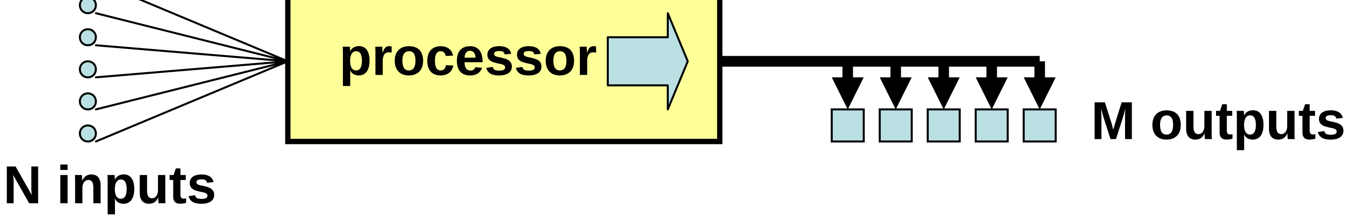
Advanced beamforming techniques with microphone arrays

A. Farina, S. Fontana, P. Martignon, A. Capra, C. Chiari
Industrial Engineering Dept., University of Parma, Italy

Journées d'étude sur la spatialisation (JES2006)

IRCAM (24/01/06) - Télécom Paris (25/01/06)

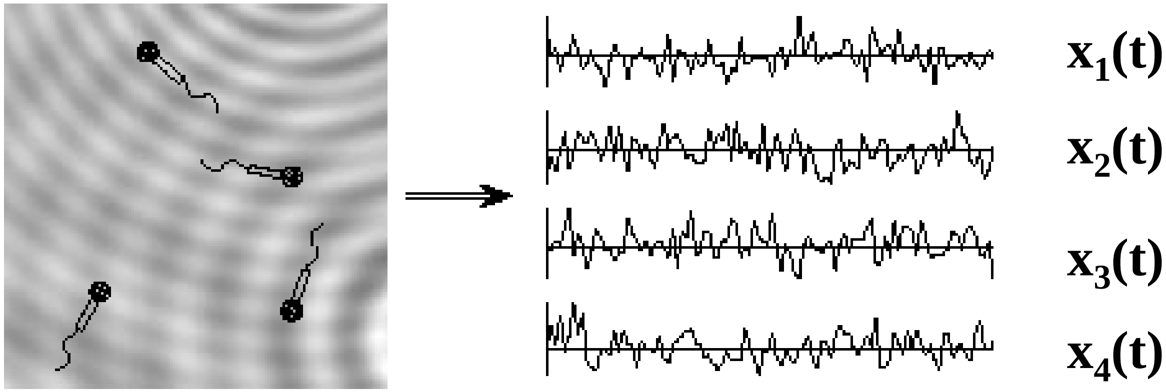
General Approach



- Whatever theory or method is chosen, we always start with N microphones, providing N signals x_i, and we derive from them M signals y_j
- And, in any case, each of these M outputs can be expressed by:

$$y_j = \sum_{i=1}^{N} h_{ij} \otimes x_i$$

Microphone arrays: target, processing



Acoustic field   Sampled acoustic field

Pre-calculated filters

$$y_j = \sum_{i=1}^{N} h_{ij} \otimes x_i$$

$y_j(t)$ Is the time-domain sampled waveform of a wave with well defined spatial characteristics, for example:

- a spherical wave centered in a precise emission point P_{source}
- a plane wave with a certain direction
- a spherical harmonic referred to a receiver point P_{rec}

Traditional approaches

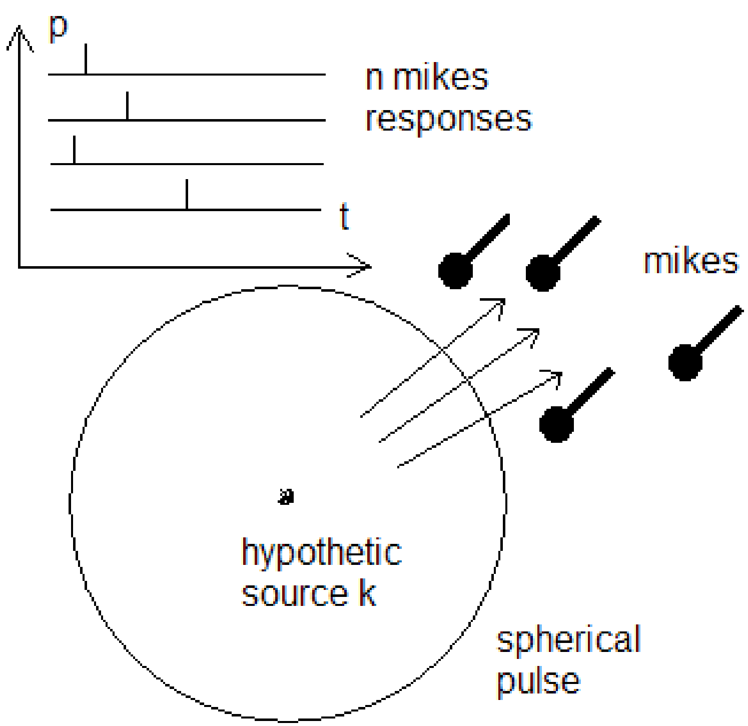
- The processing filters h_{ij} are usually computed following one of several, complex mathematical theories, based on the solution of the wave equation (often under certaing simplifications), and assuming that the microphones are ideal and identical
- In some implementations, the signal of each microphone is processed through a digital filter for compensating its deviation, at the expense of heavier computational load

Novel approach

- No theory is assumed: the set of h_{ij} filters are derived directly from a set of impulse response measurements, designed according to a least-squares principle.
- In practice, a matrix of filtering coefficients, is formed, and the matrix has to be numerically inverted (usually employing some regularization technique).
- This way, the outputs of the microphone array are maximally close to the ideal responses prescribed
- This method also inherently corrects for transducer deviations and acoustical artifacts (shielding, diffractions, reflections, etc.)

