Least-squares Distortionless Response Beamformer in Far-field Environments with Spatial Cues Preservation

Frédéric Mustière, *Integrated Device Technology Canada,* **Martin Bouchard,** *University of Ottawa*

Least-squares Beamformer Design

Basic design

- Design based **purely on knowledge of directivity vectors h** (frequency responses from source to sensors)
- **No need for noise statistics estimation** (e.g. correlation matrix), which can be difficult in complex, non-stationary environments
- Frequency domain design for a 1-D or 2-D array of M sensors, with source angle θ, desired directional response D, beamformer coefficients $\mathbf{w}$ and least-squares cost function J (frequency index is omitted for simplicity):

$$J(\mathbf{w}) = \int_0^{\pi} \left| D(\theta) - \mathbf{w}^H \mathbf{h}(\theta) \right|^2 d\theta$$

$$\frac{dJ(\mathbf{w})}{d\mathbf{w}^H} = \left[\int_0^{\pi} \mathbf{h}(\theta)\mathbf{h}^H(\theta) d\theta \right] \mathbf{w} - \int_0^{\pi} \mathbf{h}(\theta) D^*(\theta) d\theta = 0$$

$$\mathbf{Q} = \int_0^{\pi} \mathbf{h}(\theta)\mathbf{h}^H(\theta) d\theta \qquad \mathbf{p} = \int_0^{\pi} \mathbf{h}(\theta) D^*(\theta) d\theta \qquad \mathbf{w} = \mathbf{Q}^{-1}\mathbf{p}$$

- For a single target source, D is often formulated as: $D(\theta) = \begin{cases} 1, \theta_1 \le \theta \le \theta_2 \\ 0, \text{otherwise} \end{cases}$
- **In general, numerical methods are needed** to compute **Q** and **p** (no closed-form solutions). In addition, **p** is a function of the target direction.
- An arbitrary angle-dependent weighting function could also be added to the cost function.

Design with distortionless constraint $\mathbf{w}^H \mathbf{h}(\theta_{\text{target}}) = 1$:

$$J(\mathbf{w}) = \int_0^{\pi} \left| D(\theta) - \mathbf{w}^H \mathbf{h}(\theta) \right|^2 d\theta + \lambda \left(\mathbf{w}^H \mathbf{h}(\theta_{\text{target}}) - 1 \right) \qquad \frac{dJ(\mathbf{w})}{d\mathbf{w}^H}, \frac{dJ(\mathbf{w})}{d\lambda} = 0$$

$$\mathbf{w} = \mathbf{Q}^{-1}\left[\mathbf{p} - \lambda \mathbf{h}(\theta_{\text{target}}) \right] \qquad \lambda = \frac{\mathbf{h}^H(\theta_{\text{target}})\mathbf{Q}^{-1}\mathbf{p} - 1}{\mathbf{h}^H(\theta_{\text{target}})\mathbf{Q}^{-1}\mathbf{h}(\theta_{\text{target}})}$$

Special case of a linear array in free field

- With d_i ($0 \le i \le M$-1) as the distance between the i^{th} sensor and the first sensor, f as the frequency, c as the speed of sound, and θ=0 as the endfire direction, the directivity vectors **h** and the matrix **Q** become:

$$\mathbf{h}(\theta)[i] = \exp\left(-j2\pi f \frac{d_{i-1}}{c} \cos(\theta) \right) \qquad 1 \le i \le M$$

$$\mathbf{Q}[i,k] = \int_0^{\pi} \exp\left(-j2\pi f \frac{d_{i-1} - d_{k-1}}{c} \cos(\theta) \right) d\theta \qquad \mathbf{Q} = \pi J_0 \left(\frac{2\pi f}{c} \text{Toeplitz}(\mathbf{d}) \right)$$

$$1 \le i,k \le M$$

- J_0 is the Bessel function of the first kind. **Q can thus be computed analytically** in this case. **Q** also corresponds to a well-known result: the coherence function for cylindrically isotropic noise fields.
- However, **p must still be evaluated numerically** and **needs to be updated frequently** in the case of a moving target

$$\mathbf{p}[i] = \int_{\theta_1}^{\theta_2} \exp\left(-j2\pi f \frac{d_{i-1}}{c} \cos(\theta) \right) d\theta \qquad 1 \le i \le M$$

