ADVANCED AUDIO SIGNAL PROCESSING

Sound source localization

Tuesday 26th September, 2023

Sylvain Argentieri

Modelization of the problem: *from acoustics to signal*

❑ Solution

- *pour **une source ponctuelle et impulsionnelle** :* **fonction de Green**

La solution de l'équation des ondes en $\vec{r}$ à l'instant t, notée $G(\vec{r},t,\overrightarrow{r_0},t_0)$ pour une source ponctuelle et impulsionnelle placée au point $\overrightarrow{r_0}$ et allumée à l'instant t_0, $\delta(\vec{r}-\overrightarrow{r_0})\delta(t-t_0)$, est appelée fonction de Green.
Elle vérifie :

$$\left(\Delta - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right) G(\vec{r},t,\overrightarrow{r_0},t_0) = \delta(\vec{r}-\overrightarrow{r_0})\delta(t-t_0)$$

- *pour **une source quelconque***

Connaître $G(\vec{r},t,\overrightarrow{r_0},t_0)$ permet de connaître le potentiel des vitesses en tout $\vec{r}$, à tout instant t, pour une source quelconque $s(\vec{r},t)$ connue, distribuée sur le domaine $\mathcal{D}$ et allumée aux instants de $I \subset \mathbb{R}_+$.
En effet :

$$\begin{array}{rcl} \delta(\vec{r}-\overrightarrow{r_0})\delta(t-t_0) & \rightarrow & G(\vec{r},t,\overrightarrow{r_0},t_0) \\ s(\vec{r},t) = \iiint_{\mathcal{D}} \int_{\mathbb{R}_+} s(\overrightarrow{r_0},t_0)\delta(\vec{r}-\overrightarrow{r_0})\delta(t-t_0)dt_0\, d\overrightarrow{r_0} & \rightarrow & \phi(\vec{r},t) = \iiint_{\mathcal{D}} \int_{\mathbb{R}_+} s(\overrightarrow{r_0},t_0)G(\vec{r},t,\overrightarrow{r_0},t_0)dt_0\, d\overrightarrow{r_0} \end{array}$$

- *Calcul de la fonction de Green en 3D*

Méthode :
a- Transformée de Fourier spatiale et temporelle de l'équation de d'Alembert;
b- Expression de la TF de $G(\vec{r},t,\overrightarrow{r_0},t_0)$, notée $\tilde{G}(\vec{\chi},\omega)$
c- TF inverse de $\tilde{G}(\vec{\chi},\omega)$ par intégration sur les variables $\vec{\chi}$ et ω.

$$\Rightarrow G(\vec{r},t,\overrightarrow{r_0},t_0) = \frac{-1}{4\pi\|\vec{r}-\overrightarrow{r_0}\|}\,\delta\left(\frac{\|\vec{r}-\overrightarrow{r_0}\|}{c} - (t-t_0)\right)$$

Et donc $\phi(\vec{r},t) = \frac{-1}{4\pi}\iiint_{\mathcal{D}} \frac{1}{\|\vec{r}-\overrightarrow{r_0}\|}\, s\left(\overrightarrow{r_0}, t - \frac{\|\vec{r}-\overrightarrow{r_0}\|}{c}\right) d\overrightarrow{r_0}$

Notations: back to acoustics considerations

You saw during course 2 by H. Boutin that the velocity potential Φ at position r can be expressed by

$$\Phi(\mathrm{r}, t) = -\frac{1}{4\pi} \iiint_{\mathcal{D}} \frac{1}{\|\mathrm{r} - \mathrm{r}_0\|} s\left(\mathrm{r}_0, t - \frac{\|\mathrm{r} - \mathrm{r}_0\|}{c}\right) \mathrm{d}\mathrm{r}_0 \tag{1}$$

Lets now consider a punctual source, positioned at r^{S}, such that $s(\mathrm{r}_0, t) = \delta(\mathrm{r}_0 - \mathrm{r}^{\mathrm{S}})s(t)$. Then, one have

$$\begin{aligned}
\Phi(\mathrm{r}, t) &= -\frac{1}{4\pi} \iiint_{\mathcal{D}} \frac{1}{\|\mathrm{r} - \mathrm{r}_0\|} \delta(\mathrm{r}_0 - \mathrm{r}^{\mathrm{S}}) s\left(t - \frac{\|\mathrm{r} - \mathrm{r}_0\|}{c}\right) \mathrm{d}\mathrm{r}_0 \\
&= -\frac{1}{4\pi} \iiint_{\mathcal{D}} \frac{1}{\|\mathrm{r} - \mathrm{r}^{\mathrm{S}}\|} \delta(\mathrm{r}_0 - \mathrm{r}^{\mathrm{S}}) s\left(t - \frac{\|\mathrm{r} - \mathrm{r}^{\mathrm{S}}\|}{c}\right) \mathrm{d}\mathrm{r}_0 \\
&= -\frac{1}{4\pi} \frac{1}{\|\mathrm{r} - \mathrm{r}^{\mathrm{S}}\|} s\left(t - \frac{\|\mathrm{r} - \mathrm{r}^{\mathrm{S}}\|}{c}\right) = \Phi(\mathrm{r}, \mathrm{r}^{\mathrm{S}}, t)
\end{aligned}$$

Notations: back to acoustics considerations

You saw during course 2 by H. Boutin that the velocity potential Φ at position r can be expressed by

$$\Phi(\mathrm{r},t) = -\frac{1}{4\pi}\iiint_{\mathcal{D}} \frac{1}{\|\mathrm{r}-\mathrm{r}_0\|} s\left(\mathrm{r}_0, t-\frac{\|\mathrm{r}-\mathrm{r}_0\|}{c}\right) \mathrm{d}\mathrm{r}_0 \tag{1}$$

Lets now consider a punctual source, positioned at r^{s}, such that $s(\mathrm{r}_0,t) = \delta(\mathrm{r}_0-\mathrm{r}^{\mathrm{s}})s(t)$. Then, one have

$$\begin{aligned}
\Phi(\mathrm{r},t) &= -\frac{1}{4\pi}\iiint_{\mathcal{D}} \frac{1}{\|\mathrm{r}-\mathrm{r}_0\|}\delta(\mathrm{r}_0-\mathrm{r}^{\mathrm{s}}) s\left(t-\frac{\|\mathrm{r}-\mathrm{r}_0\|}{c}\right) \mathrm{d}\mathrm{r}_0 \\
&= -\frac{1}{4\pi}\iiint_{\mathcal{D}} \frac{1}{\|\mathrm{r}-\mathrm{r}^{\mathrm{s}}\|}\delta(\mathrm{r}_0-\mathrm{r}^{\mathrm{s}}) s\left(t-\frac{\|\mathrm{r}-\mathrm{r}^{\mathrm{s}}\|}{c}\right) \mathrm{d}\mathrm{r}_0 \\
&= -\frac{1}{4\pi}\frac{1}{\|\mathrm{r}-\mathrm{r}^{\mathrm{s}}\|} s\left(t-\frac{\|\mathrm{r}-\mathrm{r}^{\mathrm{s}}\|}{c}\right) = \Phi(\mathrm{r},\mathrm{r}^{\mathrm{s}},t)
\end{aligned}$$

For a purely monochromatic source with $s(t) = e^{jkct}$, this writes as

$$\Phi(\mathrm{r},\mathrm{r}^{\mathrm{s}},t) = -\frac{1}{4\pi\|\mathrm{r}-\mathrm{r}^{\mathrm{s}}\|} e^{-jk\|\mathrm{r}-\mathrm{r}^{\mathrm{s}}\|} e^{jkct}. \tag{2}$$

Notations: back to acoustics considerations

But a microphone placed at r is not sensitive to the velocity potential Φ, but rather to pressure variations p, so that

$$m(\mathrm{r}, \mathrm{r}^{\mathrm{s}}, t) = S\ a(t) * p(\mathrm{r}, \mathrm{r}^{\mathrm{s}}, t), \tag{3}$$

with S the microphone sensitivity (in V/Pa) and $a(t)$ the impulse response of the microphone, such that $a(t) = TF^{-1}[A(f)]$, with $A(f)$ the microphone frequency response.

Notations: back to acoustics considerations

But a microphone placed at r is not sensitive to the velocity potential Φ, but rather to pressure variations p, so that

$$m(\mathrm{r}, \mathrm{r}^{\mathrm{s}}, t) = S\ a(t) * p(\mathrm{r}, \mathrm{r}^{\mathrm{s}}, t), \tag{3}$$

with S the microphone sensitivity (in V/Pa) and $a(t)$ the impulse response of the microphone, such that $a(t) = TF^{-1}[A(f)]$, with $A(f)$ the microphone frequency response.

If we recall that

$$p(\mathrm{r}, \mathrm{r}^{\mathrm{s}}, t) = -\rho_0 \frac{\partial \Phi(\mathrm{r}, \mathrm{r}^{\mathrm{s}}, t)}{\partial t}, \tag{4}$$

then the pressure measured by the microphone at r for a monochromatic source at r^{s} can be expressed as

$$p(\mathrm{r}, \mathrm{r}^{\mathrm{s}}, t) = \frac{j\rho_0 kc}{4\pi \|\mathrm{r} - \mathrm{r}^{\mathrm{s}}\|} e^{-jk\|\mathrm{r}-\mathrm{r}^{\mathrm{s}}\|} e^{jkct}. \tag{5}$$

All is relative . . .

We often take the pressure $p_0(\mathrm{r}^{\mathrm{S}}, t)$ at the center O of the frame as a reference, with

$$p_0(\mathrm{r}^{\mathrm{S}}, t) = p(0, \mathrm{r}^{\mathrm{S}}, t) = \frac{j\rho_0 kc}{4\pi \|\mathrm{r}^{\mathrm{S}}\|} e^{-jk\|\mathrm{r}^{\mathrm{S}}\|} e^{jkct}. \tag{6}$$

All is relative . . .

We often take the pressure $p_0(\mathrm{r}^{\mathrm{S}}, t)$ at the center O of the frame as a reference, with

$$p_0(\mathrm{r}^{\mathrm{S}}, t) = p(0, \mathrm{r}^{\mathrm{S}}, t) = \frac{j\rho_0 kc}{4\pi \|\mathrm{r}^{\mathrm{S}}\|} e^{-jk\|\mathrm{r}^{\mathrm{S}}\|} e^{jkct}. \tag{6}$$

Then, the pressure measured by a microphone at position r and a source emitting from position r^{S} can be expressed along

$$p(\mathrm{r}, \mathrm{r}^{\mathrm{S}}, t) = \frac{\|\mathrm{r}^{\mathrm{S}}\|}{\|\mathrm{r} - \mathrm{r}^{\mathrm{S}}\|} \; e^{jk\|\mathrm{r}^{\mathrm{S}}\|} \; e^{-jk\|\mathrm{r}-\mathrm{r}^{\mathrm{S}}\|} \; p_0(\mathrm{r}^{\mathrm{S}}, t). \tag{7}$$

All is relative . . .

Then, coming back to the expression of the output signal $m(\mathrm{r}, \mathrm{r}^{\mathrm{S}}, t)$, and with the hypothesis $a(t) = \delta(t)$ (i.e. the microphone exactly reproduces the pression p up to a constant, that is $\mathrm{A}(\mathrm{f}) = 1$), then one gets

$$m(\mathrm{r}, \mathrm{r}^{\mathrm{S}}, t) = \frac{\|\mathrm{r}^{\mathrm{S}}\|}{\|\mathrm{r} - \mathrm{r}^{\mathrm{S}}\|} \, e^{jk\|\mathrm{r}^{\mathrm{S}}\|} \, e^{-jk\|\mathrm{r}-\mathrm{r}^{\mathrm{S}}\|} \, m_0(\mathrm{r}^{\mathrm{S}}, t), \tag{6}$$

where $m_0(\mathrm{r}^{\mathrm{S}}, t) = Sp_0(\mathrm{r}^{\mathrm{S}}, t)$ represents the *virtual* signal at the origin 0.

Since everything is supposed monochromatic, this virtual signal can be simply written as

$$m_0(\mathrm{r}^{\mathrm{S}}, t) = Ke^{jkct} = s^0(t), \tag{7}$$

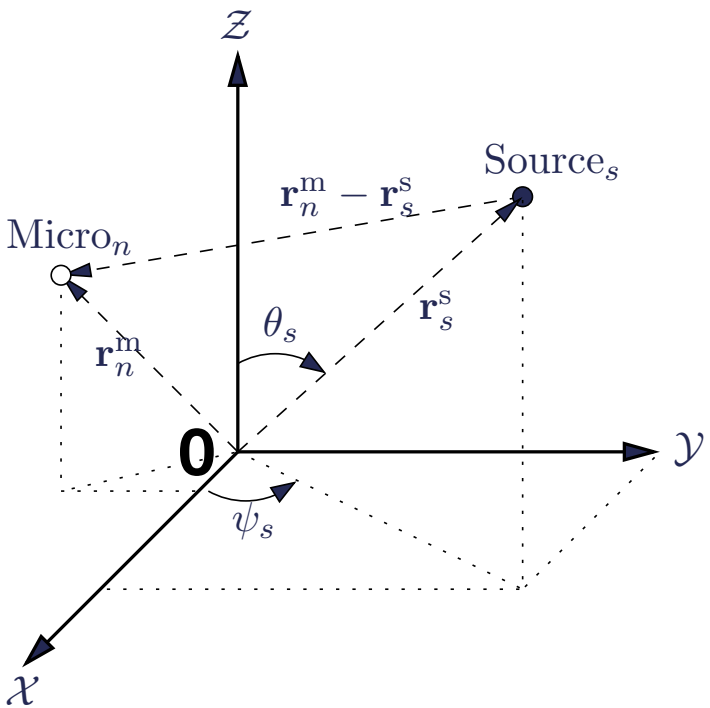
with $K = \frac{jS\rho_0 kc}{4\pi\|\mathrm{r}^{\mathrm{S}}\|} e^{-jk\|\mathrm{r}^{\mathrm{S}}\|}$ a constant.

$s^0(t)$ represents the relative contribution of the monochromatic source at position 0. It will be now considered as the reference signal.

Notations

In all the following, we now consider :

- S **punctual** monochromatic sound sources, placed at positions $\mathrm{r}_1^{\mathrm{s}}, \mathrm{r}_2^{\mathrm{s}}, \ldots, \mathrm{r}_S^{\mathrm{s}}$
- N **omnidirectional** microphones with identical frequency response $A(f)$, placed at positions $\mathrm{r}_1^{\mathrm{m}}, \mathrm{r}_2^{\mathrm{m}}, \ldots, \mathrm{r}_N^{\mathrm{m}}$.



