman and Lafon, 2006]

- Apply eigenvalue decomposition (EVD) to the matrix $\mathbf{P} \in \mathbb{R}^{n \times n}$ and obtain n eigenvalues $\{\lambda_j\}$ and n right eigenvectors $\{\boldsymbol{\varphi}_j\}$ in $\mathbb{R}^n$
- A nonlinear mapping into a new d-dimensional Euclidean space:

$$\boldsymbol{\Phi}_d : \mathbf{h}_i \mapsto [\lambda_1 \varphi_1(i), \ldots, \lambda_d \varphi_d(i)]^T$$

where $d < n$ is typically set by prior knowledge or according to a "spectral gap"



Q: In what sense the space is Euclidean?

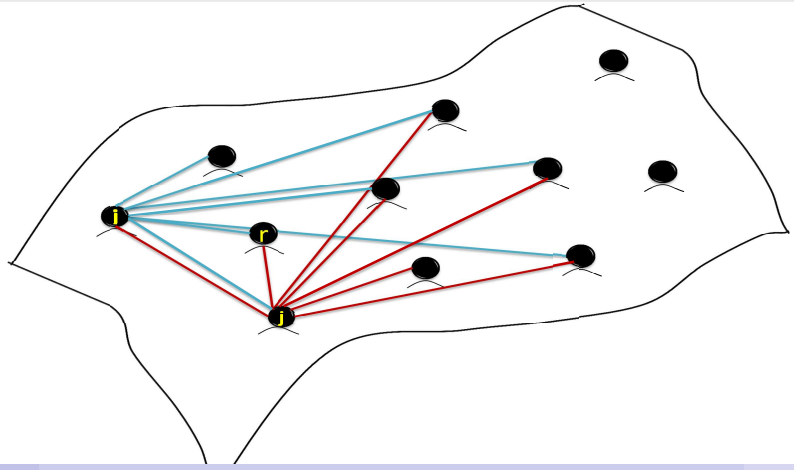
Diffusion Distance

The distance along the manifold is approximated by the diffusion distance:

$$D^2_{\text{Diff}}(\mathbf{h}_i, \mathbf{h}_j) = \sum_{r=1}^{n} \left(p\left(\mathbf{h}_i, \mathbf{h}_r\right) - p\left(\mathbf{h}_j, \mathbf{h}_r\right)\right)^2 / \phi_0^{(r)}$$

- $D^2_{\text{Diff}}(\mathbf{h}_i, \mathbf{h}_j)$ will be small if there is a large number of paths connecting $\mathbf{h}_i$ and $\mathbf{h}_j$ that is, if there is a large probability of transition between $\mathbf{h}_i$ and $\mathbf{h}_j$ and vice versa
- The diffusion distance can be well approximated by the Euclidean distance in the embedded domain:

$$D_{\text{Diff}}(\mathbf{h}_i, \mathbf{h}_j) \cong \|\mathbf{\Phi}_d(\mathbf{h}_i) - \mathbf{\Phi}_d(\mathbf{h}_j)\|$$



Measuring Smoothness over Manifold $\mathcal{M}$

- Let $\mathbf{h} \in \mathcal{M}$ and $f : \mathcal{M} \to \mathbb{R}$
- The gradient $\nabla f(\mathbf{h})$ represents amplitude and direction of variation of f around $\mathbf{h}$
- A global measure of smoothness of f on $\mathcal{M}$:

$$\|f\|_{\mathcal{M}}^2 = \int_{\mathcal{M}} \|\nabla f(\mathbf{h})\|^2 d\mu(\mathbf{h})$$

  where $\mu(\mathbf{h})$ is the probability measure of $\mathbf{h}$ on $\mathcal{M}$
- Stokes' theorem links gradient and Laplacian:

$$\int_{\mathcal{M}} \|\nabla f(\mathbf{h})\|^2 d\mu(\mathbf{h}) = \int_{\mathcal{M}} f(\mathbf{h}) \mathcal{L} f(\mathbf{h}) d\mu(\mathbf{h}) = \langle f(\mathbf{h}), \mathcal{L} f(\mathbf{h}) \rangle$$

  where $\mathcal{L} = \nabla \cdot \nabla$ is the Laplace-Beltrami ("Laplacian") operator

Smoothness on the Manifold: Discretization

- Define the graph Laplacian: $\mathbf{L} \triangleq \mathbf{D} - \mathbf{K}$
- $\mathbf{P} = \mathbf{D}^{-1}\mathbf{K}$ and $\mathbf{N} = \mathbf{D}^{-1}\mathbf{L} = \mathbf{I} - \mathbf{P}$
- Smoothness of $\mathbf{f} = [f(\mathbf{h}_1), ..., f(\mathbf{h}_n)]$ on the graph: $\mathbf{f}^T\mathbf{L}\mathbf{f} = \langle \mathbf{f}, \mathbf{L}\mathbf{f} \rangle$
- Small $\mathbf{f}^T\mathbf{L}\mathbf{f}$ implies smooth $\mathbf{f}$ on the graph
- Further insight can be obtained by:

$$\mathbf{f}^T\mathbf{L}\mathbf{f} = \frac{1}{2}\sum_{i,j=1}^{n} K_{ij}\left(f(\mathbf{h}_i) - f(\mathbf{h}_j)\right)^2$$

$\Rightarrow$ When K_{ij} is large, the mappings $f(\mathbf{h}_i)$ and $f(\mathbf{h}_j)$ are "encouraged" to be close

Outline

Data Model: The Two Microphone Case

Microphone signals:

The measured signals in the two microphones (an extension to multiple microphone pairs will be discussed later):

$$y_1(n) = a_1(n) * s(n) + u_1(n)$$
$$y_2(n) = a_2(n) * s(n) + u_2(n)$$

- $s(n)$ - the source signal
- $a_i(n),\ i = \{1, 2\}$ - the acoustic impulse responses relating the source and each of the microphones
- $u_i(n),\ i = \{1, 2\}$ - noise signals, independent of the source

Data Model: The Two Microphone Case

Microphone signals:

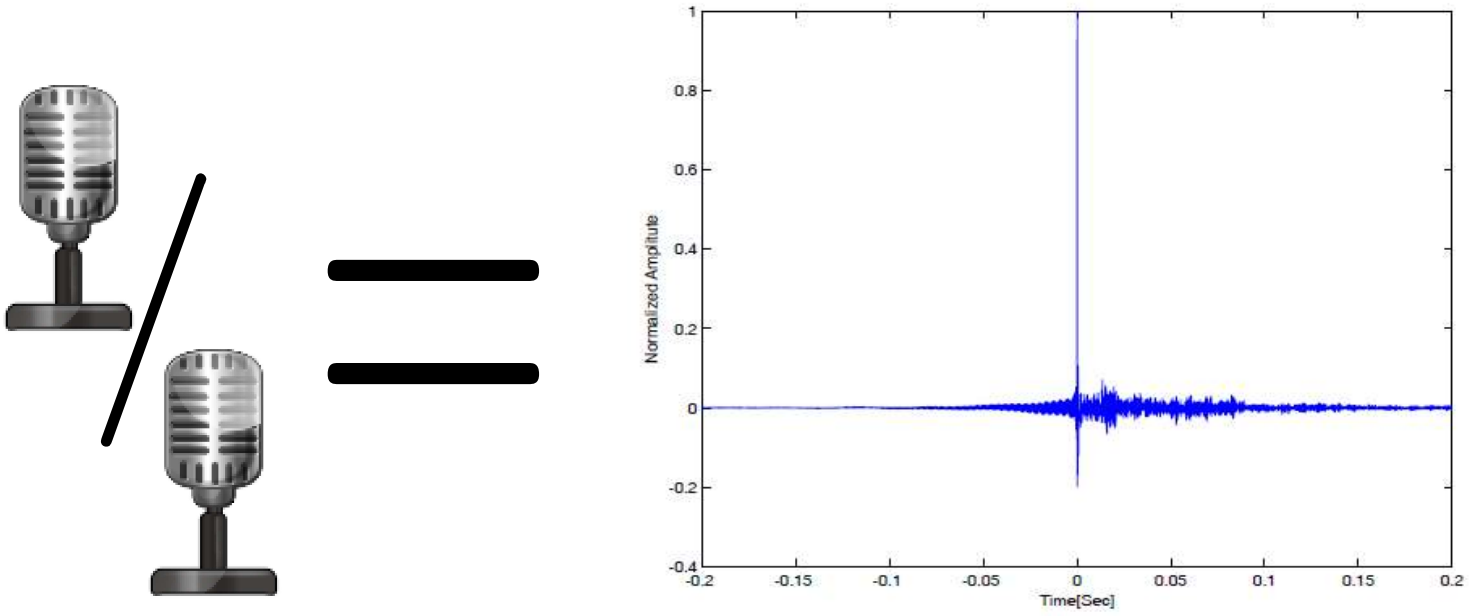
The measured signals in the two microphones:

$$y_1(n) = a_1(n) * s(n) + u_1(n)$$
$$y_2(n) = a_2(n) * s(n) + u_2(n)$$

- $s(n)$ - the source signal
- $a_i(n),\ i = \{1, 2\}$ - the acoustic impulse responses relating the source and each of the microphones
- $u_i(n),\ i = \{1, 2\}$ - noise signals, independent of the source

Find a feature vector representing the characteristics of the acoustic path and independent of the source signal

Relative Transfer Function (RTF) [Gannot et al., 2001]

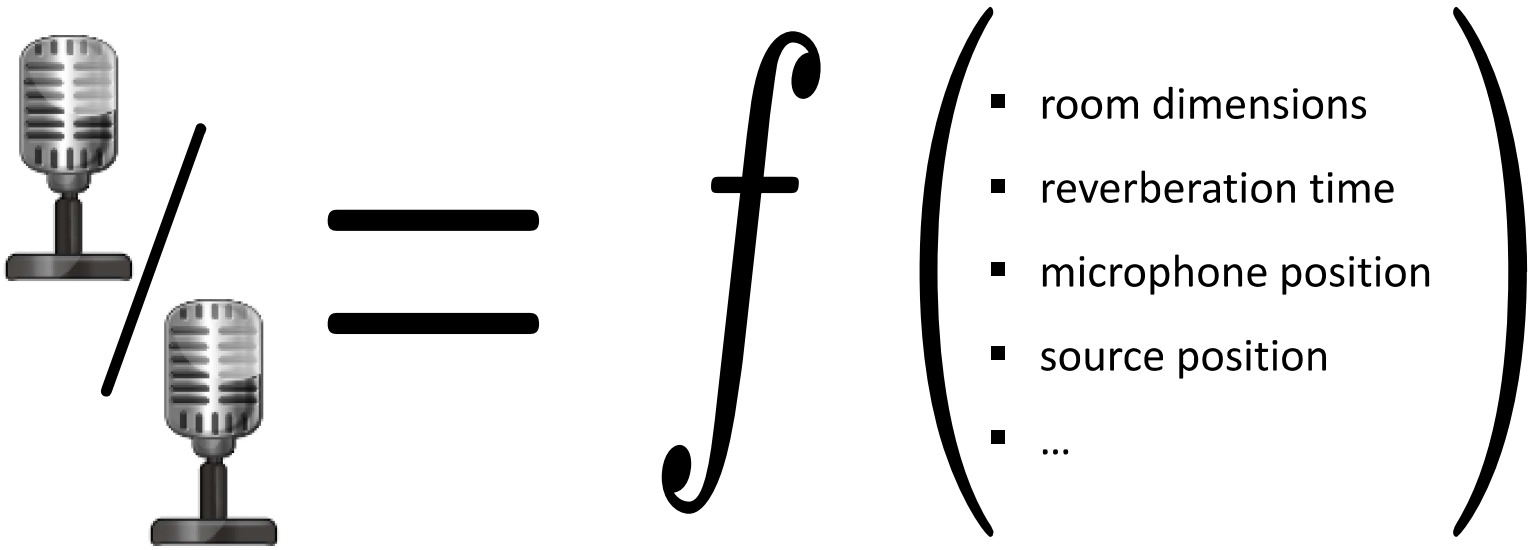


- Defined as the ratio between the transfer functions of the two mics:

$$H_{12}(k) = \frac{A_2(k)}{A_1(k)} \overset{\text{low-noise}}{\simeq} \frac{\hat{S}_{y_2 y_1}(k)}{\hat{S}_{y_1 y_1}(k)}$$

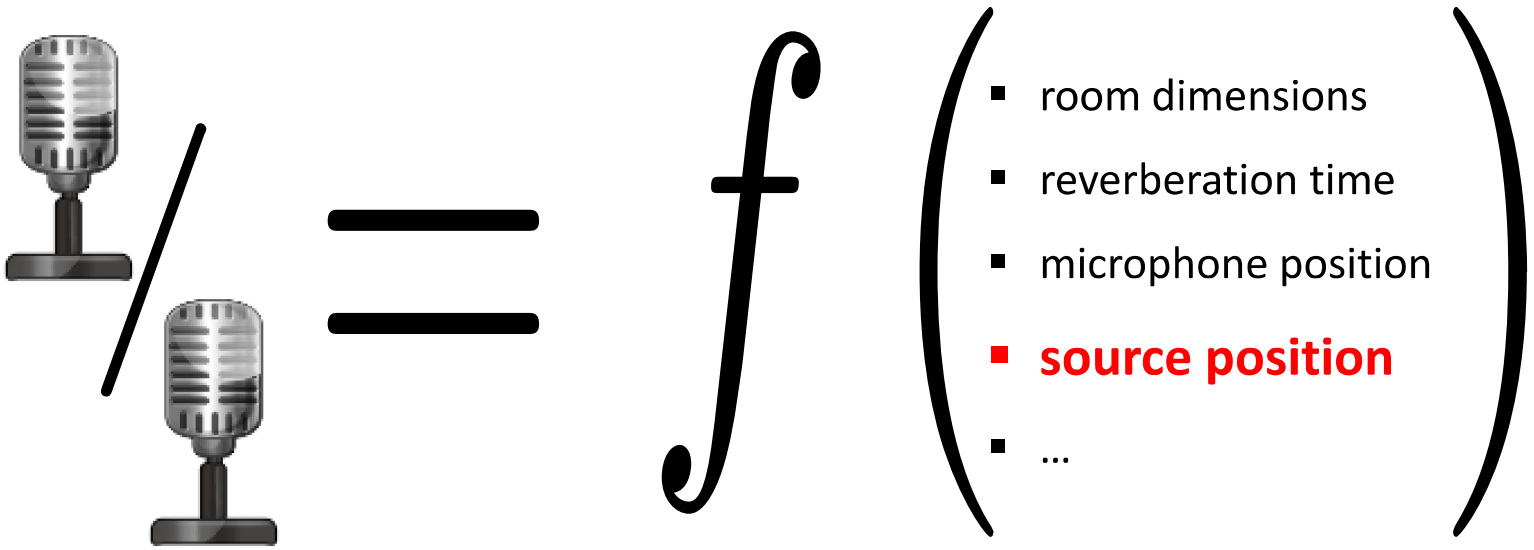
  estimated based on PSD and cross-PSD
- Define the feature vector: $\mathbf{h} = [\hat{H}_{12}(k_1), \ldots, \hat{H}_{12}(k_D)]^T$

Relative Transfer Function (RTF) [Gannot et al., 2001]



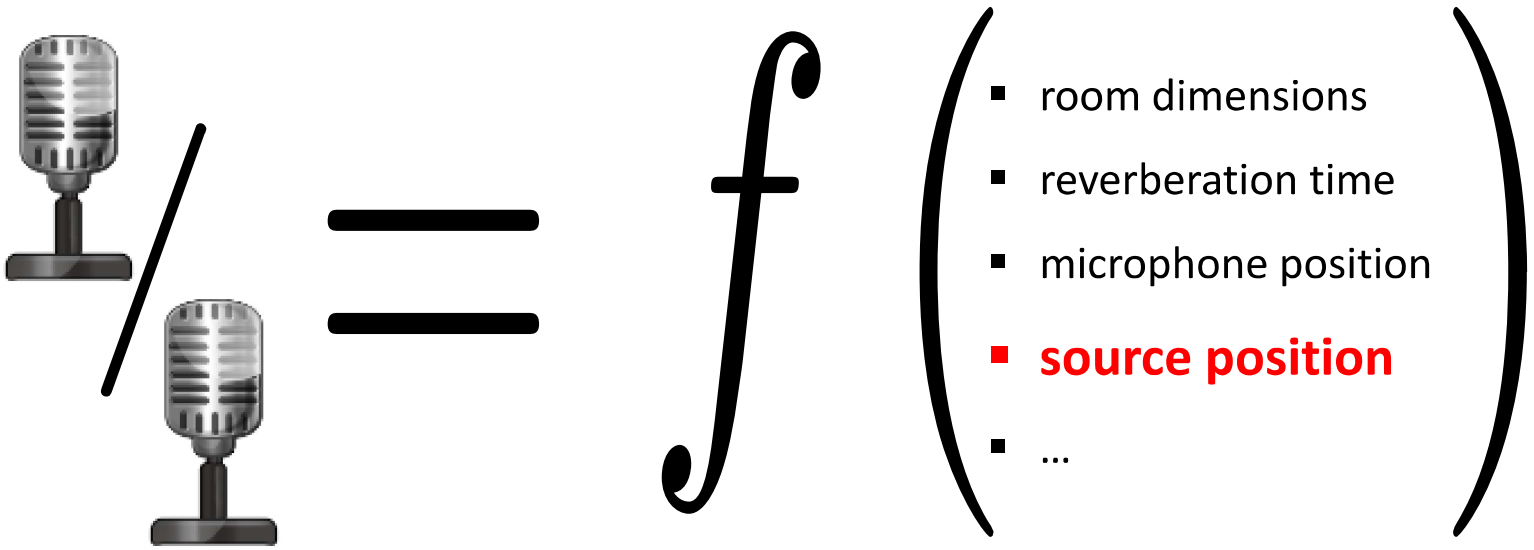
- Represents the acoustic paths and is independent of the source signal
- Generalizes the TDOA
- Depends on the physical characteristics of the environment

Relative Transfer Function (RTF) [Gannot et al., 2001]



- Represents the acoustic paths and is independent of the source signal
- Generalizes the TDOA
- Depends on the physical characteristics of the environment
- In a static environment the source position is the only varying DoF

Relative Transfer Function (RTF) [Gannot et al., 2001]

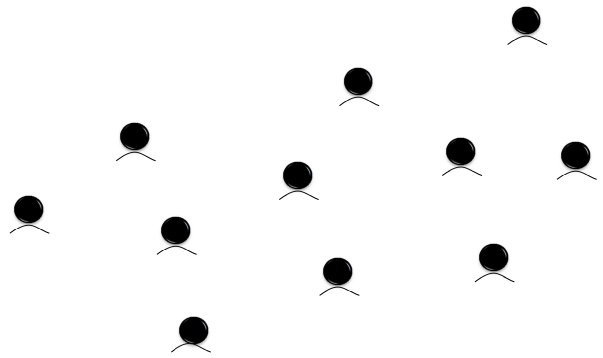


- Represents the acoustic paths and is independent of the source signal
- Generalizes the TDOA
- Depends on the physical characteristics of the environment
- In a static environment the source position is the only varying DoF
- A plethora of methods for RTF Estimation [Gannot et al., 2001]; [Markovich et al., 2009]; [Markovich-Golan et al., 2018]; [Laufer-Goldshtein et al., 2018c]

Outline

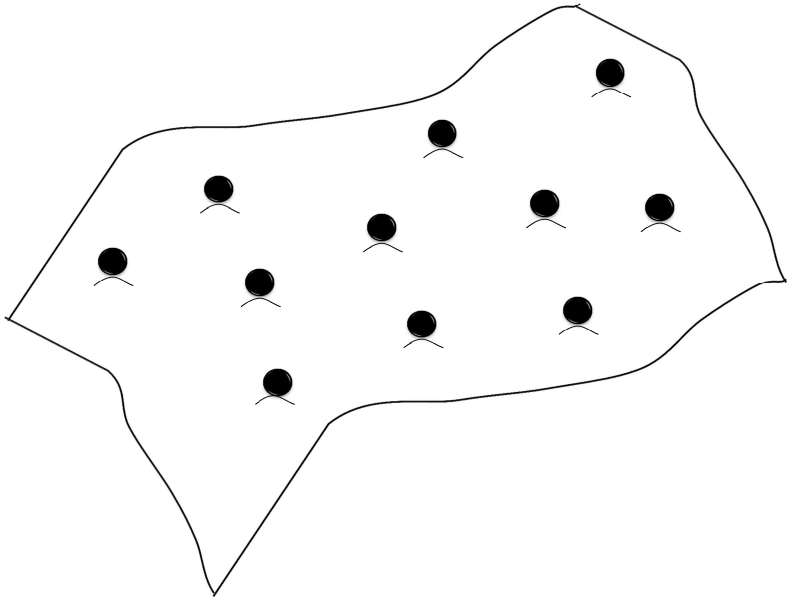
How to Measure the Affinity between Two RTF Samples? [Laufer-Goldshtein et al., 2015]

The RTFs are represented as points in a high dimensional space



How to Measure the Affinity between Two RTF Samples? [Laufer-Goldshtein et al., 2015]

The RTFs are represented as points in a high dimensional space

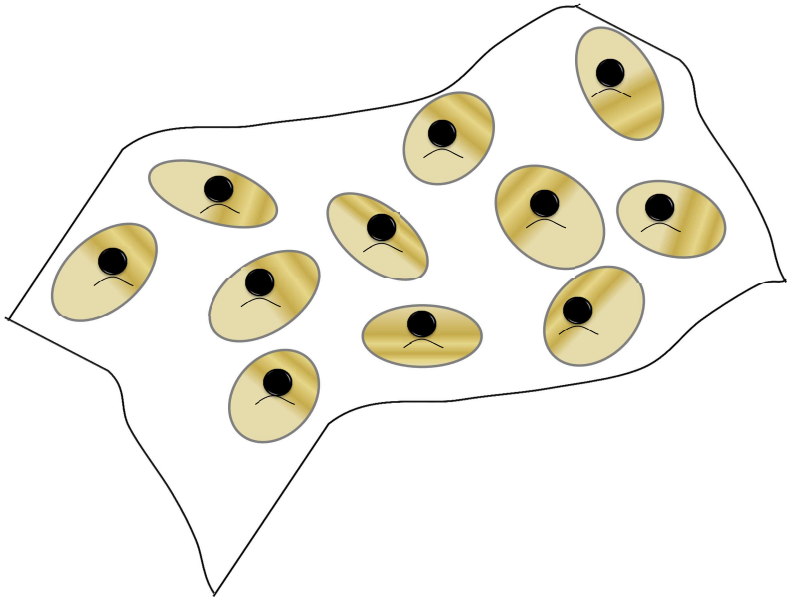


Acoustic manifold

- They lie on a low dimensional nonlinear manifold $\mathcal{M}$

How to Measure the Affinity between Two RTF Samples? [Laufer-Goldshtein et al., 2015]

The RTFs are represented as points in a high dimensional space

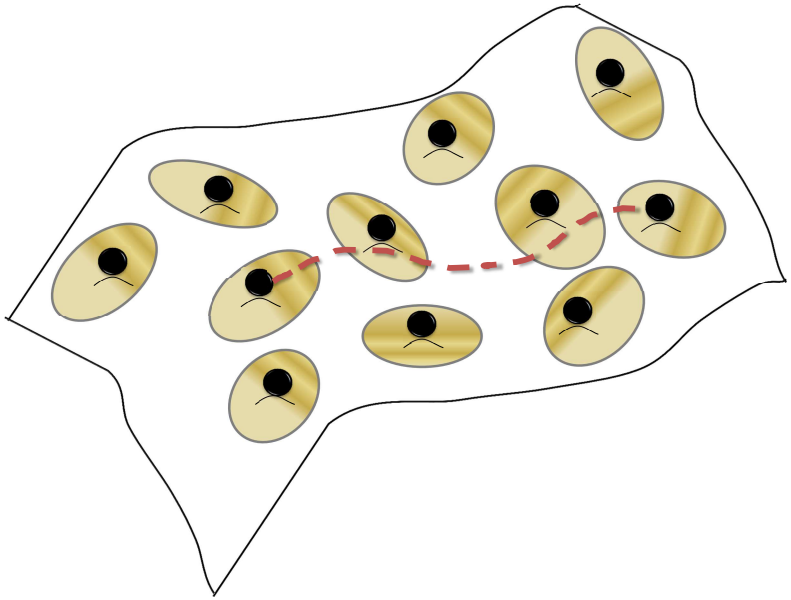


Acoustic manifold

- They lie on a low dimensional nonlinear manifold $\mathcal{M}$
- Linearity is preserved in small neighbourhoods

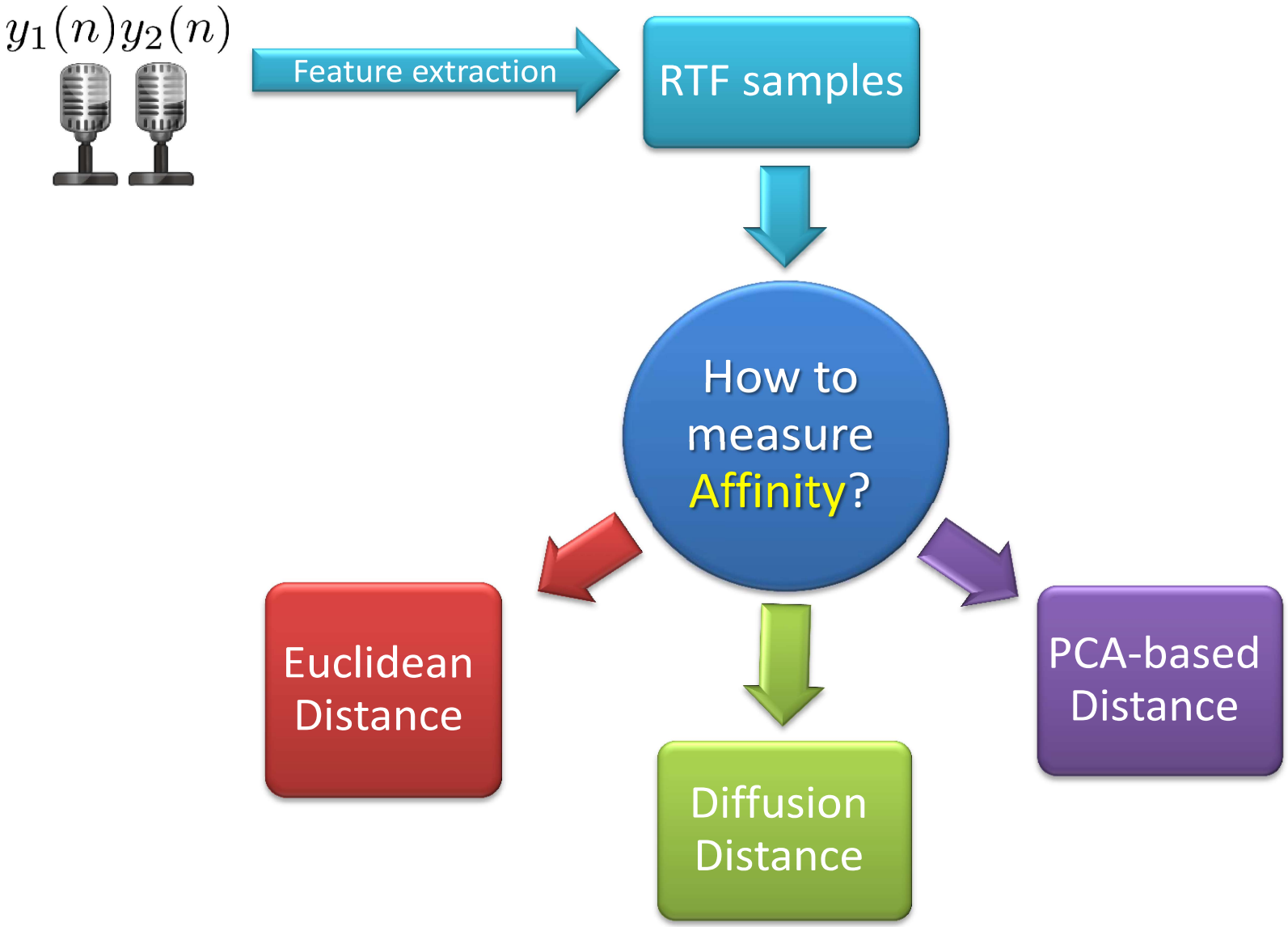
How to Measure the Affinity between Two RTF Samples? [Laufer-Goldshtein et al., 2015]

