Microphone Arrays
and Time Delay Estimation

Microphone Arrays

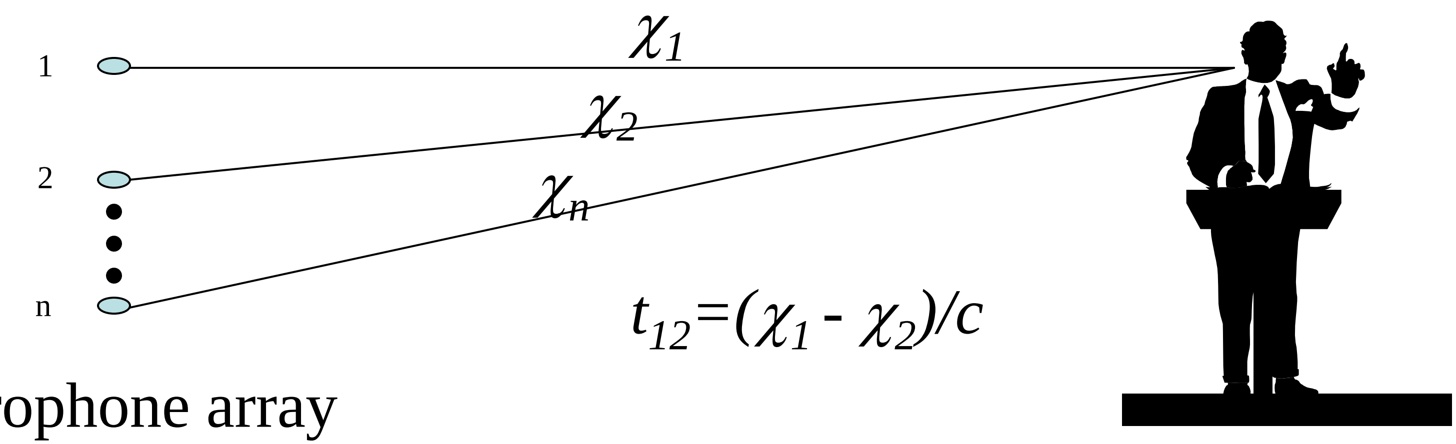
- Goals:
  - Capture sound
  - Capture sound from a particular spatial location
  - Suppress sound from other spatial locations
  - Build a spatial representation for the sound
  - Embed in some applications
- Tools
  - Time delays
    - Fourier analysis, convolution
  - Optimization
  - Statistical independence
  - Level differences

Microphone array

- Spatial distribution of microphones
- Known geometrical locations
  - If locations are unknown a calibration problem needs to be solved
  - Calibration is closely linked to the problem of source localization
- Sounds collected and processed
  - Processing is linked to the application

Time Delays

- Signal from a source arrives at different microphones at times proportional to their distance
- Measuring time differences of arrival one can compute source location and beamform signal
- Classical problem with rich literature.



Microphone array

Correlation

- The correlation *function* between two energy signals, x and y, is the area under the curve (for continuous signals) or the sum of (for discrete signals) their product *as a function of how much* y *is shifted* relative to x.

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t) y^*(t+\tau) dt \qquad R_{xy}[m] = \sum_{n=-\infty}^{\infty} x[n] y^*[n+m]$$

- Recall definition of convolution

$$x(t) * y(t) = \int_{-\infty}^{\infty} x(t-\tau) y(\tau) d\tau \qquad x[n] * y[n] = \sum_{m=-\infty}^{\infty} x[n-m] y[m]$$

- Use convolution to find the correlation of two signals
- For two similar signals that are time-shifted w.r. to each other correlation has a strong peak at the correct delay
- Delays in the digital case are integers

Autocorrelation

- A very important special case of correlation is *autocorrelation*.
  - correlation of a function with a shifted version of *itself*.

$$\mathrm{R}_{\mathrm{xx}}(\tau)=\int_{-\infty}^{\infty}\mathrm{x}(t)\mathrm{x}(t+\tau)dt \quad \text{or} \quad \mathrm{R}_{\mathrm{xx}}[m]=\sum_{n=-\infty}^{\infty}\mathrm{x}[n]\mathrm{x}[n+m]$$

- At a shift, τ or *m*, of zero,

$$\mathrm{R}_{\mathrm{xx}}(0)=\int_{-\infty}^{\infty}\mathrm{x}^2(t)dt \quad \text{or} \quad \mathrm{R}_{\mathrm{xx}}[0]=\sum_{n=-\infty}^{\infty}\mathrm{x}^2[n]$$

  which is the signal energy of the signal.