Example: focusing a point source

Considering all the $c_{k,i}$, that is the hypothetic response of the ***i***



$$C = \begin{bmatrix} c_{1,1} & c_{2,1} & c_{k,1} & c_{P,1} \\ c_{1,2} & c_{2,2} & c_{k,2} & c_{P,2} \\ c_{1,i} & c_{2,i} & c_{k,i} & c_{P,i} \\ c_{1,N} & c_{2,N} & c_{k,N} & c_{P,N} \end{bmatrix}$$

…we get the **filters** needed for the processing from the matrix inversion:

$$h_{i,k} = H = \frac{1}{C}$$

Example: focusing a point source

There are two strategies for designing the inverse matrix: local and general.
With **local** strategy, we take into account separately the k^{th} set of responses, and we are not worried about how the system will respond to the other P possible sources.
So we can invert separately each of these N responses, so that:

$$h_{i,k} \otimes \; c_{k,i} = \delta$$ (Dirac's delta function)

Once these N inverse filters are computed (for example, employing the Nelson/Kirkeby method), the output of the microphone array, focused on the k^{th} point source, will be simply:

$$y_k = x_1 \otimes h_{1,k} + x_2 \otimes h_{2,k} + \dots + x_N \otimes h_{N,k}$$

Example: focusing a point source

With **global** strategy, we design the filters in such a way that the response of the system is a Dirac's delta function for source K, and is zero for all the other P-1 source positions.
So we set up a linear equation system of P equations, imposing that:

$$\sum_{i=1}^{N} h_{i,K} \otimes c_{1,i} = 0$$

$$\sum_{i=1}^{N} h_{i,K} \otimes c_{2,i} = 0$$

....

$$\sum_{i=1}^{N} h_{i,K} \otimes c_{k,i} = \delta$$

....

$$\sum_{i=1}^{N} h_{i,K} \otimes c_{P,i} = 0$$

Lets call v_k the right-hand vector of known results.
Once this matrix of N inverse filters are computed (for example, employing the Nelson/Kirkeby method), the output of the microphone array, focused on the K point source, will again be simply:

$$y_K = \sum_{i=1}^{N} x_i \otimes h_{i,K}$$

System's least-squares inversion

- For computing the matrix of N filtering coefficients h_{ik}, a least-squares method is employed.
- A "total squared error" ε_{tot} is defined as:

$$\varepsilon_{tot} = \sum_{k=1}^{P} \left[\sum_{i=1}^{N} \left(h_{iK} \otimes c_{ki} \right) - v_k \right]^2$$

- A set of N linear equations is formed by minimising ε_{tot}, imposing that:

$$\frac{\partial \varepsilon_{tot}}{\partial h_{iK}} = 0 \qquad (i = 1 \ldots N)$$

Kirkeby's regularization

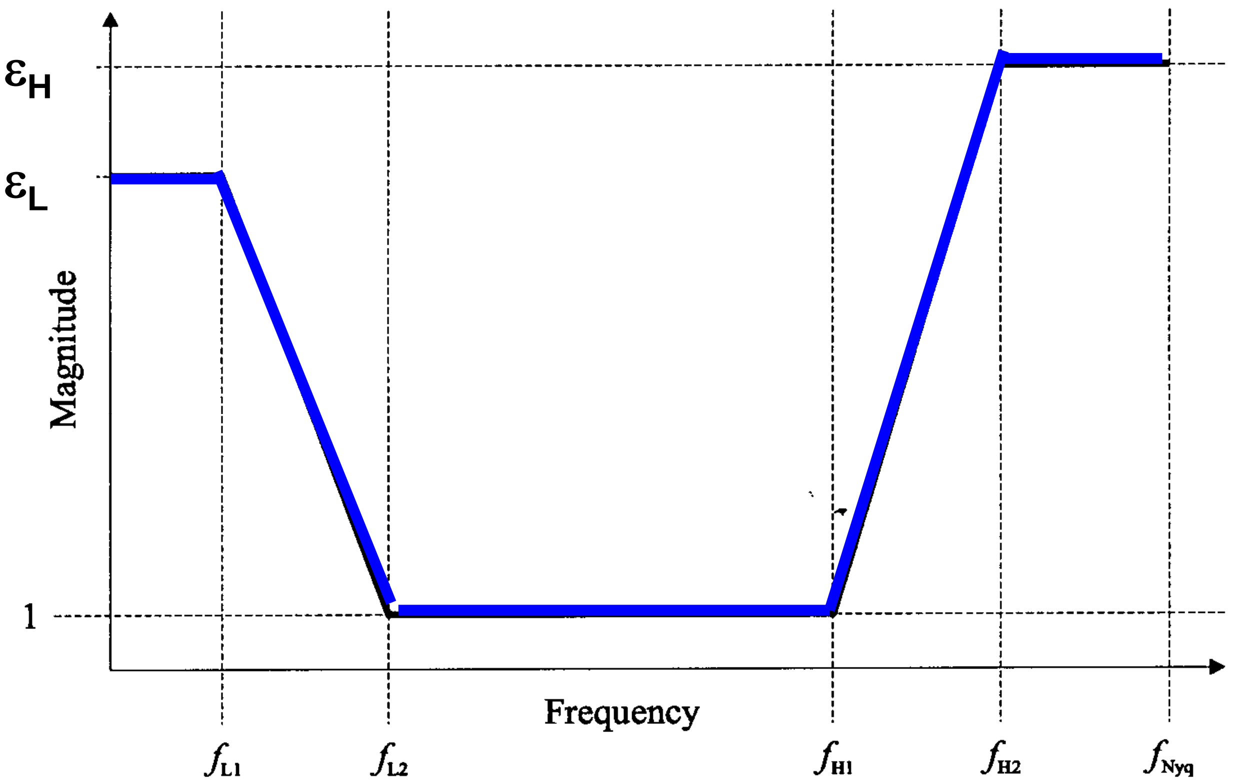
- During the computation of the inverse filter, usually operated in the frequency domain, one usually finds expressions requiring to compute a ratio between complex spectra (**H=A/D**).
- Computing the reciprocal of the denominator **D** is generally not trivial, as the inverse of a complex, mixed-phase signal is generally unstable.
- The Nelson/Kirkeby regularization method is usually employed for this task:

$$\mathrm{InvD}(\omega) = \frac{\mathrm{Conj}[\mathrm{D}(\omega)]}{\mathrm{Conj}[\mathrm{D}(\omega)] \cdot \mathrm{D}(\omega) + \varepsilon(\omega)}$$

$$\mathrm{H} = \mathrm{A} \cdot \mathrm{InvD}$$

Spectral shape of the regularization parameter $\varepsilon(\omega)$

- At very low and very high frequencies it is advisable to increase the value of ε.



Critical aspects

- LOW frequencies: wavelength longer than **array width** - no phase difference between mikes - local approach provide low spatial resolution (single, large lobe) - global approach simply fails (the linear system becomes singular)

- MID frequencies: wavelength comparable with **array width** -with local approach **secondary lobes** arise in spherical or plane wave detection (negligible if the total bandwidth is sufficiently wide) - the global approach works fine, suppressing the side lobes, and providing a narrow spot.

- HIGH frequencies: wavelength is shorter than twice the average **mike spacing** (Nyquist limit) - spatial undersampling - spatial aliasing effects – **random disposition** of microphones can help the local approach to still provide some meaningful result - the global approach fails again

Linear array

• 16 omnidirectional mikes mounted on a 1.2m aluminium beam, with exponential spacing

• 16 channels acquisition system:
2 Behringer A/D converters + RME Hammerfall digital sound card

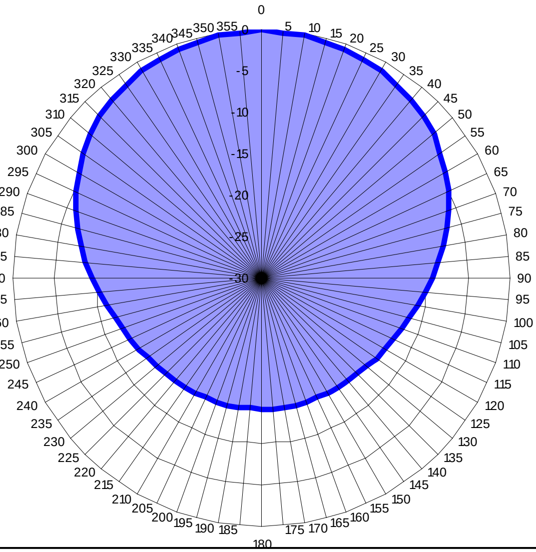
• Sound recording with Adobe Audition
• Filter calculation, off-line processing and visualization with Aurora plugins

Linear array - calibration

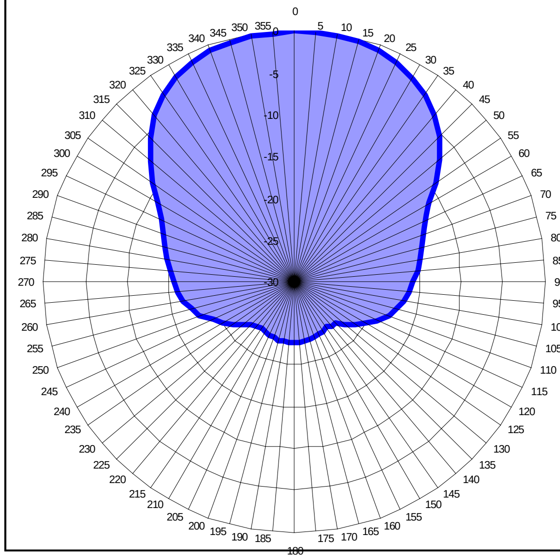
- The array was mounted on a rotating table, outdoor
- A Mackie HR24 loudspeaker was used
- A set of 72 impulse responses was measured employing Aurora plugins under Adobe Audition (log sweep method) - the sound card controls the rotating table.
- The inverse filters were designed with the local approach (separate inversion of the 16 on-axis responses, employing Aurora's "Kirkeby4" plugin)