The RTFs are represented as points in a high dimensional space



Acoustic manifold

- They lie on a low dimensional nonlinear manifold $\mathcal{M}$
- Linearity is preserved in small neighbourhoods
- Distances between RTFs should be measured along the manifold



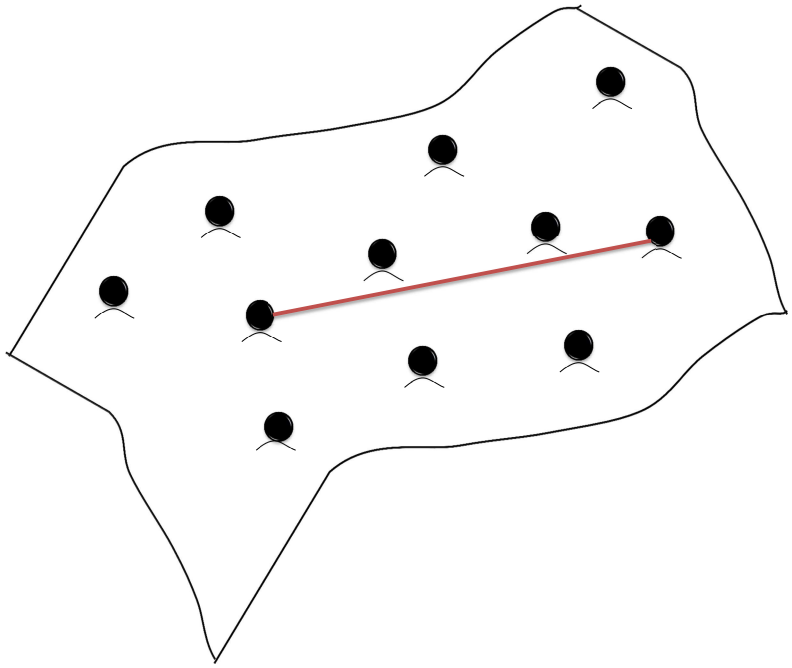
Each distance measure relies on a different hidden assumption about the underlying structure of the RTF samples

Euclidean Distance

The Euclidean distance between RTFs

$$D_{\text{Euc}}(\mathbf{h}_i, \mathbf{h}_j) = \|\mathbf{h}_i - \mathbf{h}_j\|$$

- Compares two RTFs in their original space
- Does not assume an existence of a manifold
- Respects flat manifolds



A good affinity measure only when the RTFs are uniformly scattered all over the space, or when they lie on a flat manifold

PCA-Based Distance [Pearson, 1901]

PCA algorithm

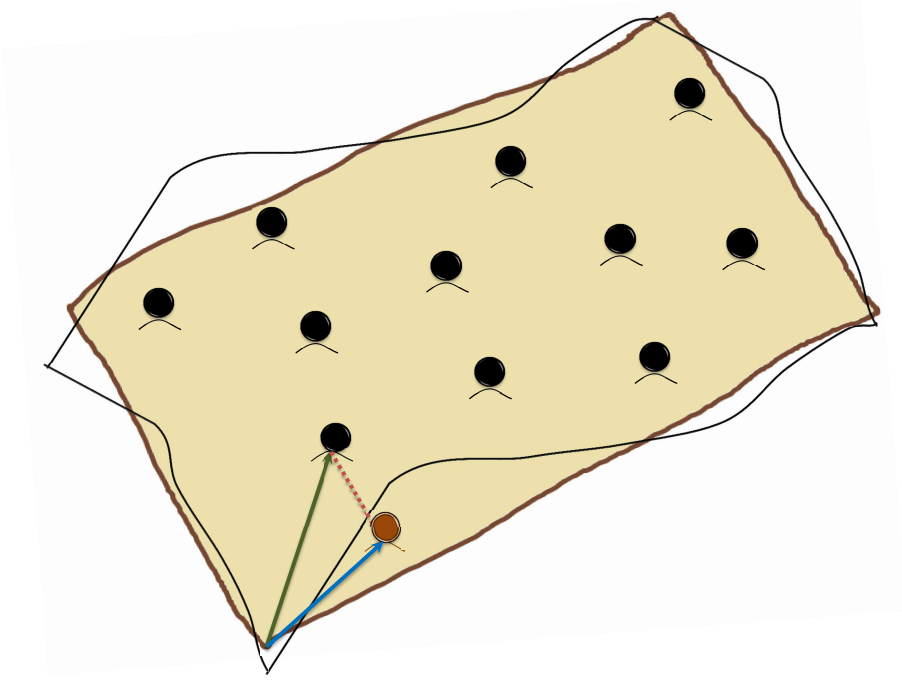
- The principal components - the d dominant eigenvectors $\{\mathbf{v}_i\}_{i=1}^{d}$ of the covariance matrix of the data
- The RTFs are linearly projected onto the principal components:

$$\boldsymbol{\nu}(\mathbf{h}_i) = [\mathbf{v}_1, \ldots \mathbf{v}_d]^T (\mathbf{h}_i - \boldsymbol{\mu})$$

PCA-based distance between RTFs

$$D_{\mathrm{PCA}}(\mathbf{h}_i, \mathbf{h}_j) = \|\boldsymbol{\nu}(\mathbf{h}_i) - \boldsymbol{\nu}(\mathbf{h}_j)\|$$

- A global approach - extracts principal directions of the entire set
- Linear projections - the manifold is assumed to be linear/flat



PCA-Based Distance [Pearson, 1901]

PCA algorithm

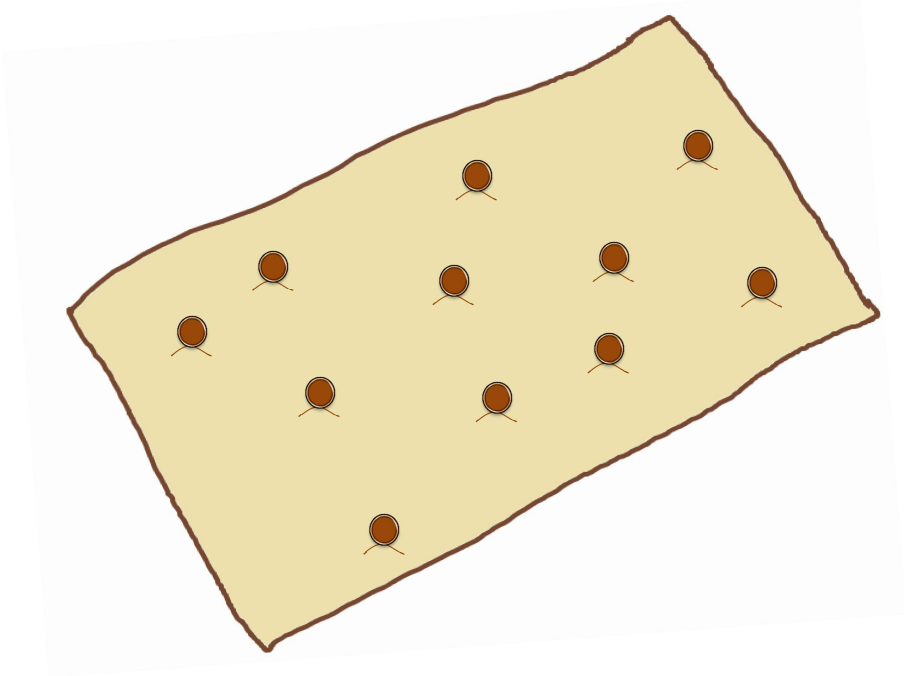
- The principal components - the d dominant eigenvectors $\{\mathbf{v}_i\}_{i=1}^{d}$ of the covariance matrix of the data
- The RTFs are linearly projected onto the principal components:

$$\boldsymbol{\nu}(\mathbf{h}_i) = [\mathbf{v}_1, \dots \mathbf{v}_d]^T (\mathbf{h}_i - \boldsymbol{\mu})$$

PCA-based distance between RTFs

$$D_{\mathrm{PCA}}(\mathbf{h}_i, \mathbf{h}_j) = \|\boldsymbol{\nu}(\mathbf{h}_i) - \boldsymbol{\nu}(\mathbf{h}_j)\|$$

- A global approach - extracts principal directions of the entire set
- Linear projections - the manifold is assumed to be linear/flat



PCA-Based Distance [Pearson, 1901]

PCA algorithm

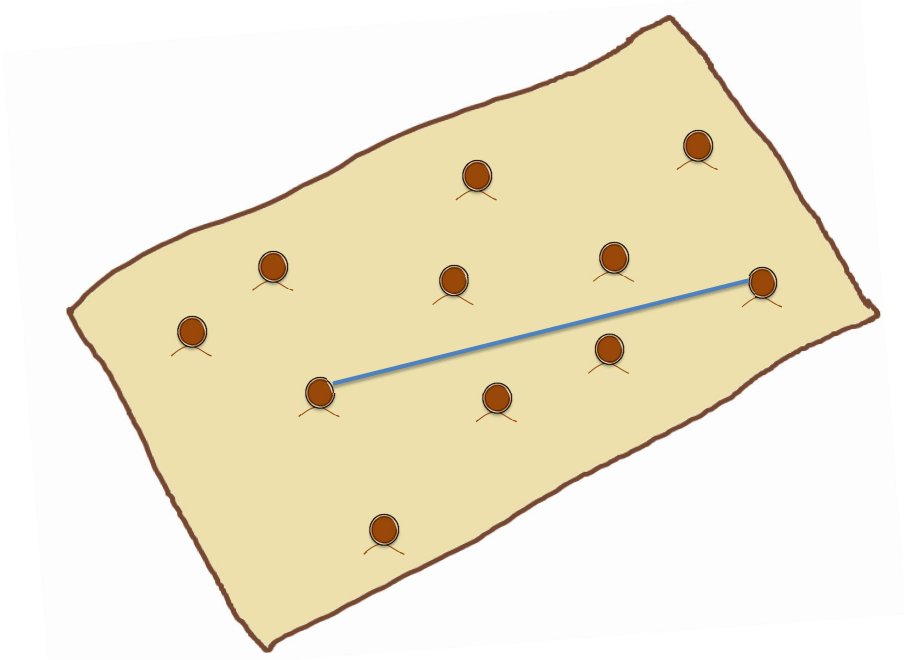
- The principal components - the d dominant eigenvectors $\{\mathbf{v}_i\}_{i=1}^{d}$ of the covariance matrix of the data
- The RTFs are linearly projected onto the principal components:

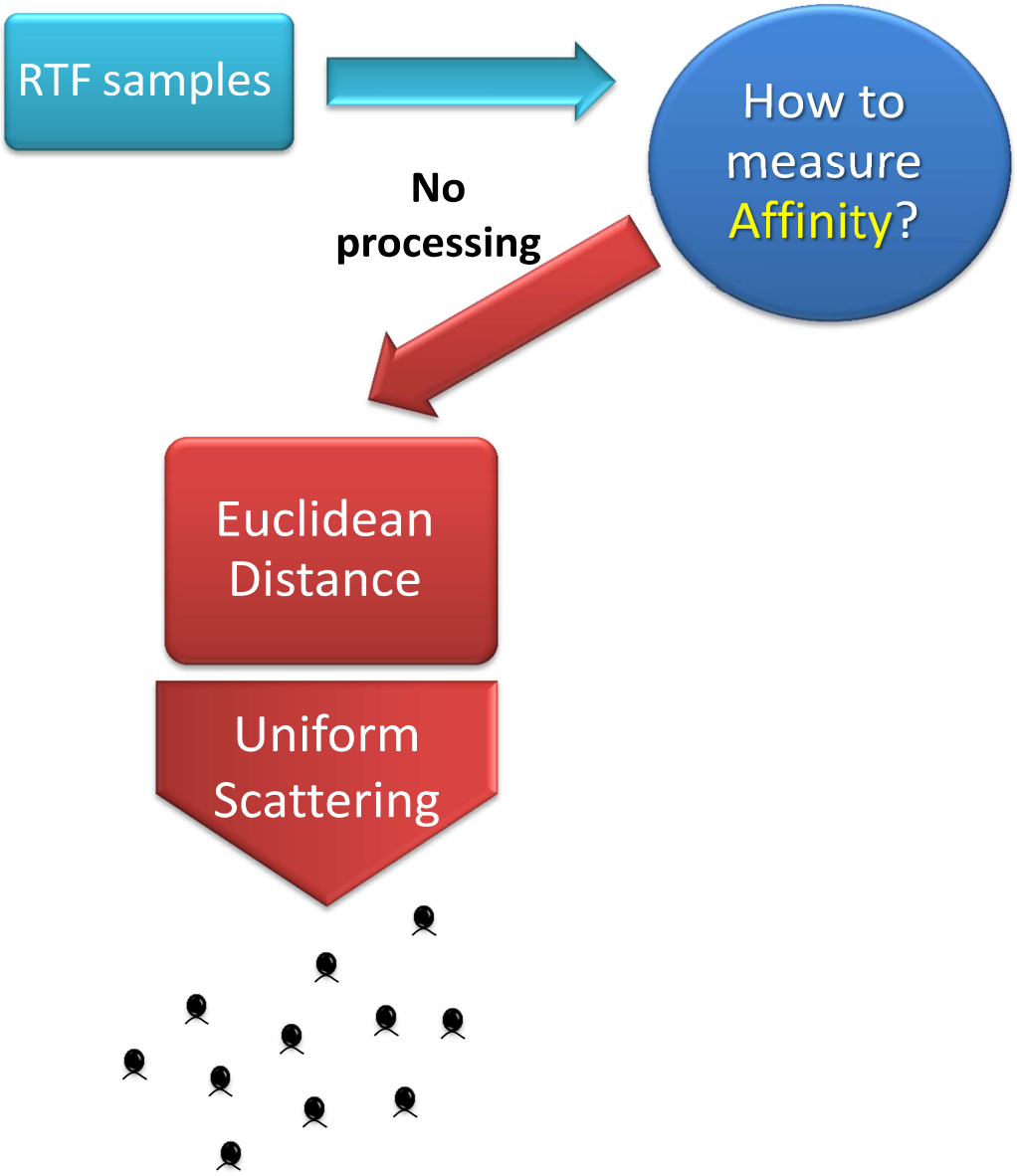
$$\boldsymbol{\nu}(\mathbf{h}_i) = [\mathbf{v}_1, \ldots \mathbf{v}_d]^T (\mathbf{h}_i - \boldsymbol{\mu})$$

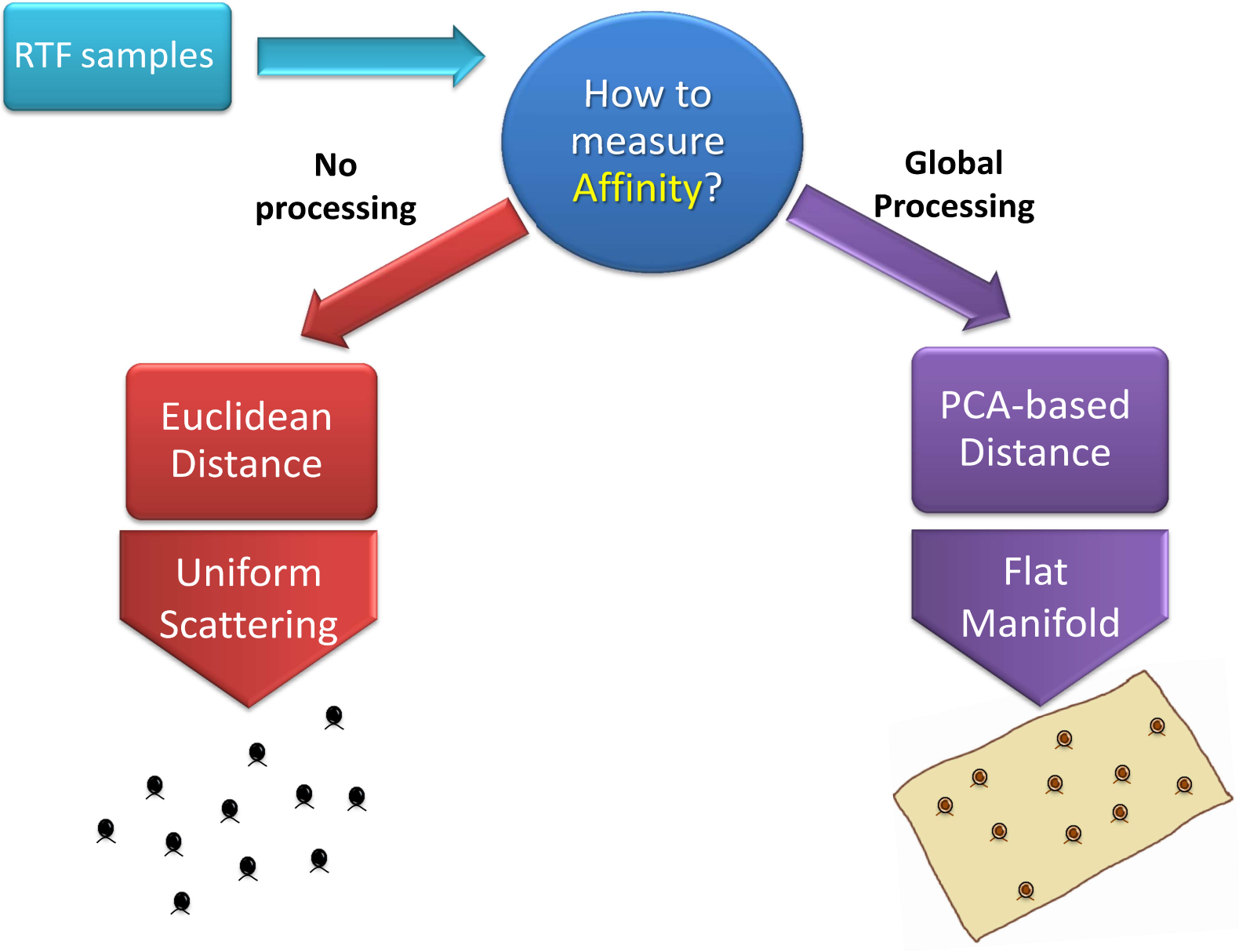
PCA-based distance between RTFs

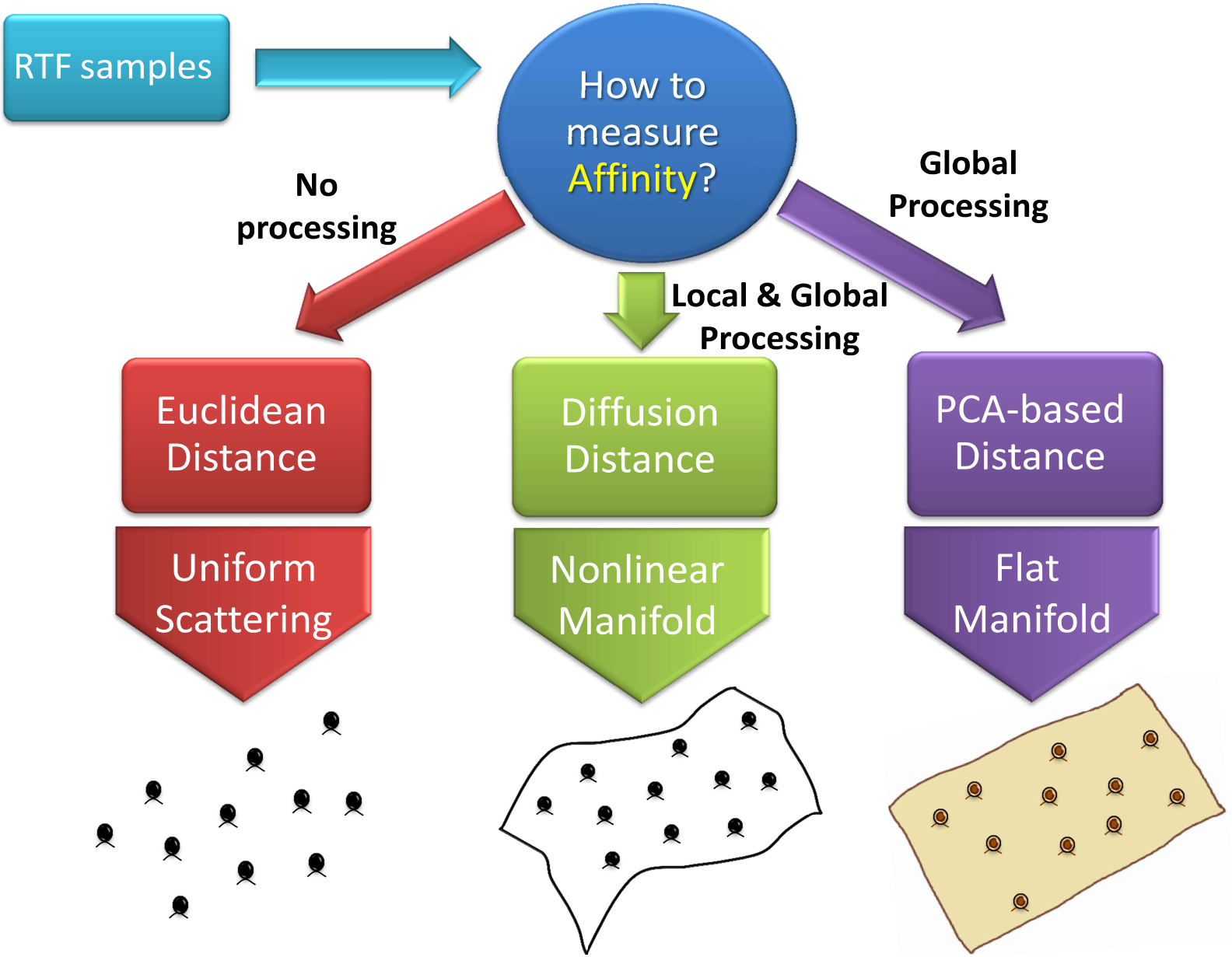
$$D_{\mathrm{PCA}}(\mathbf{h}_i, \mathbf{h}_j) = \|\boldsymbol{\nu}(\mathbf{h}_i) - \boldsymbol{\nu}(\mathbf{h}_j)\|$$

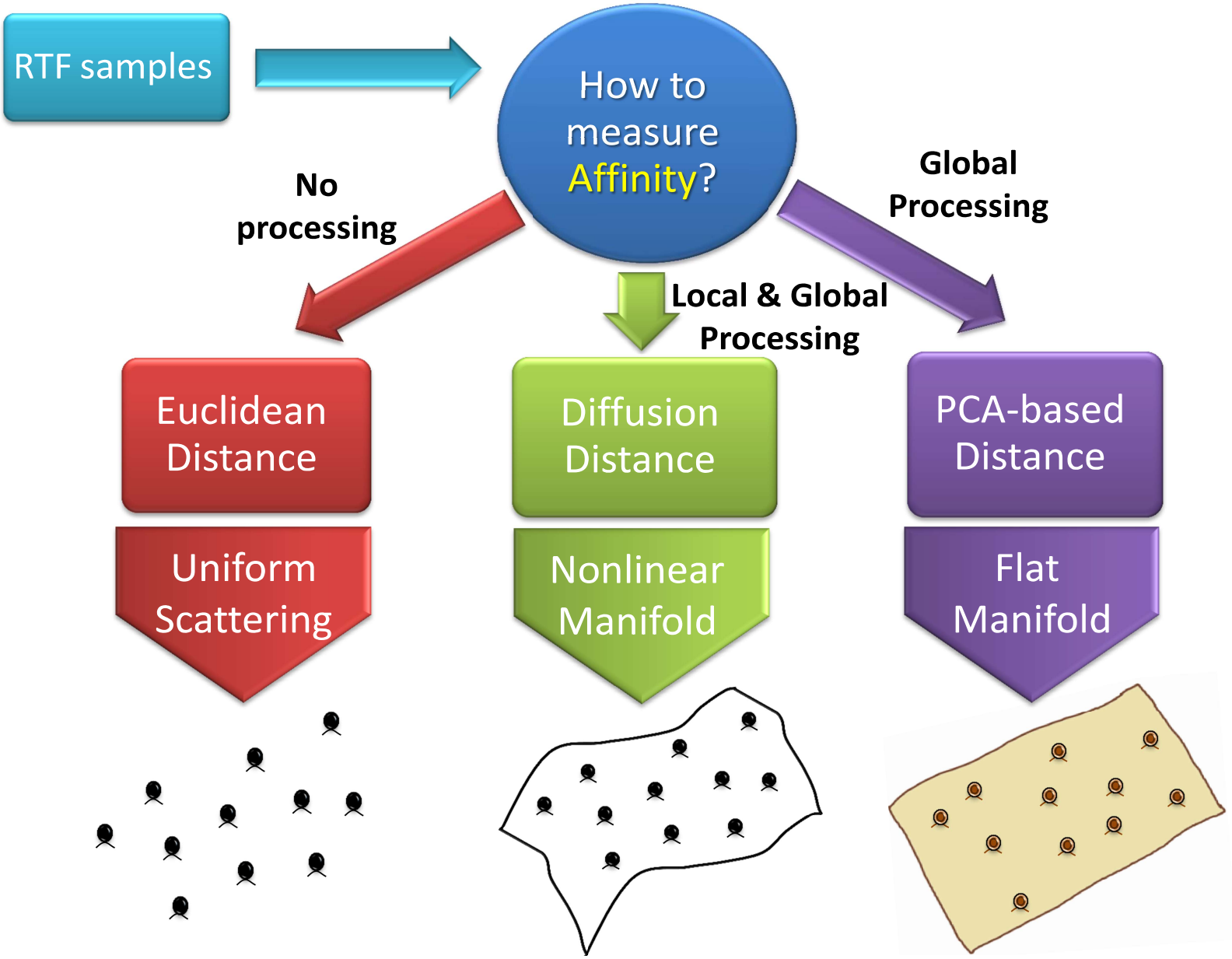
- A global approach - extracts principal directions of the entire set
- Linear projections - the manifold is assumed to be linear/flat




RTF samples
How to measure Affinity?
No processing
Euclidean Distance
Uniform Scattering


RTF samples
How to measure Affinity?
No processing
Global Processing
Euclidean Distance
Uniform Scattering
PCA-based Distance
Flat Manifold


RTF samples
How to measure Affinity?
No processing
Local & Global Processing
Global Processing
Euclidean Distance
Diffusion Distance
PCA-based Distance
Uniform Scattering
Nonlinear Manifold
Flat Manifold



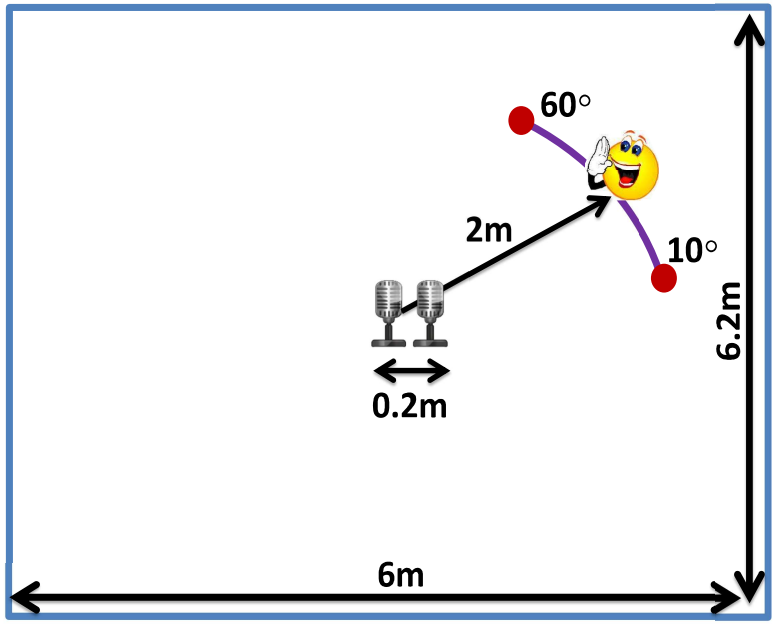
Which of the distance measures is proper?
What is the true underlying structure of the RTFs?