- Autocorrelation is an even function.

$$\mathrm{R}_{\mathrm{xx}}(\tau)=\mathrm{R}_{\mathrm{xx}}(-\tau) \quad \text{or} \quad \mathrm{R}_{\mathrm{xx}}[m]=\mathrm{R}_{\mathrm{xx}}[-m]$$

- Autocorrelation magnitude can never be larger than at zero shift.

$$\mathrm{R}_{\mathrm{xx}}(0)\geq\left|\mathrm{R}_{\mathrm{xx}}(\tau)\right| \quad \text{or} \quad \mathrm{R}_{\mathrm{xx}}[0]\geq\left|\mathrm{R}_{\mathrm{xx}}[m]\right|$$

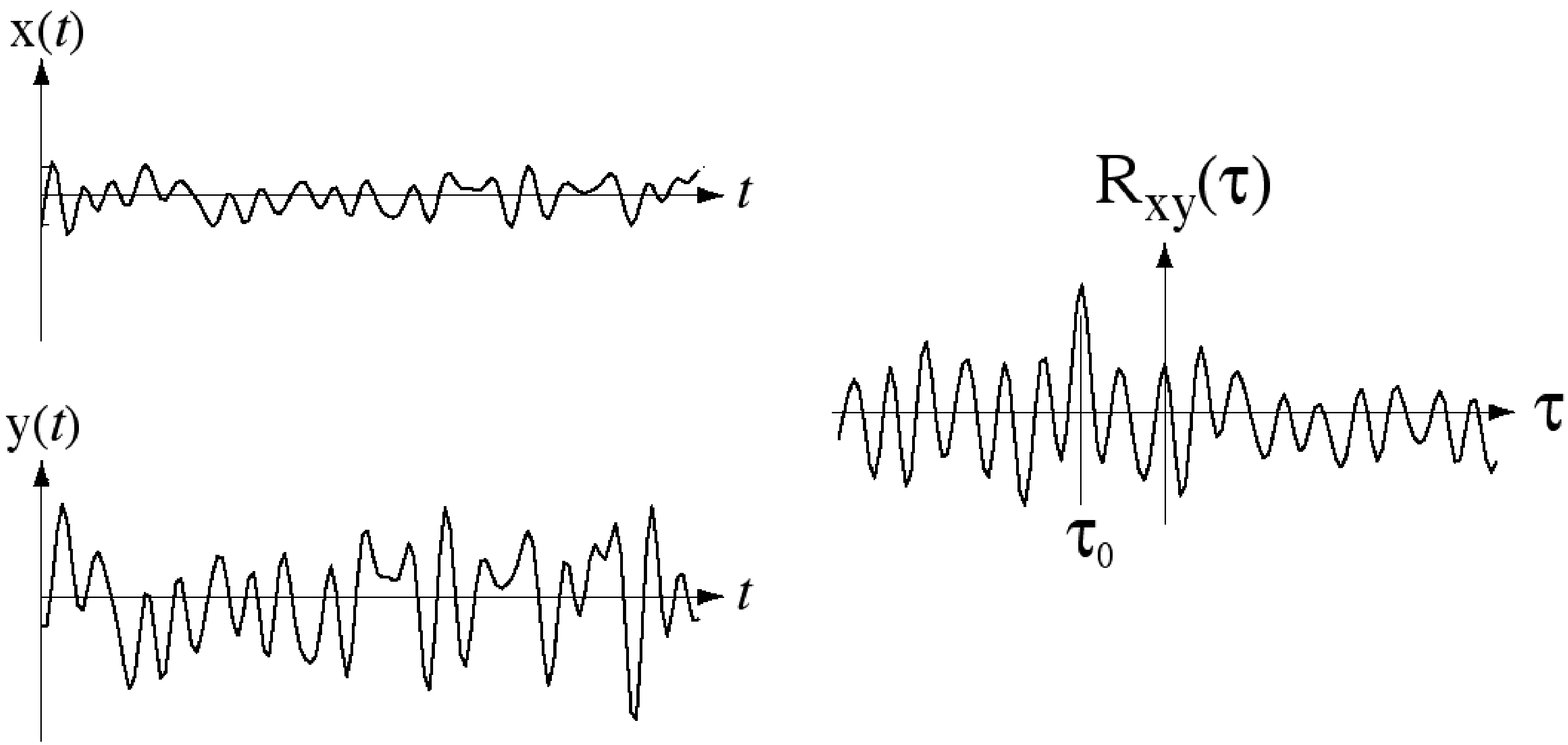
- If a signal is time shifted its autocorrelation does not change.