Linear array - polar plots

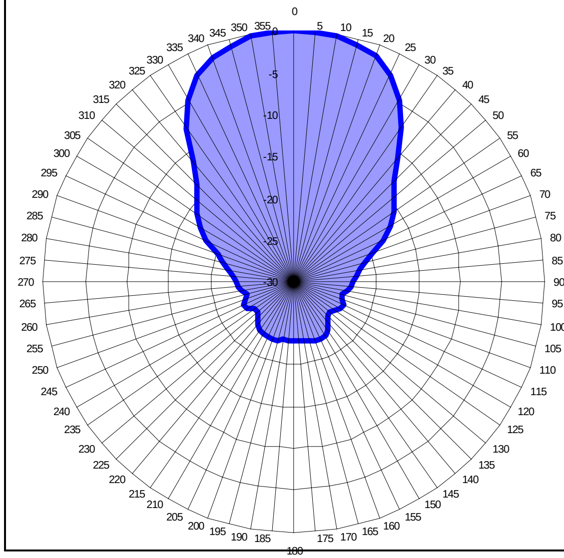
250 Hz



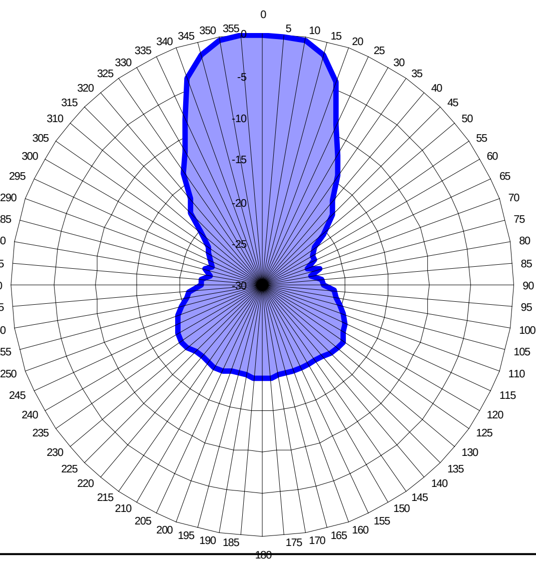
500 Hz



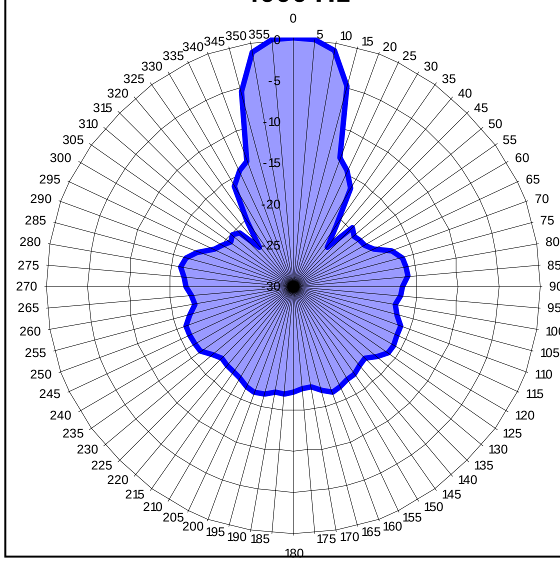
1000 Hz



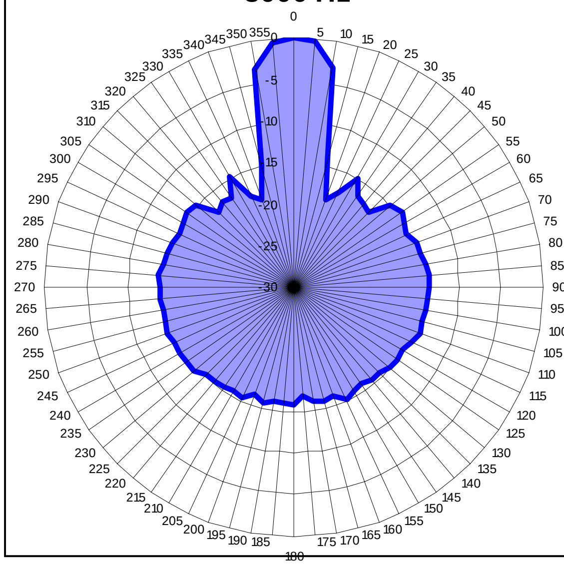
2000 Hz



4000 Hz



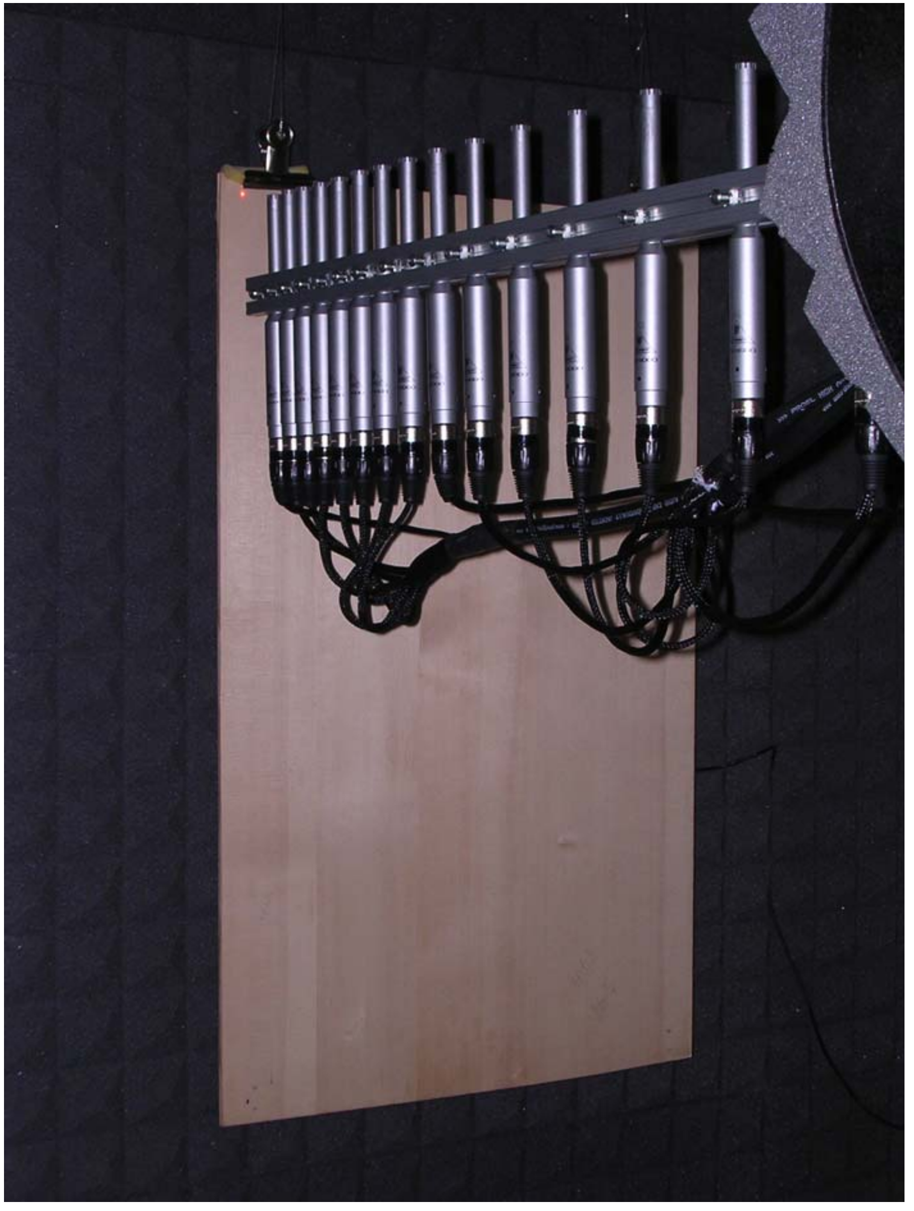
8000 Hz



Linear array - practical usage

- The array was mounted on an X-Y scanning apparatus
- a Polytec laser vibrometer is mounted along the array
- The system is used for mapping the velocity and sound pressure along a thin