Simulation Results

Room setup

Simulate a reverberant room using the image method [Allen and Berkley, 1979]:

- Room dimension $6 \times 6.2 \times 3$m
- Microphones at: $[3, 3, 1]$ and $[3.2, 3, 1]$
- The source is positioned at 2m from the mics, the azimuth angle in $10^\circ \div 60^\circ$
- $T_{60} = 150/300/500$ ms
- SNR= 20 dB

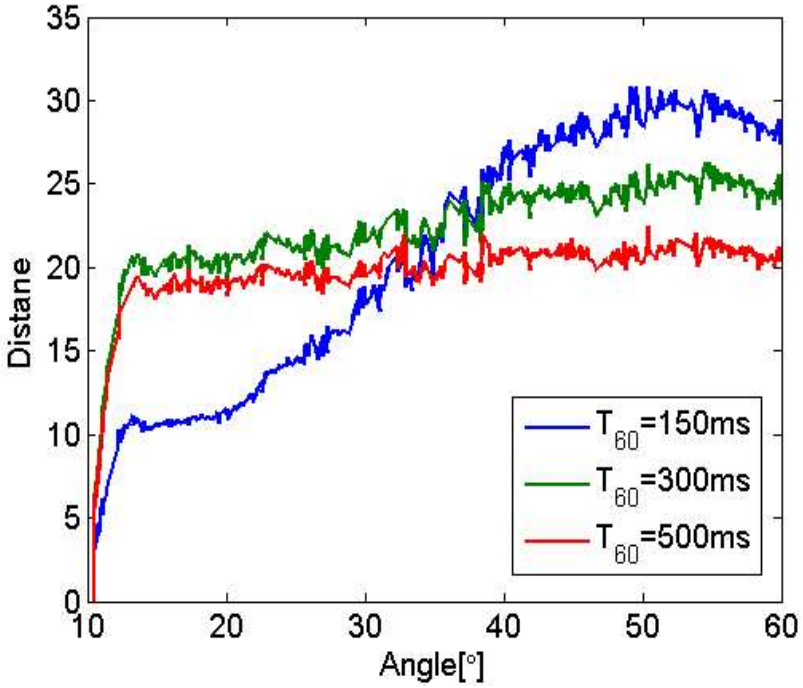


Test

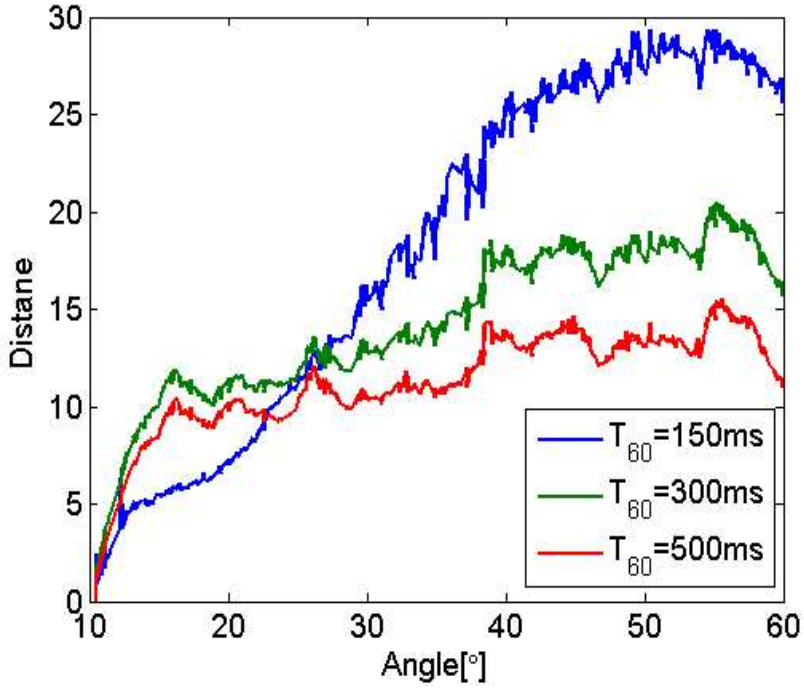
Measure the distance between each of the RTFs and the RTF corresponding to 10°:

- If monotonic with respect to the angle - proper distance
- If not monotonic with respect to the angle - improper distance

Euclidean Distance & PCA-based Distance [Laufer-Goldshtein et al., 2015]



(a) Euclidean Distance

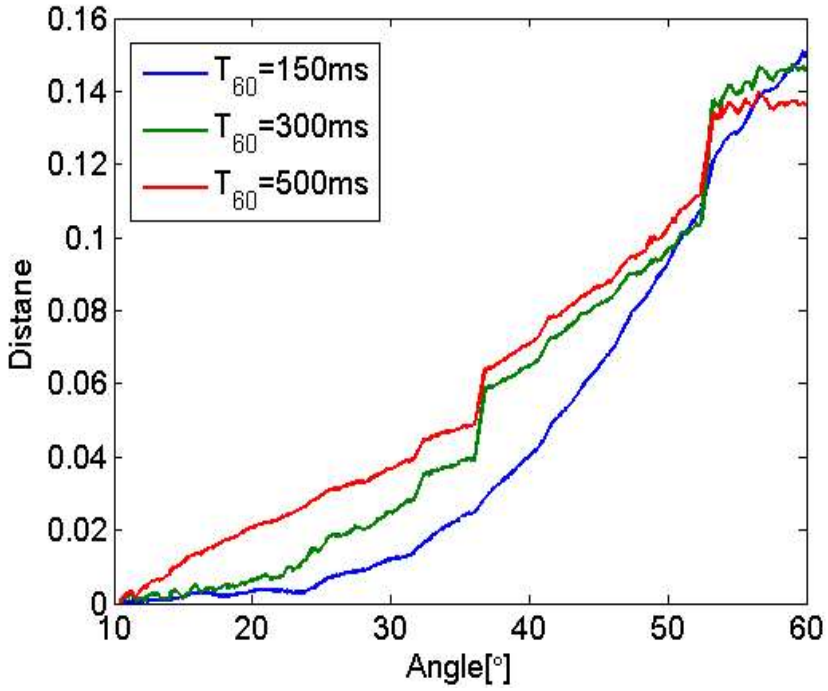


(b) PCA-based Distance

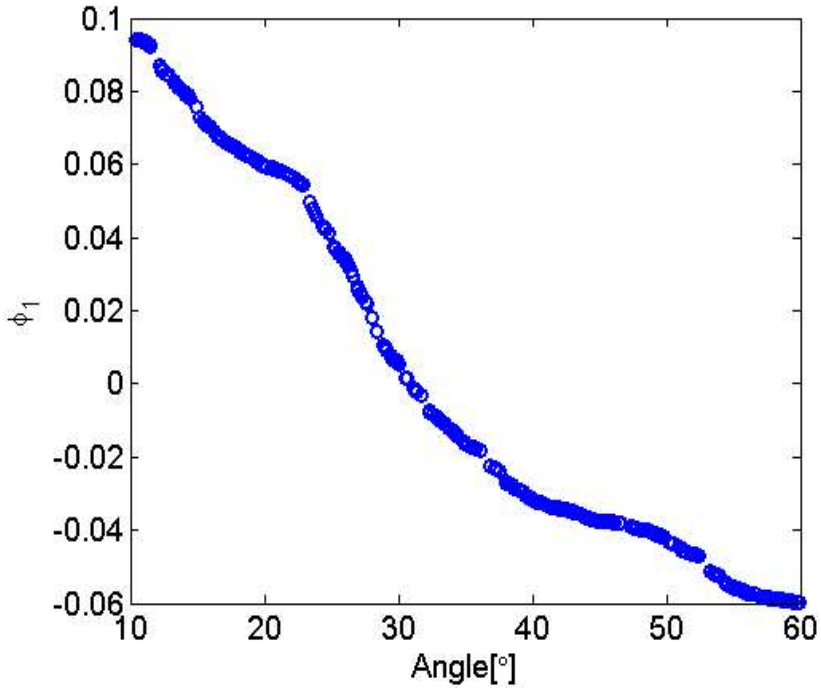
For both distance measures:

- Monotonic with respect to the angle only in a limited region
- This region becomes smaller as the reverberation time increases
- They are inappropriate for measuring angles' proximity

Diffusion Maps



(c) Diffusion Distance



(d) Diffusion Mapping

The diffusion distance:

- Monotonic with respect to the angle for almost the entire range
- It is an appropriate distance measure in terms of the source DOA
- Mapping corresponds well with angles - recovers the latent parameter

Outline

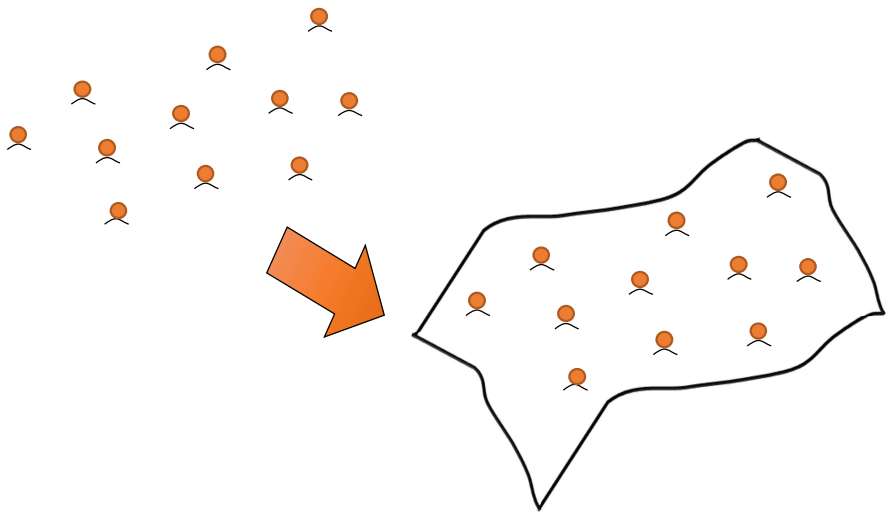
Semi-Supervised Approaches for Localization

- The existence of an acoustic manifold in a specific environment was established
- The RTF was shown to be a proper feature vector that can capture the acoustic variability as a function of the source position (alternative feature vectors [Laufer-Goldshtein et al., 2018a]; [Hu et al., 2019];[Hu et al., 2020])
- Learning paradigms:
  1. Unsupervised localization $\Rightarrow$ array constellation required
  2. Supervised localization $\Rightarrow$ many labels
  3. Semi-supervised $\Rightarrow$ utilizes a small number of labelled data and a large number of unlabelled data; array constellation not required
- Two acoustic manifold-based speaker localization methods:
  1. Diffusion Distance Search (DDS) [Talmon et al., 2011, Laufer-Goldshtein et al., 2013]
  2. Manifold Regularization for Localization (MRL) [Laufer-Goldshtein et al., 2016b]

Semi-Supervised Approaches for Localization (cont.)

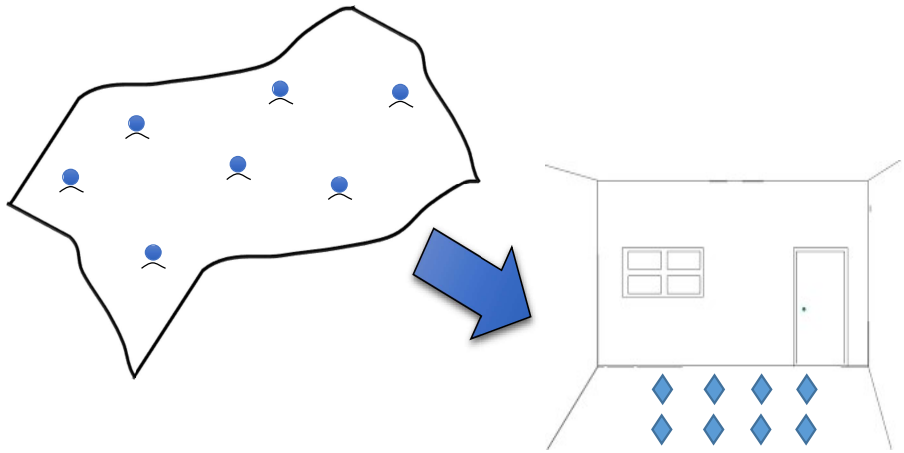
Unlabelled Samples

Recover the Manifold Structure



Labelled Samples

Anchor Points – Translate RTFs to Positions



Semi-Supervised Learning

Mixed of supervised (attached with known locations as anchors) and unsupervised (unknown locations) learning

Semi-Supervised Learning

Mixed of supervised (attached with known locations as anchors) and unsupervised (unknown locations) learning

Why using unlabeled data?

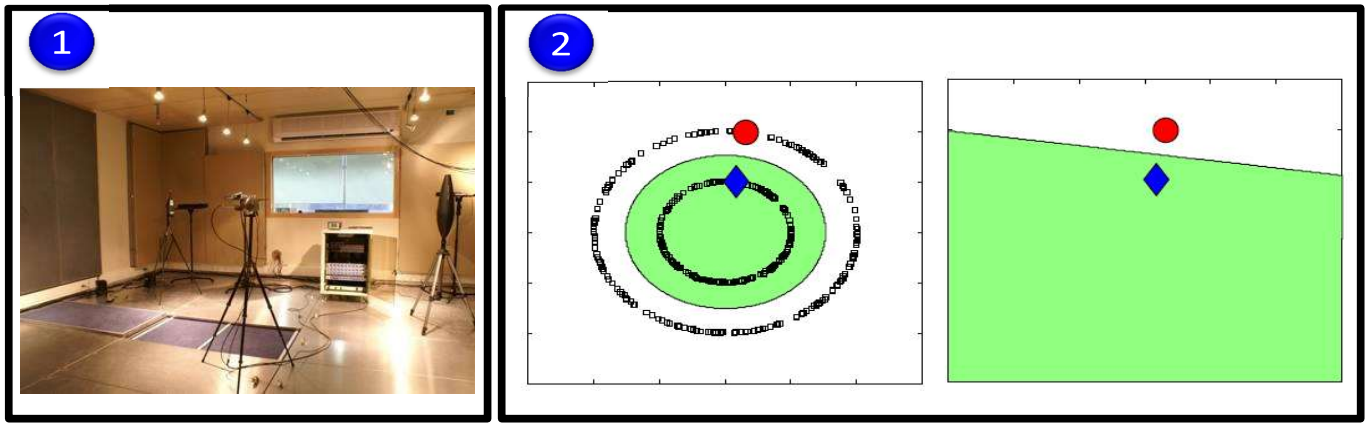
1. Localization - training should fit the specific environment of interest:
   - Cannot generate a general database for all possible acoustic scenarios
   - Generating a large amount of labelled data is cumbersome/impractical
   - Unlabelled data is freely available - whenever someone is speaking

Semi-Supervised Learning

Mixed of supervised (attached with known locations as anchors) and unsupervised (unknown locations) learning

Why using unlabeled data?

1. Localization - training should fit the specific environment of interest:
   - Cannot generate a general database for all possible acoustic scenarios
   - Generating a large amount of labelled data is cumbersome/impractical
   - Unlabelled data is freely available - whenever someone is speaking
2. Unlabelled data can be utilize to recover the manifold structure

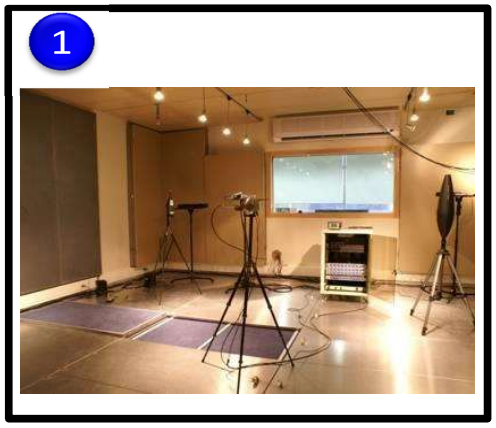


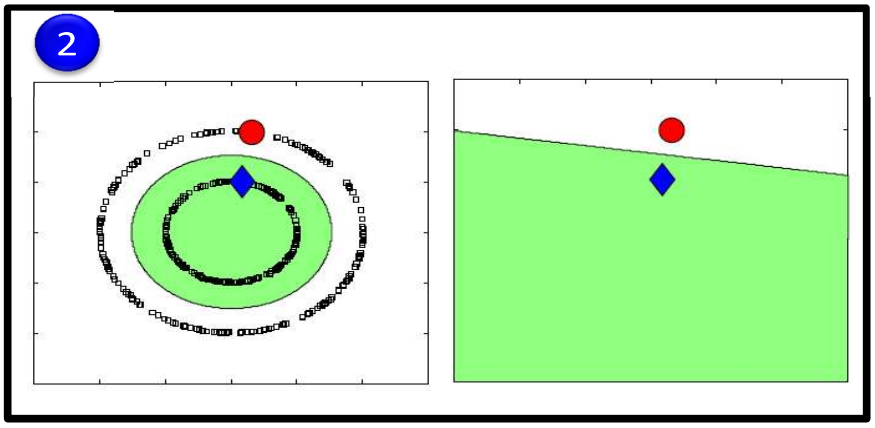
Semi-Supervised Learning

Mixed of supervised (attached with known locations as anchors) and unsupervised (unknown locations) learning

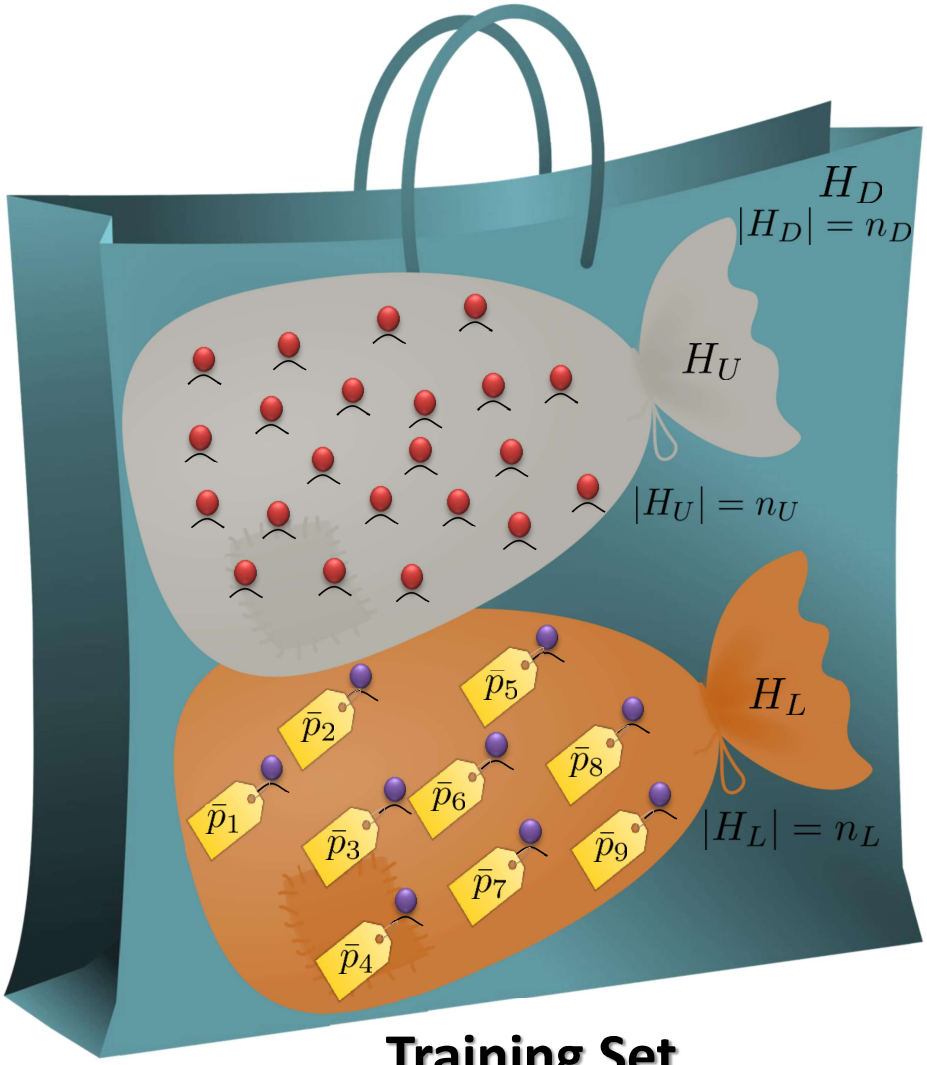
Why using unlabeled data?

1. Localization - training should fit the specific environment of interest:
   - Cannot generate a general database for all possible acoustic scenarios
   - Generating a large amount of labelled data is cumbersome/impractical
   - Unlabelled data is freely available - whenever someone is speaking
2. Unlabelled data can be utilize to recover the manifold structure
3. Semi-supervised learning is the natural setting for human learning



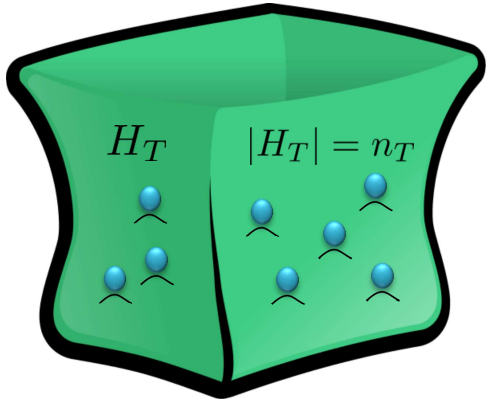


Datasets



Training Set

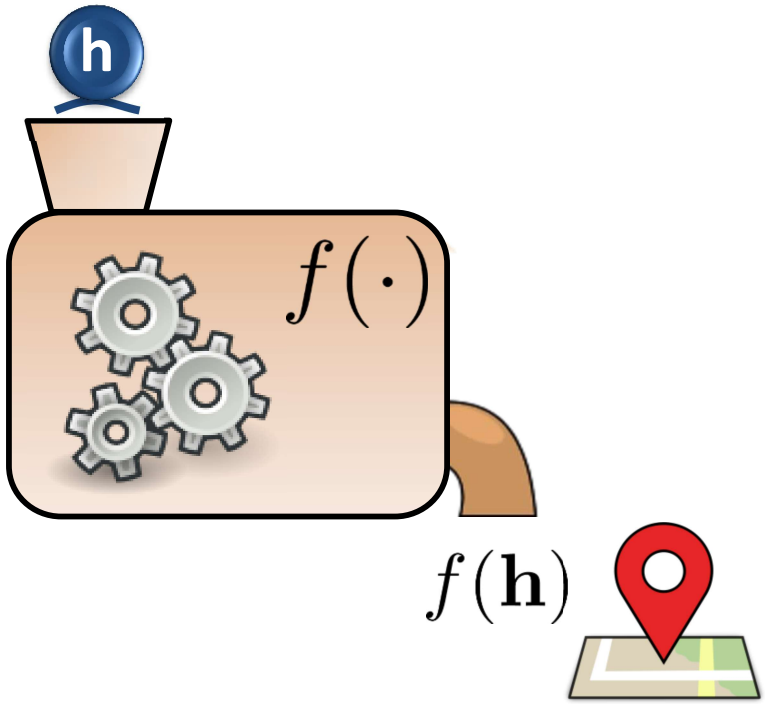
- $H_L = \{\mathbf{h}_i\}_{i=1}^{n_L}$ - n_L labelled samples
- $P_L = \{\bar{p}_i\}_{i=1}^{n_L}$ - labels/positions
- $H_U = \{\mathbf{h}_i\}_{i=n_L+1}^{n_D}$ - n_U unlabelled samples
- $H_D = H_L \cup H_U$ - entire training set
- $H_T = \{\mathbf{h}_i\}_{i=n_D+1}^{n}$ - n_T test samples



Test Set

Manifold Regularization for Localization [Laufer-Goldshtein et al., 2016b]

Goal: Recover the function f which transforms an RTF to position



Manifold Regularization for Localization [Laufer-Goldshtein et al., 2016b]

Goal: **Recover the function f which transforms an RTF to position**

Complex nonlinear relation between RTFs and positions

Infinite search space

How to prevent overfitting?

How to utilize unlabelled data?

Manifold Regularization for Localization [Laufer-Goldshtein et al., 2016b]

Goal: Recover the function f which transforms an RTF to position

Complex nonlinear relation between RTFs and positions

- Learn a data-driven model from training data

Infinite search space

- Work in a reproducing kernel Hilbert space (RKHS)

How to prevent overfitting?

- Add regularizations to control smoothness

How to utilize unlabelled data?

- Use manifold regularization

Reproducing Kernel Hilbert Space (RKHS) [Berlinet and Thomas-Agnan, 2011]

Moore-Aronszajn theorem: [Aronszajn, 1950]

A positive definite symmetric kernel k on $\mathcal{M}$, defines a unique reproducing kernel Hilbert space (RKHS) $\mathcal{H}_k$ that consists of functions on $\mathcal{M}$, satisfying:

- $k(\mathbf{h}, \cdot) \in \mathcal{H}_k, \forall \mathbf{h} \in \mathcal{M}$;
- $\text{span}\{k(\mathbf{h}, \cdot); \mathbf{h} \in \mathcal{M}\}$ is dense in $\mathcal{H}_k$;
- **The reproducing property:** $\langle f(\cdot), k(\mathbf{h}, \cdot)\rangle = f(\mathbf{h}), \forall f \in \mathcal{H}_k, \mathbf{h} \in \mathcal{M}$.

