

Microphone Array Processing: Far-field modeling, TDOA / SRP, DOA, Beamforming

Fabio Antonacci

Feb. 20, 2026

Outline

Applications and problem statement

Problem setup: geometry and far-field propagation

Time-domain viewpoint: TDOAs and SRP

Frequency-domain viewpoint: DOAs and phase differences

Beamforming from the narrowband data model

Summary and conceptual map

Applications of microphone arrays (overview)

Microphone arrays are used to:

- ▶ Improve speech capture and intelligibility (teleconferencing, smart speakers).
- ▶ Perform spatial filtering / beamforming for hearing aids and assistive devices.
- ▶ Localize sound sources (speaker tracking, surveillance, wildlife monitoring).
- ▶ Enable acoustic imaging and room mapping (audio-based RADAR-like systems).
- ▶ Support human-robot interaction and navigation (robot audition).
- ▶ Enhance speech recognition in noisy/reverberant environments.

These diverse applications share a common need: estimating where sound comes from and separating it from interference.

From applications to localization

Why localization matters

- ▶ Knowing the source direction enables steering microphones (beamforming) to improve SNR.
- ▶ Localization enables multimodal systems (audio+video) to focus attention and control actuators.
- ▶ Tracking moving sources requires low-latency, robust direction-of-arrival (DOA) estimates.

This talk: model acoustic propagation for arrays (far-field), derive TDOAs and phase differences, present SRP and SRP-PHAT (time-domain energy scanning) and narrowband beamforming (conventional and Capon/MVDR).

Array geometry and far-field assumption

Key ideas

- ▶ Consider a uniform linear array (ULA) of M microphones with inter-element spacing d .
- ▶ Far-field (plane-wave) approximation: source is sufficiently distant so wavefronts are planar.
- ▶ Source direction parameterised by angle θ (azimuth).

Geometry / notation

$$r_m = md, \quad m = 0, \dots, M - 1$$

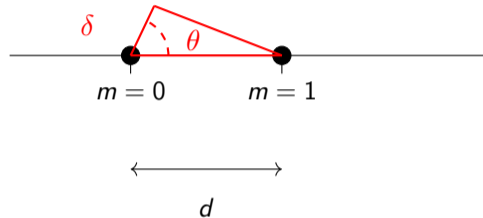
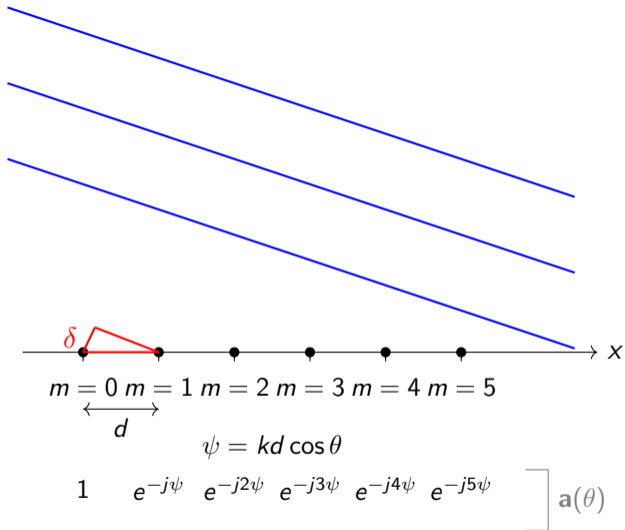
Unit direction vector in 2D:

$$\mathbf{u}(\theta) = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}$$

Time delay to microphone m (relative origin):

$$\tau_m(\theta) = \frac{md \cos \theta}{c}$$

Plane wave on ULA)



$$\delta = d \cos(\theta)$$

$$\tau = \frac{\delta}{c} = \frac{d \cos(\theta)}{c}$$

Signal model and TDOA

Time-domain model (far-field)

$$x_m(t) = s(t - \tau_m(\theta)) + v_m(t)$$

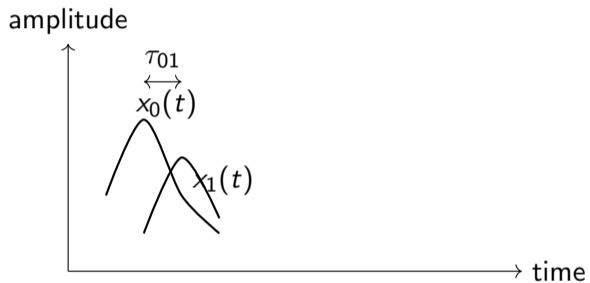
where $s(t)$ is the source signal and $v_m(t)$ is additive sensor noise.