- s^{th} source position: $\mathrm{r}_s^{\mathrm{s}}$
- n^{th} microphone position: $\mathrm{r}_n^{\mathrm{m}}$
- All microphones have a frequency response $A(f) = 1$

Notations

Then the signal outputed by the n^{th} microphone at position $\mathrm{r}_n^{\mathrm{m}}$ can be written as

$$m_n(t) = m(\mathrm{r}_n^{\mathrm{m}}, t) = \sum_{s=1}^{S} m(\mathrm{r}_n^{\mathrm{m}}, \mathrm{r}_s^{\mathrm{s}}, t) \tag{8}$$

This can be rewritten as

$$m_n(t) = \sum_{s=1}^{S} \frac{\|\mathrm{r}^{\mathrm{s}}\|}{\|\mathrm{r}_n^{\mathrm{m}} - \mathrm{r}^{\mathrm{s}}\|} \; e^{jk\|\mathrm{r}^{\mathrm{s}}\|} \; e^{-jk\|\mathrm{r}_n^{\mathrm{m}} - \mathrm{r}^{\mathrm{s}}\|} \; m_0(\mathrm{r}^{\mathrm{s}}, t) \tag{9}$$

$$= \sum_{s=1}^{S} \frac{\|\mathrm{r}^{\mathrm{s}}\|}{\|\mathrm{r}_n^{\mathrm{m}} - \mathrm{r}^{\mathrm{s}}\|} \; e^{jk\|\mathrm{r}^{\mathrm{s}}\|} \; e^{-jk\|\mathrm{r}_n^{\mathrm{m}} - \mathrm{r}^{\mathrm{s}}\|} \; s_s^0(t) \tag{10}$$

$$= \sum_{s=1}^{S} \alpha_n(\mathrm{r}_s^{\mathrm{s}}) s_s^0(t - \tau_n(\mathrm{r}_s^{\mathrm{s}})) \tag{11}$$

where $s_s^0(t)$ is the contribution of the s^{th} monochromatic source at the frame origin $\mathbf{0}$.

Notations

More generally, one writes

$$m_n(t) = \sum_{s=1}^{S} \alpha_n(\mathrm{r}_s^{\mathrm{s}}) s_s^0(t - \tau_n(\mathrm{r}_s^{\mathrm{s}})) + b_n(t), \tag{12}$$

with:

- $\alpha_n(\mathrm{r}) = \frac{\|\mathrm{r}\|}{\|\mathrm{r}_n^{\mathrm{m}} - \mathrm{r}\|}$ the relative attenuation caused by the wave propagation;
- $\tau_n(\mathrm{r}) = \frac{\|\mathrm{r}_n^{\mathrm{m}} - \mathrm{r}\|}{c} - \frac{\|\mathrm{r}\|}{c}$ the relative delay caused by the wave propagation;
- $b_n(t)$ some additive noise present on the n^{th} microphone.

And again . . . all is relative to the virtual signal "received" at the origin 0.

Notations

And everything can also be written in the frequency domain, along

$$M_n(k) = \sum_{s=1}^{S} V_n(\mathrm{r}_s^{\mathrm{s}}, k) S_s^0(k) + B_n(k), \tag{13}$$

where

$$V_n(\mathrm{r}, k) = \|\mathrm{r}\| e^{jk\|\mathrm{r}\|} \frac{e^{-jk\|\mathrm{r}_n^{\mathrm{m}} - \mathrm{r}\|}}{\|\mathrm{r}_n^{\mathrm{m}} - \mathrm{r}\|} \tag{14}$$

represents some kind of *relative frequency response* between the monochromatic signal of frequency k emitted by a source at r and the output of a microphone placed at $\mathrm{r}_n^{\mathrm{m}}$.

Notations: conclusion

Everything can be summed up into some vectorial notations ($.^T$ denotes the transposition operation):

- $\mathrm{M}(k) = (M_1(k), \ldots, M_N(k))^T$: the observation vector, of size $(N \times 1)$
- $\mathrm{V}(\mathrm{r}, k) = (V_1(\mathrm{r}, k), \ldots, V_N(\mathrm{r}, k))^T$: the **steering vector**, of size $(N \times 1)$
- $\mathrm{S}^0(k) = (S_1^0(k), \ldots, S_S^0(k))$: the source vector, of size $(S \times 1)$
- $\mathrm{B}(k) = (B_1(k), \ldots, B_N(k))^T$: the noise vector, of size $(N \times 1)$

Then, all sources/microphones can be regrouped inside one matrix:

- $\mathcal{V}(k) = (\mathrm{V}(\mathrm{r}_1^{\mathrm{s}}, k), \ldots, \mathrm{V}(\mathrm{r}_S^{\mathrm{s}}, k))$: the **steering matrix**, of size $(N \times S)$,

so that the previous equation can be rewritten as

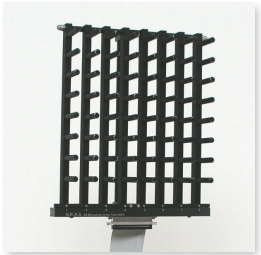
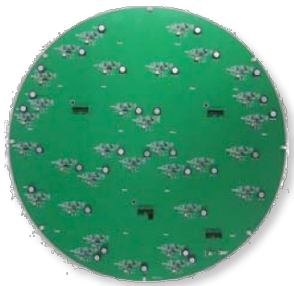
$$\mathrm{M}(k) = \mathcal{V}(k)\mathrm{S}^0(k) + \mathrm{B}(k) \tag{15}$$

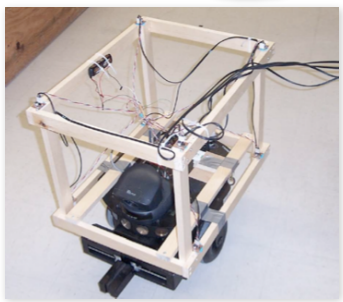
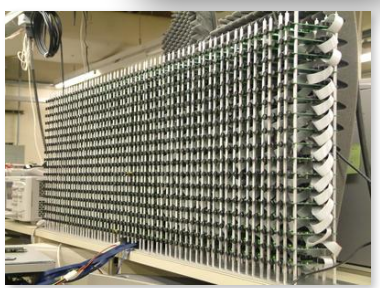
which constitutes the fundamental relation linking the S sources to the N microphones outputs.

Array processing: an example

What's a microphone array?

Some examples:

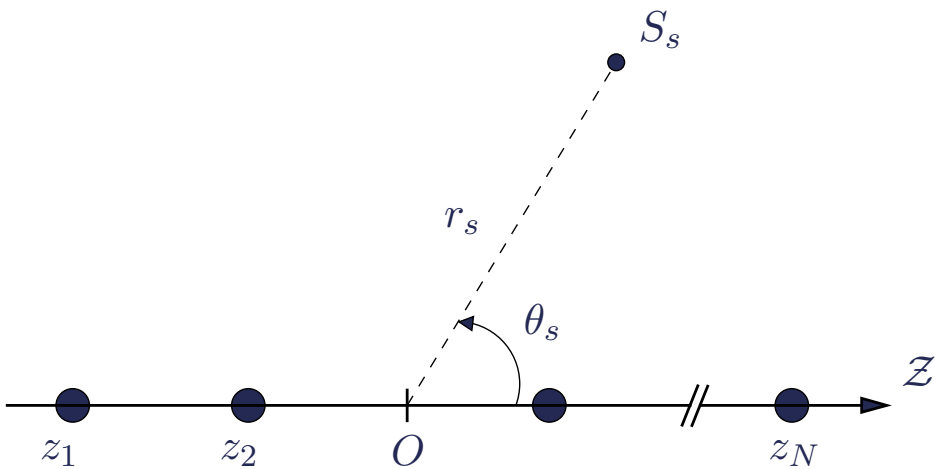




A microphone array is a fixed, geometrical, arrangement of identical (often calibrated) microphones.

Example: the linear array

Lets now consider a linear array, made of N microphones placed along a line $\mathcal{Z}$ and at positions z_n.



Taking 0 as the frame origin, a source position $\mathrm{r}_s^{\mathrm{s}}$ in space is then characterized by the vector $\mathrm{r}_s^{\mathrm{s}} = (r_s, \theta_s)$.

Consequently, one can write $\|\mathrm{r}_n^{\mathrm{m}} - \mathrm{r}_s^{\mathrm{s}}\| = \sqrt{r_s^2 + z_n^2 - 2 r_s z_n \cos\theta_s}$

Example: the linear array

Consequently, one component of the steering vector, which writes as

$$V_n(\mathrm{r}, k) = \|\mathrm{r}\| e^{jk\|\mathrm{r}\|} \frac{e^{-jk\|\mathrm{r}_n^{\mathrm{m}} - \mathrm{r}\|}}{\|\mathrm{r}_n^{\mathrm{m}} - \mathrm{r}\|}$$

. . . can be "simplified" along

$$V_n(\mathrm{r}, k) = V_n(r, \theta, k) = \frac{re^{jkr}}{\sqrt{r^2 + z_n^2 - 2rz_n \cos\theta_s}} e^{-jk\sqrt{r^2 + z_n^2 - 2rz_n \cos\theta_s}}.$$

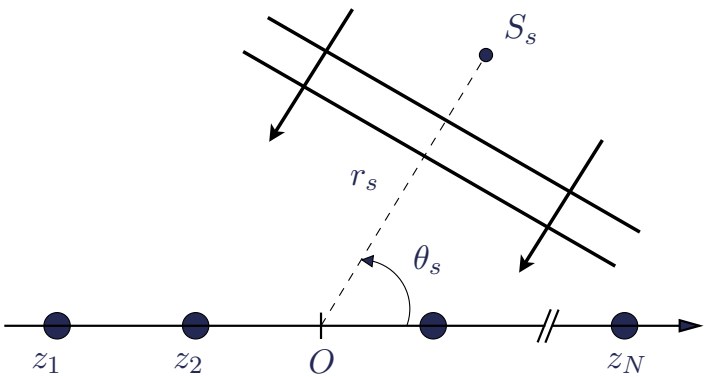
Now, if one supposes that the source is sufficiently far from the microphones (in practice, this means that $r \to +\infty$), then one gets $V_n^{\infty}(\theta, k)$, with

$$V_n^{\infty}(\theta, k) = e^{-jkz_n \cos\theta} \tag{16}$$

How can we interpret this simple equation?

Example: the linear array

Taking $r_s \to +\infty$ is the same as considering a planar wave propagation.



The source/microphone transfer relation is given by

$$V_n^{\infty}(\theta, k) = e^{-jkz_n \cos\theta}$$

. . . which represents a **pure delay**!

It then means that one can write:

$$m_n^{\infty}(t) = \sum_{s=1}^{S} s_s^0(t - \tau_n^{\infty}(\theta_s))(+b_n(t)), \tag{17}$$

with

$$\tau_n^{\infty}(\theta) = \frac{z_n}{c}\cos\theta \tag{18}$$

The signals among all microphones only differ from a pure delay $\tau_n^{\infty}(\theta_s)$ that depends on the angular source position θ_s.

To sum up

In the general case, one has:

$$m_n(t) = \sum_{s=1}^{S} \alpha_n(\mathrm{r}_s^{\mathrm{s}}) s_s^0(t - \tau_n(\mathrm{r}_s^{\mathrm{s}})) + b_n(t), \tag{19}$$

which can be rewritten in the frequency domain as

$$M_n(k) = \sum_{s=1}^{S} V_n(\mathrm{r}_s^{\mathrm{s}}, k) S_s^0(k) + B_n(k), \tag{20}$$

with

$$V_n(\mathrm{r}, k) = \|\mathrm{r}\| e^{jk\|\mathrm{r}\|} \frac{e^{-jk\|\mathrm{r}_n^{\mathrm{m}} - \mathrm{r}\|}}{\|\mathrm{r}_n^{\mathrm{m}} - \mathrm{r}\|}. \tag{21}$$

For a linear array with planar wave fronts:

$$V_n^{\infty}(\theta, k) = e^{-jkz_n \cos\theta} \tag{22}$$

TDOA estimation through correlation for sound source localization

Principle

Idea: exploit the Time Delay Of Arrival (TDOA) between each signal of a microphones array.

$$m_n(t) = \sum_{s=1}^{S} \alpha_n(\mathrm{r}_s^{\mathrm{S}}) s_s^0 \left(t - \tau_n(\mathrm{r}_s^{\mathrm{S}})\right) + b_n(t). \tag{23}$$

Principle

Idea: exploit the Time Delay Of Arrival (TDOA) between each signal of a microphones array.

$$m_n(t) = \sum_{s=1}^{S} \alpha_n(\mathrm{r}_s^{\mathrm{S}}) s_s^0 \left(t - \tau_n(\mathrm{r}_s^{\mathrm{S}})\right) + b_n(t). \tag{23}$$

Practical implementation for localization: relies on 2 successive steps:

1. first, estimation $\widehat{\Delta T_{ij}}$ of the true TDOAs $\Delta T_{ij}(\mathrm{r}_s^{\mathrm{S}}) = \tau_i(\mathrm{r}_s^{\mathrm{S}}) - \tau_j(\mathrm{r}_s^{\mathrm{S}})$ between the i^{th} and j^{th} microphone of the array;
2. and then, exploitation of these delays to actually estimate the source position $\widehat{\mathrm{r}_s^{\mathrm{S}}}$, with the hope that $\widehat{\mathrm{r}_s^{\mathrm{S}}} = \mathrm{r}_s^{\mathrm{S}}$.

Towards some TDOA estimation

You have seen during your "random signal processing" course that a delay between two signal can be estimated through the cross-correlation $R_{ij}(\tau)$:

$$R_{ij}(\tau) = \mathbb{E}\left[m_i(t)m_j(t-\tau)\right]. \tag{24}$$

Since the available signal for the n^{th} microphone writes as

$$m_n(t) = \sum_{s=1}^{S} \alpha_n(\mathrm{r}_s^{\mathrm{s}}) s_s^0(t - \tau_n(\mathrm{r}_s^{\mathrm{s}})) + b_n(t), \tag{25}$$

then for one source in the environment, $R_{ij}(\tau)$ can be written as

[1] with the hypothesis that the source signal s is a wide-sense stationary process for which $R_{ss}(\tau) \leq R_{ss}(0)$.

Towards some TDOA estimation

You have seen during your "random signal processing" course that a delay between two signal can be estimated through the cross-correlation $R_{ij}(\tau)$:

$$R_{ij}(\tau) = \mathbb{E}\left[m_i(t)m_j(t-\tau)\right]. \tag{24}$$

Since the available signal for the n^{th} microphone writes as

$$m_n(t) = \sum_{s=1}^{S} \alpha_n(\mathrm{r}_s^{\mathrm{s}}) s_s^0(t - \tau_n(\mathrm{r}_s^{\mathrm{s}})) + b_n(t), \tag{25}$$

then for one source in the environment, $R_{ij}(\tau)$ can be written as

$$R_{ij}(\tau) = \alpha_i \alpha_j R_{ss}(\tau - \Delta T_{ij}) + R_{b_i b_j}(\tau), \tag{26}$$

where R_{ss} represents the source autocorrelation, and $R_{b_i b_j}$ the noise cross-correlation.

[1] with the hypothesis that the source signal s is a wide-sense stationary process for which $R_{ss}(\tau) \leq R_{ss}(0)$.

Towards some TDOA estimation

You have seen during your "random signal processing" course that a delay between two signal can be estimated through the cross-correlation $R_{ij}(\tau)$:

$$R_{ij}(\tau) = \mathbb{E}\left[m_i(t)m_j(t-\tau)\right]. \tag{24}$$

Since the available signal for the n^{th} microphone writes as

$$m_n(t) = \sum_{s=1}^{S} \alpha_n(\mathrm{r}_s^{\mathrm{s}}) s_s^0(t - \tau_n(\mathrm{r}_s^{\mathrm{s}})) + b_n(t), \tag{25}$$

then for one source in the environment, $R_{ij}(\tau)$ can be written as

$$R_{ij}(\tau) = \alpha_i \alpha_j R_{ss}(\tau - \Delta T_{ij}) + R_{b_i b_j}(\tau), \tag{26}$$

where R_{ss} represents the source autocorrelation, and $R_{b_i b_j}$ the noise cross-correlation.