The Representer theorem: [Schölkopf et al., 2001]

$$f(\mathbf{h}) = \sum_{i=1}^{n_D} a_i k(\mathbf{h}_i, \mathbf{h})$$

where $k : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ is the reproducing kernel of $\mathcal{H}_k$

Optimization and Manifold Regularization

Optimization in a reproducing kernel Hilbert space (RKHS) [Belkin et al., 2006]:

$$f^* = \underset{f \in \mathcal{H}_k}{\operatorname{argmin}} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2 + \gamma_k \|f\|^2_{\mathcal{H}_k} + \gamma_M \|f\|^2_{\mathcal{M}}$$

Optimization and Manifold Regularization

Optimization in a reproducing kernel Hilbert space (RKHS) [Belkin et al., 2006]:

$$f^* = \underset{f \in \mathcal{H}_k}{\operatorname{argmin}} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2 + \gamma_k \|f\|^2_{\mathcal{H}_k} + \gamma_M \|f\|^2_{\mathcal{M}}$$

Cost function

$\frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2$

correspondence between function values and labels

Optimization and Manifold Regularization

Optimization in a reproducing kernel Hilbert space (RKHS) [Belkin et al., 2006]:

$$f^* = \underset{f \in \mathcal{H}_k}{\operatorname{argmin}} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2 + \gamma_k \| f \|^2_{\mathcal{H}_k} + \gamma_M \| f \|^2_{\mathcal{M}}$$

Cost function

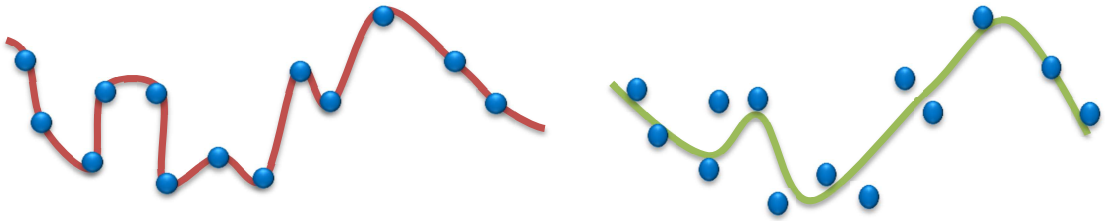
$\frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2$

Tikhonov Regularization

$\| f \|^2_{\mathcal{H}_k}$

correspondence between function values and labels

smoothness condition in the RKHS



Optimization and Manifold Regularization

Optimization in a reproducing kernel Hilbert space (RKHS) [Belkin et al., 2006]:

$$f^* = \underset{f \in \mathcal{H}_k}{\operatorname{argmin}} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2 + \gamma_k \|f\|^2_{\mathcal{H}_k} + \gamma_M \|f\|^2_{\mathcal{M}}$$

Cost function

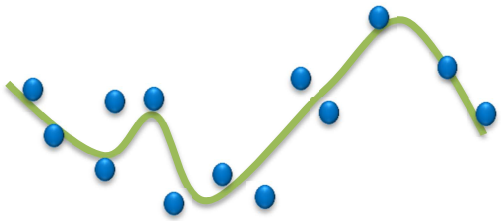
$\frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2$

correspondence between function values and labels

Tikhonov Regularization

$\|f\|^2_{\mathcal{H}_k}$

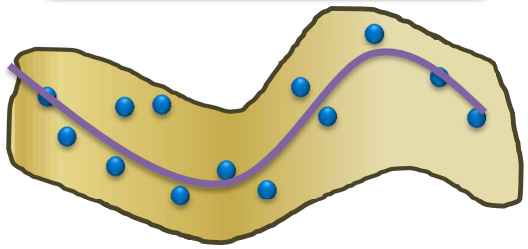
smoothness condition in the RKHS



Manifold Regularization

$\|f\|^2_{\mathcal{M}}$

smoothness penalty with respect to the manifold



Manifold Regularization

Smoothness on the manifold: A reminder

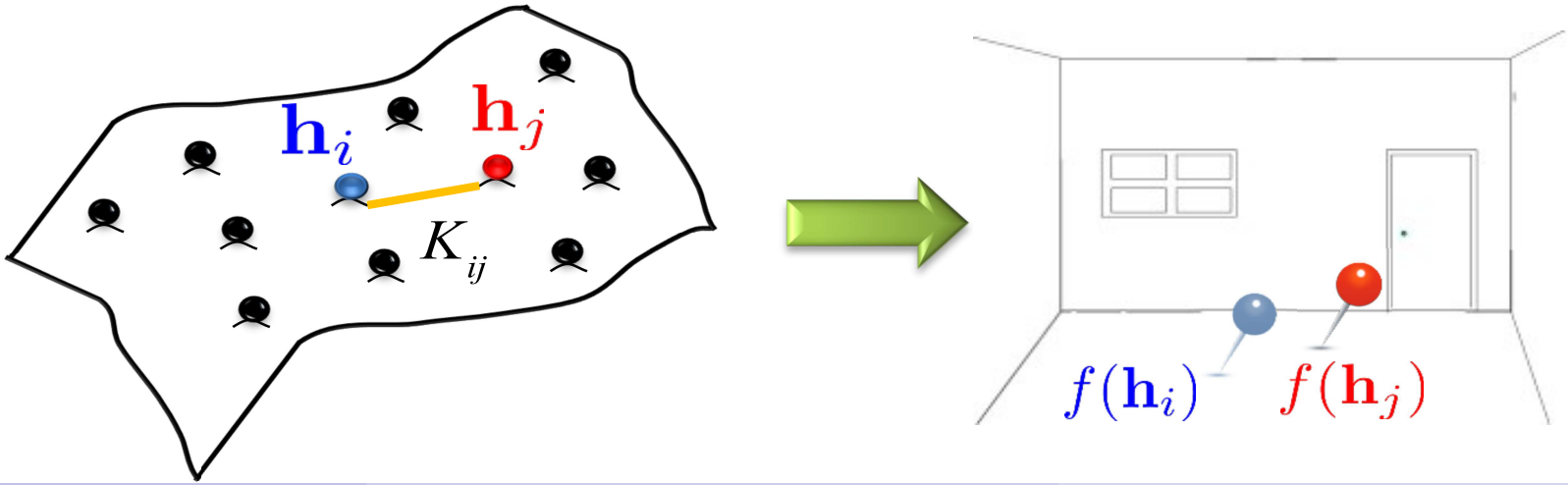
- The graph Laplacian:

$$\mathbf{L} = \mathbf{D} - \mathbf{K}$$

- Define the manifold regularization by:

$$\|f\|_{\mathcal{M}}^2 = \mathbf{f}_D^T \mathbf{L} \mathbf{f}_D = \frac{1}{2} \sum_{i,j=1}^{n_D} K_{ij} \left(f(\mathbf{h}_i) - f(\mathbf{h}_j)\right)^2$$

$\mathbf{f}_D^T = [f_1, f_2, \ldots, f_{n_D}]$ comprising labelled and unlabelled training data



Optimization and Manifold Regularization

The optimization problem can be recast as:

$$f^* = \underset{f \in \mathcal{H}_k}{\operatorname{argmin}} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2 + \gamma_k \| f \|^2_{\mathcal{H}_k} + \gamma_M \mathbf{f}_D^T \mathbf{L} \mathbf{f}_D$$

Optimization and Manifold Regularization

The optimization problem can be recast as:

$$f^* = \underset{f \in \mathcal{H}_k}{\operatorname{argmin}} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2 + \gamma_k \| f \|^2_{\mathcal{H}_k} + \gamma_M \mathbf{f}_D^T \mathbf{L} \mathbf{f}_D$$

Due to the Representer theorem, the minimizer over $\mathcal{H}_k$ of the regularized optimization is represented by:

$$f^*(\mathbf{h}) = \sum_{i=1}^{n_D} a_i k(\mathbf{h}_i, \mathbf{h})$$

with $K_{ij} = k(\mathbf{h}_i, \mathbf{h}_j) = \exp\left\{ -\frac{\|\mathbf{h}_i - \mathbf{h}_j\|^2}{\varepsilon} \right\}$

Optimization and Manifold Regularization

The optimization problem can be recast as:

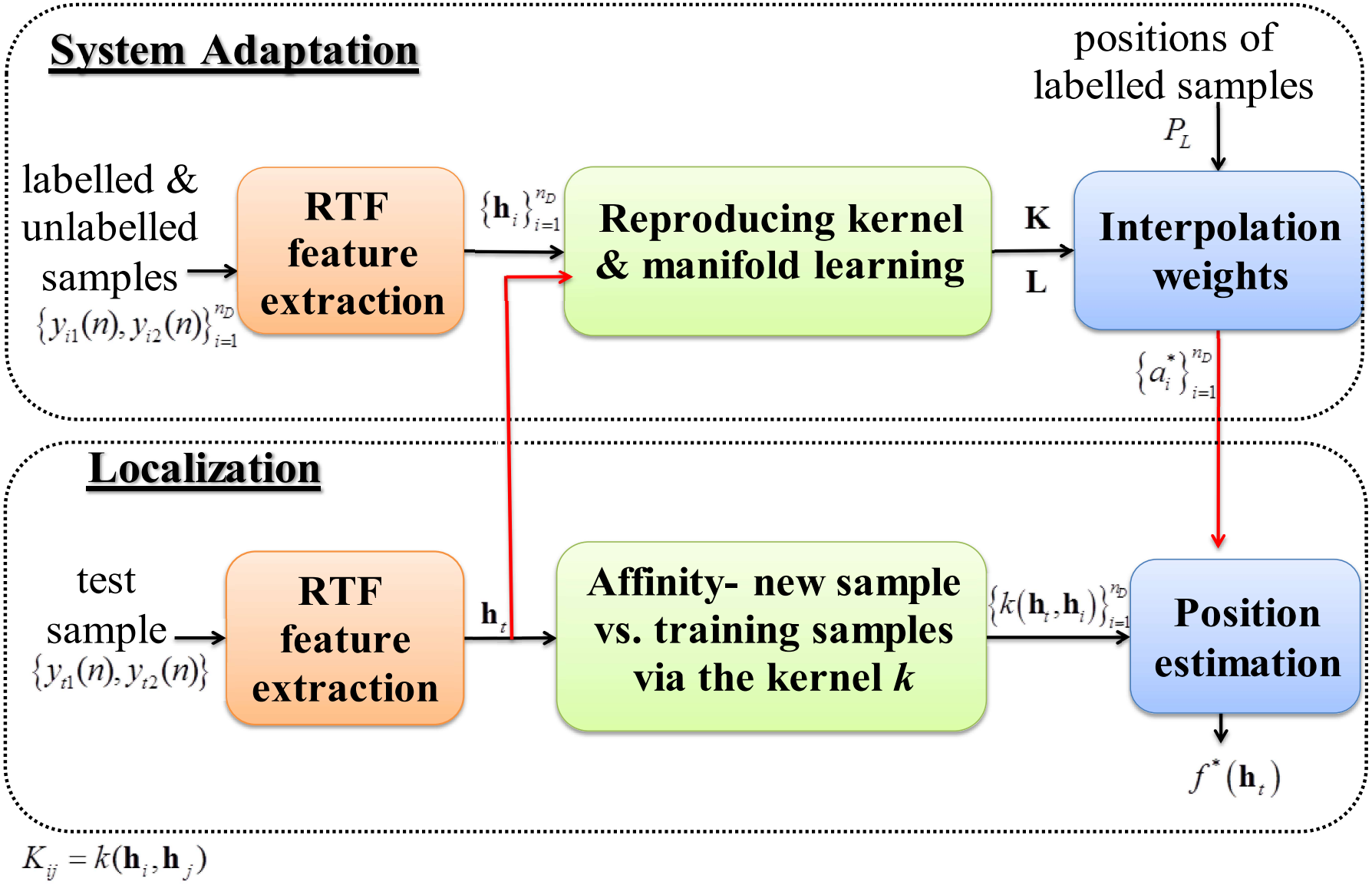
$$f^* = \underset{f \in \mathcal{H}_k}{\operatorname{argmin}} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2 + \gamma_k \|f\|^2_{\mathcal{H}_k} + \gamma_M \mathbf{f}_D^T \mathbf{L} \mathbf{f}_D$$

Substituting the function expansion in the regularized optimization yields:

$$f^*(\mathbf{h}) = \sum_{i=1}^{n_D} a_i k(\mathbf{h}_i, \mathbf{h}) \quad \Rightarrow \quad \text{closed-form solution for } \mathbf{a}^*$$

Manifold Regularization for Localization (MRL)

[Laufer-Goldshtein et al., 2017a]



Simulation Results

Setup:

- **Source positions:** angles between $10^\circ \div 60^\circ$
- **Training:** 6 labelled, 400 unlabelled (SNR=10 dB)

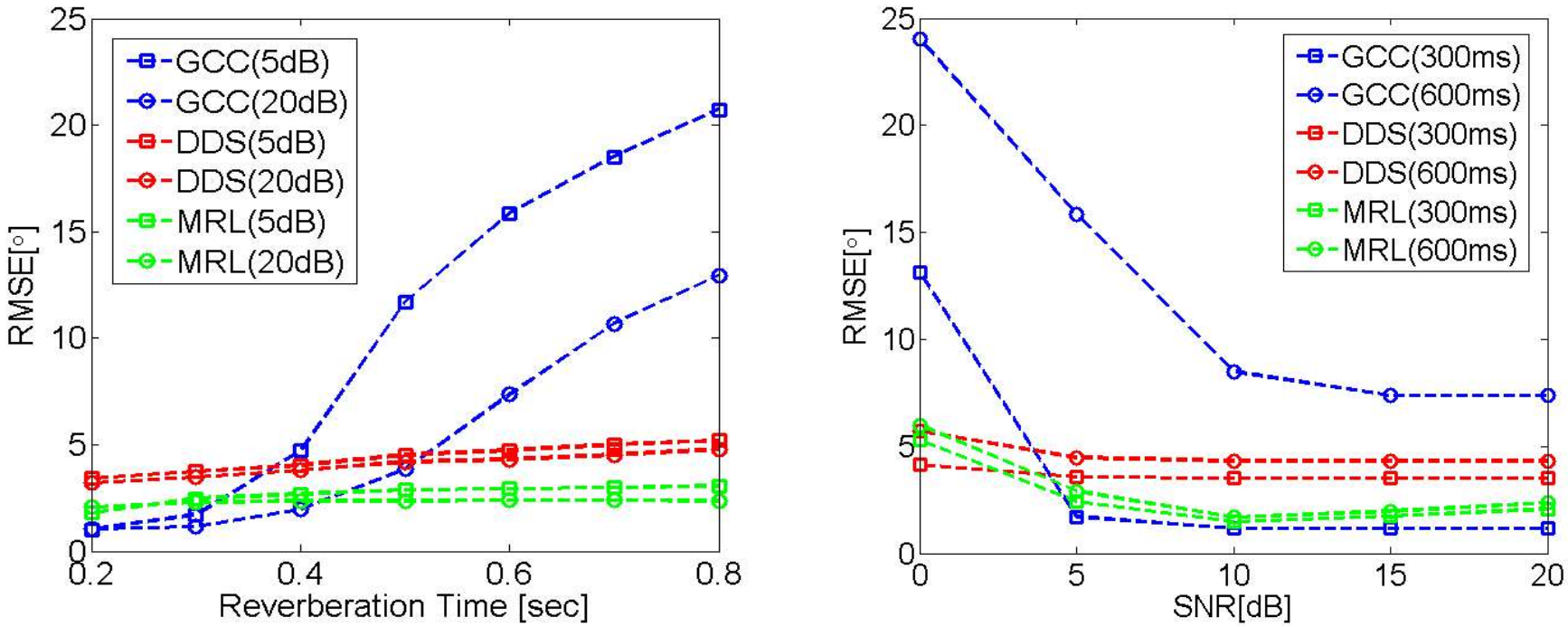


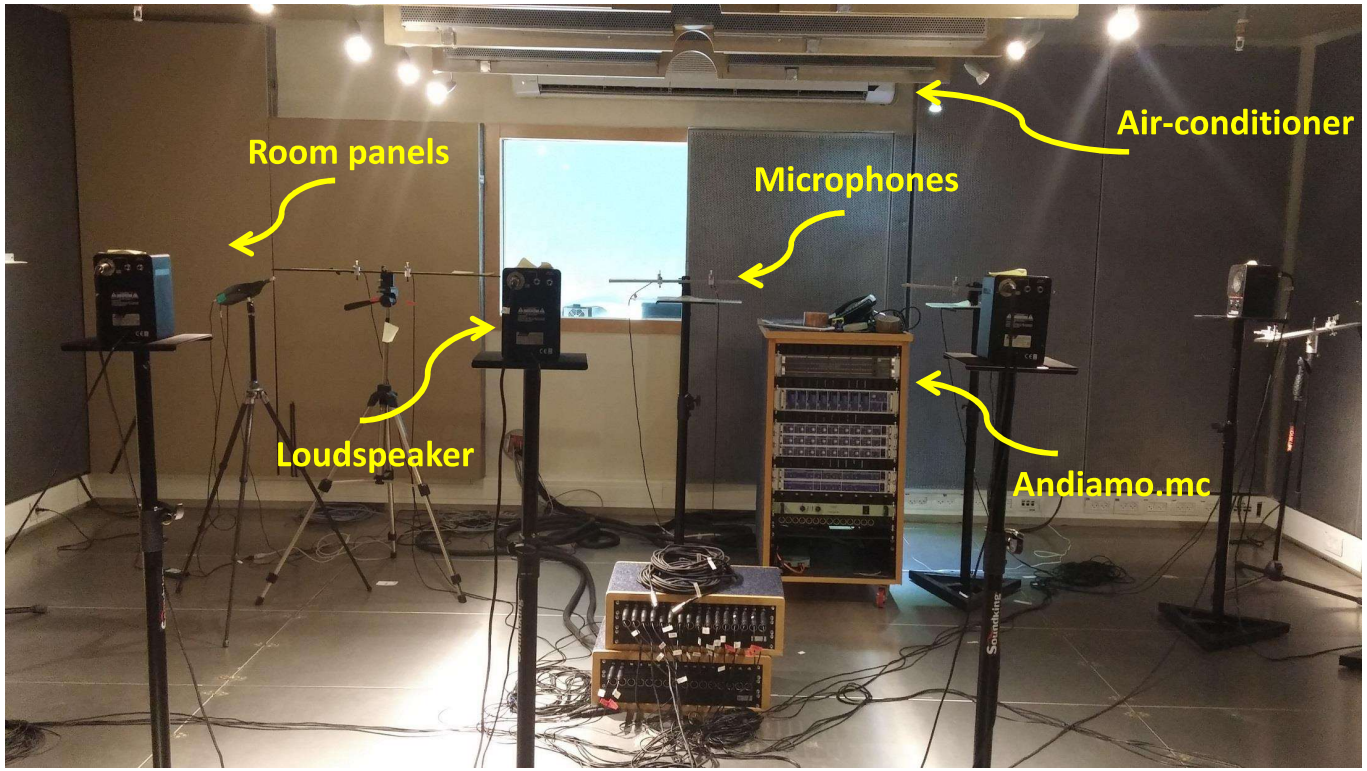
Figure: RMSEs of GCC, DDS and MRL as a function of reverberation time (left), SNR (right)

MRL achieves 2° accuracy in typical noisy and reverberant environments

Recordings setup

Setup:

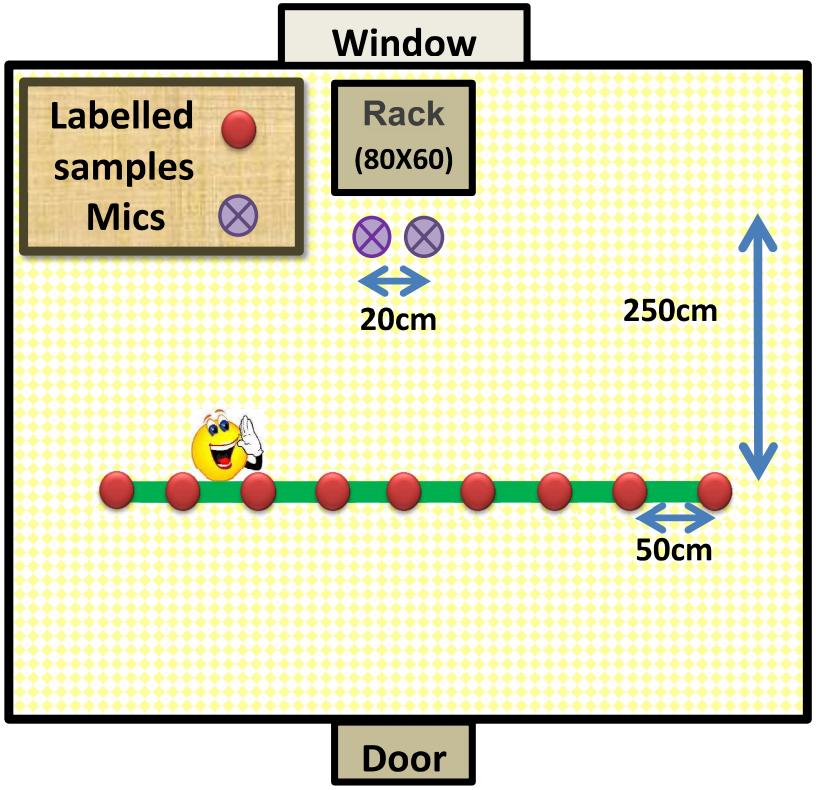
- Real recordings carried out at Bar-Ilan acoustic lab
- A $6 \times 6 \times 2.4$m room controllable reverberation time (set to 620ms)
- Region of interest: a 4m long line at 2.5m distance from the mics



Recordings setup

Setup:

- Real recordings carried out at Bar-Ilan acoustic lab
- A $6 \times 6 \times 2.4$m room controllable reverberation time (set to 620ms)
- Region of interest: a 4m long line at 2.5m distance from the mics



Experimental Results

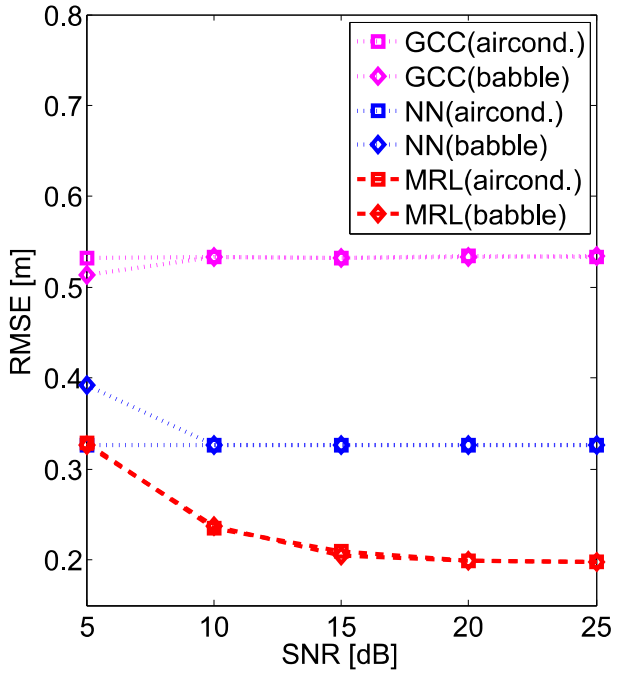
[Laufer-Goldshtein et al., 2016b]

Setup:

- Training: 5 labelled samples (1m resolution), 75 unlabelled samples
- Test: 30 random samples in the defined region
- Two noise types: air-conditioner noise and babble noise

Compare with:

- Nearest-neighbour (NN)
- Generalized cross-correlation (GCC) method [Knapp and Carter, 1976]



Experimental Results

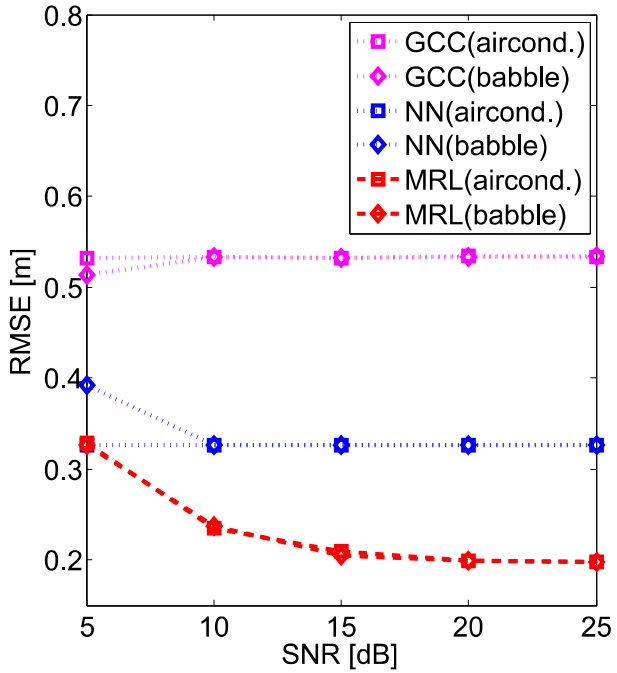
[Laufer-Goldshtein et al., 2016b]

Setup:

- Training: 5 labelled samples (1m resolution), 75 unlabelled samples
- Test: 30 random samples in the defined region
- Two noise types: air-conditioner noise and babble noise

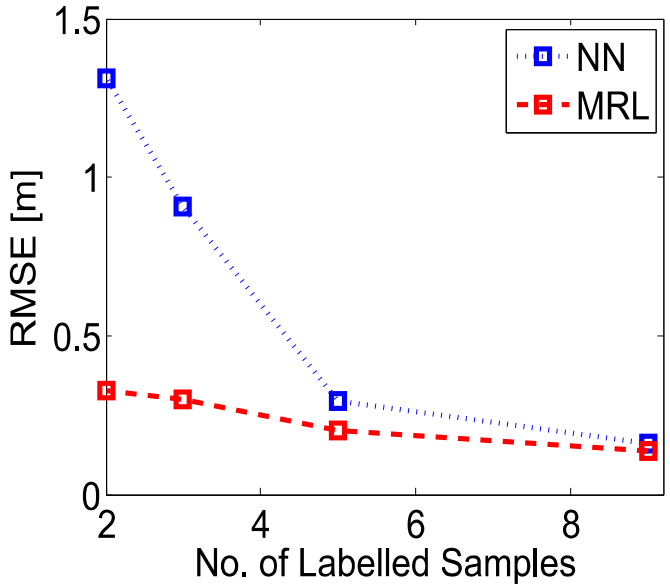
Compare with:

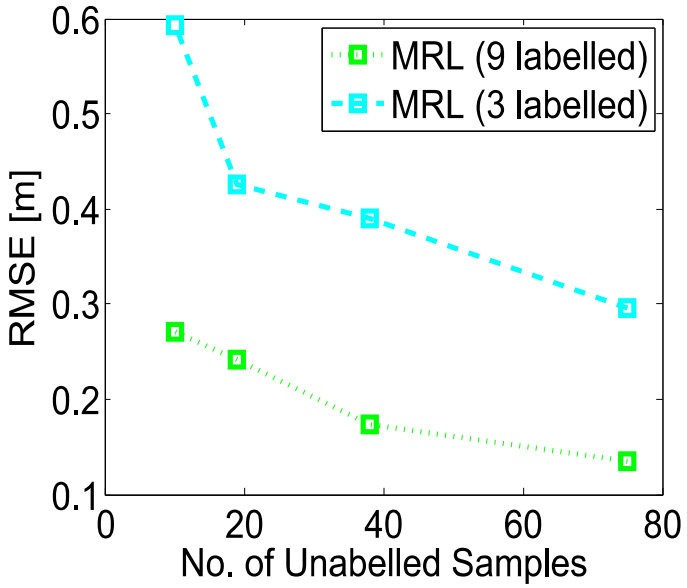
- Nearest-neighbour (NN)
- Generalized cross-correlation (GCC) method [Knapp and Carter, 1976]



The MRL algorithm outperforms the two other methods

Effect of Labelled & Unlabelled Samples





Effect of increasing the amount of labelled/unlabelled samples

→ As the size of the labelled set is reduced - performance gap increases

→ Locate the source even with few labelled samples, using unlabelled information

Outline

Manifold-Based Bayesian Inference [Sindhwani et al., 2007],[Laufer-Goldshtein et al., 2016a]

$$f^* = \underset{f \in \mathcal{H}_k}{\operatorname{argmin}} \underbrace{\frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2}_{\text{Cost Function}} + \underbrace{\gamma_k \|f\|^2_{\mathcal{H}_k}}_{\mathcal{H}_k \text{ norm}} + \underbrace{\gamma_M \mathbf{f}_D^T \mathbf{M} \mathbf{f}_D}_{\text{Manifold Regularization}}$$

Search in RKHS defined by the kernel k

Manifold-Based Bayesian Inference [Sindhwani et al., 2007],[Laufer-Goldshtein et al., 2016a]

$$f^* = \underset{f \in \mathcal{H}_k}{\operatorname{argmin}} \underbrace{\frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2}_{\text{Cost Function}} + \underbrace{\gamma_k \|f\|^2_{\mathcal{H}_k}}_{\mathcal{H}_k \text{ norm}} + \underbrace{\gamma_M \mathbf{f}_D^T \mathbf{M} \mathbf{f}_D}_{\text{Manifold Regularization}}$$

Search in RKHS defined by the kernel k

$$f^* = \underset{f \in \mathcal{H}_{\tilde{k}}}{\operatorname{argmin}} \underbrace{\frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2}_{\text{Cost Function}} + \underbrace{\gamma_k \|f\|^2_{\mathcal{H}_{\tilde{k}}}}_{\mathcal{H}_{\tilde{k}} \text{ norm}}$$

Search in RKHS defined by the kernel $\tilde{k}$

Manifold-Based Bayesian Inference [Sindhwani et al., 2007],[Laufer-Goldshtein et al., 2016a]

$$f^* = \underset{f \in \mathcal{H}_k}{\operatorname{argmin}} \underbrace{\frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2}_{\text{Cost Function}} + \underbrace{\gamma_k \|f\|^2_{\mathcal{H}_k}}_{\mathcal{H}_k \text{ norm}} + \underbrace{\gamma_M \mathbf{f}_D^T \mathbf{M} \mathbf{f}_D}_{\text{Manifold Regularization}}$$

Search in RKHS defined by the kernel k

$$f^* = \underset{f \in \mathcal{H}_{\tilde{k}}}{\operatorname{argmin}} \underbrace{\frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2}_{\text{Cost Function}} + \underbrace{\gamma_k \|f\|^2_{\mathcal{H}_{\tilde{k}}}}_{\mathcal{H}_{\tilde{k}} \text{ norm}}$$

Search in RKHS defined by the kernel $\tilde{k}$

$$\underbrace{p(f|P_L, H_L, H_U)}_{\text{Posterior}} \propto \underbrace{p(P_L|f, H_L)}_{\text{Likelihood Function}} \cdot \underbrace{p(f|H_L, H_U)}_{\text{Manifold-Based Prior}}$$

f is a Gaussian Process

Bayesian Localization

Joint probability:

- Goal: estimate the function value at some test sample $\mathbf{h}_t \in \mathcal{M}$
- The training positions $\bar{\mathbf{p}}_L = \mathrm{vec}\{P_L\}$ and $f(\mathbf{h}_t)$ are jointly Gaussian:

$$\begin{bmatrix} \bar{\mathbf{p}}_L \\ f(\mathbf{h}_t) \end{bmatrix} \Big| H_L, H_U \sim \mathcal{N}\left(\mathbf{0}_{n_L+1}, \begin{bmatrix} \tilde{\mathbf{\Sigma}}_{LL} + \sigma^2 \mathbf{I}_{n_L} & \tilde{\mathbf{\Sigma}}_{Lt} \\ \tilde{\mathbf{\Sigma}}_{Lt}^T & \tilde{\mathbf{\Sigma}}_{tt} \end{bmatrix}\right)$$

- The elements of $\tilde{\mathbf{\Sigma}}_{LL}$, $\tilde{\mathbf{\Sigma}}_{Lt}$ and $\tilde{\mathbf{\Sigma}}_{tt}$ are calculated by the manifold-regularized kernel

$$\mathrm{cov}(f(\mathbf{h}_r), f(\mathbf{h}_l)) \equiv \tilde{k}(\mathbf{h}_r, \mathbf{h}_l)$$

- Note that the relation between two RTFs is not only a function of the current samples, $\mathbf{h}_r, \mathbf{h}_l$, but also uses the information from the entire available set of RTF samples, including the unlabelled RTFs

Bayesian Localization (cont.)