Time difference of arrival (TDOA) between mic m and n :

$$\tau_{m,n}(\theta) = \tau_m(\theta) - \tau_n(\theta) = (m - n) \frac{d \cos \theta}{c}$$

This motivates delay-and-sum style processing and energy-based scanning methods.

Diagram: TDOA between channels



Steered Response Power (SRP)

Delay-and-sum steered output for direction θ :

$$y(t, \theta) = \sum_{m=0}^{M-1} x_m(t + \tau_m(\theta))$$

(note sign convention; equivalently shift by $-\tau_m$ depending on notation)

Steered Response Power:

$$P_{\text{SRP}}(\theta) = \int |y(t, \theta)|^2 dt$$

Estimate DOA by maximizing the SRP:

$$\hat{\theta} = \arg \max_{\theta} P_{\text{SRP}}(\theta)$$

Practical SRP implementations use framing and precomputed TDOA lookup tables for efficiency.

SRP-PHAT (GCC-PHAT weighting)

Motivation: In reverberant or noisy environments, pairwise GCC-PHAT emphasizes time-delay peaks (phase-only weighting) and reduces the influence of source spectrum.

GCC-PHAT between sensors i and j :

$$R_{ij}^{\text{PHAT}}(\tau) = \mathcal{F}^{-1} \left\{ \frac{X_i(f) X_j^*(f)}{|X_i(f) X_j^*(f)|} \right\} (\tau)$$

Here $X_k(f)$ is the Fourier transform of $x_k(t)$; PHAT normalizes the cross-spectrum magnitude to 1 (phase-only).

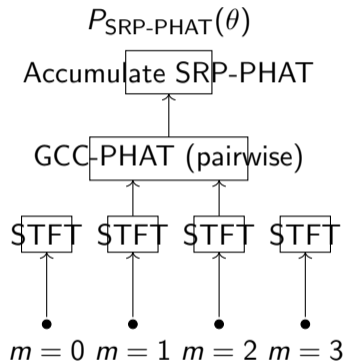
SRP-PHAT scan:

$$P_{\text{SRP-PHAT}}(\theta) = \sum_{i < j} w_{ij} R_{ij}^{\text{PHAT}}(\tau_{ij}(\theta))$$

where $\tau_{ij}(\theta)$ is the TDOA hypothesis for the microphone pair and w_{ij} are optional weights.

Estimate: $\hat{\theta} = \arg \max_{\theta} P_{\text{SRP-PHAT}}(\theta)$.

SRP-PHAT pipeline (diagram)



Monochromatic plane wave model

Assume a narrowband (monochromatic) source:

$$s(t) = Ae^{j\omega t}$$

Received signal:

$$x_m(t) = Ae^{j\omega(t-\tau_m(\theta))}$$

Phase shift:

$$\phi_m(\theta) = -\omega\tau_m(\theta)$$

Phase difference and DOA

For two microphones separated by distance d (along array axis) and a plane wave:

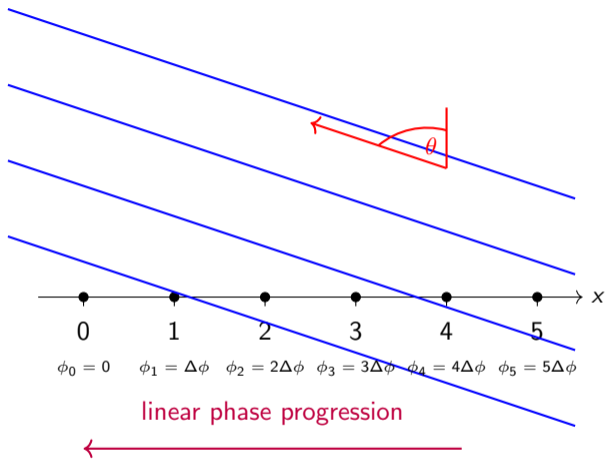
$$\Delta\phi = \omega \frac{d}{c} \cos \theta$$

Hence:

$$\theta = \arccos \left(\frac{c \Delta\phi}{\omega d} \right)$$

Note: phase is observed modulo 2π , so unwrapping / ambiguity arises.