Then $\Delta T_{ij} = \tau_i(\mathrm{r}_s^{\mathrm{s}}) - \tau_j(\mathrm{r}_s^{\mathrm{s}})$ can be estimated by detecting the position of the maximum[1] in $R_{ij}(\tau)$, i.e.

$$\widehat{\Delta T_{ij}} = \arg_\tau \max R_{ij}(\tau) \tag{27}$$

[1] with the hypothesis that the source signal s is a wide-sense stationary process for which $R_{ss}(\tau) \leq R_{ss}(0)$.

Estimation of the cross-correlation function

In practice, only an estimation $\widehat{R}_{ij}(\tau)$ of the cross-correlation $R_{ij}(\tau)$ can be obtained from a unique realization of the two signals $m_i(t)$ and $m_j(t)$ on a finite time window $[-T/2; T/2]$ of length T, i.e.

$$\widehat{R}_{ij}(\tau) = \frac{1}{T} \int_{-\infty}^{+\infty} m_{1,T}(t) m_{2,T}(t-\tau) \mathrm{d}t, \tag{28}$$

One can show that this estimator is a biased estimator of the cross-correlation $R_{ij}(\tau)$.

Estimation of the cross-correlation function

In practice, only an estimation $\widehat{R}_{ij}(\tau)$ of the cross-correlation $R_{ij}(\tau)$ can be obtained from a unique realization of the two signals $m_i(t)$ and $m_j(t)$ on a finite time window $[-T/2; T/2]$ of length T, i.e.

$$\widehat{R}_{ij}(\tau) = \frac{1}{T}\int_{-\infty}^{+\infty} m_{1,T}(t) m_{2,T}(t-\tau)\mathrm{d}t, \tag{28}$$

One can show that this estimator is a biased estimator of the cross-correlation $R_{ij}(\tau)$.

How to choose the length T?
If we consider a planar wave propagation, we have shown that $\tau_n(\mathrm{r}_s^{\mathrm{S}}) = \tau_n^{\infty}(\theta_s) = \frac{z_n}{c}\cos\theta_s$, then one have

$$\Delta T_{ij} = \tau_i(\mathrm{r}_s^{\mathrm{S}}) - \tau_j(\mathrm{r}_s^{\mathrm{S}}) = \tau_i^{\infty}(\theta_s) - \tau_j^{\infty}(\theta_s) = \frac{z_i - z_j}{c}\cos\theta_s, \tag{29}$$

so that the maximal delay between two microphones is $D_{\max} = \frac{\max z_i - z_j}{c}$.

$\rightarrow$ *Then $\widehat{R}_{ij}(\tau)$ is shown to have a small variance and bias if $T >> D_{max}$.*

Estimation of the cross-correlation function

In practice we often compute the cross-correlation in the frequency domain, along

$$R_{ij}(\tau) = \int_{-\infty}^{+\infty} S_{m_i m_j}(f) e^{j2\pi f \tau} \mathrm{d}f, \tag{30}$$

where $S_{m_i m_j}(f)$ is the cross-power spectral density of the two signals m_i and m_j.

Estimation of the cross-correlation function

In practice we often compute the cross-correlation in the frequency domain, along

$$R_{ij}(\tau) = \int_{-\infty}^{+\infty} S_{m_i m_j}(f) e^{j2\pi f \tau} \mathrm{d}f, \tag{30}$$

where $S_{m_i m_j}(f)$ is the cross-power spectral density of the two signals m_i and m_j.

Again, all these quantities must be estimated from a finite time observation of length T, so that one works in the end with

Generalized cross-correlation

$$\widehat{R}_{ij}(\tau) = \int_{-\infty}^{+\infty} \Psi(f) \widehat{S}_{m_{i,T} m_{j,T}}(f) e^{j2\pi f \tau} \mathrm{d}f, \tag{31}$$

with $\Psi(f)$ **a frequency weight function**, and $\widehat{S}_{m_{i,T} m_{j,T}}(f)$ the estimated cross-power spectral density of one realization of the two signals $m_{i,T}$ and $m_{j,T}$ (e.g. by a mean cross-periodogram approach).

Generalized cross-correlation: the PHAse Transform weight

As an example, a frequently used weight function is given by

$$\Psi^{\text{PHAT}}(f) = \frac{1}{\|S_{m_i m_j}(f)\|}, \tag{32}$$

which defines the Cross-power Spectrum Phase (CSP) or PHAse Transform (PHAT).

Generalized cross-correlation: the PHAse Transform weight

As an example, a frequently used weight function is given by

$$\Psi^{\text{PHAT}}(f) = \frac{1}{\|S_{m_i m_j}(f)\|}, \tag{32}$$

which defines the Cross-power Spectrum Phase (CSP) or PHAse Transform (PHAT).

Interest:
If $M_1(f) = \alpha_1 S(f) e^{-j2\pi f \tau_1}$ and $M_2(f) = \alpha_2 S(f) e^{-j2\pi f \tau_2}$, then

$$S_{m_1 m_2}(f) = M_1(f) M_2^*(f) = \alpha_1 \alpha_2 |S(f)|^2 e^{-j2\pi f(\tau_1 - \tau_2)} \tag{33}$$

$$\Psi^{\text{PHAT}}(f) = \frac{1}{\alpha_1 \alpha_2 |S(f)|^2} \tag{34}$$

Then:

$$\begin{aligned} R_{ij}^{\text{PHAT}}(\tau) &= \int_{-\infty}^{+\infty} \Psi^{\text{PHAT}}(f) S_{m_i m_j}(f) e^{j2\pi f \tau} \mathrm{d}f = \int_{-\infty}^{+\infty} e^{-j2\pi f(\tau_1 - \tau_2)} e^{j2\pi f \tau} \mathrm{d}f \\ &= \delta\left(\tau - (\tau_1 - \tau_2)\right). \end{aligned} \tag{35}$$

Generalized cross-correlation: the PHAse Transform weight

One then have:

$$\begin{aligned} R_{ij}^{\text{PHAT}}(\tau) &= \delta\left(\tau - (\tau_1 - \tau_2)\right), \\ &= \delta\left(\tau - \Delta T_{ij}\right), \end{aligned} \tag{36}$$

which means that the cross-correlation function is a Dirac centered at ΔT_{ij}!

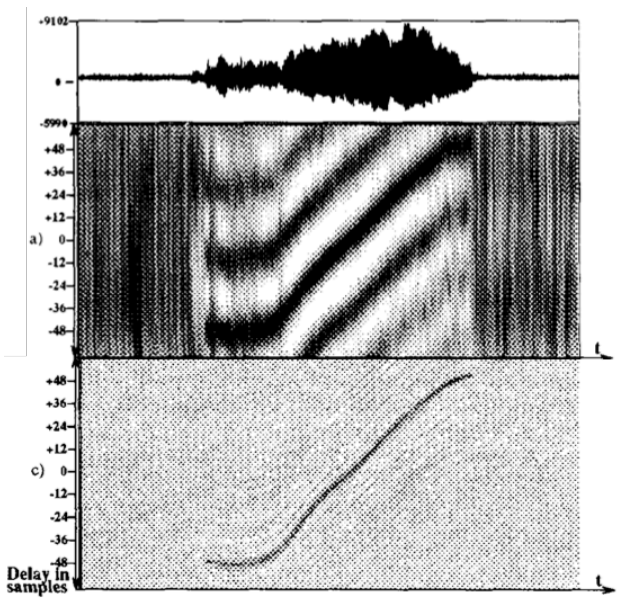
Generalized cross-correlation: the PHAse Transform weight

One then have:

$$\begin{aligned} R_{ij}^{\text{PHAT}}(\tau) &= \delta\left(\tau - (\tau_1 - \tau_2)\right), \\ &= \delta\left(\tau - \Delta T_{ij}\right), \end{aligned} \tag{36}$$

which means that the cross-correlation function is a Dirac centered at ΔT_{ij}!

Example[2]:



$\rightarrow$ Signal

$\rightarrow$ "Traditional" cross-correlation

$\rightarrow$ CSP/PHAT

[2] Taken from (Omologo and Svaizer, 1994).

Generalized cross-correlation: the PHAse Transform weight

One then have:

$$\begin{aligned} R_{ij}^{\text{PHAT}}(\tau) &= \delta\left(\tau - (\tau_1 - \tau_2)\right), \\ &= \delta\left(\tau - \Delta T_{ij}\right), \end{aligned} \tag{36}$$

which means that the cross-correlation function is a Dirac centered at ΔT_{ij}!

Lets now do some code!

Generalized cross-correlation

Other weighting functions are proposed in the literature:

- the Roth function (Roth, 1971): $\Psi^{\mathrm{ROTH}}(f) = \dfrac{1}{S_{m_1 m_2}(f)}$

Generalized cross-correlation

Other weighting functions are proposed in the literature:

- the Roth function (Roth, 1971): $\Psi^{\mathrm{ROTH}}(f) = \dfrac{1}{S_{m_1 m_2}(f)}$
- the SCoT function (Smoothed COherence Transform) (Carter, Nuttall, and Cable, 1973): $\Psi^{\mathrm{SCoT}}(f) = \dfrac{1}{\sqrt{S_{m_1 m_1}(f) S_{m_2 m_2}(f)}}$

Generalized cross-correlation

Other weighting functions are proposed in the literature:

- the Roth function (Roth, 1971): $\Psi^{\mathrm{ROTH}}(f) = \dfrac{1}{S_{m_1 m_2}(f)}$
- the SCoT function (Smoothed COherence Transform) (Carter, Nuttall, and Cable, 1973): $\Psi^{\mathrm{SCoT}}(f) = \dfrac{1}{\sqrt{S_{m_1 m_1}(f) S_{m_2 m_2}(f)}}$
- the HT function (Hannan and Thomson, 1973):
  $\Psi^{\mathrm{HT}}(f) = \dfrac{|\gamma_{m_1 m_2}(f)|^2}{|S_{m_1 m_2}(f)| \left(1 - |\gamma_{m_1 m_2}(f)|^2\right)}$, with
  $\gamma_{m_1 m_2}(f) = \dfrac{S_{m_1 m_2}(f)}{\sqrt{S_{m_1 m_1}(f) S_{m_2 m_2}(f)}}$ the Magnitude Squared Coherence (MSC).

Generalized cross-correlation

Other weighting functions are proposed in the literature:

- the Roth function (Roth, 1971): $\Psi^{\mathrm{ROTH}}(f) = \dfrac{1}{S_{m_1 m_2}(f)}$
- the SCoT function (Smoothed COherence Transform) (Carter, Nuttall, and Cable, 1973): $\Psi^{\mathrm{SCoT}}(f) = \dfrac{1}{\sqrt{S_{m_1 m_1}(f) S_{m_2 m_2}(f)}}$
- the HT function (Hannan and Thomson, 1973):
  $\Psi^{\mathrm{HT}}(f) = \dfrac{|\gamma_{m_1 m_2}(f)|^2}{|S_{m_1 m_2}(f)| \left(1 - |\gamma_{m_1 m_2}(f)|^2\right)}$, with
  $\gamma_{m_1 m_2}(f) = \dfrac{S_{m_1 m_2}(f)}{\sqrt{S_{m_1 m_1}(f) S_{m_2 m_2}(f)}}$ the Magnitude Squared Coherence (MSC).
- ...and others ad-hoc functions, like the Reliability-Weighted PHAT (Valin et al., 2003), etc.

Generalized cross-correlation

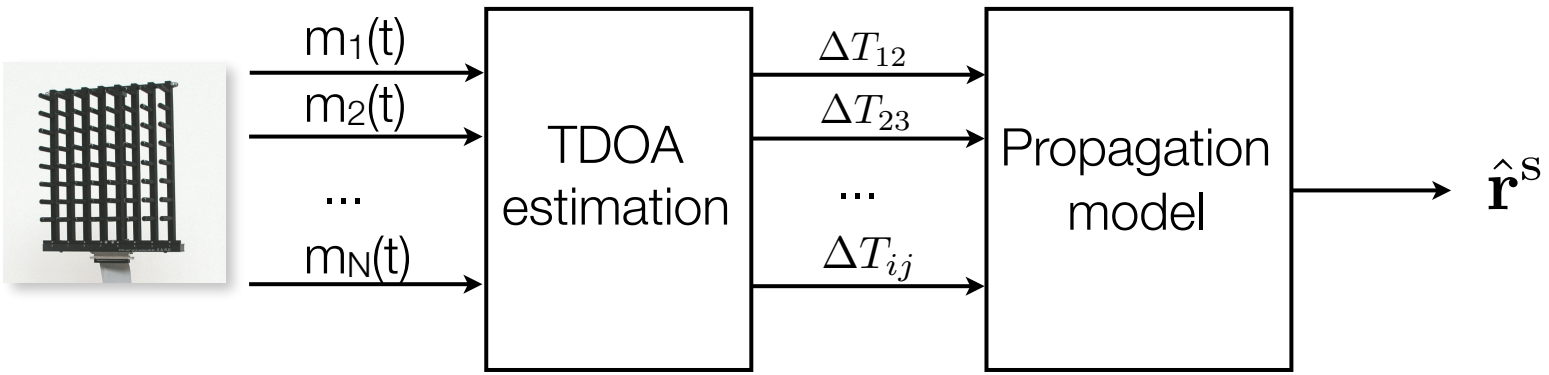
Other weighting functions are proposed in the literature:

- the Roth function (Roth, 1971): $\Psi^{\mathrm{ROTH}}(f) = \dfrac{1}{S_{m_1 m_2}(f)}$
- the SCoT function (Smoothed COherence Transform) (Carter, Nuttall, and Cable, 1973): $\Psi^{\mathrm{SCoT}}(f) = \dfrac{1}{\sqrt{S_{m_1 m_1}(f) S_{m_2 m_2}(f)}}$
- the HT function (Hannan and Thomson, 1973):
  $\Psi^{\mathrm{HT}}(f) = \dfrac{|\gamma_{m_1 m_2}(f)|^2}{|S_{m_1 m_2}(f)| \left(1 - |\gamma_{m_1 m_2}(f)|^2\right)}$, with
  $\gamma_{m_1 m_2}(f) = \dfrac{S_{m_1 m_2}(f)}{\sqrt{S_{m_1 m_1}(f) S_{m_2 m_2}(f)}}$ the Magnitude Squared Coherence (MSC).
- ...and others ad-hoc functions, like the Reliability-Weighted PHAT (Valin et al., 2003), etc.

Most of them tends to increase/decrease the contribution of certain frequencies (on the basis in their high/low SNR for instance) to the overall cross-correlation (Portello, 2013).

From TDOA to source position

Recall that we have a 2-step approach:



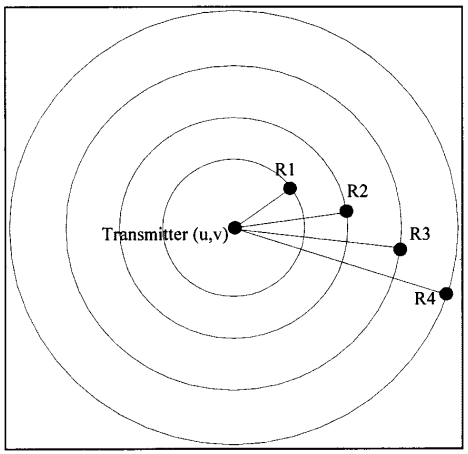
What's a "propagation model"?