MAP/MMSE estimator:

- The posterior

$$p(f(\mathbf{h}_t)|P_L, H_L, H_U) \sim \mathcal{N}(\hat{f}(\mathbf{h}_t), \mathrm{var}(\hat{f}(\mathbf{h}_t)))$$

is a multivariate Gaussian, where:

- The MAP/MMSE estimator of $f(\mathbf{h}_t)$ is given by:

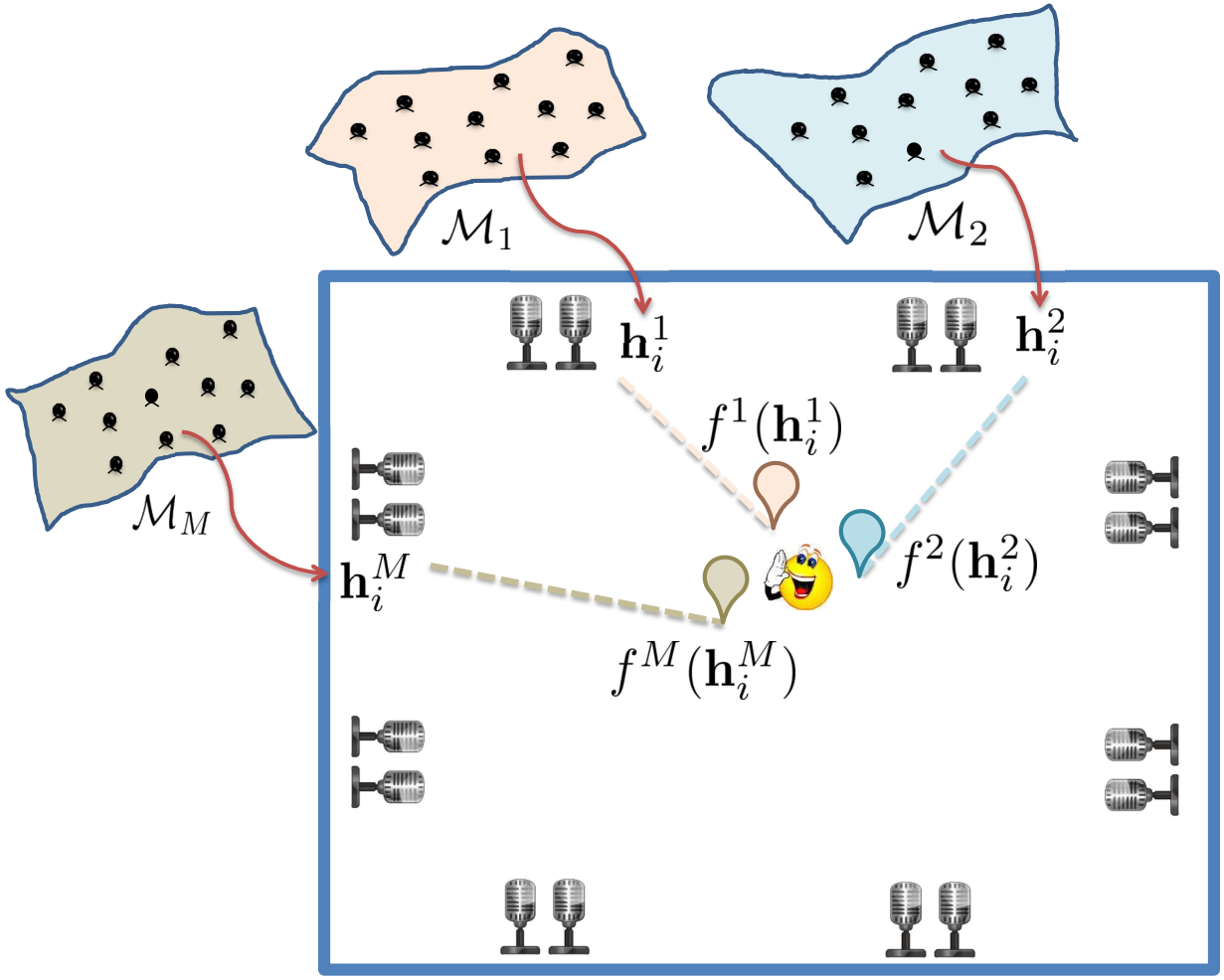
$$\hat{f}(\mathbf{h}_t) = \tilde{\mathbf{\Sigma}}_{Lt}^T \left(\tilde{\mathbf{\Sigma}}_{LL} + \sigma^2 \mathbf{I}_{n_L}\right)^{-1} \bar{\mathbf{p}}_L$$

- The estimation confidence:

$$\mathrm{var}(\hat{f}(\mathbf{h}_t)) = \tilde{\Sigma}_{tt} - \tilde{\mathbf{\Sigma}}_{Lt}^T \left(\tilde{\mathbf{\Sigma}}_{LL} + \sigma^2 \mathbf{I}_{n_L}\right)^{-1} \tilde{\mathbf{\Sigma}}_{Lt}$$

Outline

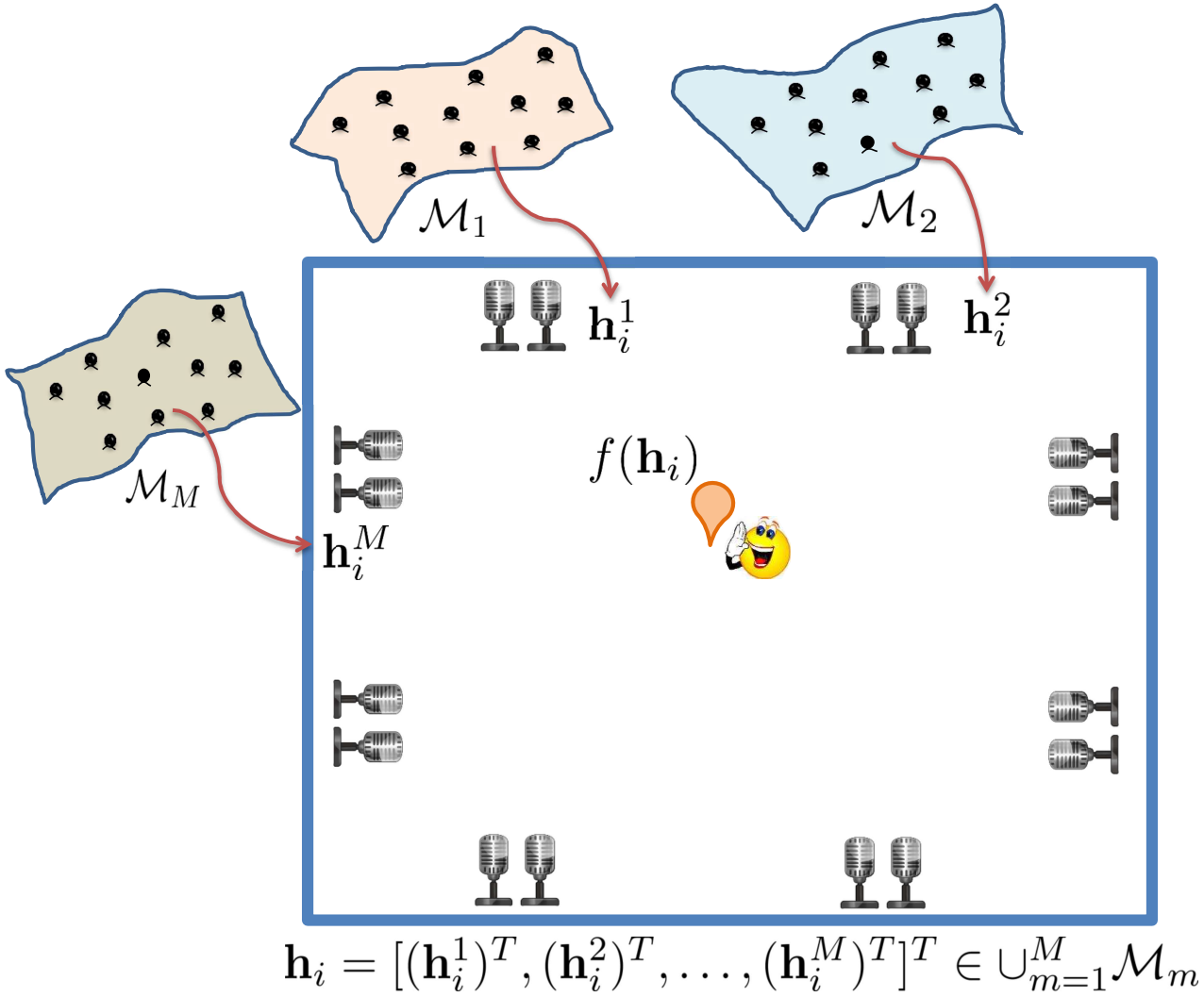
Source Localization with Ad Hoc Array [Laufer-Goldshtein et al., 2017a]



Each node

- Represents a different view point on the same acoustic event
- Induces relations between RTFs according to the associated manifold

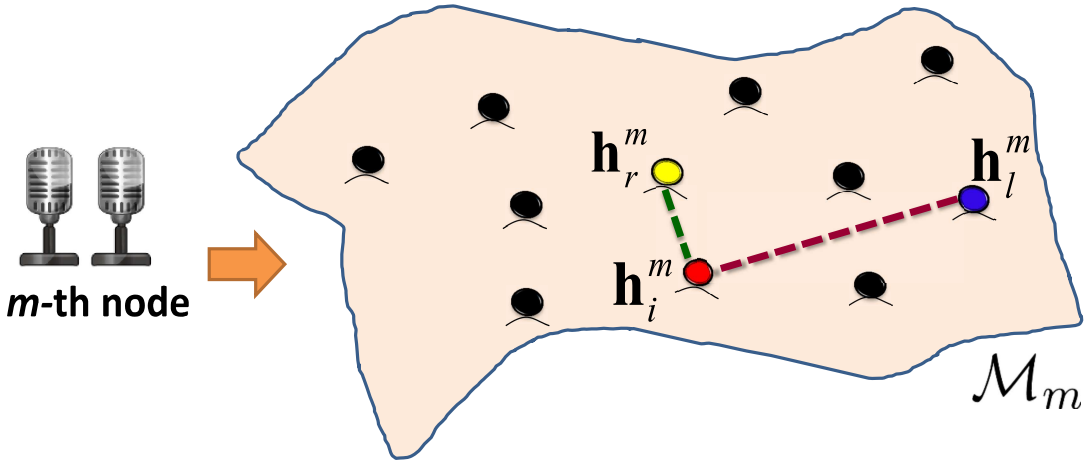
Source Localization with Ad Hoc Array [Laufer-Goldshtein et al., 2017a]



How to fuse the different views in a unified mapping $f : \cup_{m=1}^M \mathcal{M}_m \mapsto \mathbb{R}$?

Intra-Manifold Relations

The mapping follows a Gaussian process $f^m(\mathbf{h}^m) \sim \mathcal{GP}(0, \tilde{k}_m(\mathbf{h}^m, \mathbf{h}_i^m))$



Covariance function

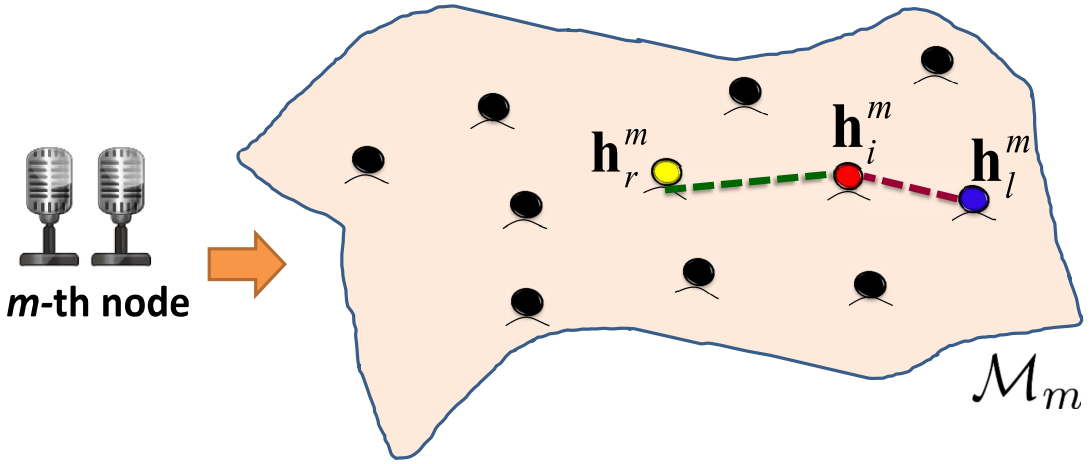
Defined by a new manifold-based covariance function:

$$\operatorname{cov}\left(f^m(\mathbf{h}_r^m), f^m(\mathbf{h}_l^m)\right) \equiv \tilde{k}_m(\mathbf{h}_r^m, \mathbf{h}_l^m) = \sum_{i=1}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

$$= 2k_m(\mathbf{h}_r^m, \mathbf{h}_l^m) + \sum_{\substack{i=1 \\ i \neq l,r}}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

Intra-Manifold Relations

The mapping follows a Gaussian process $f^m(\mathbf{h}^m) \sim \mathcal{GP}(0, \tilde{k}_m(\mathbf{h}^m, \mathbf{h}_i^m))$



Covariance function

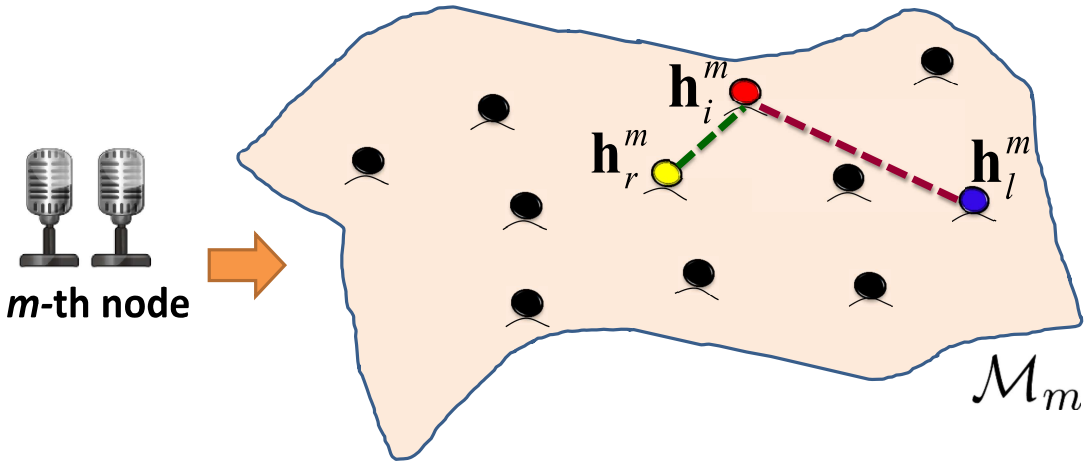
Defined by a new manifold-based covariance function:

$$\operatorname{cov}\left(f^m(\mathbf{h}_r^m), f^m(\mathbf{h}_l^m)\right) \equiv \tilde{k}_m(\mathbf{h}_r^m, \mathbf{h}_l^m) = \sum_{i=1}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

$$= 2k_m(\mathbf{h}_r^m, \mathbf{h}_l^m) + \sum_{\substack{i=1 \\ i \neq l,r}}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

Intra-Manifold Relations

The mapping follows a Gaussian process $f^m(\mathbf{h}^m) \sim \mathcal{GP}(0, \tilde{k}_m(\mathbf{h}^m, \mathbf{h}_i^m))$



Covariance function

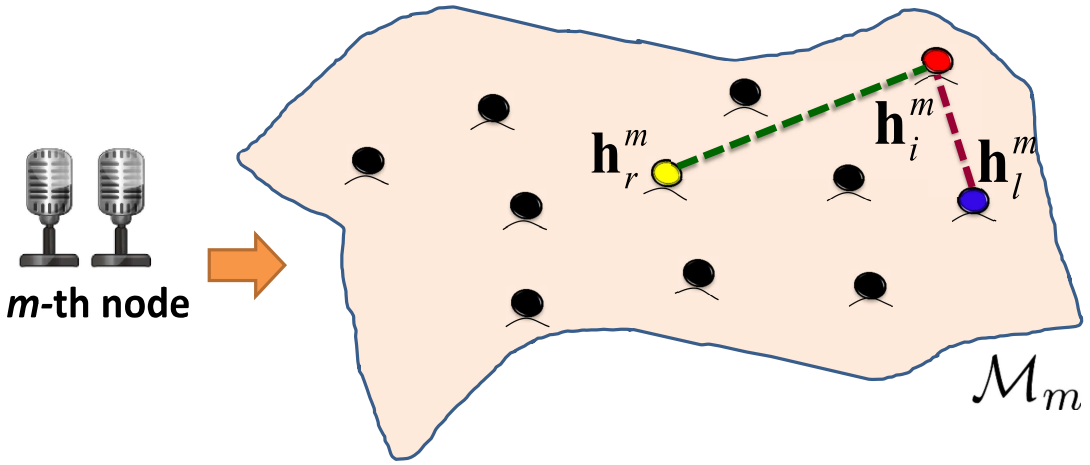
Defined by a new manifold-based covariance function:

$$\operatorname{cov}\left(f^m(\mathbf{h}_r^m), f^m(\mathbf{h}_l^m)\right) \equiv \tilde{k}_m(\mathbf{h}_r^m, \mathbf{h}_l^m) = \sum_{i=1}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

$$= 2k_m(\mathbf{h}_r^m, \mathbf{h}_l^m) + \sum_{\substack{i=1 \\ i \neq l,r}}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

Intra-Manifold Relations

The mapping follows a Gaussian process $f^m(\mathbf{h}^m) \sim \mathcal{GP}(0, \tilde{k}_m(\mathbf{h}^m, \mathbf{h}_i^m))$



Covariance function

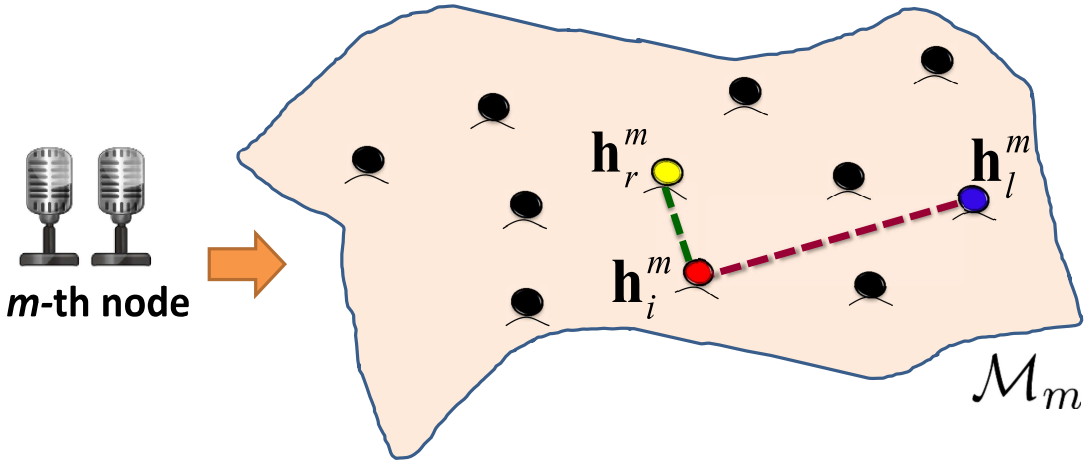
Defined by a new manifold-based covariance function:

$$\operatorname{cov}\left(f^m(\mathbf{h}_r^m), f^m(\mathbf{h}_l^m)\right) \equiv \tilde{k}_m(\mathbf{h}_r^m, \mathbf{h}_l^m) = \sum_{i=1}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

$$= 2k_m(\mathbf{h}_r^m, \mathbf{h}_l^m) + \sum_{\substack{i=1 \\ i \neq l,r}}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

Intra-Manifold Relations

The mapping follows a Gaussian process $f^m(\mathbf{h}^m) \sim \mathcal{GP}(0, \tilde{k}_m(\mathbf{h}^m, \mathbf{h}_i^m))$



Covariance function

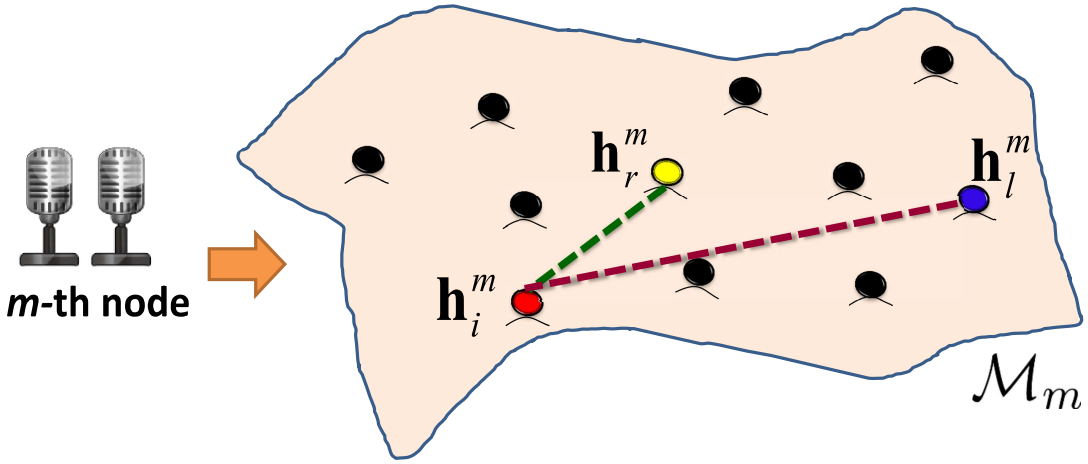
Defined by a new manifold-based covariance function:

$$\operatorname{cov}\left(f^m(\mathbf{h}_r^m), f^m(\mathbf{h}_l^m)\right) \equiv \tilde{k}_m(\mathbf{h}_r^m, \mathbf{h}_l^m) = \sum_{i=1}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

$$= 2k_m(\mathbf{h}_r^m, \mathbf{h}_l^m) + \sum_{\substack{i=1 \\ i \neq l,r}}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

Intra-Manifold Relations

The mapping follows a Gaussian process $f^m(\mathbf{h}^m) \sim \mathcal{GP}(0, \tilde{k}_m(\mathbf{h}^m, \mathbf{h}_i^m))$



Covariance function

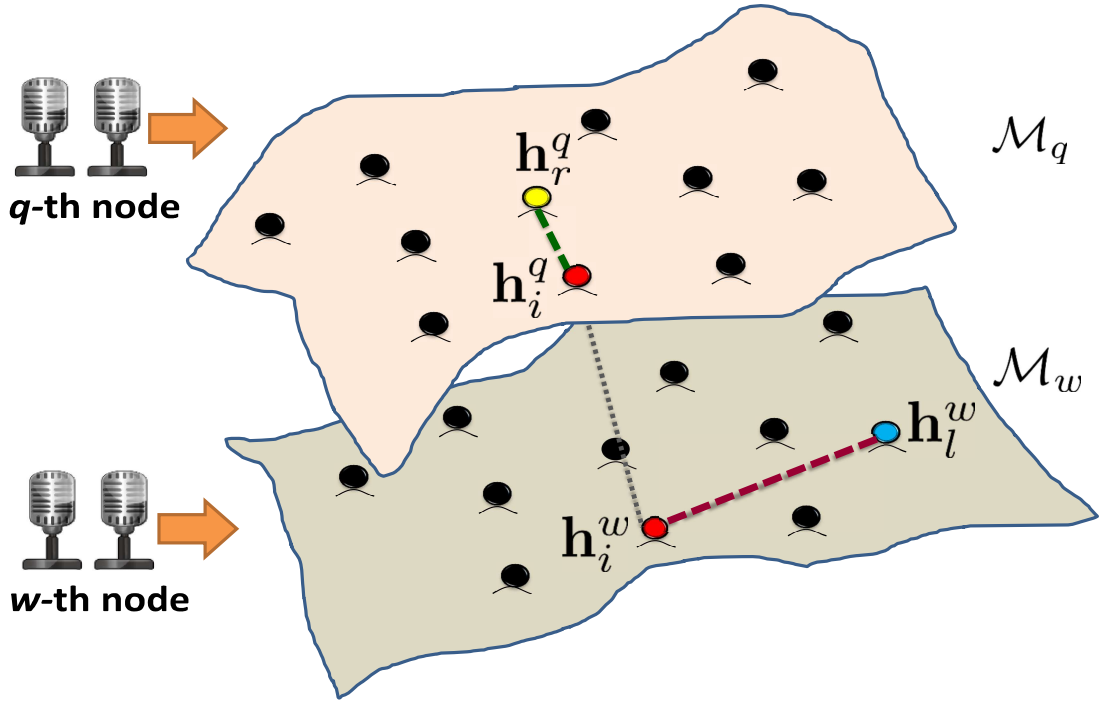
Defined by a new manifold-based covariance function:

$$\operatorname{cov}\left(f^m(\mathbf{h}_r^m), f^m(\mathbf{h}_l^m)\right) \equiv \tilde{k}_m(\mathbf{h}_r^m, \mathbf{h}_l^m) = \sum_{i=1}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

$$= 2k_m(\mathbf{h}_r^m, \mathbf{h}_l^m) + \sum_{\substack{i=1 \\ i \neq l,r}}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

Inter-Manifold Relations

How to measure relations between RTFs from different nodes?