Diagram: phase progression across array



ULA (spacing d), mic index $m = 0, 1, \dots, 5$

Where spatial aliasing comes from (sampling a spatial sinusoid)

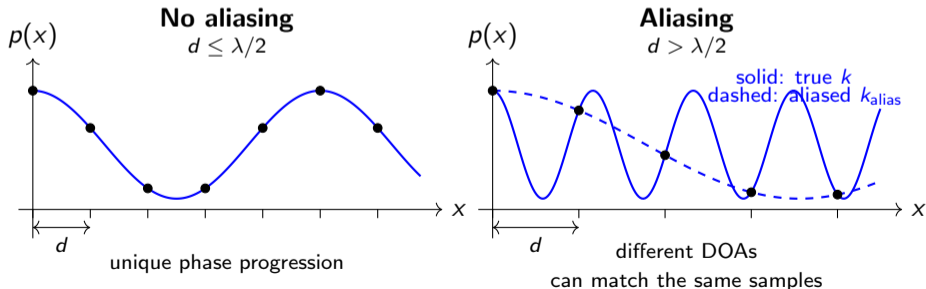
For a narrowband plane wave, the pressure along the array axis is a spatial sinusoid:

$$p(x) = A \cos(kx + \phi_0), \quad k = \frac{2\pi}{\lambda}.$$

A ULA samples it at $x_m = md$. If d is too large, another wavenumber

$$k_{\text{alias}} = k - \frac{2\pi}{d}$$

produces the *same* samples because $k_{\text{alias}} md = k md - 2\pi m$.



Spatial aliasing

Phase ambiguity:

$$\Delta\phi \equiv \Delta\phi + 2\pi k$$

To avoid spatial aliasing we require

$$d \leq \frac{\lambda}{2} \quad (\text{Nyquist spatial sampling})$$

where $\lambda = \frac{c}{f}$. Thus maximum unambiguous frequency for spacing d :

$$f_{\max} = \frac{c}{2d}$$

If $d > \lambda/2$ multiple DOAs produce the same phase pattern.

Narrowband array data model

For a narrowband (monochromatic) source at angular frequency ω :

$$\mathbf{x}(t) = \mathbf{a}(\theta) s(t) + \mathbf{v}(t)$$

Propagation delay for a ULA:

$$\tau_m(\theta) = \frac{md \cos \theta}{c}$$

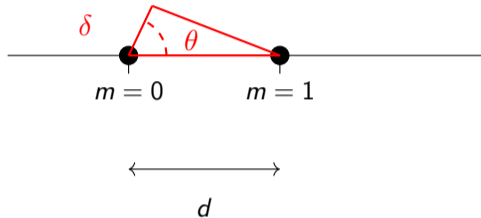
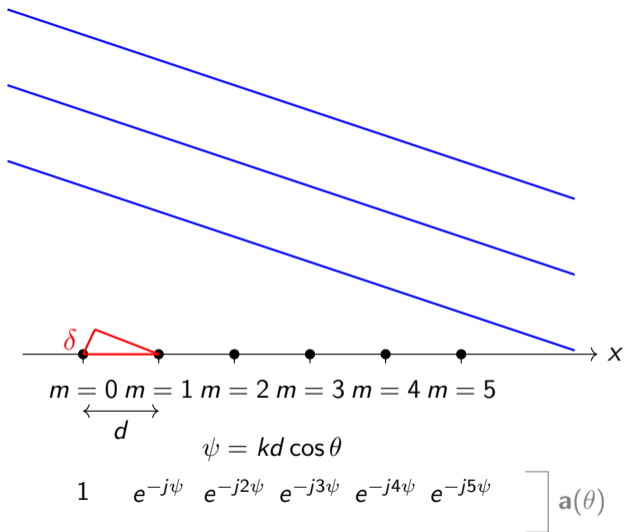
Wavenumber definition:

$$k = \frac{\omega}{c} = \frac{2\pi}{\lambda}$$

Steering vector (ULA):

$$\mathbf{a}(\theta) = \begin{bmatrix} 1 \\ e^{-jkd \cos \theta} \\ e^{-j2kd \cos \theta} \\ \vdots \\ e^{-j(M-1)kd \cos \theta} \end{bmatrix}$$