Matched Filters

- If we seek a signal of a particular type, convolution with a template signal of that type can be used
- E.g., image edge detectors
- Peak finding by convolution with narrow gaussian
- Fourier coefficients can be interpreted as filter response to a filter of shape of the Fourier basis function

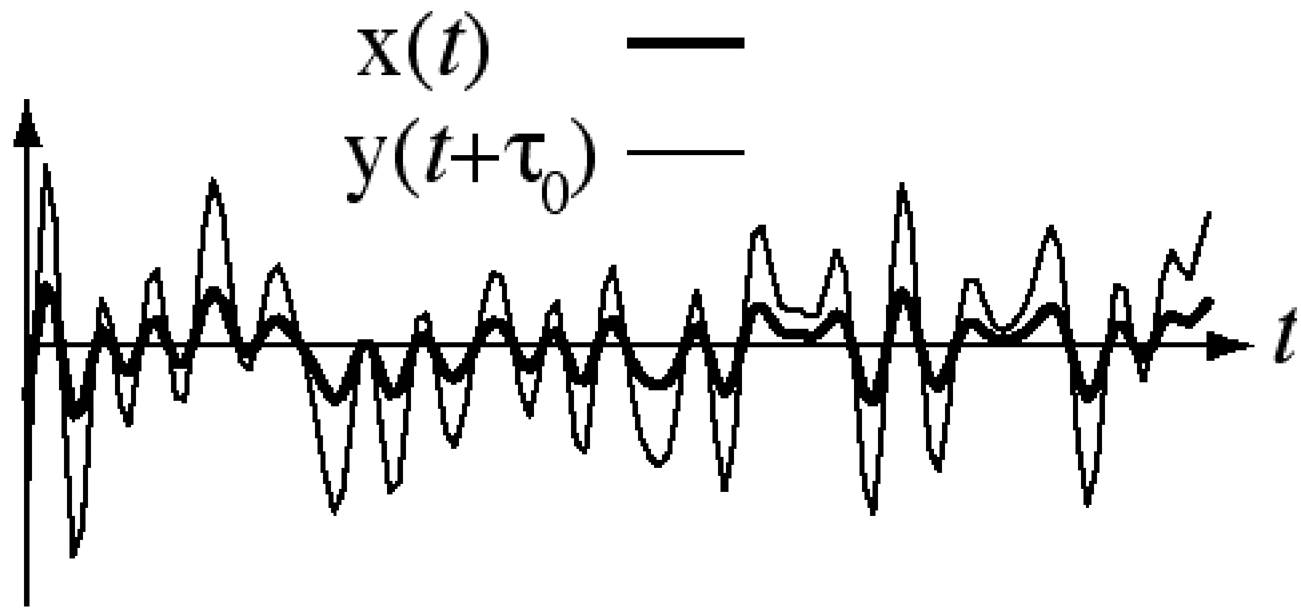
Cross Correlation

Cross correlation is really just "correlation" in the cases in which the two signals being compared are different. The name is commonly used to distinguish it from autocorrelation.



Cross Correlation

A comparison of x and y with y shifted for maximum correlation.



Generalized cross correlation

- Is there someway to enhance the peak, to simplify the delay detection task?
- GCC-PHAT, GCC-ML, …
- Look up in Brandstein’s thesis

Source localization

- Knowing time delays how do we detect sources?

N microphones located at points $\mathbf{m}_i = (x_i, y_i, z_i)$,
Source at $\mathbf{s} = (x_s, y_s, z_s)$.
speed of sound c
Distances between microphone i and source is χ_i,

$$\chi_i = \sqrt{(x_i - x_s)^2 + (y_i - y_s)^2 + (z_i - z_s)^2}$$

Time delays between i and j provide a linear relationship

$$\chi_i - \chi_j = ct_{ij}$$

For N microphones $C(N, 2)$ measurements
$N - 1$ are independent

- 2, 3, 4 microphones

  embed a spherical coordinate system at first microphone
  axis directed towards the second microphone
  source is at (r, θ, ϕ)
  microphone 2 at $(r_2, 0, 0)$
  microphone 3 at $(r_3, \theta_3, 0)$
  microphone 4 at (r_4, θ_4, ϕ_4).