board of “resonance” wood (Abete della val di Fiemme, the wood employed for building high-quality musical instruments)
- A National Instruments board controls the step motors through a Labview interface
- The system is currently in usage at IVALSA (CNR laboratory on wood, San Michele all’Adige, Trento, Italy)

Linear array - practical usage



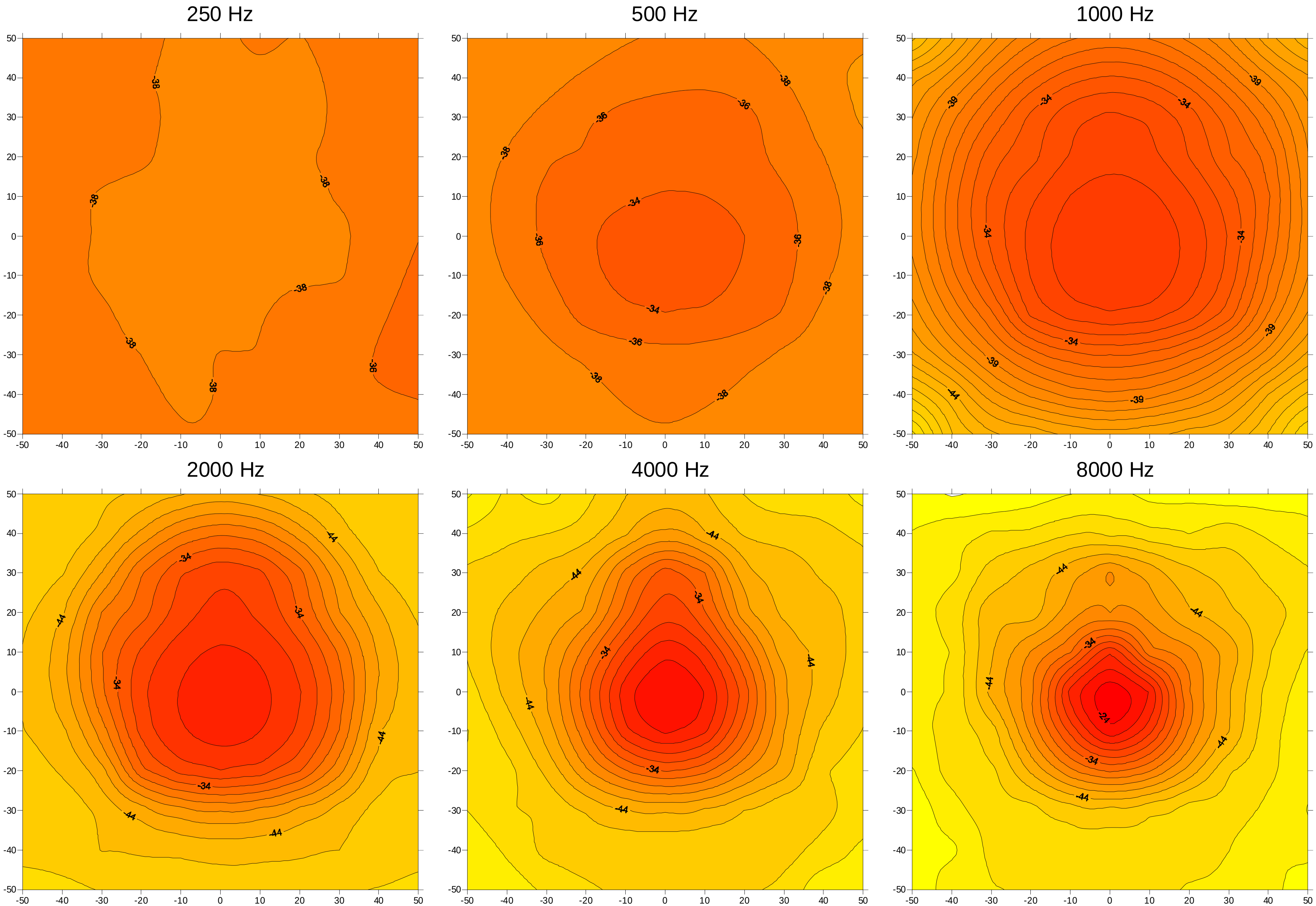
• The wood panel is excited by a small piezoelectric transducer

• When scanning a wood panel, two types of results are obtained:

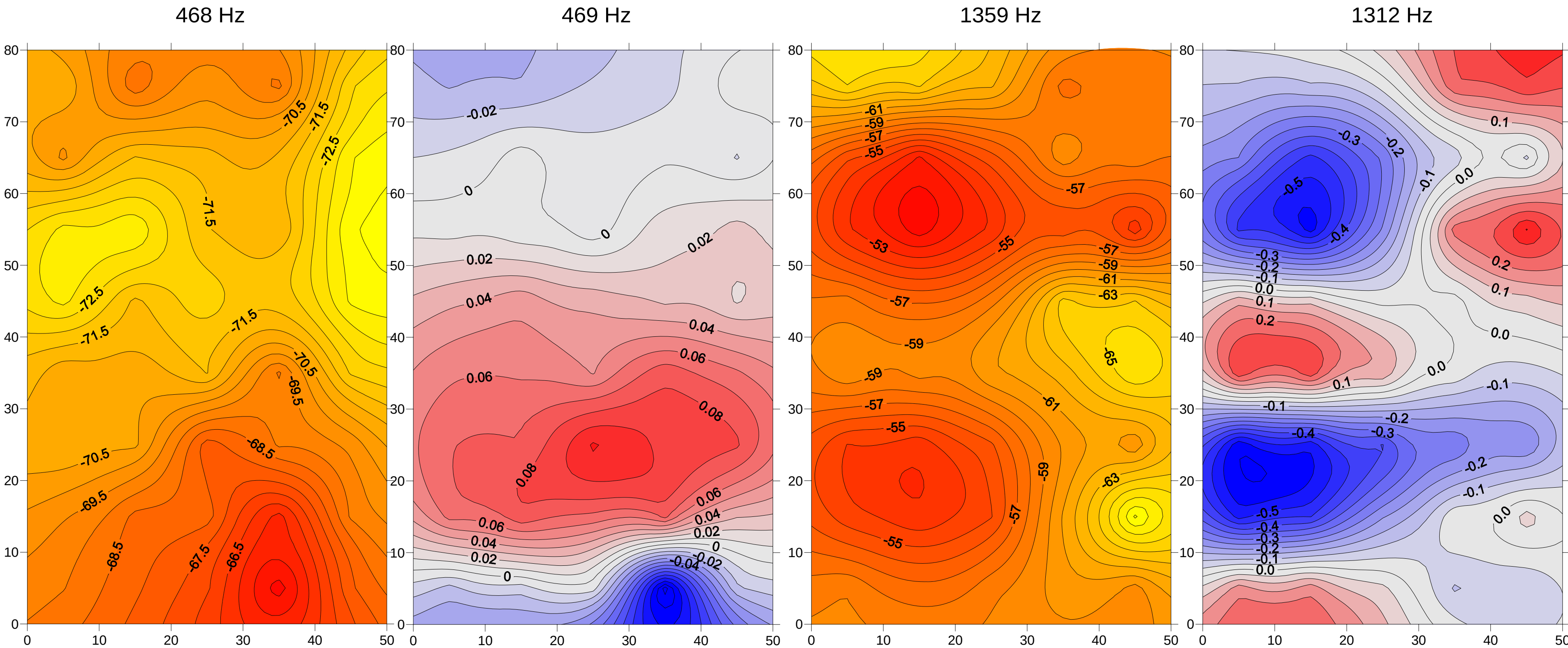
• A spatially-averaged spectrum of either radiated pressure, vibration velocity, or of their product (which provides an estimate of the radiated sound power)

• A colour map of the radiated pressure or of the vibration velocity at each resonance frequency of the board

Linear array - test results (small loudspeaker)

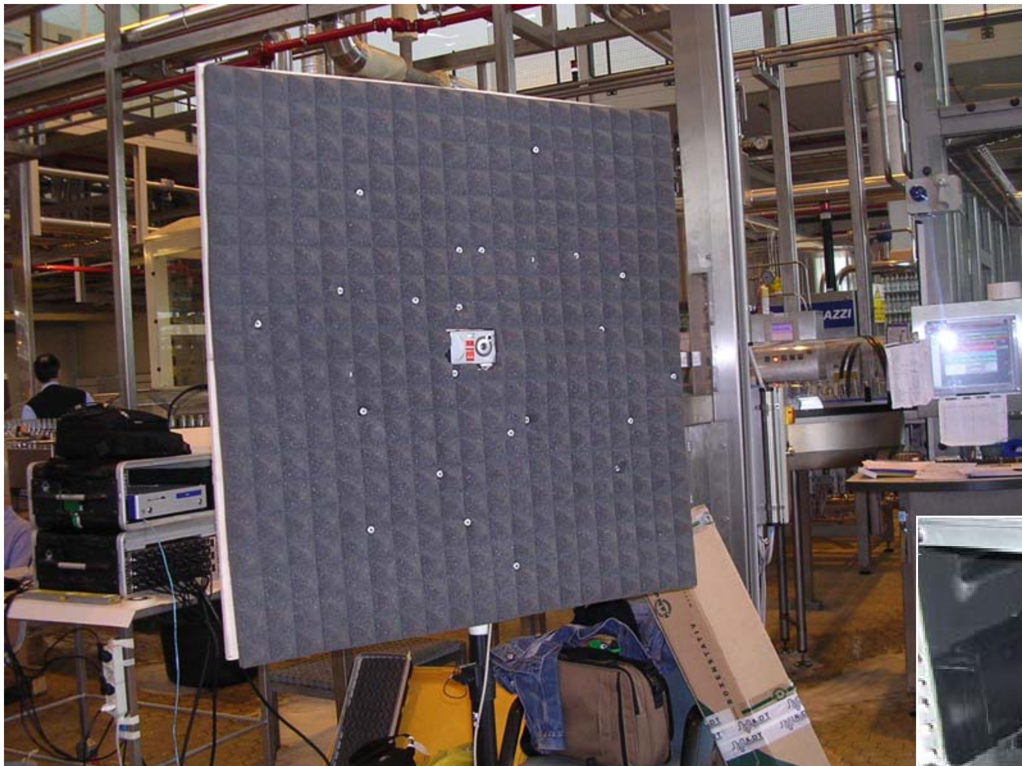


Linear array - test results (rectangular wood panel)