It relates the estimated TDOA to the source position thanks to the array geometry.

For instance: for only 2 microphones spaced by d in the farfield, one have

$$\Delta T_{12} = \frac{d}{c} \sin\theta_s \Longrightarrow \widehat{\theta_s} = \sin^{-1}\left(\frac{c\widehat{\Delta T_{12}}}{d}\right) \tag{37}$$

From TDOA to source position



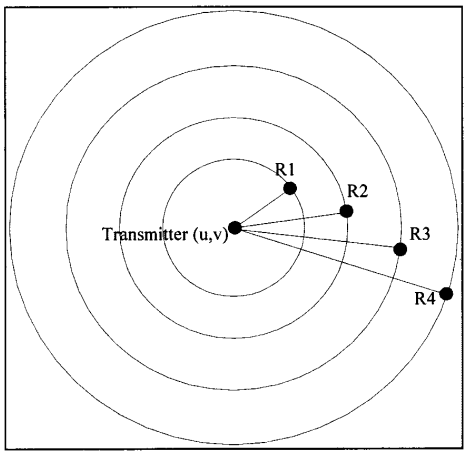
We only focus in the TDOAs $\Delta T_{1i} = \tau_1 - \tau_i$.

For a spherical propagation, the wave fronts write as, for all $i \in [1, \ldots, N]$

$$(x_i^{\mathrm{m}} - x^{\mathrm{s}})^2 + (y_i^{\mathrm{m}} - y^{\mathrm{s}})^2 + (z_i^{\mathrm{m}} - z^{\mathrm{s}})^2 = (d + c\Delta T_{1i})^2,$$

with d the distance between $R1$ and the source.

From TDOA to source position



We only focus in the TDOAs $\Delta T_{1i} = \tau_1 - \tau_i$.

For a spherical propagation, the wave fronts write as, for all $i \in [1, \dots, N]$

$$(x_i^{\mathrm{m}} - x^{\mathrm{s}})^2 + (y_i^{\mathrm{m}} - y^{\mathrm{s}})^2 + (z_i^{\mathrm{m}} - z^{\mathrm{s}})^2 = (d + c\Delta T_{1i})^2,$$

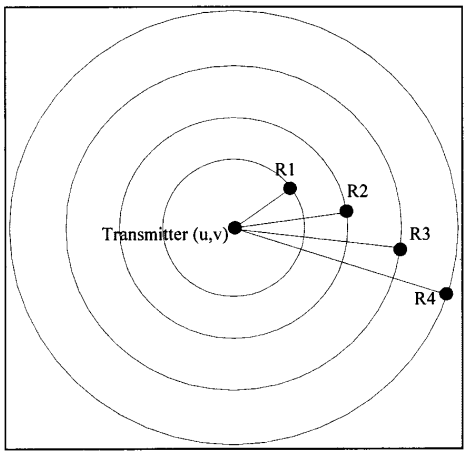
with d the distance between $R1$ and the source.

This equation can be rewritten with matrices, along

$$\begin{pmatrix} 2(x_1^{\mathrm{m}} - x_2^{\mathrm{m}}) & 2(y_1^{\mathrm{m}} - y_2^{\mathrm{m}}) & 2(z_1^{\mathrm{m}} - z_2^{\mathrm{m}}) & -2c\Delta T_{12} \\ 2(x_1^{\mathrm{m}} - x_3^{\mathrm{m}}) & 2(y_1^{\mathrm{m}} - y_3^{\mathrm{m}}) & 2(z_1^{\mathrm{m}} - z_3^{\mathrm{m}}) & -2c\Delta T_{13} \\ \vdots & \vdots & \vdots & \vdots \\ 2(x_1^{\mathrm{m}} - x_N^{\mathrm{m}}) & 2(y_1^{\mathrm{m}} - y_N^{\mathrm{m}}) & 2(z_1^{\mathrm{m}} - z_N^{\mathrm{m}}) & -2c\Delta T_{1N} \end{pmatrix} \begin{pmatrix} x^{\mathrm{s}} \\ y^{\mathrm{s}} \\ z^{\mathrm{s}} \\ d \end{pmatrix} =$$

$$\begin{pmatrix} c^2\Delta T_{12}^2 + (x_1^{\mathrm{m}})^2 + (y_1^{\mathrm{m}})^2 + (z_1^{\mathrm{m}})^2 - (x_2^{\mathrm{m}})^2 - (y_2^{\mathrm{m}})^2 - (z_2^{\mathrm{m}})^2 \\ c^2\Delta T_{13}^2 + (x_1^{\mathrm{m}})^2 + (y_1^{\mathrm{m}})^2 + (z_1^{\mathrm{m}})^2 - (x_3^{\mathrm{m}})^2 - (y_3^{\mathrm{m}})^2 - (z_3^{\mathrm{m}})^2 \\ \vdots \\ c^2\Delta T_{1N}^2 + (x_1^{\mathrm{m}})^2 + (y_1^{\mathrm{m}})^2 + (z_1^{\mathrm{m}})^2 - (x_N^{\mathrm{m}})^2 - (y_N^{\mathrm{m}})^2 - (z_N^{\mathrm{m}})^2 \end{pmatrix} \quad (38)$$

From TDOA to source position



We only focus in the TDOAs $\Delta T_{1i} = \tau_1 - \tau_i$.

For a spherical propagation, the wave fronts write as, for all $i \in [1, \ldots, N]$

$$(x_i^{\mathrm{m}} - x^{\mathrm{s}})^2 + (y_i^{\mathrm{m}} - y^{\mathrm{s}})^2 + (z_i^{\mathrm{m}} - z^{\mathrm{s}})^2 = (d + c\Delta T_{1i})^2,$$

with d the distance between $R1$ and the source.

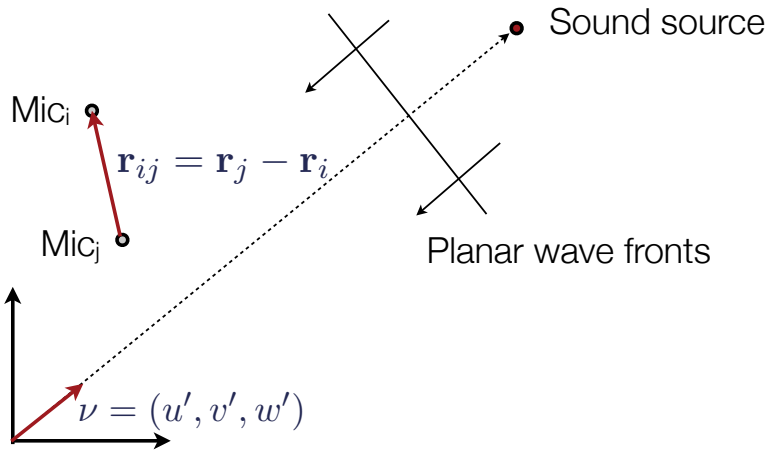
but ...

$$\begin{pmatrix} 2(x_1^{\mathrm{m}} - x_2^{\mathrm{m}}) & 2(y_1^{\mathrm{m}} - y_2^{\mathrm{m}}) & 2(z_1^{\mathrm{m}} - z_2^{\mathrm{m}}) & -2c\Delta T_{12} \\ 2(x_1^{\mathrm{m}} - x_3^{\mathrm{m}}) & 2(y_1^{\mathrm{m}} - y_3^{\mathrm{m}}) & 2(z_1^{\mathrm{m}} - z_3^{\mathrm{m}}) & -2c\Delta T_{13} \\ \vdots & \vdots & \vdots & \vdots \\ 2(x_1^{\mathrm{m}} - x_N^{\mathrm{m}}) & 2(y_1^{\mathrm{m}} - y_N^{\mathrm{m}}) & 2(z_1^{\mathrm{m}} - z_N^{\mathrm{m}}) & -2c\Delta T_{1N} \end{pmatrix}$$

has to be (pseudo-)inversed, and its conditioning actually depends on ΔT_{1i} ...

From TDOA to source position

Instead, the following approach might be preferred:



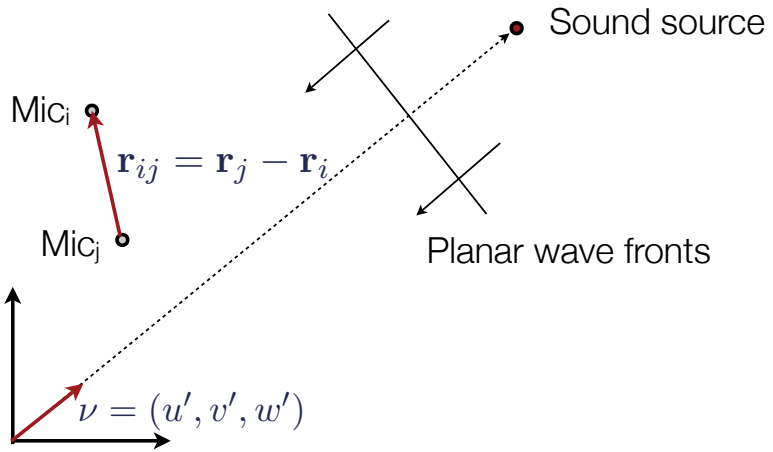
The wave fronts being supposed planar, we can write

$$\boldsymbol{\nu}.\mathrm{r}_{i1}^{\mathrm{m}} = c\Delta T_{i1}, \qquad (39)$$

with $\boldsymbol{\nu} = (u, v, w)$ the vector pointing towards the source direction and . the scalar product.

From TDOA to source position

Instead, the following approach might be preferred:



The wave fronts being supposed planar, we can write

$$\boldsymbol{\nu}.\mathrm{r}_{i1}^{\mathrm{m}} = c\Delta T_{i1}, \tag{39}$$

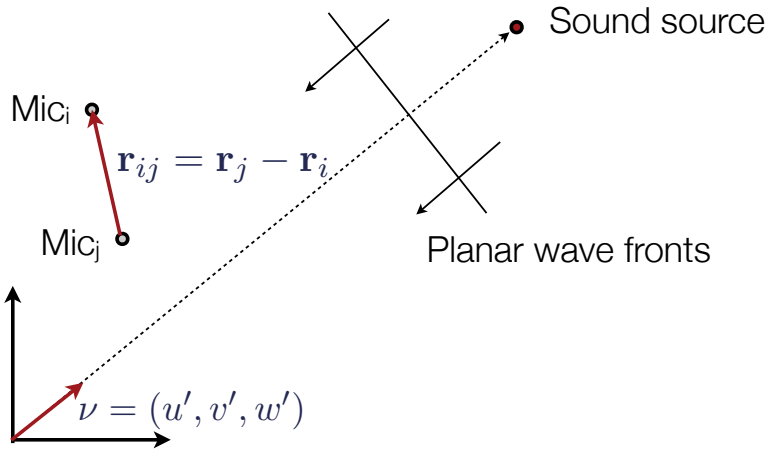
with $\boldsymbol{\nu} = (u, v, w)$ the vector pointing towards the source direction and . the scalar product.

This can be also written, for all microphones,

$$\begin{pmatrix} x_2^{\mathrm{m}} - x_1^{\mathrm{m}} & y_2^{\mathrm{m}} - y_1^{\mathrm{m}} & z_2^{\mathrm{m}} - z_1^{\mathrm{m}} \\ x_3^{\mathrm{m}} - x_1^{\mathrm{m}} & y_3^{\mathrm{m}} - y_1^{\mathrm{m}} & z_3^{\mathrm{m}} - z_1^{\mathrm{m}} \\ \vdots & \vdots & \vdots \\ x_N^{\mathrm{m}} - x_1^{\mathrm{m}} & y_N^{\mathrm{m}} - y_1^{\mathrm{m}} & z_N^{\mathrm{m}} - z_1^{\mathrm{m}} \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} c\Delta T_{12} \\ c\Delta T_{13} \\ \vdots \\ c\Delta T_{1N} \end{pmatrix} \tag{40}$$

From TDOA to source position

Instead, the following approach might be preferred:



The wave fronts being supposed planar, we can write

$$\boldsymbol{\nu}.\mathrm{r}_{i1}^{\mathrm{m}} = c\Delta T_{i1}, \tag{39}$$

with $\boldsymbol{\nu} = (u, v, w)$ the vector pointing towards the source direction and . the scalar product.

This can be also written, for all microphones,

$$\begin{pmatrix} x_2^{\mathrm{m}} - x_1^{\mathrm{m}} & y_2^{\mathrm{m}} - y_1^{\mathrm{m}} & z_2^{\mathrm{m}} - z_1^{\mathrm{m}} \\ x_3^{\mathrm{m}} - x_1^{\mathrm{m}} & y_3^{\mathrm{m}} - y_1^{\mathrm{m}} & z_3^{\mathrm{m}} - z_1^{\mathrm{m}} \\ \vdots & \vdots & \vdots \\ x_N^{\mathrm{m}} - x_1^{\mathrm{m}} & y_N^{\mathrm{m}} - y_1^{\mathrm{m}} & z_N^{\mathrm{m}} - z_1^{\mathrm{m}} \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} c\Delta T_{12} \\ c\Delta T_{13} \\ \vdots \\ c\Delta T_{1N} \end{pmatrix} \tag{40}$$

. . . which now only depends on the microphone position/geometry!

Correlation techniques: conclusion

Pros:

- relatively "simple" approach, only relying a cross-correlation computation
- well-chosen pre-filtering functions can better the delay estimation performances

Cons:

- moving from the TDOAs to the estimated source position might be tricky (e.g. how to deal with scattering elements in the array?)
- poor performances of correlation strategies in the presence of noise, with low SNR conditions . . .
- . . . or in the presence of reverberations, which should then carefully taken into account (strategies not discussed here)

Beamforming approaches to localization

Principle

Basic idea:

Polarization of a microphones array towards a specific direction so as to

- amplify the signal coming from a given direction of interest,
- and reduce everything coming from other directions.

Principle

Basic idea:
Polarization of a microphones array towards a specific direction so as to

- amplify the signal coming from a given direction of interest,
- and reduce everything coming from other directions.

This amplification/reduction of signals as a function of position is entirely characterized by the **array pattern**, or **beam pattern**, $D(\mathrm{r}, k)$ which represents the spatial *and* frequential filtering operated by the microphone array.

Principle

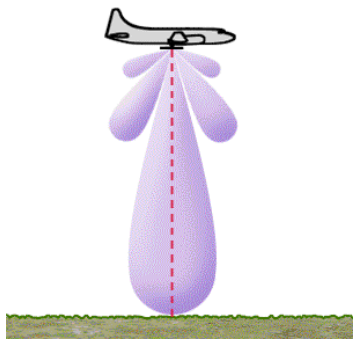
Basic idea:
Polarization of a microphones array towards a specific direction so as to

- amplify the signal coming from a given direction of interest,
- and reduce everything coming from other directions.

This amplification/reduction of signals as a function of position is entirely characterized by the **array pattern**, or **beam pattern**, $D(\mathrm{r}, k)$ which represents the spatial *and* frequential filtering operated by the microphone array.