Alternative formulation leading to simpler, closed-form solution

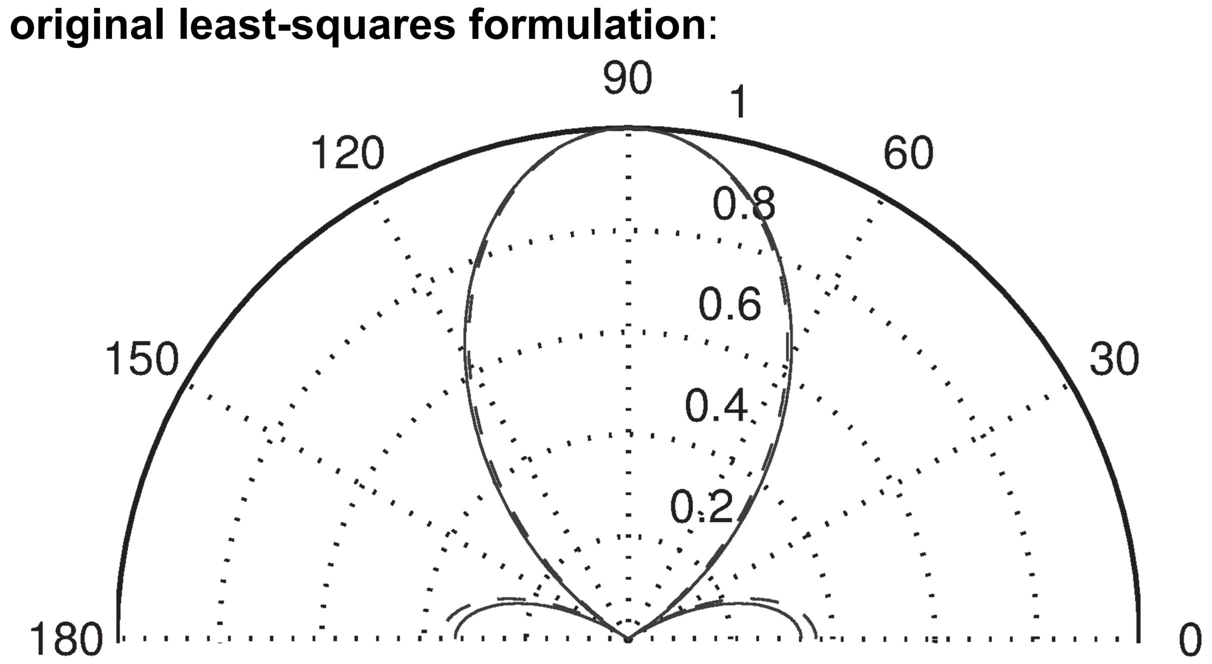
- Consider the use of the following alternative desired directional response, leading to a simplified expression for **p**, which **no longer requires the use of numerical methods**:

$$D(\theta) = \delta(\theta - \theta_{\text{target}}) \qquad \mathbf{p} = \mathbf{h}(\theta_{\text{target}})$$

- The resulting least-squares beamformer with distortionless constraint is:

$$\mathbf{w} = \mathbf{Q}^{-1}\left[\mathbf{p} - \frac{\mathbf{h}^H(\theta_{\text{target}})\mathbf{Q}^{-1}\mathbf{p} - 1}{\mathbf{h}^H(\theta_{\text{target}})\mathbf{Q}^{-1}\mathbf{h}(\theta_{\text{target}})} \mathbf{h}(\theta_{\text{target}}) \right] = \frac{\mathbf{Q}^{-1}\mathbf{h}(\theta_{\text{target}})}{\mathbf{h}^H(\theta_{\text{target}})\mathbf{Q}^{-1}\mathbf{h}(\theta_{\text{target}})}$$

- For the **special case of a linear array in free field**, **Q** corresponds to the Bessel function of the first kind as previously described (thus **Q** is available analytically). The above resulting least-squares design with distortionless constraint becomes **exactly the same as a MVDR beamformer design for cylindrically isotropic noise conditions**.
- The alternative least-squares formulation can produce **results similar to the original least-squares formulation**:

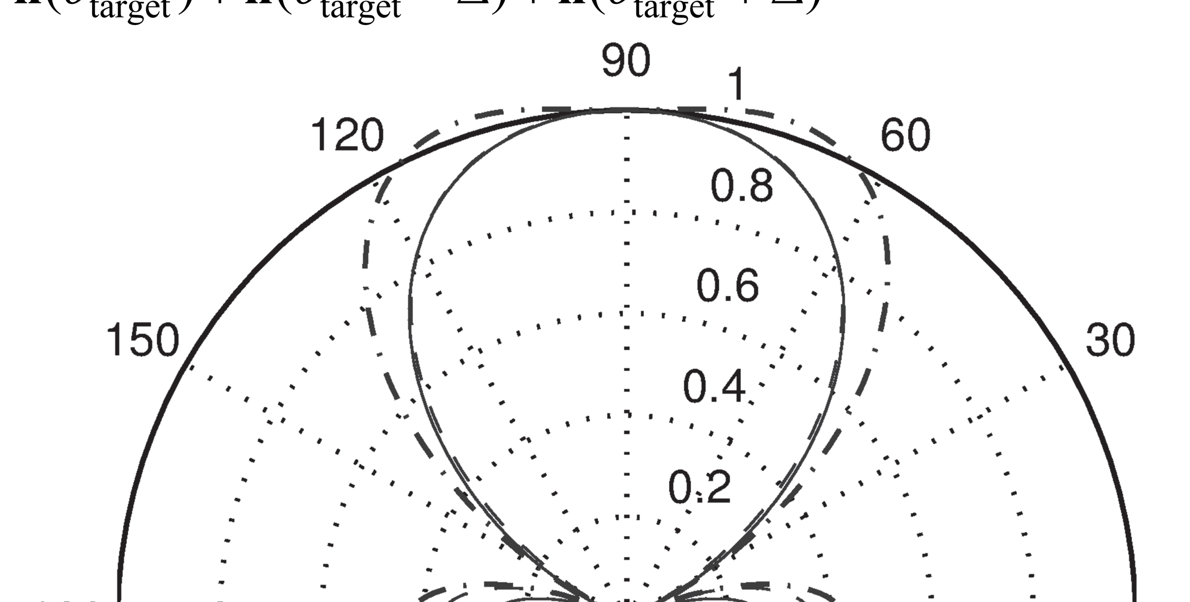


1.5 kHz, 4 microphones, linear array, 3 cm spacing, free field. Solid line: original formulation with desired directional response D set to 1.0 **between 70 and 110 degrees**. Dashed line: alternative formulation with desired directional response D set as frontal target.

- It is also **easy to obtain a wider main lobe** with the alternative formulation, using the following form where Δ determines the width of the beam:

$$D(\theta) = \delta(\theta - \theta_{\text{target}}) + \delta(\theta - \theta_{\text{target}} - \Delta) + \delta(\theta - \theta_{\text{target}} + \Delta)$$

$$\mathbf{p} = \mathbf{h}(\theta_{\text{target}}) + \mathbf{h}(\theta_{\text{target}} - \Delta) + \mathbf{h}(\theta_{\text{target}} + \Delta)$$



Solid line: original formulation with desired directional response D set to 1.0 **between 50 and 130 degrees**. Dashed line: alternative formulation with Δ adjusted to closely fit the response of the original formulation. Dashed-dot line: alternative formulation with a larger Δ value.

- Note that this is **no longer equivalent to a MVDR design for cylindrically isotropic noise (it is a new type of design)**
- It is also possible to develop a distortionless design for all of the specified directions ($\theta_{\text{target}}, \theta_{\text{target}} - \Delta, \theta_{\text{target}} + \Delta$), but in the previous figure the distortionless condition was only applied for the original direction θ_{target}.

Application for a challenging acoustic scenario with spatial cues preservation

- In some scenarios **perfect preservation of the spatial cues** is preferred and a simple solution is to apply to all input signals a **common gain** G obtained from a beamformer solution:

$$G = \frac{\left| \mathbf{w}^H \mathbf{z} \right|}{\left\| \mathbf{z} \right\|_1} \qquad \mathbf{w}^H \mathbf{h}(\theta_{\text{target}}) = \left\| \mathbf{h}(\theta_{\text{target}}) \right\|_1$$

- To obtain a unity gain G when the input signal is the target signal, it is required to perform the beamformer design with the standard target unity gain constraint replaced by the above constraint, leading to:

$$\mathbf{w} = \mathbf{Q}^{-1}\left[\mathbf{p} - \frac{\mathbf{h}^H(\theta_{\text{target}})\mathbf{Q}^{-1}\mathbf{p} - \left\| \mathbf{h}(\theta_{\text{target}}) \right\|_1}{\mathbf{h}^H(\theta_{\text{target}})\mathbf{Q}^{-1}\mathbf{h}(\theta_{\text{target}})} \mathbf{h}(\theta_{\text{target}}) \right]$$

- **Modulation of the enhancement strength** can be performed by using a modified gain, where larger α values will remove more noise but generate more distortion: $G^{\alpha}\ 0 \le \alpha \le 1$
- **Experiments** for a **challenging acoustic scenario** were performed:
  - 4-channel recordings of 4 different speakers accompanied with background subway noise (test data from the SISEC 2010 evaluation campaign, but with an additional speaker added).
  - The DOAs of each speaker are 20, 85, 115, and 140 degrees, assumed fixed over the 10-seconds long recordings.
  - For the experiments the DOAs are estimated using an SNR-based angular spectra method.
  - The targets with the first two experimentally measured DOAs (at 19.2 and 88.3 degrees) were selected for extraction, using the alternative formulation of the least-squares beamformer with distortionless constraint.
- Since we had no access to the clean speech files at DOAs 20 and 85 degrees, no objective results are reported here; nevertheless the reader is invited to **listen to the results** available at: www.eecs.uottawa.ca/%7Ebouchard/papers/Icassp13%5FLSB%5Fresults.zip
- The two **extracted speakers are significantly more intelligible** than in the crowded, noisy mixture, confirming the usefulness of the above approach in complex situations.
- For further noise reduction, we have also included results with an additional common gain post-processing derived from a log-MMSE criterion.
- The **two main advantages** of the method used here are:
  - the design is done the same way regardless of the noise complexity, and it is mostly invariant to noise changes
  - the design is fast and efficient.