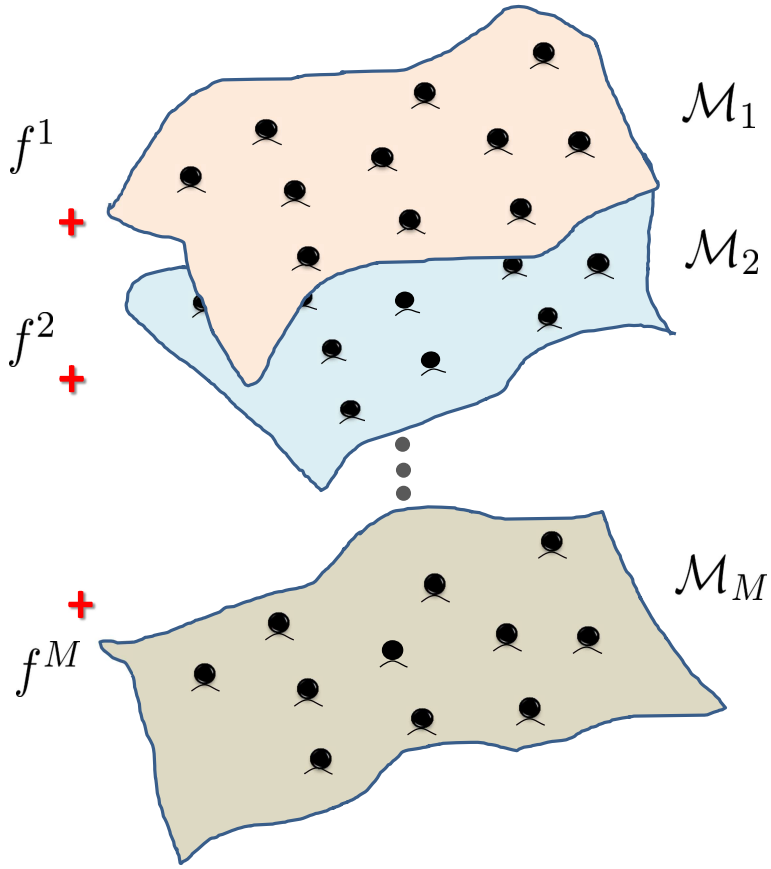
Multi-node covariance

The covariance between $f^q(\mathbf{h}_r^q)$ and $f^w(\mathbf{h}_r^w)$:

$$\operatorname{cov}\left(f^q(\mathbf{h}_r^q), f^w(\mathbf{h}_r^w)\right) = \sum_{i=1}^{n_D} k_q(\mathbf{h}_r^q, \mathbf{h}_i^q) k_w(\mathbf{h}_l^w, \mathbf{h}_i^w)$$

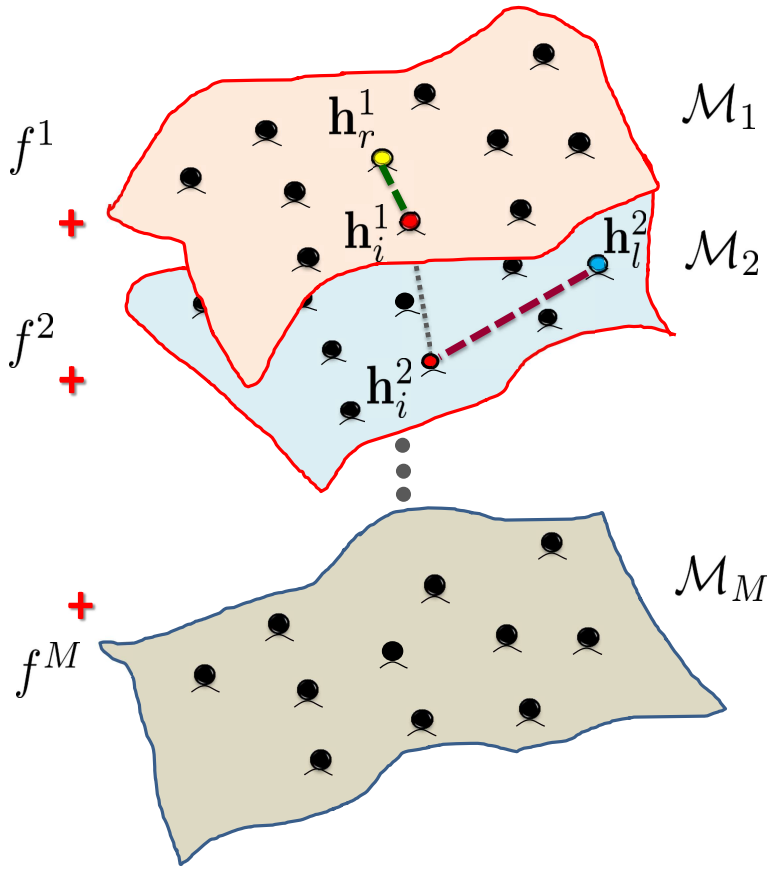
Multiple Manifold Gaussian Process (MMGP)

Define the average process $f = \frac{1}{M}(f^1 + f^2 + \ldots + f^M) \sim \mathcal{GP}(0, \tilde{k})$



Multiple Manifold Gaussian Process (MMGP)

Define the average process $f = \frac{1}{M}(f^1 + f^2 + \ldots + f^M) \sim \mathcal{GP}(0, \tilde{k})$

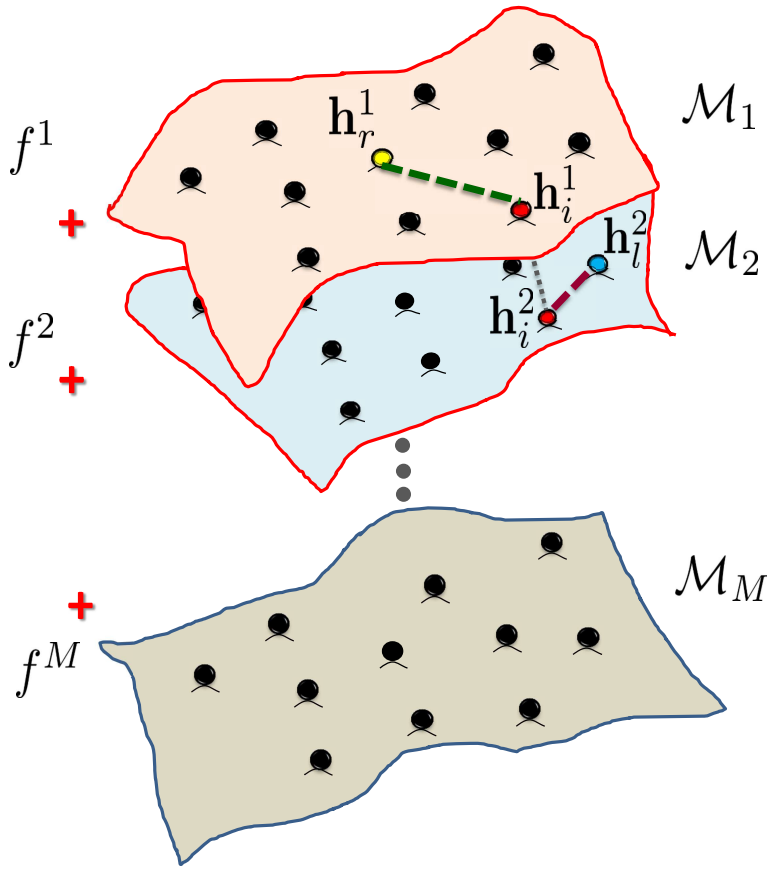


The covariance between $f(\mathbf{h}_r)$ and $f(\mathbf{h}_l)$

$$\text{cov}(f(\mathbf{h}_r), f(\mathbf{h}_l)) \equiv \tilde{k}(\mathbf{h}_r, \mathbf{h}_l) = \frac{1}{M^2} \sum_{q,w=1}^{M} \sum_{i=1}^{n_D} k_q(\mathbf{h}_r^q, \mathbf{h}_i^q) k_w(\mathbf{h}_l^w, \mathbf{h}_i^w)$$

Multiple Manifold Gaussian Process (MMGP)

Define the average process $f = \frac{1}{M}(f^1 + f^2 + \ldots + f^M) \sim \mathcal{GP}(0, \tilde{k})$

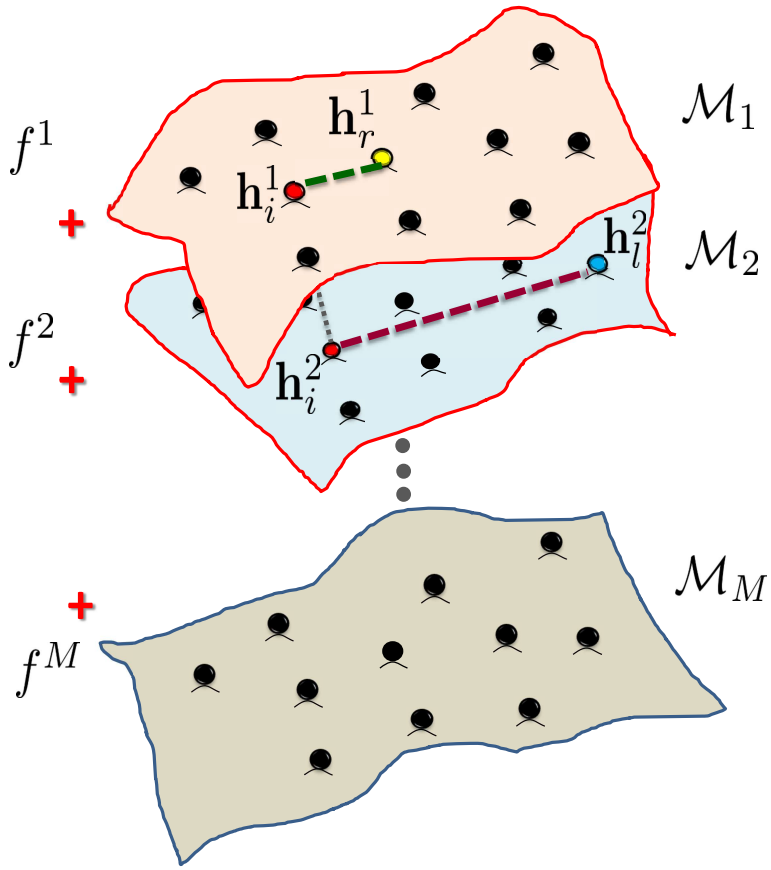


The covariance between $f(\mathbf{h}_r)$ and $f(\mathbf{h}_l)$

$$\text{cov}(f(\mathbf{h}_r), f(\mathbf{h}_l)) \equiv \tilde{k}(\mathbf{h}_r, \mathbf{h}_l) = \frac{1}{M^2} \sum_{q,w=1}^{M} \sum_{i=1}^{n_D} k_q(\mathbf{h}_r^q, \mathbf{h}_i^q) k_w(\mathbf{h}_l^w, \mathbf{h}_i^w)$$

Multiple Manifold Gaussian Process (MMGP)

Define the average process $f = \frac{1}{M}(f^1 + f^2 + \ldots + f^M) \sim \mathcal{GP}(0, \tilde{k})$

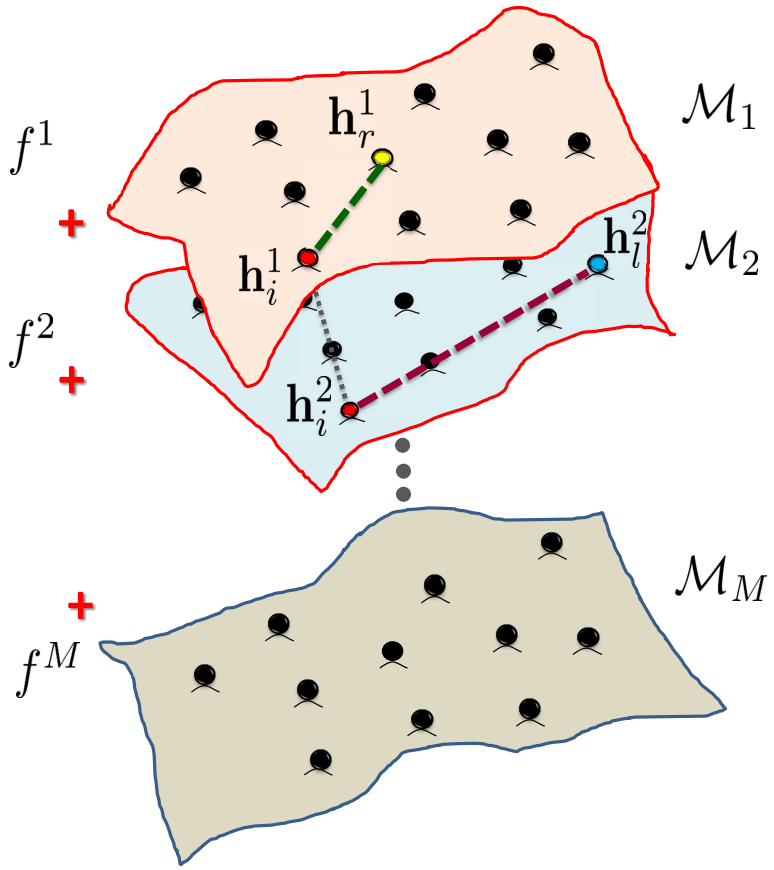


The covariance between $f(\mathbf{h}_r)$ and $f(\mathbf{h}_l)$

$$\text{cov}(f(\mathbf{h}_r), f(\mathbf{h}_l)) \equiv \tilde{k}(\mathbf{h}_r, \mathbf{h}_l) = \frac{1}{M^2} \sum_{q,w=1}^{M} \sum_{i=1}^{n_D} k_q(\mathbf{h}_r^q, \mathbf{h}_i^q) k_w(\mathbf{h}_l^w, \mathbf{h}_i^w)$$

Multiple Manifold Gaussian Process (MMGP)

Define the average process $f = \frac{1}{M}(f^1 + f^2 + \ldots + f^M) \sim \mathcal{GP}(0, \tilde{k})$

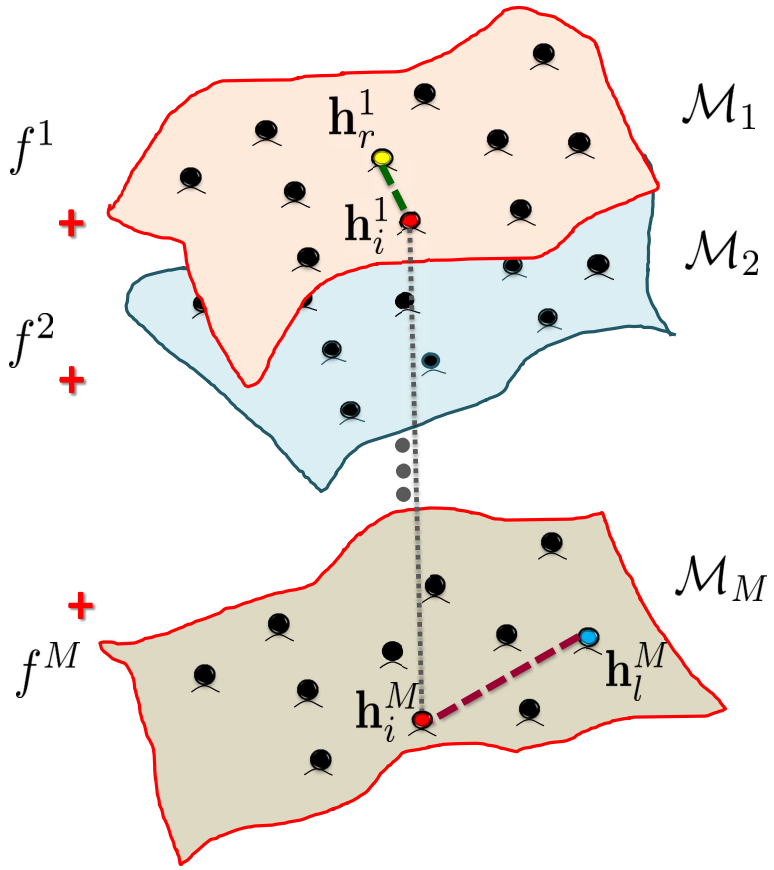


The covariance between $f(\mathbf{h}_r)$ and $f(\mathbf{h}_l)$

$$\text{cov}(f(\mathbf{h}_r), f(\mathbf{h}_l)) \equiv \tilde{k}(\mathbf{h}_r, \mathbf{h}_l) = \frac{1}{M^2} \sum_{q,w=1}^{M} \sum_{i=1}^{n_D} k_q(\mathbf{h}_r^q, \mathbf{h}_i^q) k_w(\mathbf{h}_l^w, \mathbf{h}_i^w)$$

Multiple Manifold Gaussian Process (MMGP)

Define the average process $f = \frac{1}{M}(f^1 + f^2 + \ldots + f^M) \sim \mathcal{GP}(0, \tilde{k})$

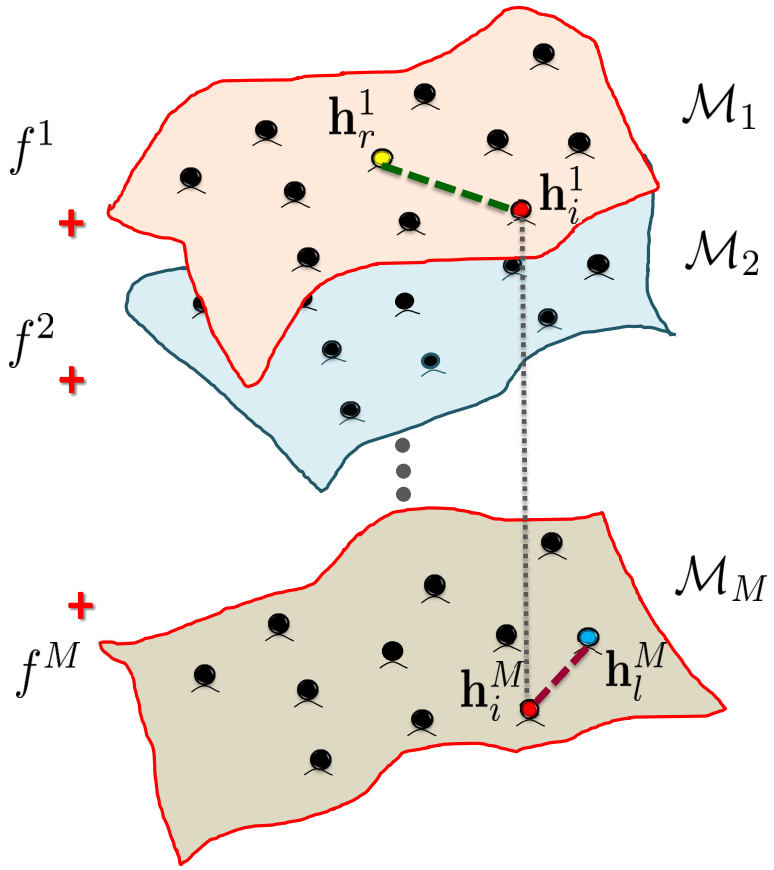


The covariance between $f(\mathbf{h}_r)$ and $f(\mathbf{h}_l)$

$$\mathrm{cov}(f(\mathbf{h}_r), f(\mathbf{h}_l)) \equiv \tilde{k}(\mathbf{h}_r, \mathbf{h}_l) = \frac{1}{M^2} \sum_{q,w=1}^{M} \sum_{i=1}^{n_D} k_q(\mathbf{h}_r^q, \mathbf{h}_i^q) k_w(\mathbf{h}_l^w, \mathbf{h}_i^w)$$

Multiple Manifold Gaussian Process (MMGP)

Define the average process $f = \frac{1}{M}(f^1 + f^2 + \ldots + f^M) \sim \mathcal{GP}(0, \tilde{k})$

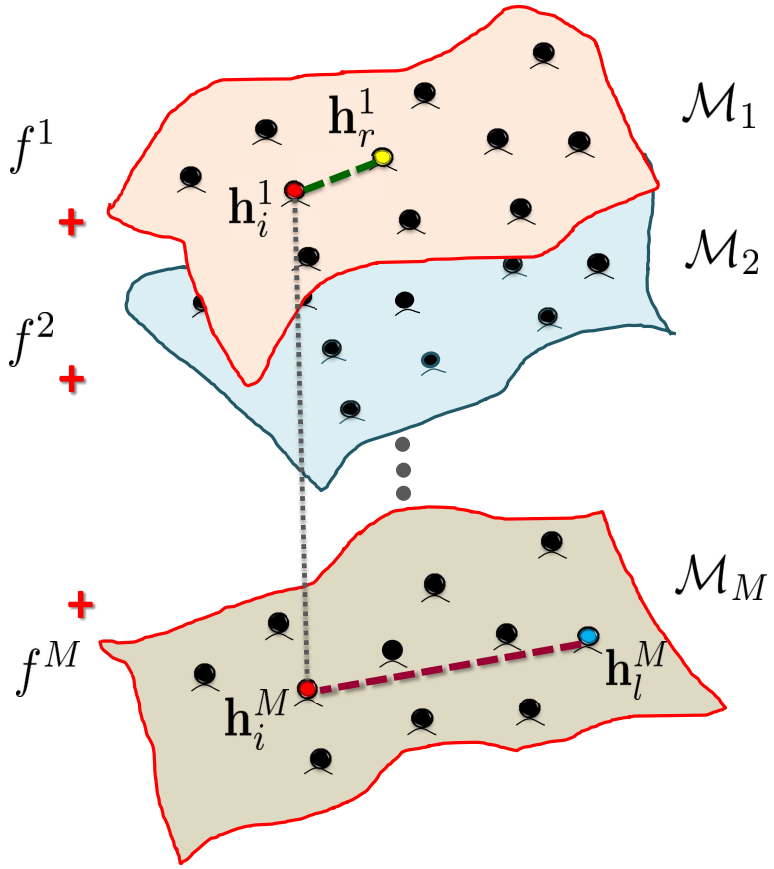


The covariance between $f(\mathbf{h}_r)$ and $f(\mathbf{h}_l)$

$$\text{cov}(f(\mathbf{h}_r), f(\mathbf{h}_l)) \equiv \tilde{k}(\mathbf{h}_r, \mathbf{h}_l) = \frac{1}{M^2} \sum_{q,w=1}^{M} \sum_{i=1}^{n_D} k_q(\mathbf{h}_r^q, \mathbf{h}_i^q) k_w(\mathbf{h}_l^w, \mathbf{h}_i^w)$$

Multiple Manifold Gaussian Process (MMGP)

Define the average process $f = \frac{1}{M}(f^1 + f^2 + \ldots + f^M) \sim \mathcal{GP}(0, \tilde{k})$

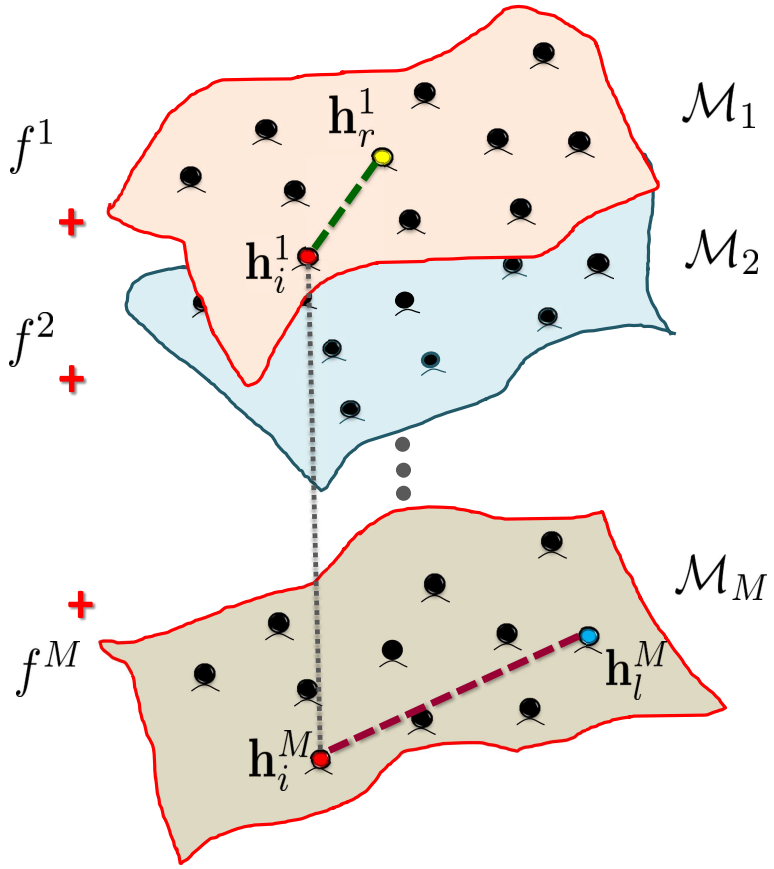


The covariance between $f(\mathbf{h}_r)$ and $f(\mathbf{h}_l)$

$$\mathrm{cov}(f(\mathbf{h}_r), f(\mathbf{h}_l)) \equiv \tilde{k}(\mathbf{h}_r, \mathbf{h}_l) = \frac{1}{M^2} \sum_{q,w=1}^{M} \sum_{i=1}^{n_D} k_q(\mathbf{h}_r^q, \mathbf{h}_i^q) k_w(\mathbf{h}_l^w, \mathbf{h}_i^w)$$

Multiple Manifold Gaussian Process (MMGP)

Define the average process $f = \frac{1}{M}(f^1 + f^2 + \ldots + f^M) \sim \mathcal{GP}(0, \tilde{k})$



The covariance between $f(\mathbf{h}_r)$ and $f(\mathbf{h}_l)$

$$\mathrm{cov}(f(\mathbf{h}_r), f(\mathbf{h}_l)) \equiv \tilde{k}(\mathbf{h}_r, \mathbf{h}_l) = \frac{1}{M^2} \sum_{q,w=1}^{M} \sum_{i=1}^{n_D} k_q(\mathbf{h}_r^q, \mathbf{h}_i^q) k_w(\mathbf{h}_l^w, \mathbf{h}_i^w)$$

Bayesian Multi-View Localization

MAP/MMSE estimator:

- The posterior

$$p(f(\mathbf{h}_t)|P_L, H_L, H_U) \sim \mathcal{N}(\hat{f}(\mathbf{h}_t), \mathrm{var}(\hat{f}(\mathbf{h}_t)))$$

is a multivariate Gaussian, where:

- The MAP/MMSE estimator of $f(\mathbf{h}_t)$ is given by:

$$\hat{f}(\mathbf{h}_t) = \tilde{\mathbf{\Sigma}}_{Lt}^T \left(\tilde{\mathbf{\Sigma}}_{LL} + \sigma^2 \mathbf{I}_{n_L}\right)^{-1} \bar{\mathbf{p}}_L$$

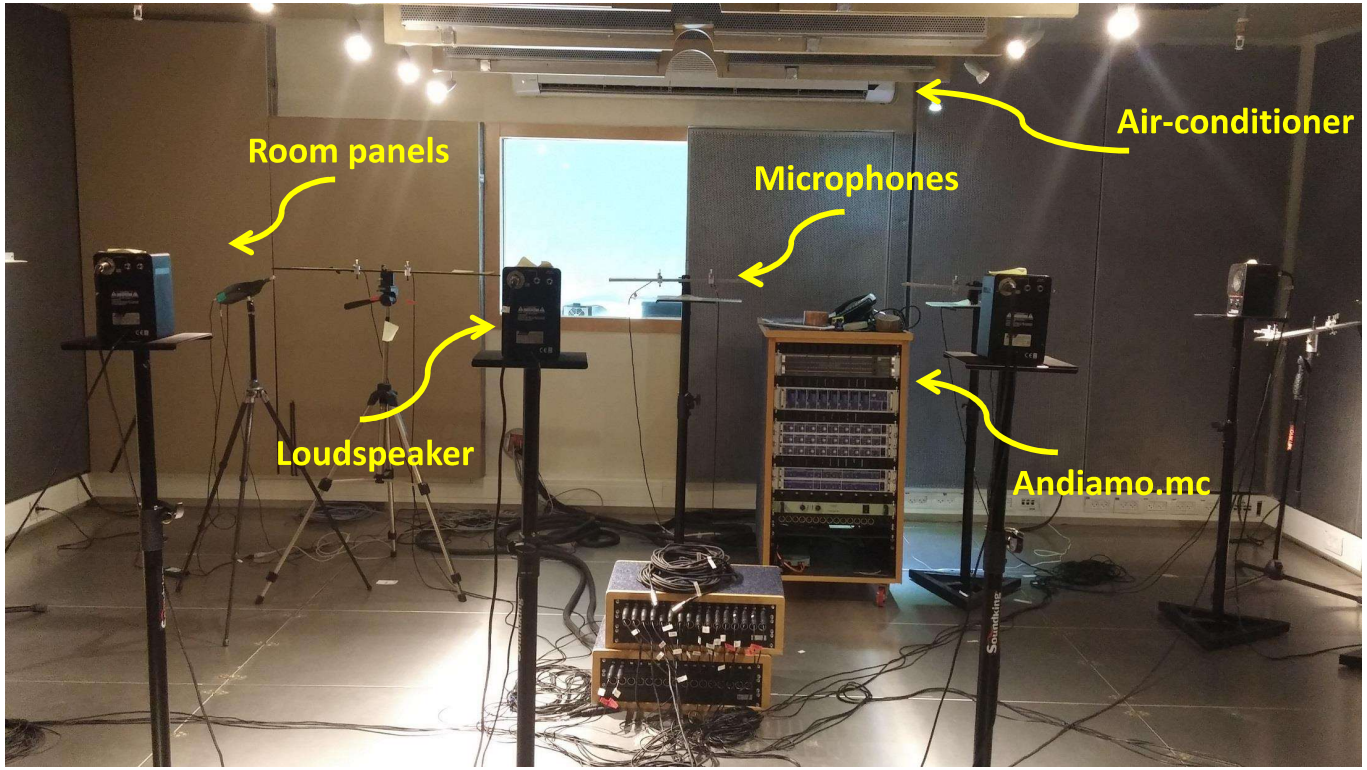
- The estimation confidence

$$\mathrm{var}(\hat{f}(\mathbf{h}_t)) = \tilde{\Sigma}_{tt} - \tilde{\mathbf{\Sigma}}_{Lt}^T \left(\tilde{\mathbf{\Sigma}}_{LL} + \sigma^2 \mathbf{I}_{n_L}\right)^{-1} \tilde{\mathbf{\Sigma}}_{Lt}$$