$$r - \sqrt{(r\cos\theta - r_2)^2 + r^2\sin^2\theta} = r - \sqrt{r^2 - r_2^2 - 2rr_2\cos\theta} = ct_{12}$$

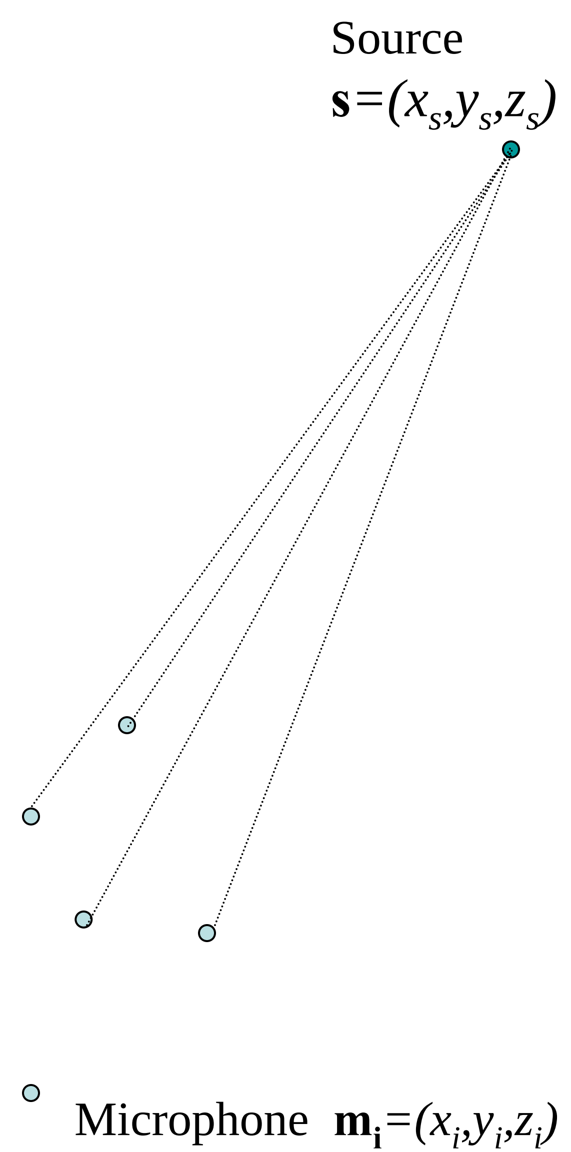
$$r - \sqrt{(r\cos\theta - r_3\cos\theta_3)^2 + (r\sin\theta\cos\phi - r_3\sin\theta_3)^2 + (r\sin\theta\sin\phi)^2} = ct_{13}$$

$$r - \sqrt{r^2 + r_3^2 - 2rr_3\left(\cos\theta\cos\theta_3 + \sin\theta\sin\theta_3\cos\phi\right)} = ct_{14}$$

Previous Solutions

- Closed Form solutions based on least-squares
  - Smith and Abel 1987
  - Schau and Robinson

- Solutions based on particular geometric arrangements of multiple microphones
  - Brandstein and Silverman (1994)
- These solutions are "numerical" as opposed to exact.
- Given time differences and $\mathbf{m_i}$ get $\boldsymbol{s}$



Exact Solution

- Seek to find 3 unknown source coordinates $\mathbf{s}=(x_s,y_s,z_s)$ given microphone coordinates and time differences.
- 4 microphones $\mathbf{m_i}=(x_i,y_i,z_i)$ provide 6 time differences. 3 of them are unique.
- Each time difference equation involves the distances from the source to the microphones.
- If we had a fourth relation we can get a solution for the distances.
- Knowing the distances we can get a solution for the coordinates.

$$\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ u & v & w & 1 \end{bmatrix} \begin{bmatrix} \chi_2 \\ \chi_3 \\ \chi_4 \\ \chi_1 \end{bmatrix} = \begin{bmatrix} ct_{12} \\ ct_{13} \\ ct_{14} \\ A \end{bmatrix}$$

$$\chi_1 = \frac{uct_{12} + vct_{13} + wct_{14} + A}{u + v + w + 1}$$

$$\begin{bmatrix} \chi_2 \\ \chi_1 \\ \chi_1 \end{bmatrix} = \begin{bmatrix} ct_{12} + \chi_1 \\ ct_{13} + \chi_1 \\ ct_{14} + \chi_1 \end{bmatrix}$$