Steering vector on a ULA (phase pattern)



$$\delta = d \cos(\theta)$$

$$\tau = \frac{\delta}{c} = \frac{d \cos(\theta)}{c}$$

Conventional (Delay-and-Sum) beamformer

$$y(t) = \mathbf{w}^H \mathbf{x}(t)$$

Simple steering (conventional beamformer) chooses weights proportional to the steering vector:

$$\mathbf{w}_{\text{CBF}} = \frac{1}{M} \mathbf{a}(\theta)$$

$$B(\theta) = \mathbf{w}^H \mathbf{a}(\theta)$$

Capon (MVDR) beamformer

Optimization

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_{xx} \mathbf{w} \quad \text{s.t.} \quad \mathbf{w}^H \mathbf{a}(\theta) = 1$$

Solution

$$\mathbf{w}_{\text{MVDR}} = \frac{\mathbf{R}_{xx}^{-1} \mathbf{a}(\theta)}{\mathbf{a}^H(\theta) \mathbf{R}_{xx}^{-1} \mathbf{a}(\theta)}$$

Capon adapts to the interference covariance, producing nulls toward interferers while keeping unity gain in look direction.

From spatial response to spatial spectrum

Narrowband beamformer output:

$$y(t) = \mathbf{w}^H \mathbf{x}(t)$$

Spatial (steering) response to a plane wave from angle θ :

$$B(\theta; \mathbf{w}) \triangleq \mathbf{w}^H \mathbf{a}(\theta)$$

Steered output power (“spatial spectrum” associated to weights \mathbf{w}):

$$P_y(\theta) \triangleq \mathbb{E}\{|y(t)|^2\} = \mathbf{w}^H(\theta) \mathbf{R}_{xx} \mathbf{w}(\theta)$$

where $\mathbf{R}_{xx} = \mathbb{E}\{\mathbf{x}(t)\mathbf{x}^H(t)\}$.

In practice, \mathbf{R}_{xx} is estimated from data (sample covariance) and $P_y(\theta)$ is evaluated on an angle grid.

Pseudospectra for DAS and MVDR (Capon)

Delay-and-sum (DAS): choose weights steered to θ

$$\mathbf{w}_{\text{DAS}}(\theta) = \frac{1}{M} \mathbf{a}(\theta)$$

DAS spatial spectrum (steered power):

$$P_{\text{DAS}}(\theta) = \mathbf{w}_{\text{DAS}}^H(\theta) \mathbf{R}_{xx} \mathbf{w}_{\text{DAS}}(\theta) = \frac{1}{M^2} \mathbf{a}^H(\theta) \mathbf{R}_{xx} \mathbf{a}(\theta)$$

MVDR / Capon: minimize output power while enforcing unity gain in look direction

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_{xx} \mathbf{w} \quad \text{s.t.} \quad \mathbf{w}^H \mathbf{a}(\theta) = 1$$

Solution:

$$\mathbf{w}_{\text{MVDR}}(\theta) = \frac{\mathbf{R}_{xx}^{-1} \mathbf{a}(\theta)}{\mathbf{a}^H(\theta) \mathbf{R}_{xx}^{-1} \mathbf{a}(\theta)}$$

Substitute into $P_y(\theta) = \mathbf{w}^H \mathbf{R}_{xx} \mathbf{w}$:

$$P_{\text{MVDR}}(\theta) = \frac{1}{\mathbf{a}^H(\theta) \mathbf{R}_{xx}^{-1} \mathbf{a}(\theta)}$$

Conceptual map

- ▶ **Time-domain branch:** TDOA \rightarrow delay-and-sum \rightarrow SRP (energy scanning) \rightarrow SRP-PHAT (robust variant).
- ▶ **Frequency-domain branch:** narrowband phase differences \rightarrow DOA (with ambiguity) \rightarrow beamforming (CBF, MVDR).
- ▶ **Key practical issues:** spatial aliasing, noise and reverberation, covariance estimation, computational cost for scanning.

End

Thank you!