Recordings Setup

Setup:

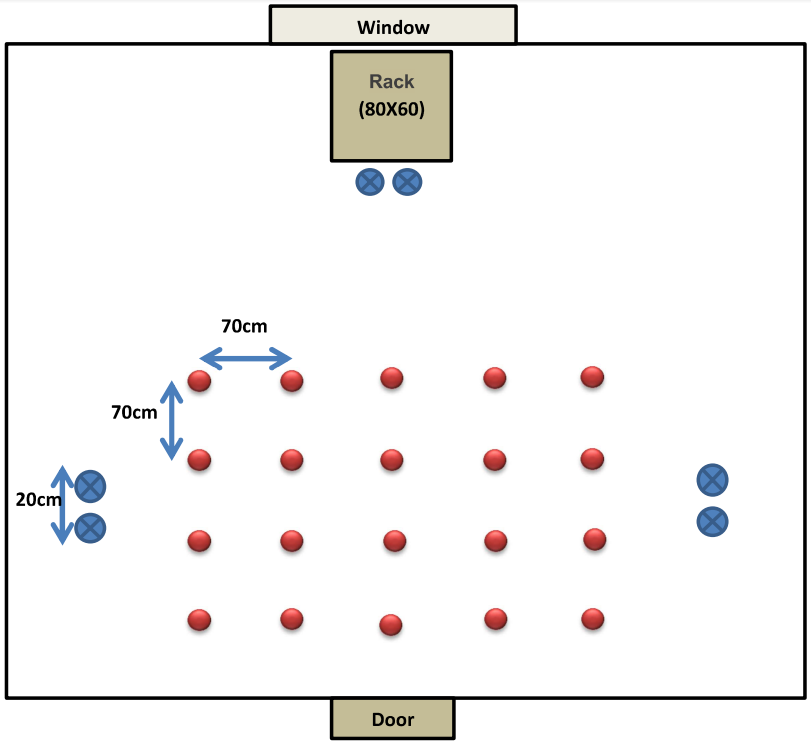
- Real recordings carried out at Bar-Ilan acoustic lab
- A $6 \times 6 \times 2.4$m room controllable reverberation time (set to 620ms)
- Region of interest: Source position is confined to a 2.8×2.1m area
- 3 microphone pairs with inter-distance of 0.2m



Recordings Setup

Setup:

- Real recordings carried out at Bar-Ilan acoustic lab
- A $6 \times 6 \times 2.4$m room controllable reverberation time (set to 620ms)
- Region of interest: Source position is confined to a 2.8×2.1m area
- 3 microphone pairs with inter-distance of 0.2m



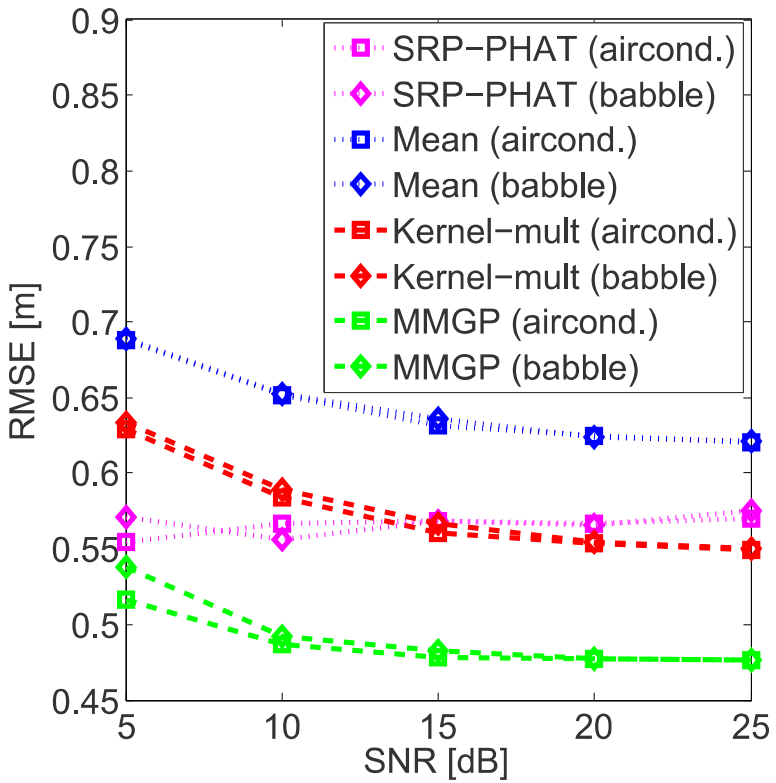
Experimental Results [Laufer-Goldshtein et al., 2017a]

Setup:

- Training: 20 labelled samples (0.7m resolution), 50 unlabelled samples
- Test: 25 random samples in the defined region
- Two noise types: air-conditioner noise and babble noise

Compare with:

- Concatenated independent measurements (Kernel-mult)
- Average of single-node estimates (Mean)
- Beamformer scanning (SRP-PHAT [DiBiase et al., 2001])



Outline

Conclusions

Summary

- Acoustic reflection patterns (RTF) pertain to a low-dimensional manifold controlled by source position
- Data-driven, manifold learning algorithms for source localization and tracking were presented, using regularized optimization in RKHS (or an equivalent Bayesian inference), with highlights:
  - Successful application to both simulations and real-life recordings
  - Array constellation not required, but instead labelled RTFs
  - Multi-view fusion of several manifolds
  - Linearized Kalman filter with propagation and observation models inferred from the manifold structure to address dynamic scenarios

  [Laufer-Goldshtein et al., 2017b]; hybrid approach [Laufer-Goldshtein et al., 2018b] Bayesian Tracking

Challenges and Perspectives

Challenges

- Robustness to environmental changes:
  - Mismatch between train and test conditions
  - Displacement of microphones
- Multiple concurrent speakers
  - A preliminary study using a Mixture of Gaussian Model with Manifold-based Centroids [Bross and Gannot, 2020]
- Source extraction problems are even more complex, as they target enhanced speech rather than only its location
  - A first attempt using projections of beamformer weights on the inferred manifold [Talmon and Gannot, 2013]

Structured List of Algorithms

Single-step

- MUSIC [Schmidt, 1986]; used as a baseline for LOCATA challenge [Löllmann et al., 2018]
- ESPRIT [Roy and Kailath, 1989]; applied to speech signals (e.g. [Teutsch and Kellermann, 2005]) or as features for subsequent spatial processing (e.g. [Thiergart et al., 2014])
- Steered-response beamformer phase transform (SRP-PHAT) [DiBiase et al., 2001, Do et al., 2007]; can also be used as features for subsequent spatial processing (e.g. [Madhu and Martin, 2018, Hadad and Gannot, 2018])
- Maximum-Likelihood (e.g. [Yao et al., 2002])

Structured List of Algorithms (cont.)

TDOA estimation and tracking

- Generalized cross-correlation (GCC) [Knapp and Carter, 1976]
- Subspace methods

  [Benesty, 2000, Doclo and Moonen, 2003]
- Relative transfer function (RTF)-based

  [Dvorkind and Gannot, 2005]

Geometric intersections

- Linear intersections

  [Brandstein et al., 1997]
- Spherical intersections

  [Schau and Robinson, 1987]
- Spherical interpolation

  [Smith and Abel, 1987]
- One-step least squares (OSLS) [Huang et al., 2000]
- Linear-correction least-squares [Huang et al., 2001]

Structured List of Algorithms (cont.)

Bayesian

- Extended, Unscented and Iterated-Extended Kalman filter

  [Gannot and Dvorkind, 2006, Faubel et al., 2009, Klee et al., 2006]

- Particle filters (PF), Rao-Blackwellised Monte-Carlo

  [Ward et al., 2003, Lehmann and Williamson, 2006, Zhong and Hopgood, 2008, Levy et al., 2011]

- Variational Bayes [Ban et al., 2019, Soussana and Gannot, 2019]
- Probability hypothesis density (PHD) filters [Evers and Naylor, 2017]
- Viterbi algorithm for Hidden Markov model (HMM) [Roman et al., 2003]

Structured List of Algorithms (cont.)

Non-Bayesian

- Mixture of Gaussians (MoG) clustering of SRP outputs with expectation-maximization (EM) [Madhu et al., 2008]; using binaural cues and MoG clustering with predefined grid positions as Gaussian centroids [Mandel et al., 2007, Mandel et al., 2010]; using mixture of von Mises distribution [Brendel et al., 2018]
- RANdom SAmple Consensus (RANSAC) and EM [Traa and Smaragdis, 2014]
- Recursive [Schwartz and Gannot, 2013] and distributed [Dorfan and Gannot, 2015, Dorfan et al., 2018] EM MoG clustering with predefined grid positions as Gaussian centroids
- EM with spectrogram clustering [Dorfan et al., 2016, Schwartz et al., 2017, Weisberg et al., 2019]

Structured List of Algorithms (cont.)

Learning-based methods

- Probabilistic piece-wise affine mapping based on smooth binaural manifolds of low dimensions

  [Deleforge and Horaud, 2012, Deleforge et al., 2013, Deleforge et al., 2015]

- MoG clustering of binaural cues using multi-condition training

  [May et al., 2011]

- Gaussian processes inference to map coherent-to-diffuse power ratio and source distance [Brendel and Kellermann, 2019]

- Deep learning for classifying feature vectors to candidate positions: Fully connected [Xiao et al., 2015]; convolutional neural networks (CNN) [Takeda and Komatani, 2016, Chakrabarty and Habets, 2019], convolutional recurrent neural network (CRNN) [Adavanne et al., 2018, Perotin et al., 2019]

- Deep ranking using triplet loss [Opochinsky et al., 2019]

Back to main

Outline

Dynamic Scenario

Received Signals

$$y^{mi}(n) = \sum_k a_n^{mi}(k) s(n-k) + u^{mi}(n); \quad m = 1, \ldots, M, i = 1, 2$$

- a_n^{mi} - a time-varying AIR at node m, microphone i in time n
- $\mathbf{h}^m(t)$ - the instantaneous RTF (iRTF) vector at node m in the STFT frame t
- $\mathbf{h}(t) = \left[[\mathbf{h}^1(t)]^T, \ldots, [\mathbf{h}^M(t)]^T\right]^T$ - a concatenation of the iRTF vectors from all nodes
- $p_c(t) = f(\mathbf{h}(t)), c \in \{x, y, z\}$ - mapping of the concatenated iRTF vector to position (for brevity $p_c(t) \equiv p(t)$)

Reminder: The covariance between $p_r = f(\mathbf{h}_r)$ and $p_l = f(\mathbf{h}_l)$

$$\text{cov}(f(\mathbf{h}_r), f(\mathbf{h}_l)) \equiv \tilde{k}(\mathbf{h}_r, \mathbf{h}_l) = \frac{1}{M^2} \sum_{q,w=1}^{M} \sum_{i=1}^{n_D} k_q(\mathbf{h}_r^q, \mathbf{h}_i^q) k_w(\mathbf{h}_l^w, \mathbf{h}_i^w)$$

Bayesian Inference for Source Tracking

Standard (Nonlinear) State-Space Model

$$p(t) = b_t(p(t-1)) + \xi_t$$
$$q_t = c_t(p(t)) + \zeta_t$$

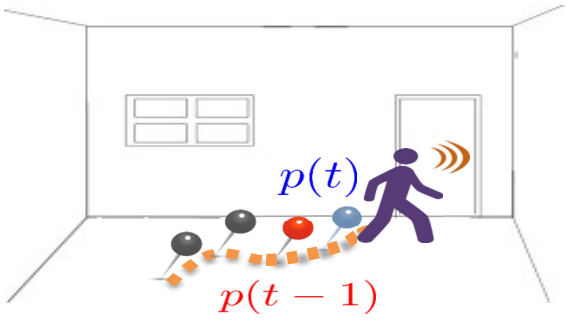
Bayesian Inference for Source Tracking

Standard (Nonlinear) State-Space Model

$$p(t) = b_t(p(t-1)) + \xi_t$$
$$q_t = c_t(p(t)) + \zeta_t$$

Propagation Model

- Relate current and previous positions arbitrarily using random walk or Langevin
- Independent of measurements
- Noise statistics is unknown



Bayesian Inference for Source Tracking

Standard (Nonlinear) State-Space Model

$$p(t) = b_t(p(t-1)) + \xi_t$$

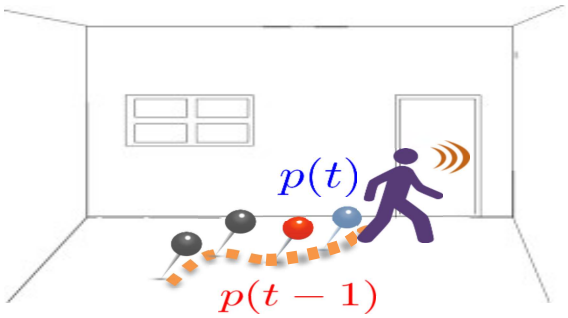
$$q_t = c_t(p(t)) + \zeta_t$$

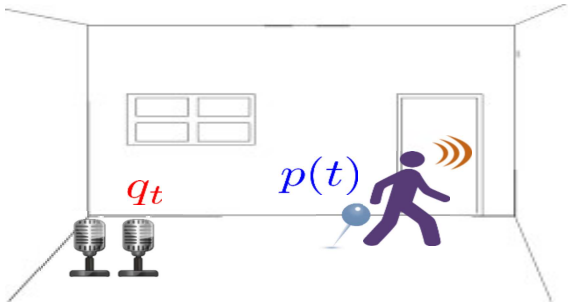
Propagation Model

- Relate current and previous positions arbitrarily using random walk or Langevin
- Independent of measurements
- Noise statistics is unknown

Observation Model

- Relate current position to measurements
- Examples: TDOA or steered response power readings
- Noise statistics is unknown

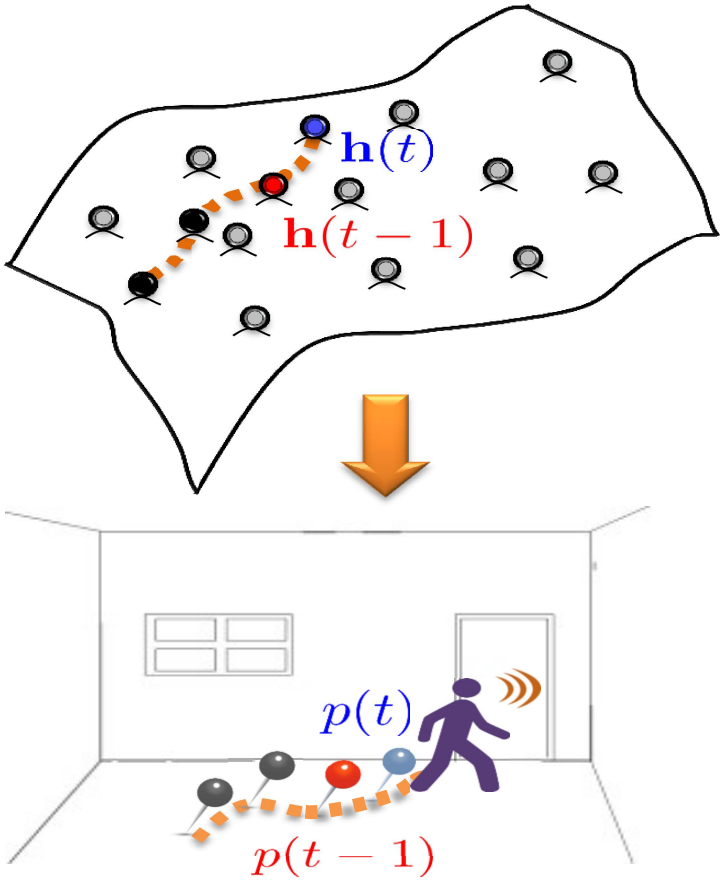




Tracking on the Manifold [Laufer-Goldshtein et al., 2017b]

Propagation Model - Local

Transform nonlinear regression of high-dimensional RTFs to linear transition of source positions



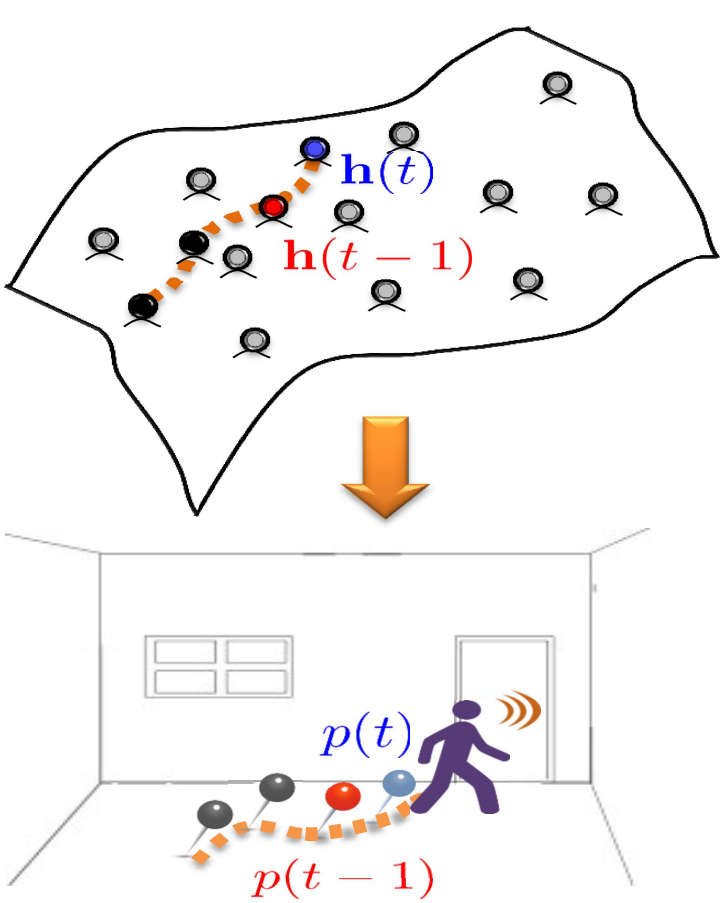
Tracking on the Manifold [Laufer-Goldshtein et al., 2017b]

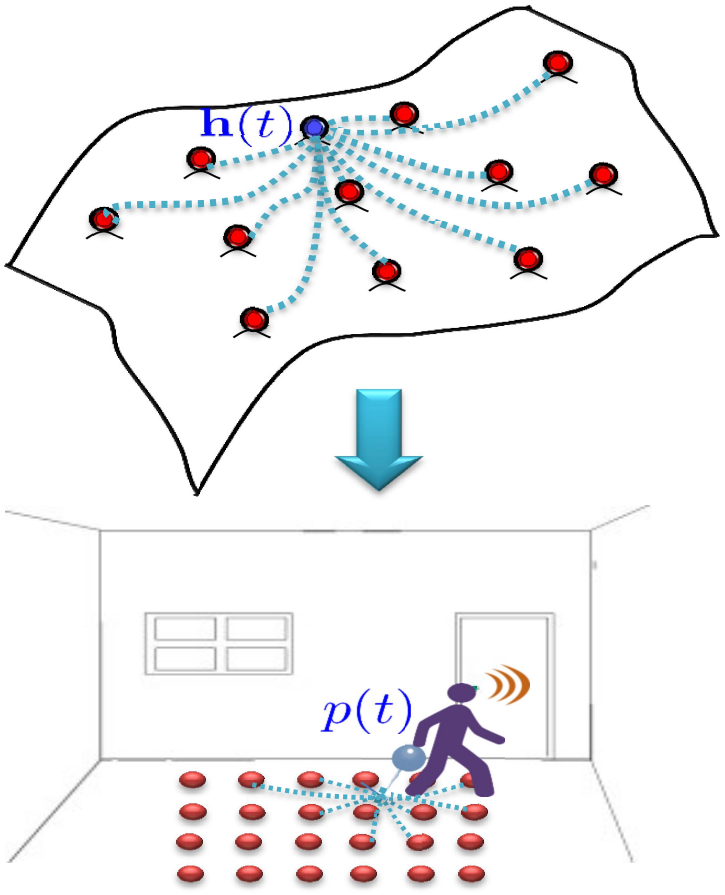
Propagation Model - Local

Transforms nonlinear regression of high-dimensional RTFs to linear transition of source positions

Observation model - Global

Formed by a regression of training positions according to relations on the manifold





State Space Representation (1)

Probabilistic Motion Model:

- Current and previous positions, $p(t) = f(\mathbf{h}(t))$ and $p(t-1) = f(\mathbf{h}(t-1))$, are jointly GP:

$$\begin{bmatrix} p(t) \\ p(t-1) \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \tilde{\Sigma}_{t,t} & \tilde{\Sigma}_{t,t-1} \\ \tilde{\Sigma}_{t,t-1} & \tilde{\Sigma}_{t-1,t-1} \end{bmatrix}\right)$$

- Their conditional probability is given by:

$$p(t)|p(t-1) \sim \mathcal{N}\left(\frac{\tilde{\Sigma}_{t,t-1}}{\tilde{\Sigma}_{t-1,t-1}} p(t-1), \tilde{\Sigma}_{t,t} - \frac{\tilde{\Sigma}^2_{t,t-1}}{\tilde{\Sigma}_{t-1,t-1}}\right)$$

where $\tilde{\Sigma}_{t,\tau} \equiv \tilde{k}\left(\mathbf{h}(t), \mathbf{h}(\tau)\right)$

State Space Representation (2)

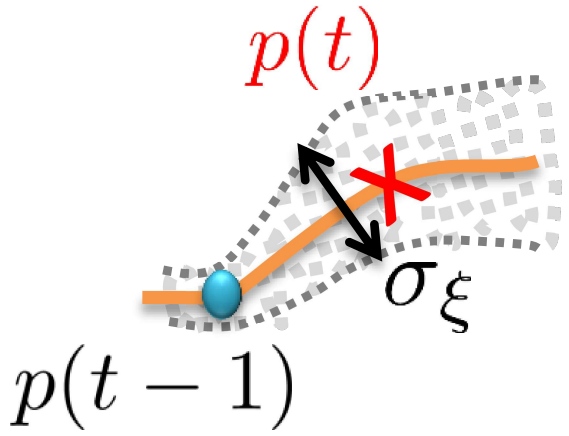
Propagation Model:

Induces a linear propagation equation with an additive Gaussian noise ξ_t:

$$p(t) = b_t \cdot p(t-1) + \xi_t$$

with

- $b_t = \frac{\tilde{\Sigma}_{t,t-1}}{\tilde{\Sigma}_{t-1,t-1}}$ - The Wiener filter
- $\xi_t \sim \mathcal{N}\left(0, \sigma_\xi^2\right)$ with $\sigma_\xi^2 = \tilde{\Sigma}_{t,t} - \frac{\tilde{\Sigma}_{t,t-1}^2}{\tilde{\Sigma}_{t-1,t-1}}$, the corresponding variance



State Space Representation (3)

Probabilistic Observation Model:

- $\bar{\mathbf{p}}_L = [\bar{p}_1, \ldots, \bar{p}_{n_L}]^T$ - measured positions of the labelled set
- $\bar{p}_i = p_i + \eta_i$ - noisy versions of the actual position p_i
- η_i - independent Gaussian noise with variance σ^2
- $p(t) = f(\mathbf{h}(t))$ and $\bar{\mathbf{p}}_L$ are jointly GP:

$$\begin{bmatrix} p(t) \\ \bar{\mathbf{p}}_L \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \tilde{\Sigma}_{t,t} & \tilde{\mathbf{\Sigma}}_{Lt} \\ \tilde{\mathbf{\Sigma}}_{Lt} & \tilde{\mathbf{\Sigma}}_{LL} + \sigma^2 \mathbf{I}_{n_L} \end{bmatrix}\right)$$

- Their conditional probability is given by:

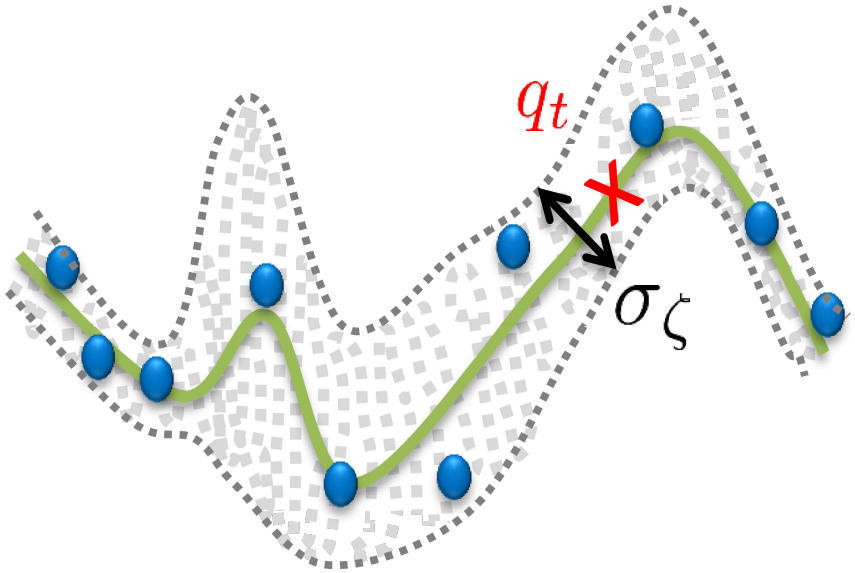
$$p(t)|\bar{\mathbf{p}}_L \sim \mathcal{N}\left(\tilde{\mathbf{\Sigma}}_{Lt}^H \left(\tilde{\mathbf{\Sigma}}_L + \sigma^2 \mathbf{I}_{n_L}\right)^{-1} \bar{\mathbf{p}}_L, \tilde{\Sigma}_{t,t} - \tilde{\mathbf{\Sigma}}_{Lt}^H \left(\tilde{\mathbf{\Sigma}}_L + \sigma^2 \mathbf{I}_{n_L}\right)^{-1} \tilde{\mathbf{\Sigma}}_{Lt}\right)$$

State-Space Representation (4)

Observation model:

- Induces a noisy *artificial observation* q_t that represents a linear regression on the training set:

$$q_t = \tilde{\mathbf{\Sigma}}_{Lt}^H \left(\tilde{\mathbf{\Sigma}}_{LL} + \sigma^2 \mathbf{I}_{n_L} \right)^{-1} \bar{\mathbf{p}}_L$$



The corresponding observation model:

$$q_t = p(t) + \zeta_t$$

where $\zeta_t \sim \mathcal{N}\left(0, \sigma_\zeta^2\right)$ with $\sigma_\zeta^2 = \tilde{\Sigma}_{t,t} - \tilde{\mathbf{\Sigma}}_{Lt}^H \left(\tilde{\mathbf{\Sigma}}_{LL} + \mathbf{I}_{n_L} \right)^{-1} \tilde{\mathbf{\Sigma}}_{Lt}$.