Linear relation

- Does a linear relationship of the form desired exist?
  - **necessary and sufficient condition for a linear relation is that the four microphones lie in a plane.**
  - W.l.o.g. let the microphone coordinates be $(0,0,0)$, $(r_2,0,0)$, $(r_3, \theta_3,0)$ and $(r_4,\theta_4,0)$

$$u = \frac{r_3 r_4 t_{12} \sin(\theta_3 - \theta_4)}{D},$$

$$v = \frac{r_2 r_4 t_{13} \sin(\theta_4)}{D}$$

$$w = \frac{-r_2 r_3 t_{14} \sin(\theta_4)}{D},$$

- Then u,v,w,A are as beside.

$$A = \frac{r_2 r_3 r_4 t_{12}\left(r_4 \sin\theta_3 - r_3 \sin\theta_4 - r_2 \sin(\theta_3 - \theta_4)\right)}{D}$$

$$D = c\left(r_3 r_4 t_{12} \sin(\theta_3 - \theta_4) + r_2\left(r_4 t_{13} \sin\theta_4 - r_3 t_{14} \sin\theta_3\right)\right)$$

- Relation was derived from the planarity constraint

$$\mathbf{s} = [x_s \quad y_s \quad 0]^T = \frac{1}{2}\left(\mathbf{S}^T\mathbf{S}\right)^{-1}\mathbf{S}^T(\boldsymbol{\delta} - 2\chi_1 \mathbf{d})$$

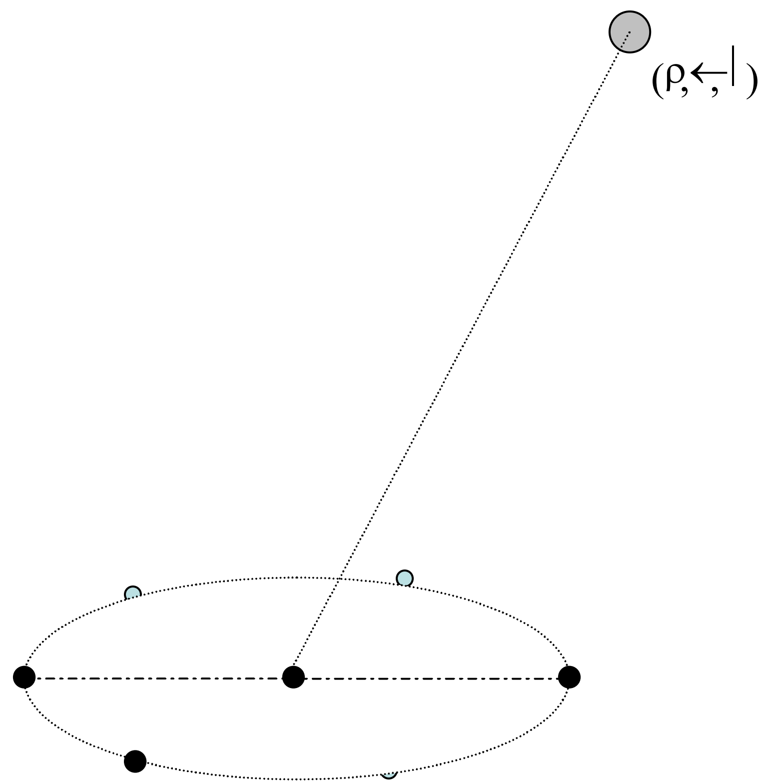
- Knowing the four χ_i can solve for source coordinates using

$$z_s = \sqrt{{\chi_1}^2 - x_s^2 - y_s^2}$$

where $\mathbf{S} = \begin{bmatrix} x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \\ x_4 & y_4 & z_4 \end{bmatrix}$ $\mathbf{d} = \begin{bmatrix} \chi_1 - \chi_2 \\ \chi_1 - \chi_3 \\ \chi_1 - \chi_4 \end{bmatrix}$ $\boldsymbol{\delta} = \begin{bmatrix} r_2^2 - d_1^2 \\ r_3^2 - d_2^2 \\ r_4^2 - d_3^2 \end{bmatrix}$

Special Cases

- 3 microphones in a line
  - In this case we can solve for the range and the coordinate along the line
- 5 microphones in a cross
  - Apply equations for 3 microphones in a line to get range and source coordinates
- Multiple microphones in a plane (e.g. as in a proposed army weapon). Can form ${}^{7}C_{4}$ solutions



In practice

- Time delays are noisy
- Sound may be absent in some frames
- Several strategies
- Apply delay estimation over various bands