Applications:

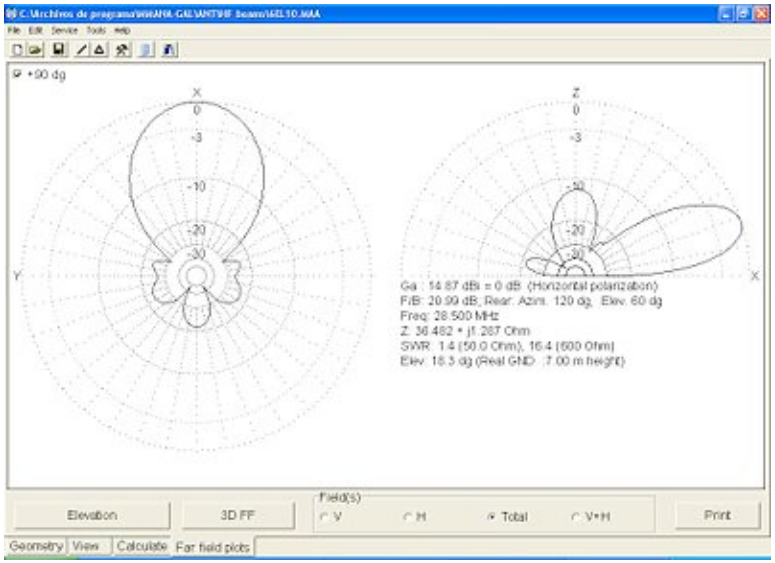
Telecommunications

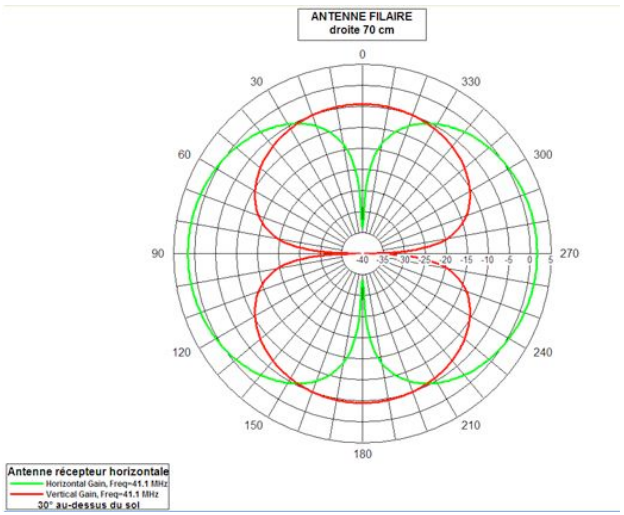


Radars

Principle

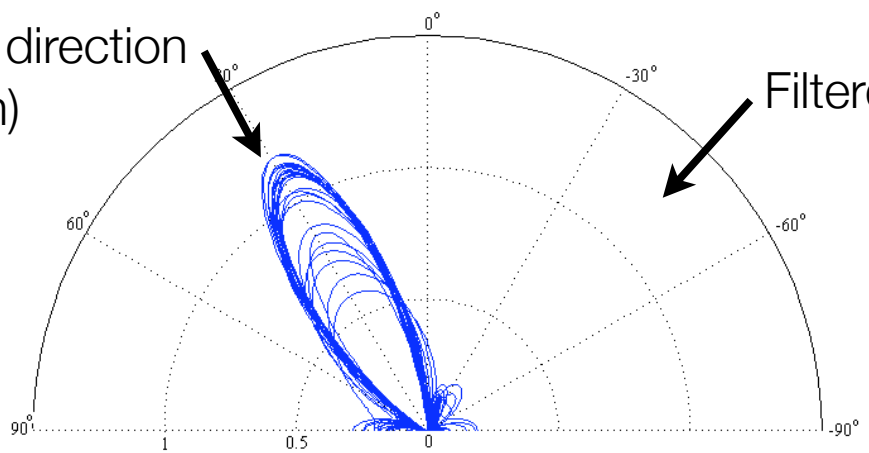
Some beampatterns:





Priviledged direction (polarization)

Filtered directions

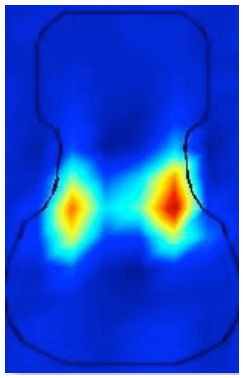


Principle

And what for?

- **Localization**

The array is successively polarized in every candidate positions. For each of them, the signal level can be computed and reported on an energy map.

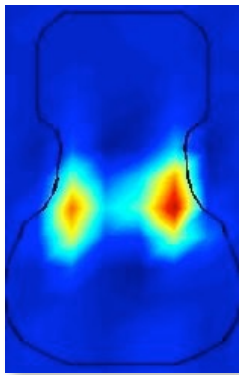


Principle

And what for?

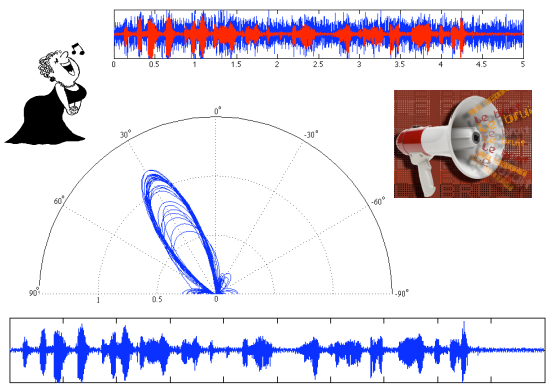
- **Localization**

The array is successively polarized in every candidate positions. For each of them, the signal level can be computed and reported on an energy map.



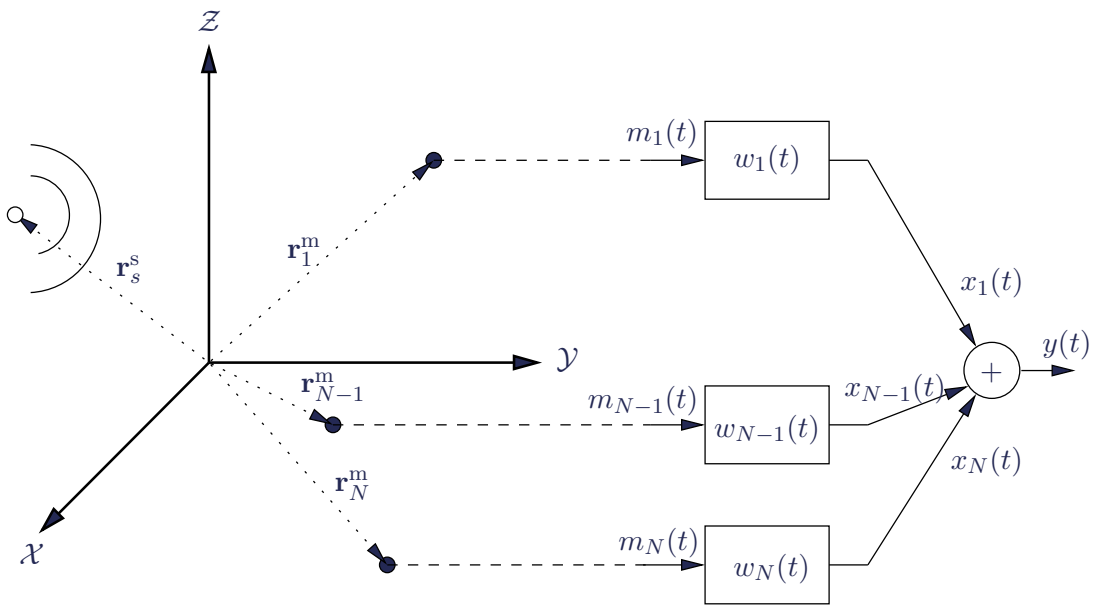
- **Extraction**

The array is polarized in some direction of interest, where a sound source is present. The signal at the array output is expected to mainly reproduce the focused sound source, while all other sounds are filtered out.



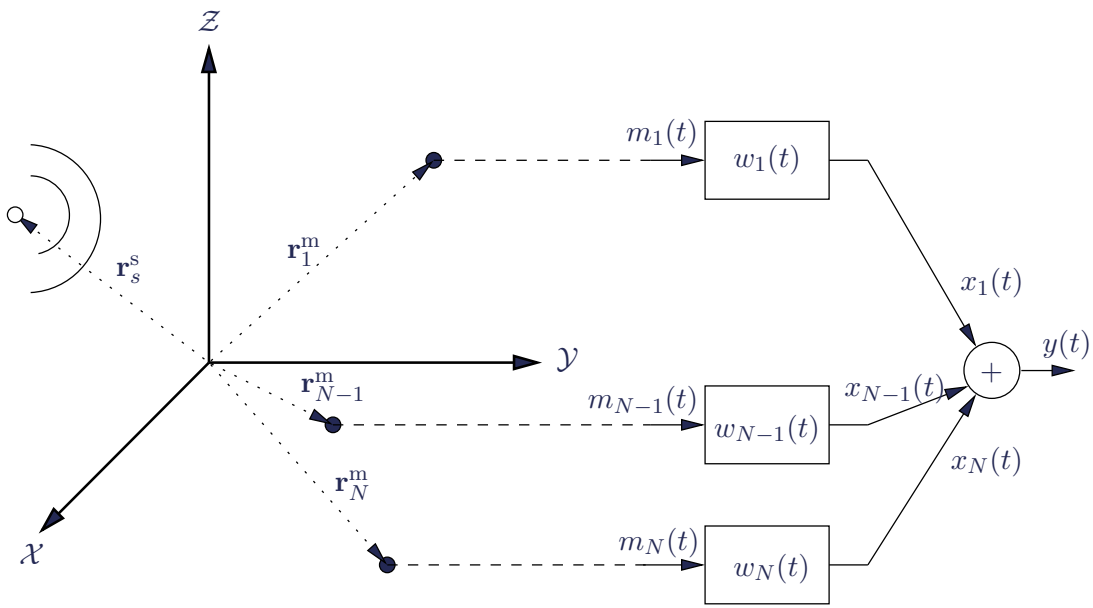
Formalization

Architecture: the N microhones outputs $m_n(t)$ are first processed with some adequate linear filters with impulse response $w_n(t)$, and their outputs are then combined to form the beamformer output $y(t)$.



Formalization

Architecture: the N microhones outputs $m_n(t)$ are first processed with some adequate linear filters with impulse response $w_n(t)$, and their outputs are then combined to form the beamformer output $y(t)$.



One can then write

$$y(t) = \sum_{n=1}^{N} w_n(t) * m_n(t) \tag{41}$$

Formalization

We saw earlier that

$$M_n(k) = \sum_{s=1}^{S} V_n(\mathrm{r}_s^{\mathrm{s}}, k) S_s^0(k) + B_n(k),$$

So we can write, in the frequency domain

Formalization

We saw earlier that

$$M_n(k) = \sum_{s=1}^{S} V_n(\mathrm{r}_s^{\mathrm{s}}, k) S_s^0(k) + B_n(k),$$

So we can write, in the frequency domain

$$Y(k) = \sum_{s=1}^{S} D(\mathrm{r}_s^{\mathrm{s}}, k) S_s^0(k) + \sum_{n=1}^{N} W_n(k) B_n(k), \tag{42}$$

where $W_n(k)$ is the frequency response of the n^{th} filter placed after the n^{th} microphone, and $D(\mathrm{r}, k)$ the beampattern evaluated at source position $\mathrm{r}_s^{\mathrm{s}}$, with

$$D(\mathrm{r}, k) = \sum_{n=1}^{N} W_n(k) V_n(\mathrm{r}, k). \tag{43}$$

Formalization

We saw earlier that

$$M_n(k) = \sum_{s=1}^{S} V_n(\mathrm{r}_s^{\mathrm{S}}, k) S_s^0(k) + B_n(k),$$

So we can write, in the frequency domain

$$Y(k) = \sum_{s=1}^{S} D(\mathrm{r}_s^{\mathrm{S}}, k) S_s^0(k) + \sum_{n=1}^{N} W_n(k) B_n(k), \tag{42}$$

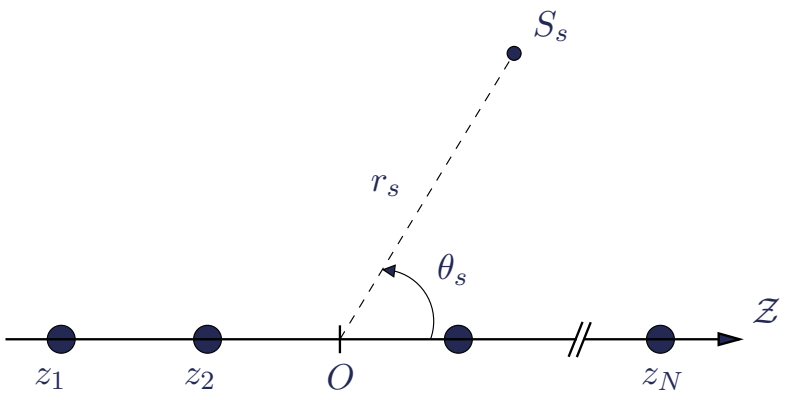
where $W_n(k)$ is the frequency response of the n^{th} filter placed after the n^{th} microphone, and $D(\mathrm{r}, k)$ the beampattern evaluated at source position $\mathrm{r}_s^{\mathrm{S}}$, with

$$D(\mathrm{r}, k) = \sum_{n=1}^{N} W_n(k) V_n(\mathrm{r}, k). \tag{43}$$

Again, if one considers the front waves case, one can define the **farfield beampattern** D^∞ as $D^\infty(\theta, \Psi, k) = \lim_{r \to +\infty} D(\mathrm{r}, k)$

For a linear microphones array and planar waves

Lets consider the following array geometry (already seen in Sec. 1):



We saw earlier:

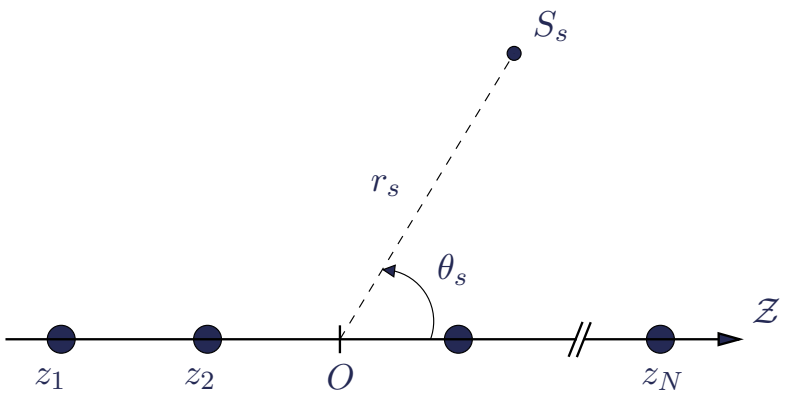
$$V_n^{\infty}(\theta, \Psi, k) = \lim_{r \to +\infty} V_n(\mathrm{r}, k),$$

and for a linear array

$$V_n^{\infty}(\theta, k) = e^{-jkz_n \cos\theta}.$$

For a linear microphones array and planar waves

Lets consider the following array geometry (already seen in Sec. 1):



We saw earlier:

$$V_n^{\infty}(\theta, \Psi, k) = \lim_{r \to +\infty} V_n(\mathrm{r}, k),$$

and for a linear array

$$V_n^{\infty}(\theta, k) = e^{-jkz_n \cos\theta}.$$

Then, the array beampattern of a linear microphones array writes as

$$D^{\infty}(\theta, k) = \sum_{n=1}^{N} W_n(k) e^{-jkz_n \cos\theta} \tag{44}$$

For a linear microphones array and planar waves

In the following, we shall consider a linear array with N microphones evenly placed along the $\mathcal{Z}$ axis at abscissa z_n such that

$$z_n = \left(n - \frac{N+1}{2} \right) d, \tag{45}$$

with d the interspace between two successive microphones. Then, the array size is $L = (N-1)d$.