SPL (dB) **velocity (m/s)** **SPL (dB)** **velocity (m/s)**

Planar array (“acoustic camera”)

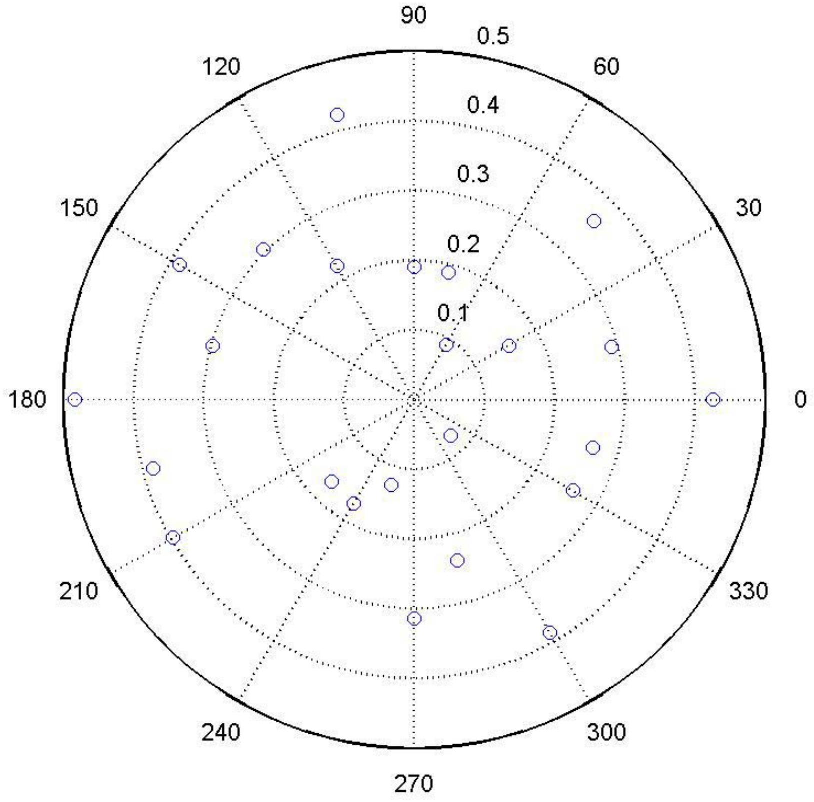


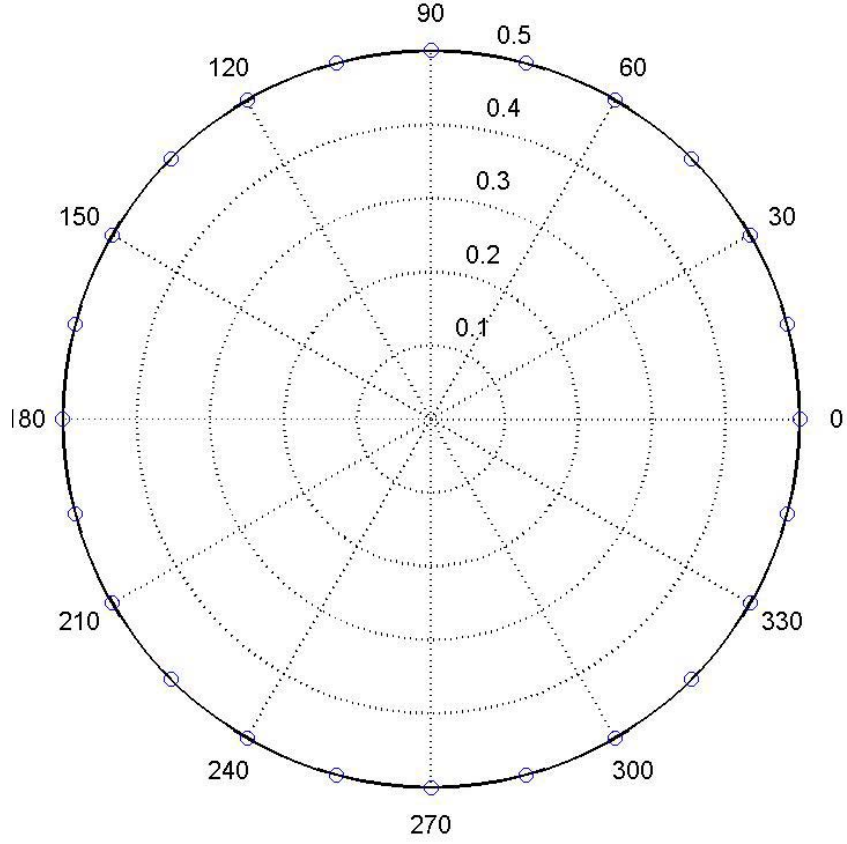
- 24 omnidirectional mikes mounted on a 1m square foldable wooden baffle with an optimized random disposition

- 24 channels acquisition system: Behringer A/D converters + RME Hammerfall digital audio card