Tracking Algorithm

Space-State Representation:

The proposed state-space model is given by:

$$p(t) = b_t \cdot p(t-1) + \xi_t$$

$$q_t = p(t) + \zeta_t$$

Kalman Filter

Time Update

- Predicted Position:

$$\hat{f}(t|t-1) = g_t \cdot \hat{f}(t-1|t-1)$$

- Predicted Covariance:

$$\gamma(t|t-1) = \gamma(t-1|t-1) + \sigma_\xi^2$$

Measurement Update

- Kalman Gain:

$$\kappa(t) \equiv \frac{\gamma(t|t-1)}{\gamma(t|t-1)+\sigma_\zeta^2}$$

- Updated position estimate:

$$\hat{f}(t|t) = \hat{f}(t|t-1) + \kappa(t)\left(y_t - \hat{f}(t|t-1)\right)$$

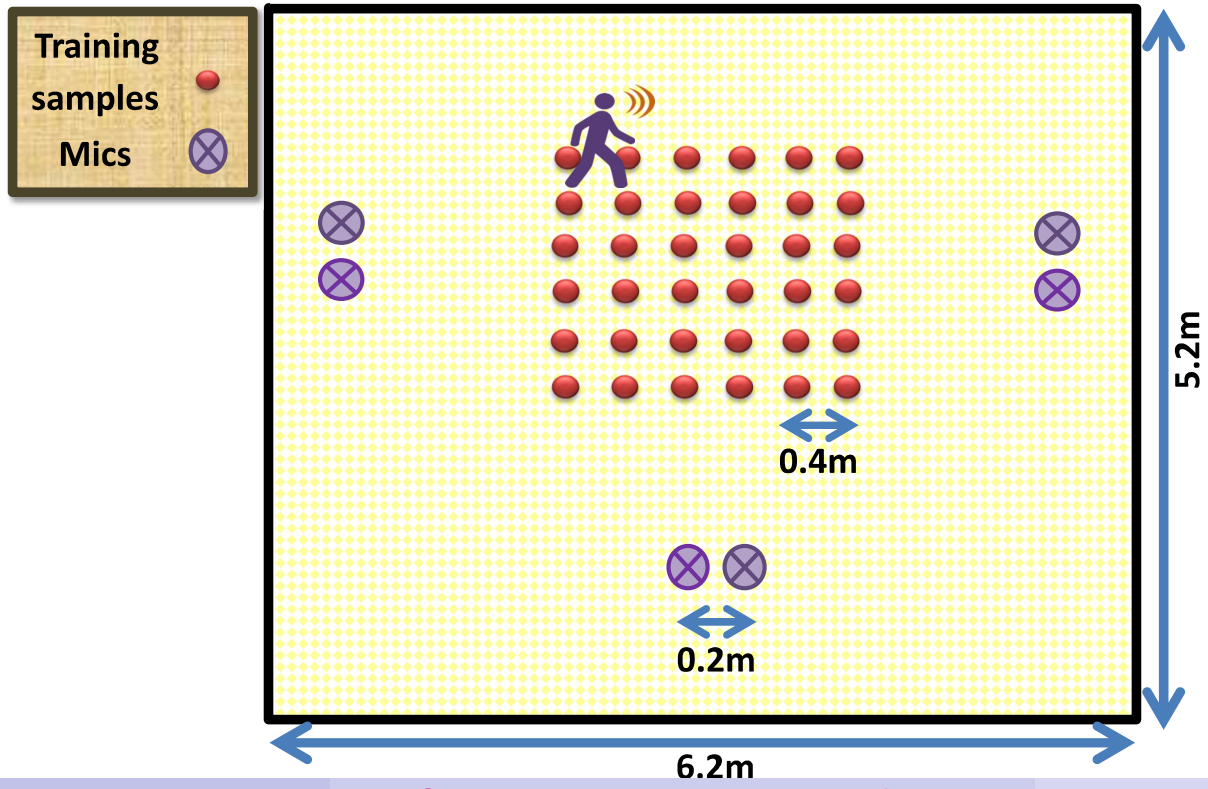
- Updated Covariance:

$$\gamma(t|t) = (1-\kappa(t))\,\gamma(t|t-1)$$

Experimental Results

Setup:

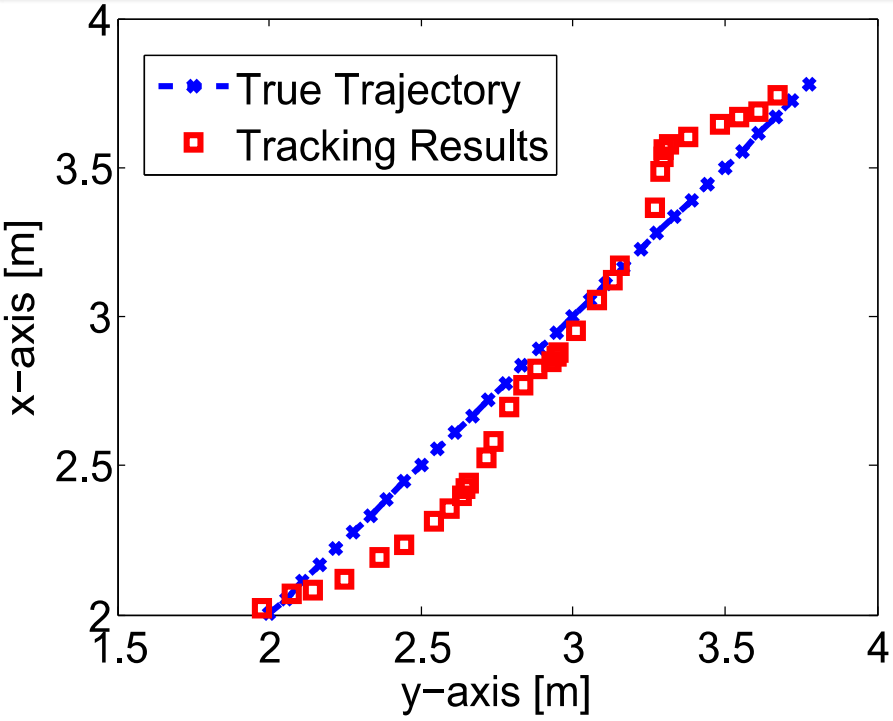
- A $5.2 \times 6.2 \times 3$m room with $T_{60} = 300$ms
- $M = 4$ nodes with 0.2m distance between microphones
- **Region of interest:** a 2×2m square region
- **Training:** 36 samples (0.4m resolution)

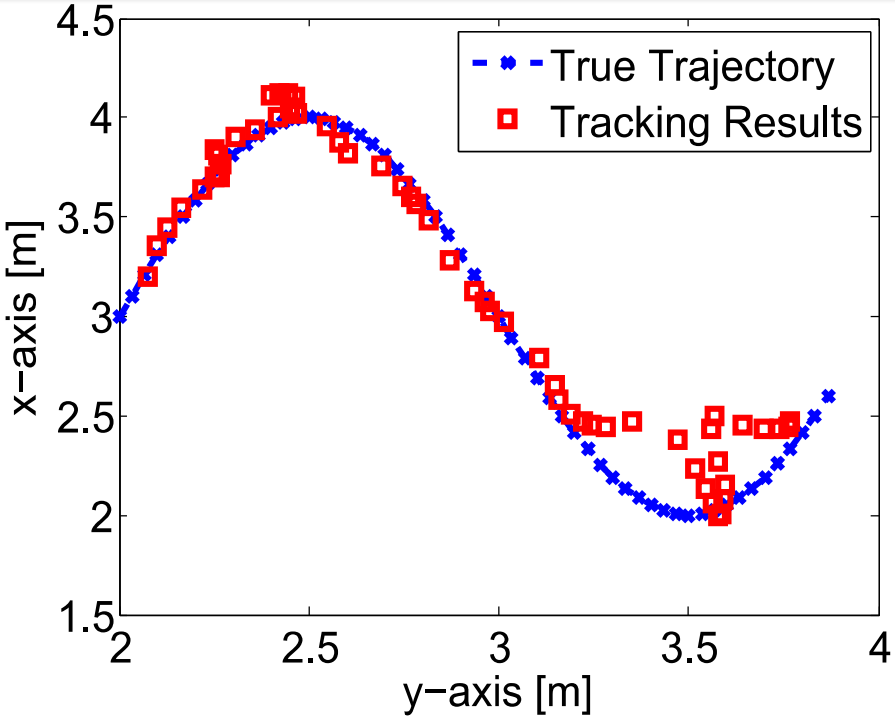


Results

Test:

- **Trajectories:** straight line (for 3s) and sinusoidal movement (for 5s).
- **Velocity:** approximately 1m/s





RMSE: 13cm for straight line and 17cm for sinusoidal movement.

Back to Conclusions

References

[Adavanne et al., 2018] Adavanne, S., Politis, A., and Virtanen, T. (2018).
Direction of arrival estimation for multiple sound sources using convolutional recurrent neural network.
In *Proc. of European Signal Processing Conference (EUSIPCO)*, pages 1462–1466.

[Allen and Berkley, 1979] Allen, J. B. and Berkley, D. A. (1979).
Image method for efficiently simulating small-room acoustics.
The Journal of the Acoustical Society of America, 65(4):943–950.

[Aronszajn, 1950] Aronszajn, N. (1950).
Theory of reproducing kernels.
Transactions of the American mathematical society, 68(3):337–404.

[Ban et al., 2019] Ban, Y., Alameda-Pineda, X., Evers, C., and Horaud, R. (2019).
Tracking multiple audio sources with the von Mises distribution and variational EM.
IEEE Signal Processing Letters, 26(6):798–802.

[Belkin et al., 2006] Belkin, M., Niyogi, P., and Sindhwani, V. (2006).
Manifold regularization: A geometric framework for learning from labeled and unlabeled examples.
Journal of machine learning research, 7(Nov):2399–2434.

[Benesty, 2000] Benesty, J. (2000).
Adaptive eigenvalue decomposition algorithm for passive acoustic source localization.
The Journal of the Acoustical Society of America, 107(1):384–391.

[Berlinet and Thomas-Agnan, 2011] Berlinet, A. and Thomas-Agnan, C. (2011).
Reproducing kernel Hilbert spaces in probability and statistics.
Springer Science & Business Media.

[Brandstein et al., 1997] Brandstein, M. S., Adcock, J. E., and Silverman, H. F. (1997).
A closed-form location estimator for use with room environment microphone arrays.
IEEE transactions on Speech and Audio Processing, 5(1):45–50.

References (cont.)

[Brendel et al., 2018] Brendel, A., Gannot, S., and Kellermann, W. (2018).
Localization of multiple simultaneously active speakers in an acoustic sensor network.
In *IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Sheffield, United Kingdom (Great Britain).

[Brendel and Kellermann, 2019] Brendel, A. and Kellermann, W. (2019).
Distributed source localization in acoustic sensor networks using the coherent-to-diffuse power ratio.
IEEE Journal of Selected Topics in Signal Processing, 13(1):61–75.

[Bross and Gannot, 2020] Bross, A. and Gannot, S. (2020).
Multiple speaker localization using mixture of Gaussian model with manifold-based centroids.
In *28th European Signal Processing Conference (EUSIPCO)*, Amsterdam, The Netherlands.
submitted.

[Chakrabarty and Habets, 2019] Chakrabarty, S. and Habets, E. A. (2019).
Multi-speaker doa estimation using deep convolutional networks trained with noise signals.
IEEE Journal of Selected Topics in Signal Processing.

[Coifman and Lafon, 2006] Coifman, R. R. and Lafon, S. (2006).
Diffusion maps.
Applied and Computational Harmonic Analysis, 21(1):5–30.

[Dal Degan and Prati, 1988] Dal Degan, N. and Prati, C. (1988).
Acoustic noise analysis and speech enhancement techniques for mobile radio applications.
Signal Processing, 15(1):43–56.

[Deleforge et al., 2013] Deleforge, A., Forbes, F., and Horaud, R. (2013).
Variational EM for binaural sound-source separation and localization.
In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 76–80.

References (cont.)

[Deleforge et al., 2015] Deleforge, A., Forbes, F., and Horaud, R. (2015).
Acoustic space learning for sound-source separation and localization on binaural manifolds.
International journal of neural systems, 25(1).

[Deleforge and Horaud, 2012] Deleforge, A. and Horaud, R. (2012).
2D sound-source localization on the binaural manifold.
In *Proc. of IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6, Santander, Spain.

[DiBiase et al., 2001] DiBiase, J. H., Silverman, H. F., and Brandstein, M. S. (2001).
Robust localization in reverberant rooms.
In *Microphone Arrays*, pages 157–180. Springer.

[Do et al., 2007] Do, H., Silverman, H. F., and Yu, Y. (2007).
A real-time SRP-PHAT source location implementation using stochastic region contraction (SRC) on a large-aperture microphone array.
In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 1, pages 121–124.

[Doclo and Moonen, 2003] Doclo, S. and Moonen, M. (2003).
Robust adaptive time delay estimation for speaker localization in noisy and reverberant acoustic environments.
EURASIP Journal on Applied Signal Processing, 2003:1110–1124.

[Dorfan and Gannot, 2015] Dorfan, Y. and Gannot, S. (2015).
Tree-based recursive expectation-maximization algorithm for localization of acoustic sources.
IEEE/ACM Transactions on Audio, Speech and Language Processing, 23(10):1692–1703.

[Dorfan et al., 2018] Dorfan, Y., Plinge, A., Hazan, G., and Gannot, S. (2018).
Distributed expectation-maximization algorithm for speaker localization in reverberant environments.
IEEE/ACM Transactions on Audio, Speech and Language Processing, 26(3):682–695.

References (cont.)

[Dorfan et al., 2016] Dorfan, Y., Schwartz, O., Schwartz, B., Habets, E. A., and Gannot, S. (2016).
Multiple DOA estimation and blind source separation using expectation-maximization algorithm.
In *International conference on the science of electrical engineering (ICSEE)*, Eilat, Israel.

[Dvorkind and Gannot, 2005] Dvorkind, T. and Gannot, S. (2005).
Time difference of arrival estimation of speech source in a noisy and reverberant environment.
Signal Processing, 85(1):177–204.

[Evers and Naylor, 2017] Evers, C. and Naylor, P. A. (2017).
Optimized self-localization for slam in dynamic scenes using probability hypothesis density filters.
IEEE Transactions on Signal Processing, 66(4):863–878.

[Faubel et al., 2009] Faubel, F., McDonough, J., and Klakow, D. (2009).
The split and merge unscented gaussian mixture filter.
IEEE Signal Processing Letters, 16(9):786–789.

[Gannot et al., 2001] Gannot, S., Burshtein, D., and Weinstein, E. (2001).
Signal enhancement using beamforming and nonstationarity with applications to speech.
IEEE Transactions on Signal Processing, 49(8):1614–1626.

[Gannot and Dvorkind, 2006] Gannot, S. and Dvorkind, T. G. (2006).
Microphone array speaker localizers using spatial-temporal information.
EURASIP Journal on Advances in Signal Processing, 2006(1):1–17.

[Habets and Gannot, 2007] Habets, E. and Gannot, S. (2007).
Generating sensor signals in isotropic noise fields.
The Journal of the Acoustical Society of America, 122:3464–3470.

[Hadad and Gannot, 2018] Hadad, E. and Gannot, S. (2018).
Multi-speaker direction of arrival estimation using SRP-PHAT algorithm with a weighted histogram.
In *IEEE International Conference on the Science of Electrical Engineering in Israel (ICSEE)*.

References (cont.)

[Hu et al., 2020] Hu, Y., Samarasinghe, P., Abhayapala, T., and Gannot, S. (2020).
Unsupervised multiple source localization using relative harmonic coefficient.
In *IEEE International Conference on Audio and Acoustic Signal Processing (ICASSP)*, Barcelona, Spain.

[Hu et al., 2019] Hu, Y., Samarasinghe, P. N., and Abhayapala, T. D. (2019).
Sound source localization using relative harmonic coefficients in modal domain.
In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, USA.

[Huang et al., 2000] Huang, Y., Benesty, J., and Elko, G. W. (2000).
Passive acoustic source localization for video camera steering.
In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 909–912.

[Huang et al., 2001] Huang, Y., Benesty, J., Elko, G. W., and Mersereati, R. M. (2001).
Real-time passive source localization: A practical linear-correction least-squares approach.
IEEE transactions on Speech and Audio Processing, 9(8):943–956.

[Jot et al., 1997] Jot, J.-M., Cerveau, L., and Warusfel, O. (1997).
Analysis and synthesis of room reverberation based on a statistical time-frequency model.
In *Audio Engineering Society Convention 103*. Audio Engineering Society.

[Klee et al., 2006] Klee, U., Gehrig, T., and McDonough, J. (2006).
Kalman filters for time delay of arrival-based source localization.
EURASIP Journal on Applied Signal Processing, 2006:167–167.

[Knapp and Carter, 1976] Knapp, C. and Carter, G. (1976).
The generalized correlation method for estimation of time delay.
IEEE Transactions on Acoustics, Speech, and Signal Processing, 24(4):320–327.

[Laufer-Goldshtein et al., 2018a] Laufer-Goldshtein, B., Talmon, R., Cohen, I., and Gannot, S. (2018a).
Multi-view source localization based on power ratios.
In *IEEE International Conference on Audio and Acoustic Signal Processing (ICASSP)*, Calgary, Alberta, Canada.

References (cont.)

[Laufer-Goldshtein et al., 2013] Laufer-Goldshtein, B., Talmon, R., and Gannot, S. (2013).
Relative transfer function modeling for supervised source localization.
In *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 1–4.

[Laufer-Goldshtein et al., 2015] Laufer-Goldshtein, B., Talmon, R., and Gannot, S. (2015).
Study on manifolds of acoustic responses.
In *Proc. of Interntional Conference on Latent Variable Analysis and Signal Seperation (LVA/ICA)*, pages 203–210.

[Laufer-Goldshtein et al., 2016a] Laufer-Goldshtein, B., Talmon, R., and Gannot, S. (2016a).
Manifold-based Bayesian inference for semi-supervised source localization.
In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6335–6339.

[Laufer-Goldshtein et al., 2016b] Laufer-Goldshtein, B., Talmon, R., and Gannot, S. (2016b).
Semi-supervised sound source localization based on manifold regularization.
IEEE Transactions on Audio, Speech, and Language Processing, 24(8):1393–1407.

[Laufer-Goldshtein et al., 2017a] Laufer-Goldshtein, B., Talmon, R., and Gannot, S. (2017a).
Semi-supervised source localization on multiple-manifolds with distributed microphones.
IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25(7):1477–1491.

[Laufer-Goldshtein et al., 2017b] Laufer-Goldshtein, B., Talmon, R., and Gannot, S. (2017b).
Speaker tracking on multiple-manifolds with distributed microphones.
In *Proc. International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*, pages 59–67.

[Laufer-Goldshtein et al., 2018b] Laufer-Goldshtein, B., Talmon, R., and Gannot, S. (2018b).
A hybrid approach for speaker tracking based on TDOA and data-driven models.
IEEE/ACM Transactions on Audio, Speech and Language Processing, 26(4):725–735.

[Laufer-Goldshtein et al., 2018c] Laufer-Goldshtein, B., Talmon, R., and Gannot, S. (2018c).
Source counting and separation based on simplex analysis.
IEEE Transactions on Signal Processing, 66(24):6458–6473.

References (cont.)

[Lehmann and Williamson, 2006] Lehmann, E. A. and Williamson, R. C. (2006).
Particle filter design using importance sampling for acoustic source localisation and tracking in reverberant environments.
EURASIP Journal on Applied Signal Processing, 2006:168–168.

[Levy et al., 2011] Levy, A., Gannot, S., and Habets, E. A. (2011).
Multiple-hypothesis extended particle filter for acoustic source localization in reverberant environments.
IEEE Transactions on Audio, Speech, and Language Processing, 19(6):1540–1555.

[Löllmann et al., 2018] Löllmann, H. W., Evers, C., Schmidt, A., Mellmann, H., Barfuss, H., Naylor, P. A., and Kellermann, W. (2018).
The LOCATA challenge data corpus for acoustic source localization and tracking.
In *IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pages 410–414.

[Madhu and Martin, 2018] Madhu, N. and Martin, R. (2018).
Source number estimation for multi-speaker localisation and tracking.
In *Workshop on speech processing for voice, speech and hearing disorders (WSPD)*.

[Madhu et al., 2008] Madhu, N., Martin, R., Heute, U., and Antweiler, C. (2008).
Acoustic source localization with microphone arrays.
Advances in Digital Speech Transmission, pages 135–170.

[Mandel et al., 2007] Mandel, M. I., Ellis, D. P., and Jebara, T. (2007).
An EM algorithm for localizing multiple sound sources in reverberant environments.
In *Proc. of Advances in neural information processing systems*, pages 953–960.

[Mandel et al., 2010] Mandel, M. I., Weiss, R. J., and Ellis, D. P. W. (2010).
Model-based expectation-maximization source separation and localization.
IEEE Transactions on Audio, Speech, and Language Processing, 18(2):382–394.

References (cont.)

[Markovich et al., 2009] Markovich, S., Gannot, S., and Cohen, I. (2009).
Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals.
IEEE Transactions on Audio, Speech, and Language Processing, 17(6):1071–1086.

[Markovich-Golan et al., 2018] Markovich-Golan, S., Gannot, S., and Kellermann, W. (2018).
Performance analysis of the Covariance-Whitening and the Covariance-Subtraction methods for estimating the relative transfer function.
In *The 26th European Signal Processing Conference (EUSIPCO)*, Rome, Italy.

[May et al., 2011] May, T., van de Par, S., and Kohlrausch, A. (2011).
A probabilistic model for robust localization based on a binaural auditory front-end.
IEEE Transactions on Audio, Speech, and Language Processing, 19(1):1–13.

[Opochinsky et al., 2019] Opochinsky, R., Laufer, B., Gannot, S., and Chechik, G. (2019).
Deep Ranking-Based sound source localization.
In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, USA.

[Pearson, 1901] Pearson, K. (1901).
Principal components analysis.
The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 6(2):559.

[Perotin et al., 2019] Perotin, L., Serizel, R., Vincent, E., and Guerin, A. (2019).
Crnn-based multiple doa estimation using acoustic intensity features for ambisonics recordings.
IEEE Journal of Selected Topics in Signal Processing.

[Peterson, 1986] Peterson, P. M. (1986).
Simulating the response of multiple microphones to a single acoustic source in a reverberant room.
The Journal of the Acoustical Society of America, 80(5):1527–1529.

References (cont.)

[Polack, 1993] Polack, J.-D. (1993).
Playing billiards in the concert hall: The mathematical foundations of geometrical room acoustics.
Applied Acoustics, 38(2):235–244.

[Roman et al., 2003] Roman, N., Wang, D., and Brown, G. J. (2003).
Speech segregation based on sound localization.
The Journal of the Acoustical Society of America, 114(4):2236–2252.

[Roweis and Saul, 2000] Roweis, S. T. and Saul, L. K. (2000).
Nonlinear dimensionality reduction by locally linear embedding.
science, 290(5500):2323–2326.

[Roy and Kailath, 1989] Roy, R. and Kailath, T. (1989).
ESPRIT-estimation of signal parameters via rotational invariance techniques.
IEEE Transactions on Acoustics, Speech and Signal Processing, 37(7):984–995.

[Schau and Robinson, 1987] Schau, H. and Robinson, A. (1987).
Passive source localization employing intersecting spherical surfaces from time-of-arrival differences.
IEEE Transactions on Acoustics, Speech, and Signal Processing, 35(8):1223–1225.

[Schmidt, 1986] Schmidt, R. O. (1986).
Multiple emitter location and signal parameter estimation.
IEEE Transactions on Antennas and Propagation, 34(3):276–280.

[Schölkopf et al., 2001] Schölkopf, B., Herbrich, R., and Smola, A. J. (2001).
A generalized representer theorem.
In *Proc. of The 14th Annual Conference on Computational learning theory (COLT)*, pages 416–426. Springer.

[Schroeder, 1996] Schroeder, M. R. (1996).
The "schroeder frequency" revisited.
The Journal of the Acoustical Society of America, 99(5):3240–3241.

References (cont.)

[Schwartz et al., 2017] Schwartz, O., Dorfan, Y., Taseska, M., Habets, E. A., and Gannot, S. (2017).
DOA estimation in noisy environment with unknown noise power using the EM algorithm.
In *The 5th Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, San-Francisco, CA, USA.

[Schwartz and Gannot, 2013] Schwartz, O. and Gannot, S. (2013).
Speaker tracking using recursive EM algorithms.
IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(2):392–402.

[Sindhwani et al., 2007] Sindhwani, V., Chu, W., and Keerthi, S. S. (2007).
Semi-supervised Gaussian process classifiers.
In *Proc. of International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1059–1064.

[Smith and Abel, 1987] Smith, J. and Abel, J. (1987).
Closed-form least-squares source location estimation from range-difference measurements.
IEEE Transactions on Acoustics, Speech, and Signal Processing, 35(12):1661–1669.

[Soussana and Gannot, 2019] Soussana, Y. and Gannot, S. (2019).
Variational inference for DOA estimation in reverberant conditions.
In *27th European Signal Processing Conference (EUSIPCO)*, A Coruña, Spain.

[Takeda and Komatani, 2016] Takeda, R. and Komatani, K. (2016).
Sound source localization based on deep neural networks with directional activate function exploiting phase information.
In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 405–409.

[Talmon and Gannot, 2013] Talmon, R. and Gannot, S. (2013).
Relative transfer function identification on manifolds for supervised gsc beamformers.
In *Proc. of 21st European Signal Processing Conference (EUSIPCO)*, pages 1–5.

References (cont.)

[Talmon et al., 2011] Talmon, R., Kushnir, D., Coifman, R. R., Cohen, I., and Gannot, S. (2011).
Parametrization of linear systems using diffusion kernels.
IEEE Transactions on Signal Processing, 60(3):1159–1173.

[Tenenbaum et al., 2000] Tenenbaum, J. B., De Silva, V., and Langford, J. C. (2000).
A global geometric framework for nonlinear dimensionality reduction.
science, 290(5500):2319–2323.

[Teutsch and Kellermann, 2005] Teutsch, H. and Kellermann, W. (2005).
EB-ESPRIT: 2D localization of multiple wideband acoustic sources using eigen-beams.
In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 89–92.

[Thiergart et al., 2014] Thiergart, O., Taseska, M., and Habets, E. A. P. (2014).
An informed parametric spatial filter based on instantaneous direction-of-arrival estimates.
IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(12):2182–2196.

[Traa and Smaragdis, 2014] Traa, J. and Smaragdis, P. (2014).
Multichannel source separation and tracking with ransac and directional statistics.
IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(12):2233–2243.

[Ward et al., 2003] Ward, D. B., Lehmann, E. A., and Williamson, R. C. (2003).
Particle filtering algorithms for tracking an acoustic source in a reverberant environment.
IEEE Transactions on speech and audio processing, 11(6):826–836.

[Weisberg et al., 2019] Weisberg, K., Gannot, S., and Schwartz, O. (2019).
An online multiple-speaker DOA tracking using the Cappé-Moulines recursive expectation-maximization algorithm.
In *IEEE International Conference on Audio and Acoustic Signal Processing (ICASSP)*, pages 656–660.

[Xiao et al., 2015] Xiao, X., Zhao, S., Zhong, X., Jones, D. L., Chng, E. S., and Li, H. (2015).
A learning-based approach to direction of arrival estimation in noisy and reverberant environments.
In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 76–80.

References (cont.)

[Yao et al., 2002] Yao, K., Chen, J. C., and Hudson, R. E. (2002).
Maximum-likelihood acoustic source localization: experimental results.
In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 2949–2952.

[Zhong and Hopgood, 2008] Zhong, X. and Hopgood, J. R. (2008).
Nonconcurrent multiple speakers tracking based on extended Kalman particle filter.
In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 293–296.