For a linear microphones array and planar waves

In the following, we shall consider a linear array with N microphones evenly placed along the $\mathcal{Z}$ axis at abscissa z_n such that

$$z_n = \left(n - \frac{N+1}{2}\right) d, \tag{45}$$

with d the interspace between two successive microphones. Then, the array size is $L = (N-1)d$.

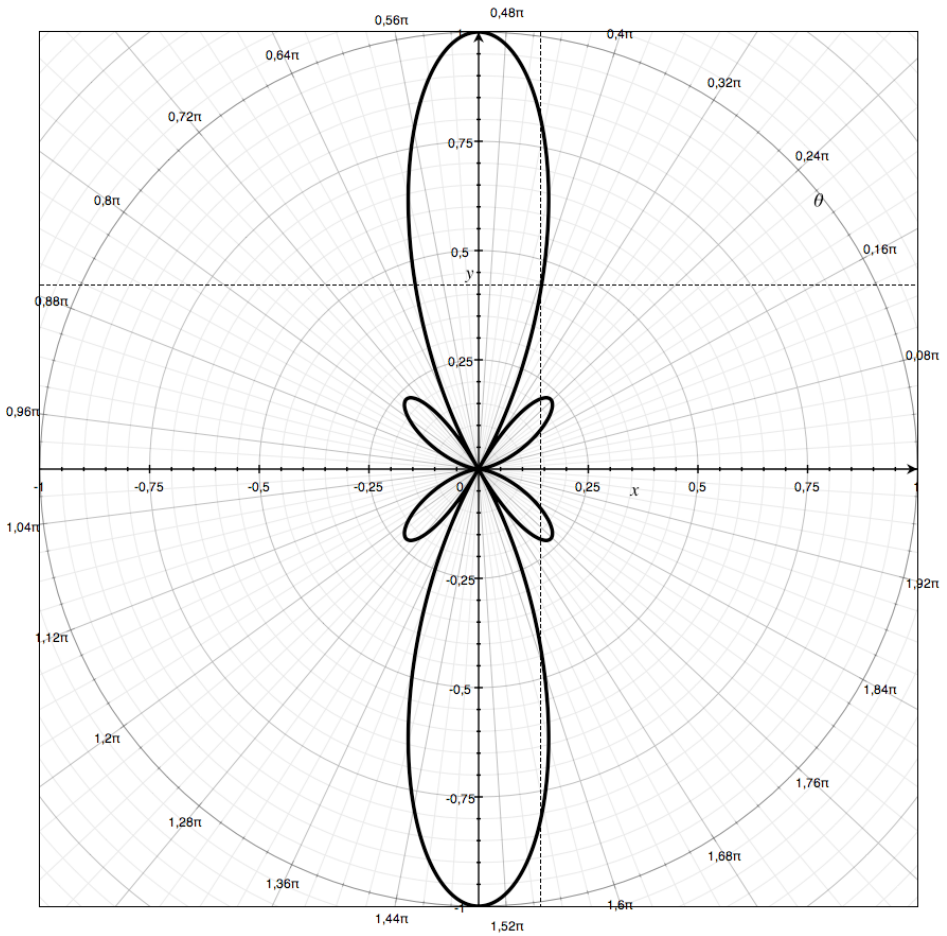
For now, lets suppose that each filter of the beamformer has the same frequency response $W_n(k) = 1$. One then have

$$D^{\infty}(\theta, k) = \sum_{n=1}^{N} e^{-jk\left(n - \frac{N+1}{2}\right) d\cos\theta} = \frac{\sin\left(k\frac{Nd}{2}\cos\theta\right)}{\sin\left(k\frac{d}{2}\cos\theta\right)} \tag{46}$$

$$= \frac{\sin\left(\frac{\pi f}{c} Nd\cos\theta\right)}{\sin\left(\frac{\pi f}{c} d\cos\theta\right)}. \tag{47}$$

For a linear microphones array and planar waves

We can now represent the farfield beampattern $D^{\infty}(\theta, k) = \frac{\sin\left(\frac{\pi f}{c} Nd \cos\theta\right)}{\sin\left(\frac{\pi f}{c} d \cos\theta\right)}$:

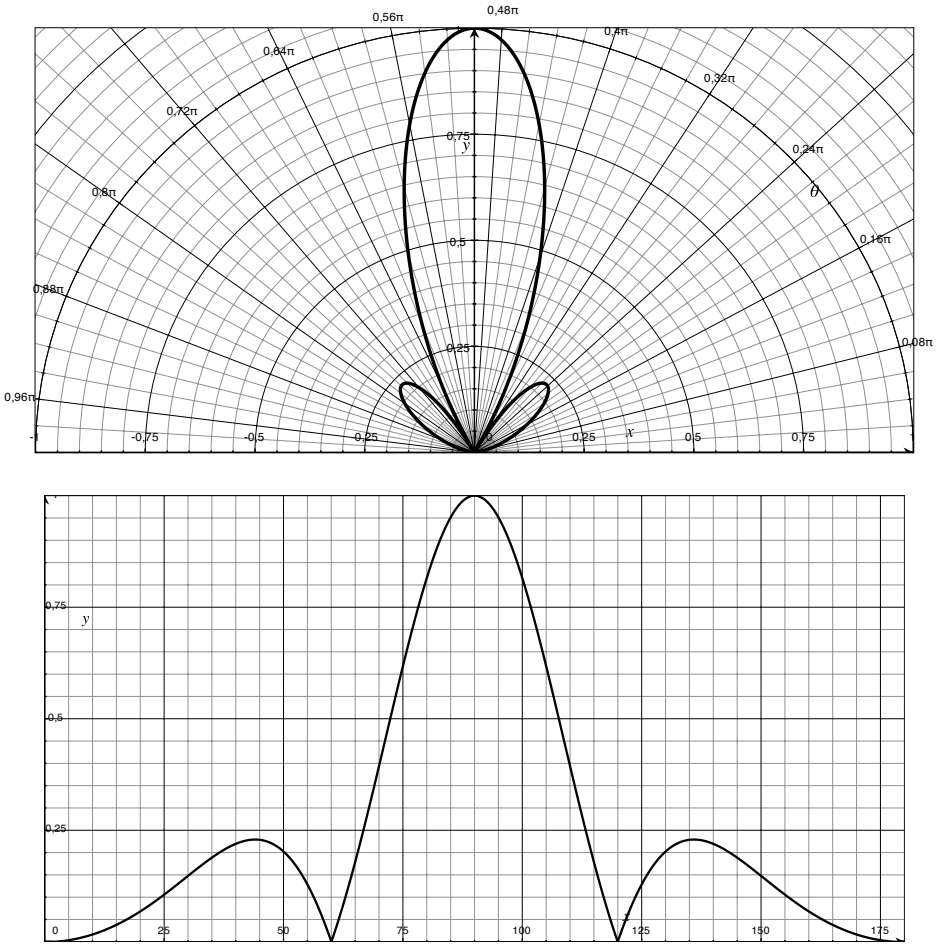


- The array pattern is symetric w.r.t. the array axis;

Plot obtained for $N = 8$, $f = 1500Hz$, $d = 5.6cm$, so that $L = 39.6cm$.

For a linear microphones array and planar waves

We can now represent the farfield beampattern $D^{\infty}(\theta, k) = \frac{\sin\left(\frac{\pi f}{c} Nd \cos\theta\right)}{\sin\left(\frac{\pi f}{c} d \cos\theta\right)}$:

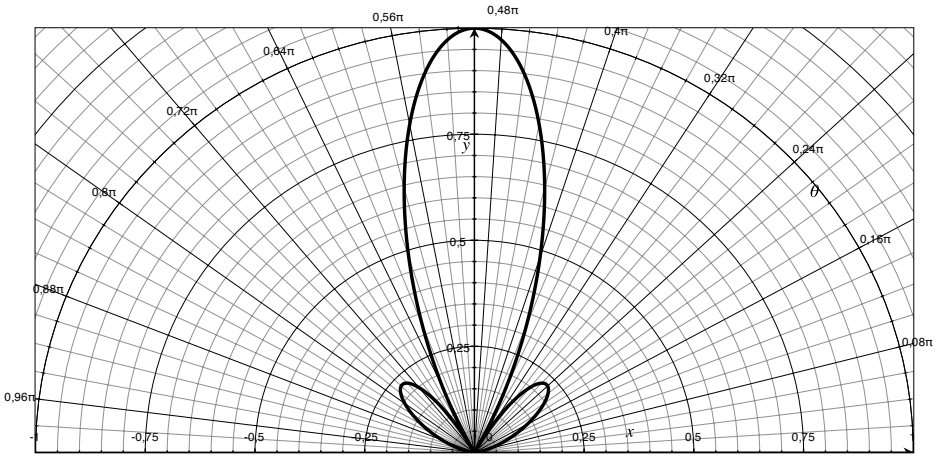


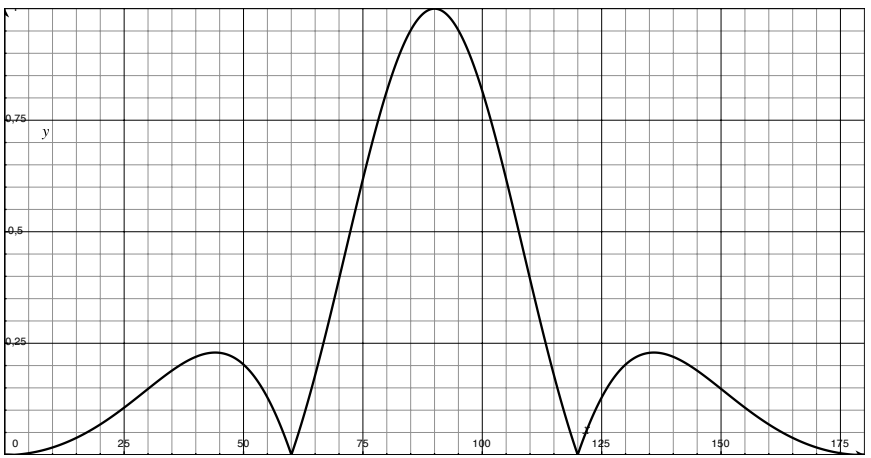
- The array pattern is symetric w.r.t. the array axis;
- Choosing $W_n(k) = 1$ makes the array pointing towards the frontal direction $\theta = \pi/2$;

Plot obtained for $N = 8$, $f = 1500Hz$,
$d = 5.6cm$, so that $L = 39.6cm$.

For a linear microphones array and planar waves

We can now represent the farfield beampattern $D^{\infty}(\theta, k) = \frac{\sin\left(\frac{\pi f}{c} Nd \cos\theta\right)}{\sin\left(\frac{\pi f}{c} d \cos\theta\right)}$:



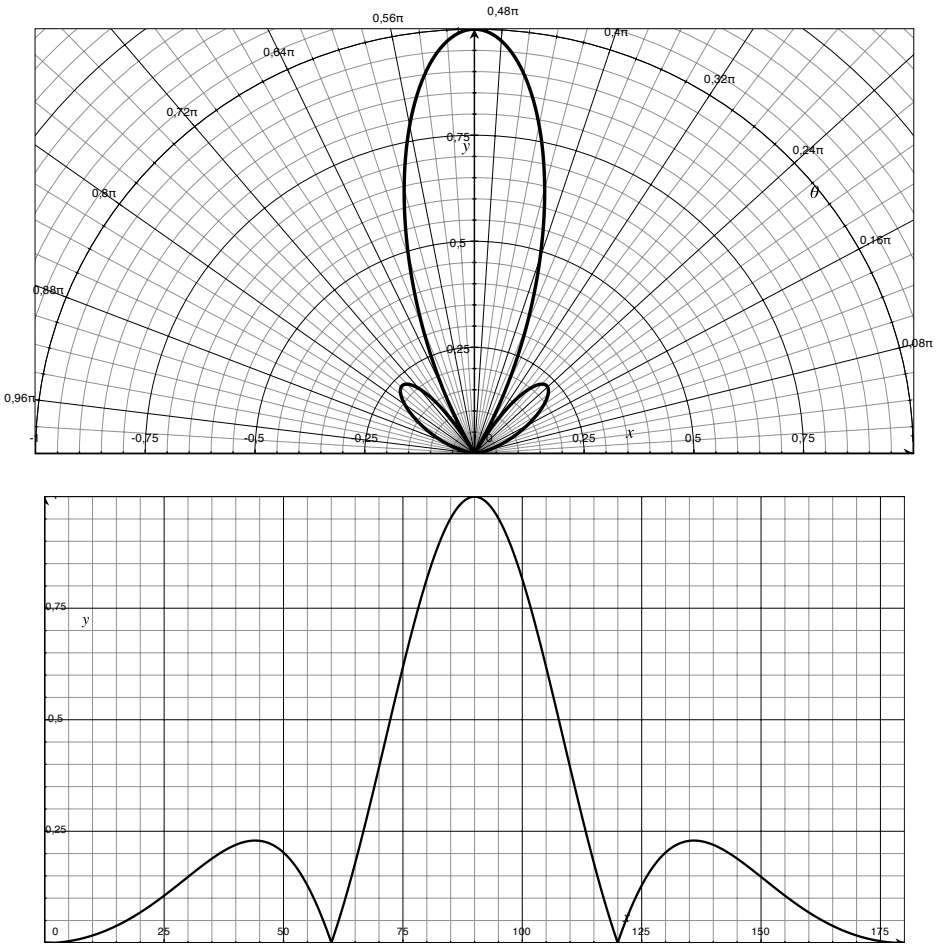


Plot obtained for $N = 8$, $f = 1500Hz$, $d = 5.6cm$, so that $L = 39.6cm$.

- The array pattern is symetric w.r.t. the array axis;
- Choosing $W_n(k) = 1$ makes the array pointing towards the frontal direction $\theta = \pi/2$;
- The main lobe is significantly wide ...

For a linear microphones array and planar waves

We can now represent the farfield beampattern $D^{\infty}(\theta, k) = \frac{\sin\left(\frac{\pi f}{c} Nd \cos\theta\right)}{\sin\left(\frac{\pi f}{c} d \cos\theta\right)}$:



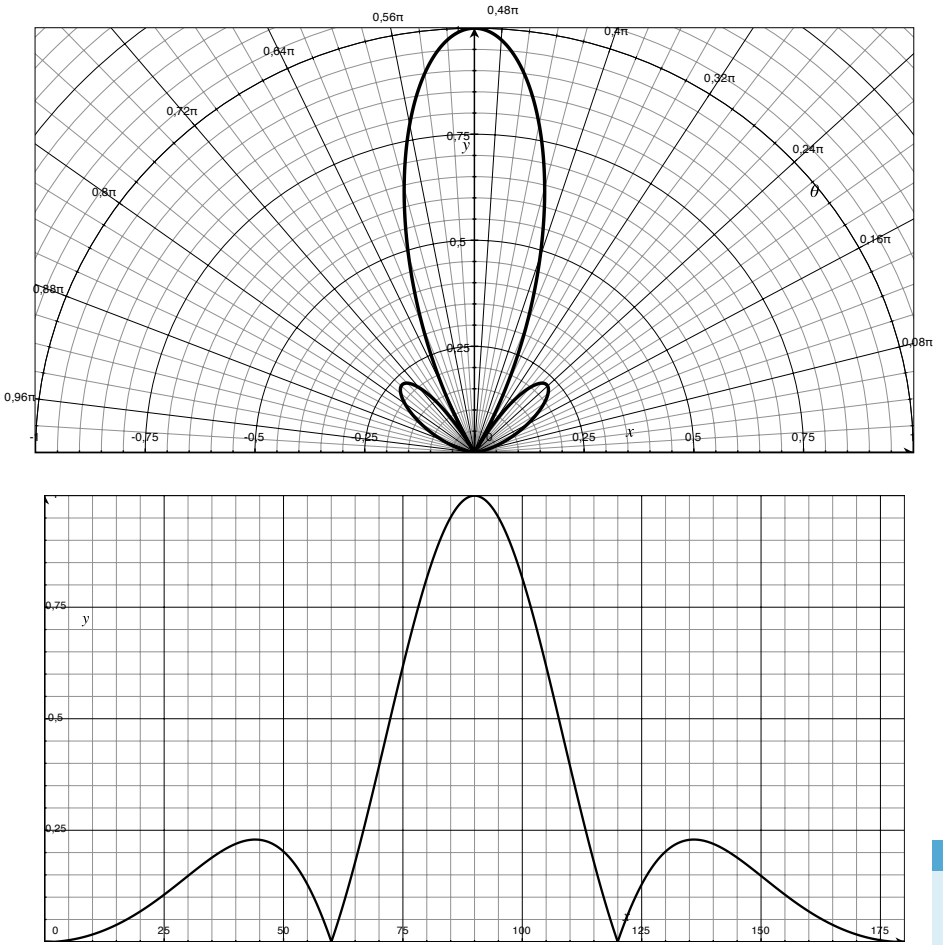
Plot obtained for $N = 8$, $f = 1500Hz$, $d = 5.6cm$, so that $L = 39.6cm$.

- The array pattern is symetric w.r.t. the array axis;
- Choosing $W_n(k) = 1$ makes the array pointing towards the frontal direction $\theta = \pi/2$;
- The main lobe is significantly wide ...
- ... and secondary lobes also appear in the beampattern.

$\rightarrow$ the spatial filtering is not perfect!

For a linear microphones array and planar waves

We can now represent the farfield beampattern $D^{\infty}(\theta, k) = \frac{\sin\left(\frac{\pi f}{c} Nd \cos\theta\right)}{\sin\left(\frac{\pi f}{c} d \cos\theta\right)}$:



Plot obtained for $N = 8$, $f = 1500Hz$, $d = 5.6cm$, so that $L = 39.6cm$.

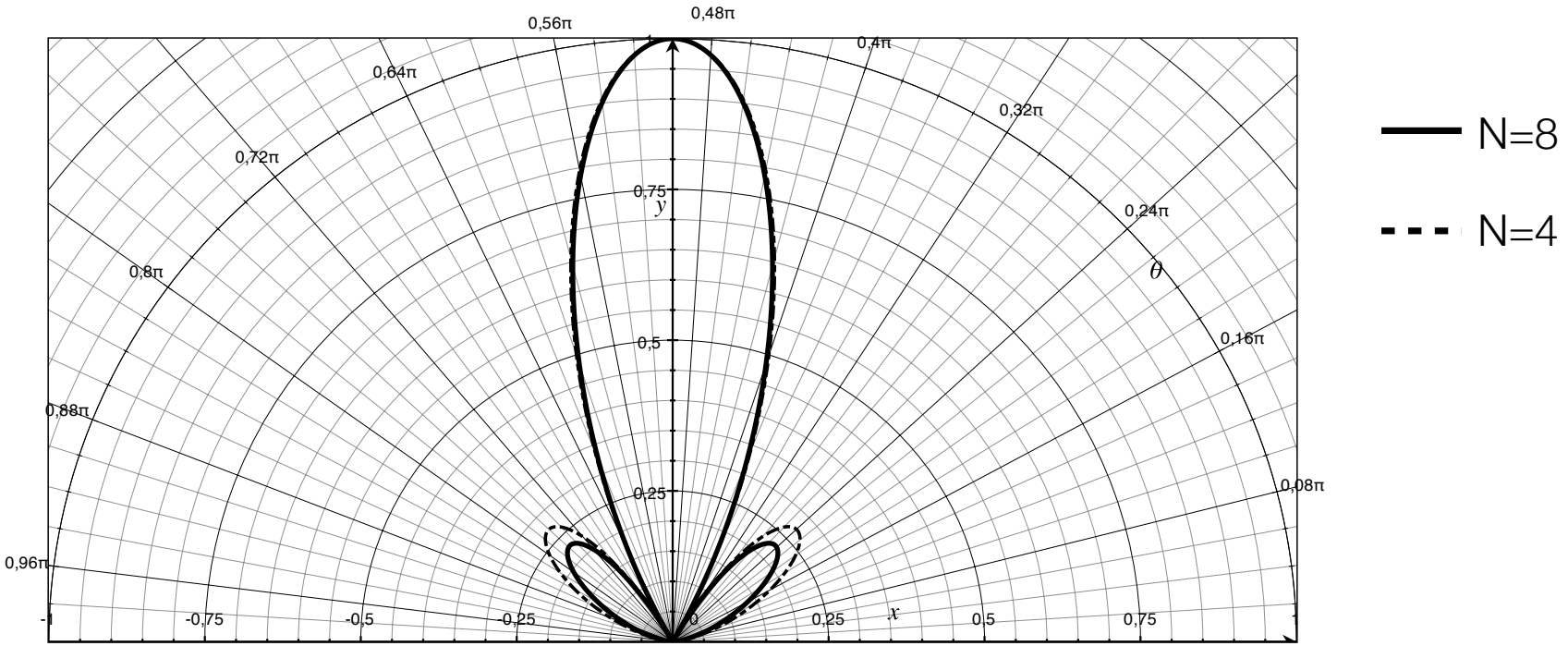
- The array pattern is symetric w.r.t. the array axis;
- Choosing $W_n(k) = 1$ makes the array pointing towards the frontal direction $\theta = \pi/2$;
- The main lobe is significantly wide ...
- ... and secondary lobes also appear in the beampattern.

$\rightarrow$ the spatial filtering is not perfect!

What is the influence of the array geometry (N, L parameters) and of the frequency f?

Influence of the array geometry and frequency on the beampattern

→ Array size $L = 39.6$cm is constant, and $f = 1500$Hz.

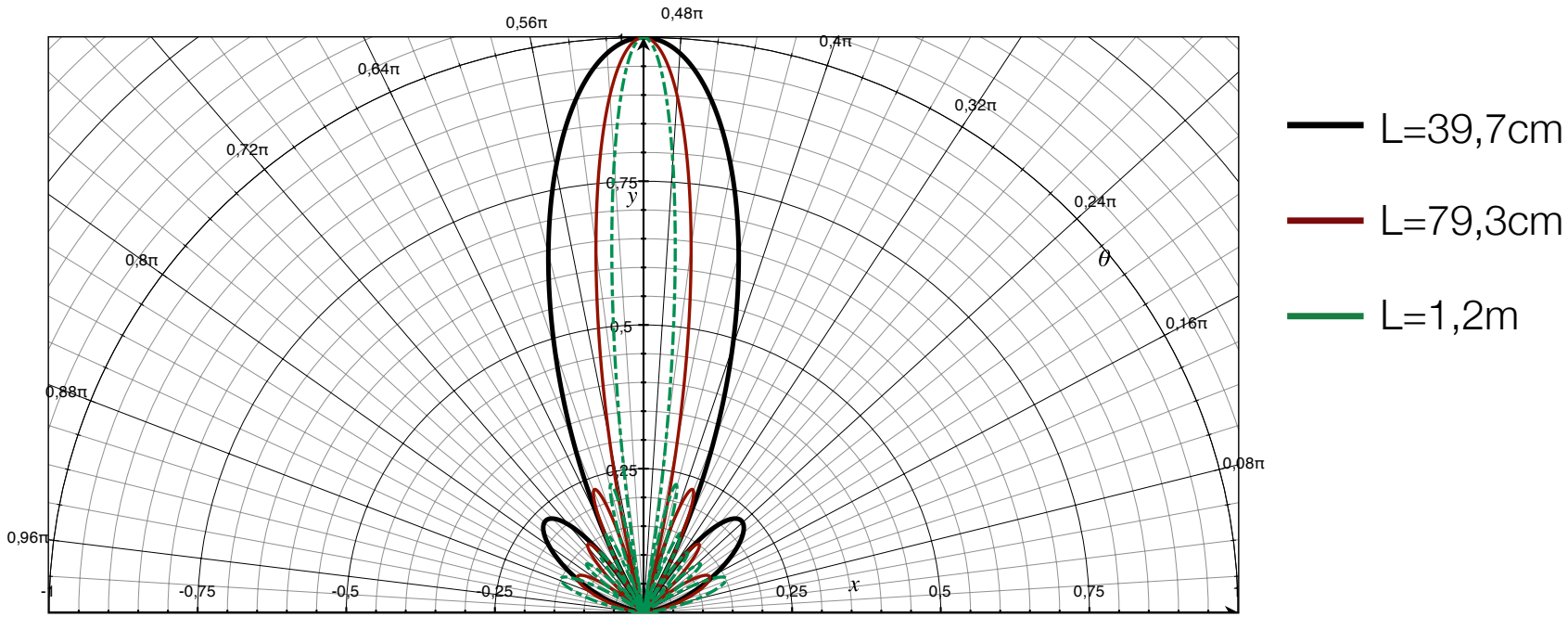


Influence of N

Sidelobes level is reduced (for a constant array length).

Influence of the array geometry and frequency on the beampattern

$\rightarrow$ $N = 8$ microphones, and $f = 1500$Hz.

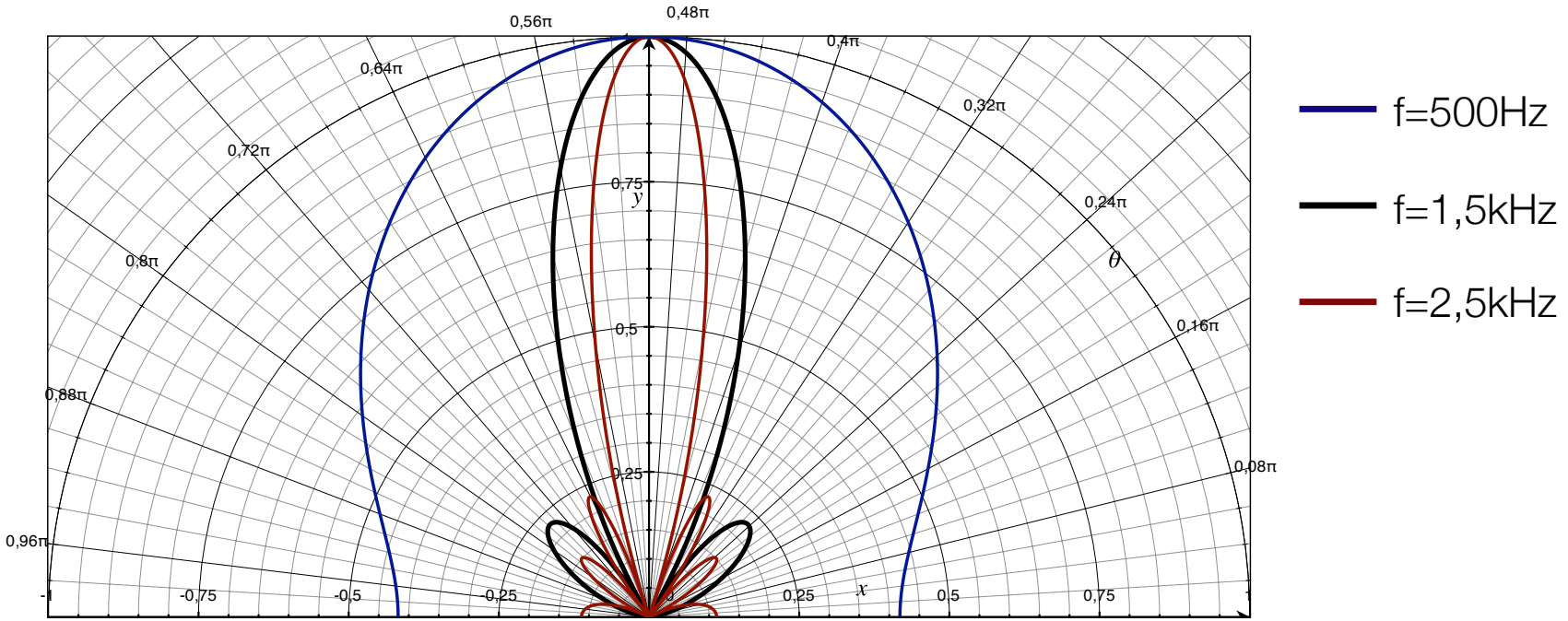


Influence of L

The spatial filtering is more selective for a wide array.

Influence of the array geometry and frequency on the beampattern

$\rightarrow$ $N = 8$ microphones, and $L = 39.6$cm.

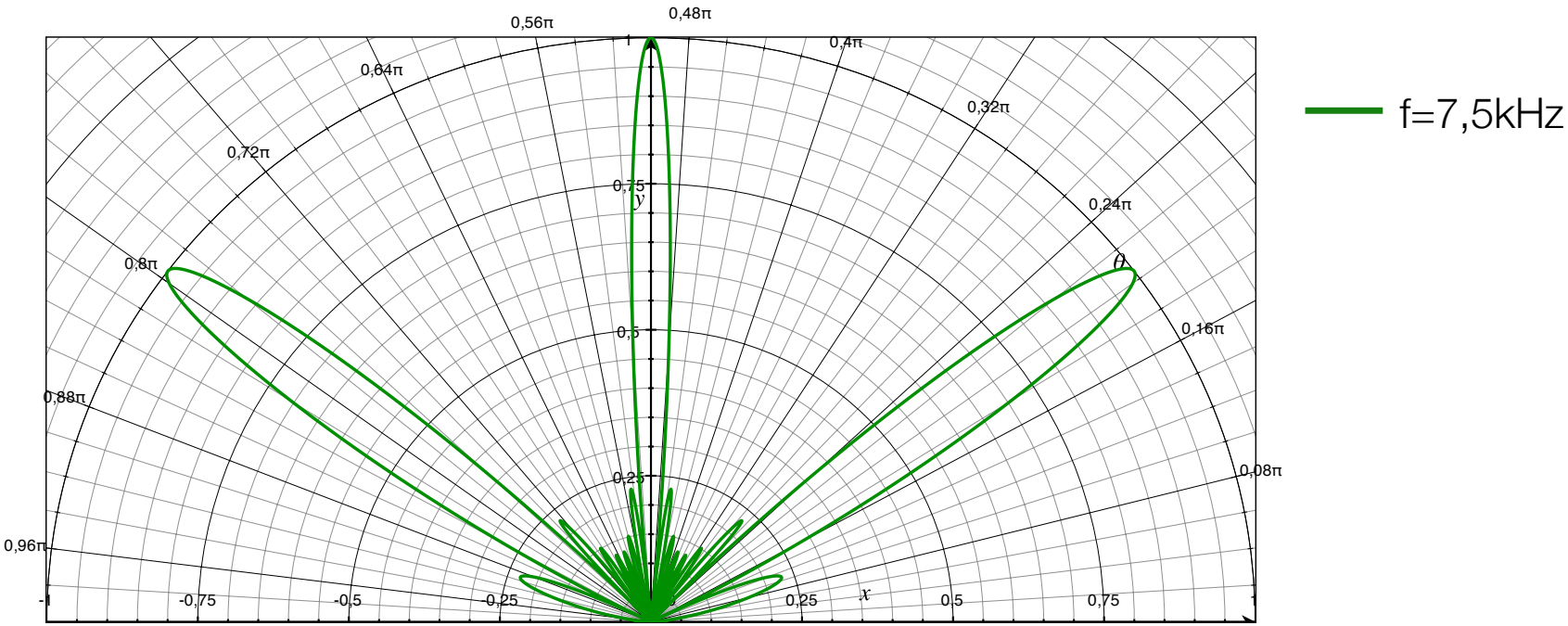


Influence of f

The spatial selectivity is (far) lower for low frequencies.

Influence of the array geometry and frequency on the beampattern

$\rightarrow$ $N = 8$ microphones, and $L = 39.6$cm.

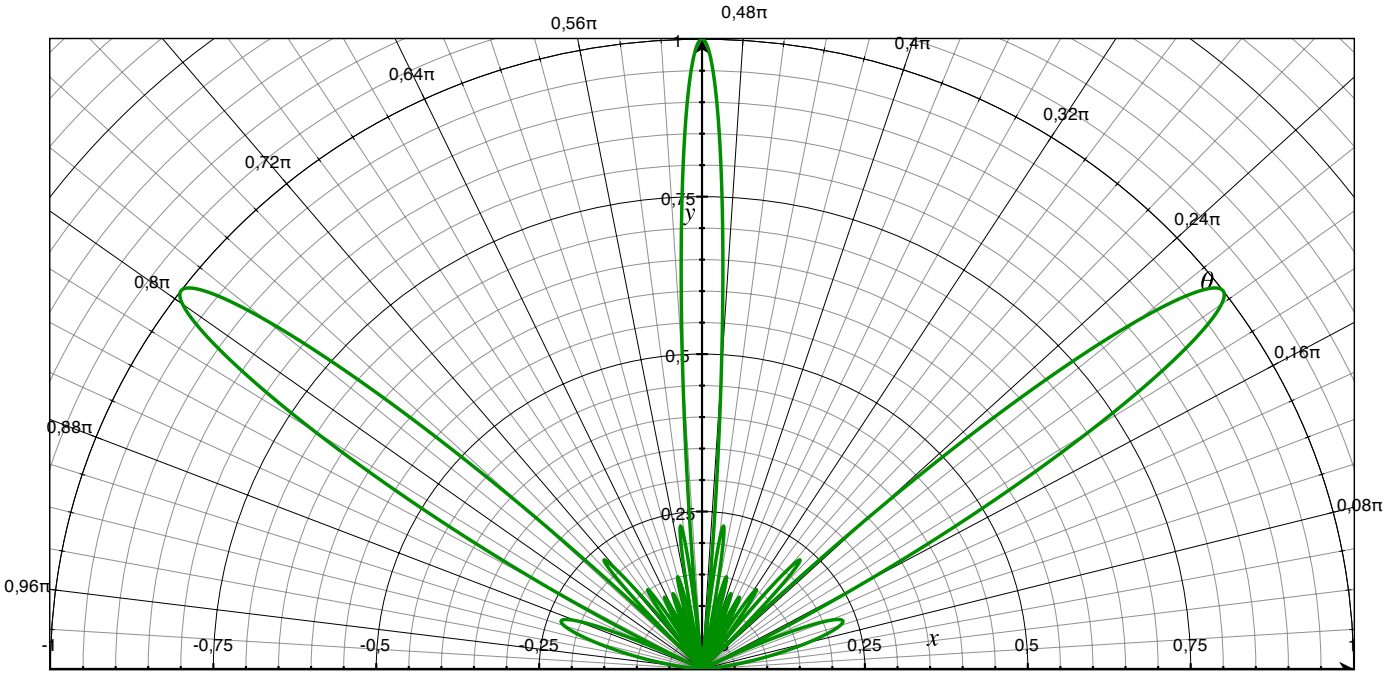


Influence of f

Be careful with spatial aliasing!

Influence of the array geometry and frequency on the beampattern

$\rightarrow$ $N = 8$ microphones, and $L = 39.6$cm.

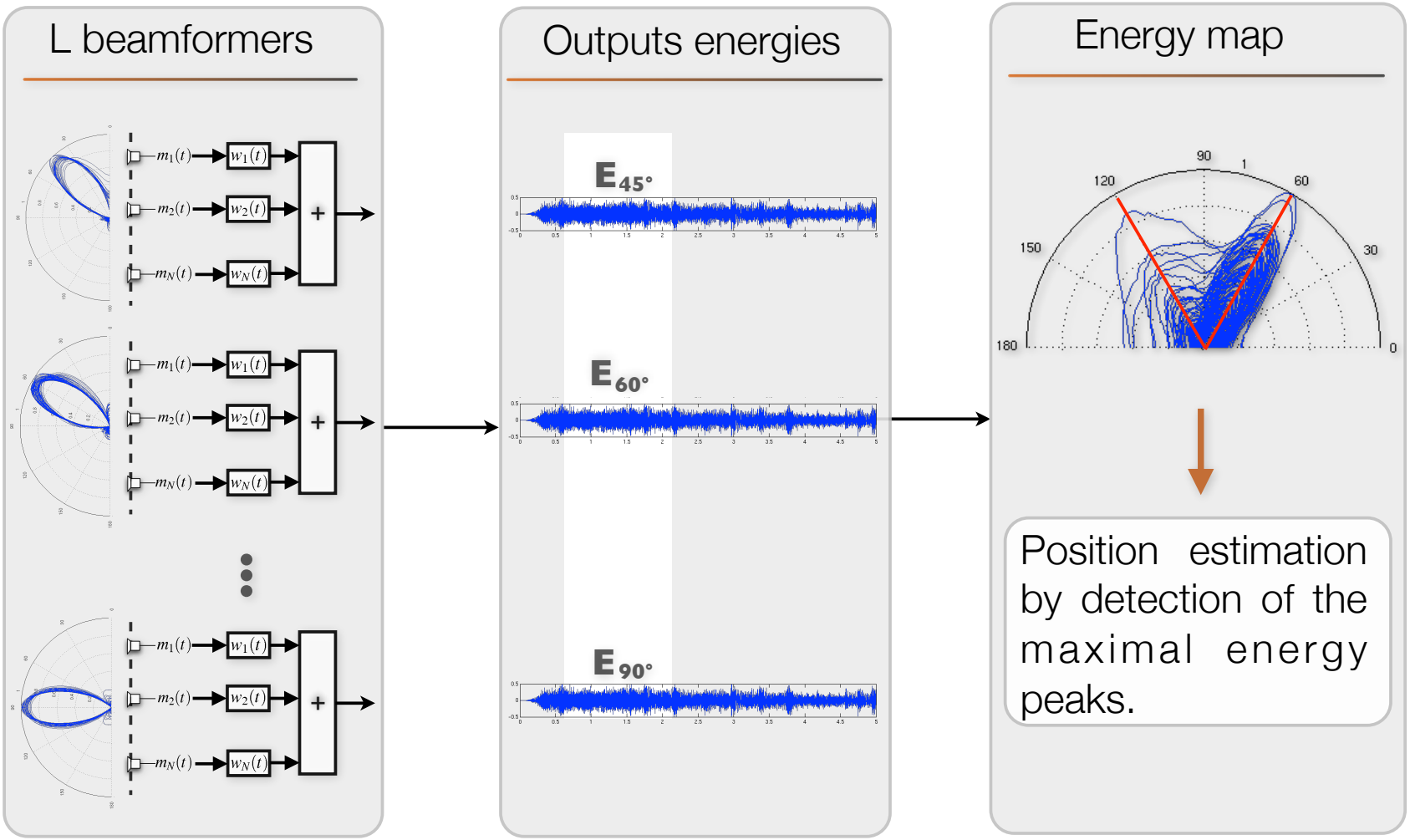


- array size $\leftrightarrow$ observation window of a signal
- microphone interspace d $\leftrightarrow$ sampling time for a discrete signal

Sampling theorem

$$d < d_{\text{max}} = \frac{\lambda_{\text{min}}}{2} = \frac{c}{2f_{\text{max}}}. \tag{48}$$

Using a "traditional" beamformer for localization



→ **How can we polarize an array in such directions?**

Using a "traditional" beamformer for localization

Using filters $W_n(k) = 1$ made an array mainly sensitive to frontal positions: **why is that?**

$\rightarrow$ it allows a *constructive* sum of the filters outputs only when a source is coming from the front, i.e. only when there is **no delay** between all microphones outputs.

To polarize the (linear) array towards the direction θ_0, one needs to compensate for the delays caused by the propagation from direction θ_0. Then:

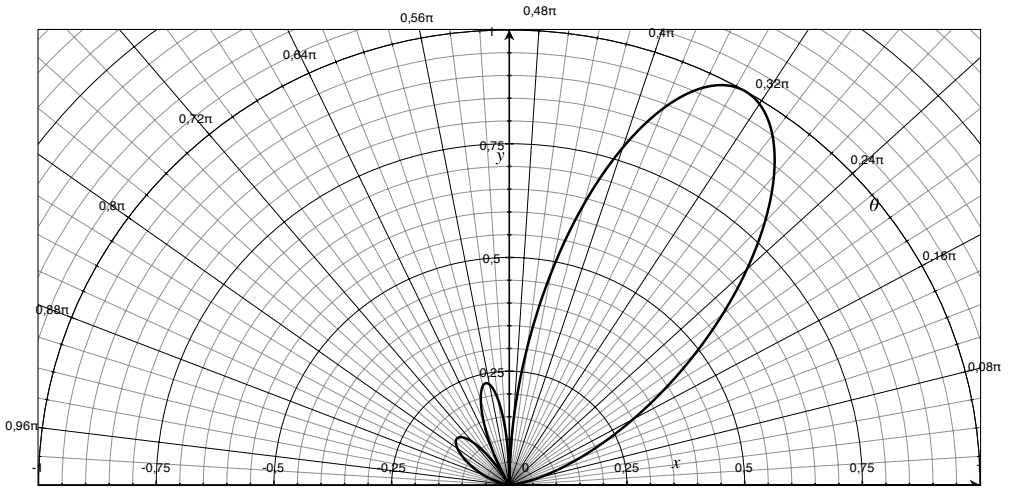
$$W_n(k) = e^{jkz_N \cos\theta}, \tag{49}$$

which defines a pure phase shifter. The farfield beampattern, for an evenly spaced linear array, then writes as

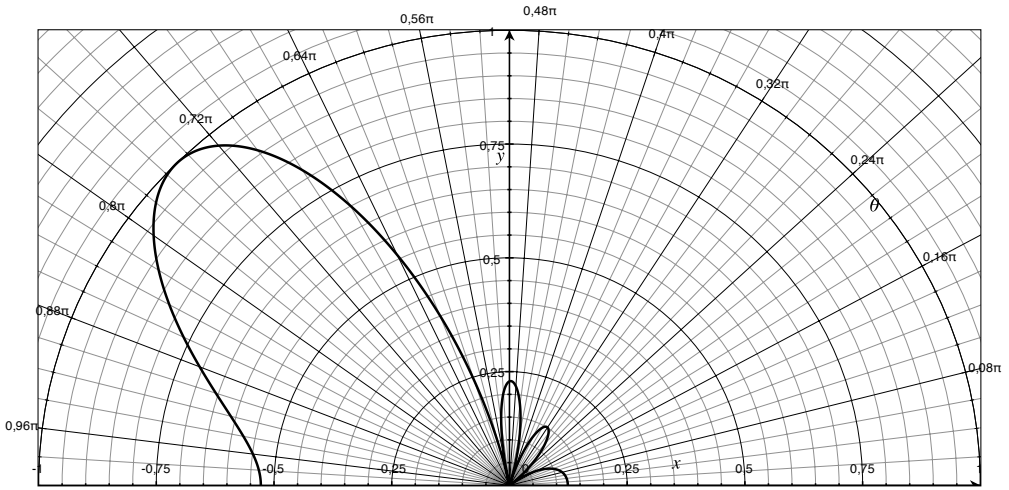
$$D^{\infty}(\theta) = \frac{\sin\left(\frac{\pi f}{c} Nd(\cos\theta_0 - \cos\theta)\right)}{\sin\left(\frac{\pi f}{c} d(\cos\theta_0 - \cos\theta)\right)} \tag{50}$$

Using a "traditional" beamformer for localization

Some examples, for $N = 8$ microphones array of length $L = 39.6$cm at frequency $f = 1500$Hz:



$\theta_0 = 60°$



$\theta_0 = 135°$

Beamforming techniques: conclusion

SCIENCES SORBONNE UNIVERSITÉ

Pros:

- simple approach, only relying on linear filters placed at the output of each microphone
- limited computational requirements, well-adapted to low-powered / specialized hardware (FPGA)
- can easily be used in real time
- can also be used for sound source extraction (naive approach)

Cons:

- the spatial filtering is really effective only for very wide arrays
- spatial filtering performances very poor for low frequencies
- generally requires a high number of microphones

(See Van Trees, 2002 for (far!) more details).

References

References I

Carter, G. C., A. H. Nuttall, and P. G. Cable (1973). "The smoothed coherence transform". In: *Proceedings of the IEEE* 61.10, pp. 1497–1498.

Hannan, E. J. and P. J. Thomson (1973). "Estimating Group Delay". In: *Biometrika* 60.2, pp. 241–253. issn: 00063444. url: http://www.jstor.org/stable/2334536.

Omologo, M. and P. Svaizer (1994). "Acoustic event localization using a crosspower-spectrum phase based technique". In: *Proceedings of ICASSP '94. IEEE International Conference on Acoustics, Speech and Signal Processing*. Vol. ii, II/273–II/276 vol.2.

Portello, Alban (Dec. 2013). "localisation binaurale active de sour es sonores en robotique humanoïde". PhD thesis. Université Paul Sabatier.

Roth, P. R. (1971). "Effective measurements using digital signal analysis". In: *IEEE Spectrum* 8.4, pp. 62–70.

Valin, J. -. et al. (2003). "Robust sound source localization using a microphone array on a mobile robot". In: *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*. Vol. 2, 1228–1233 vol.2.

Van Trees, Harry L. (2002). *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. Wiley. isbn: 978-0-47109390-9. doi: 10.1002/0471221104. url: https://doi.org/10.1002/0471221104.