- Filters calculation, off-line processing and visualization with MATLAB

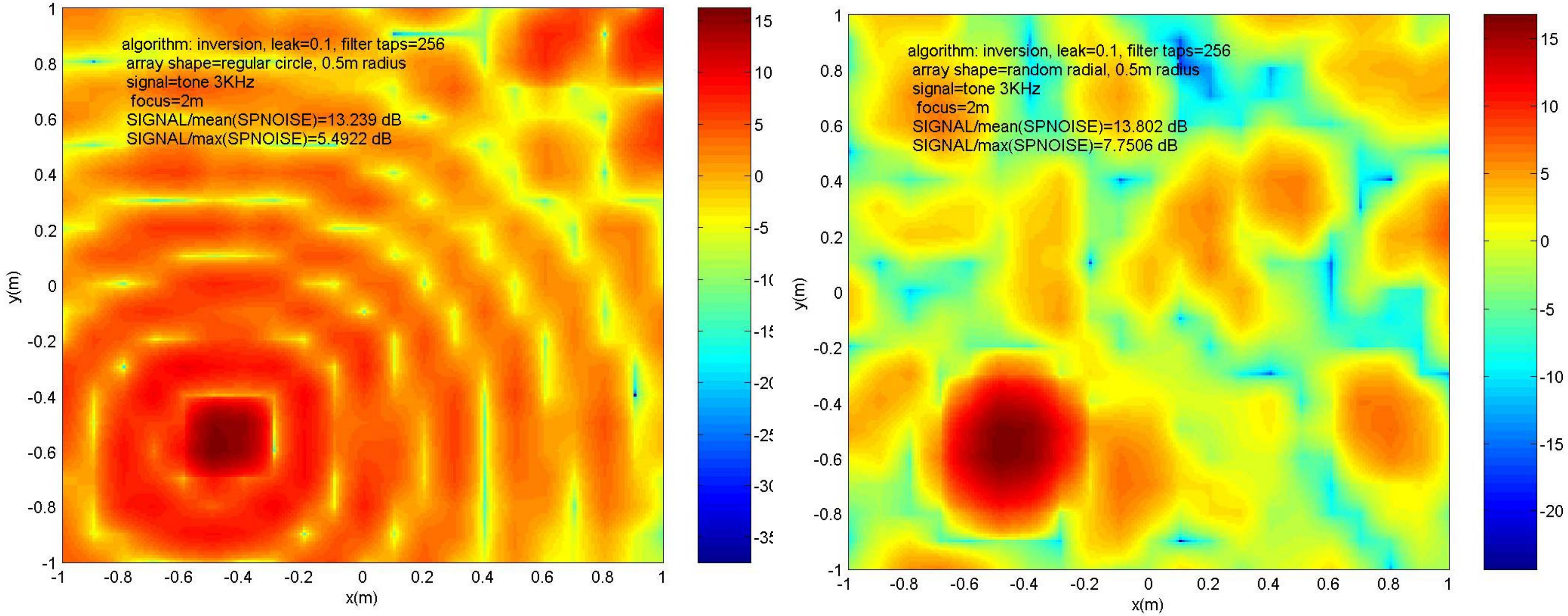
Random array vs. Circular array





Circular array vs. Random array

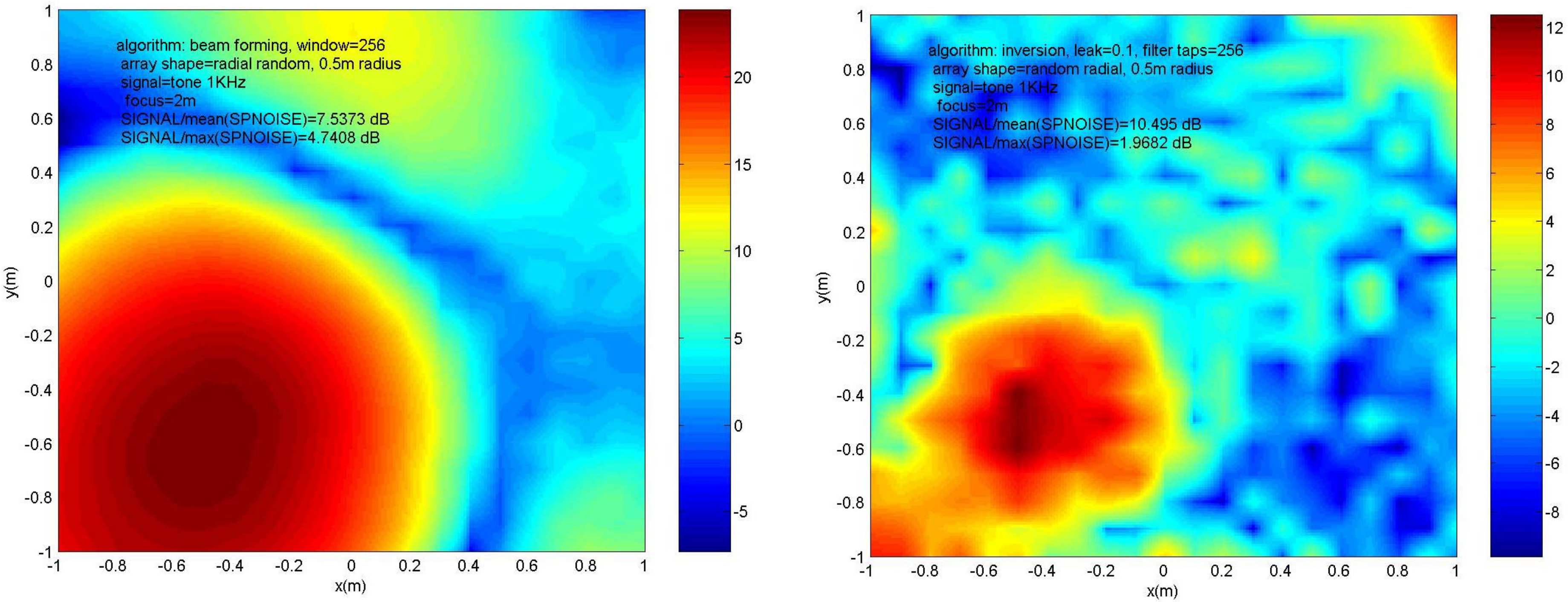
Simulated matrix inversion - 3 kHz



The optimal randomized positions were found by running 10000 Matlab simulations, and chosing the one providing the better peak-to-noise ratio

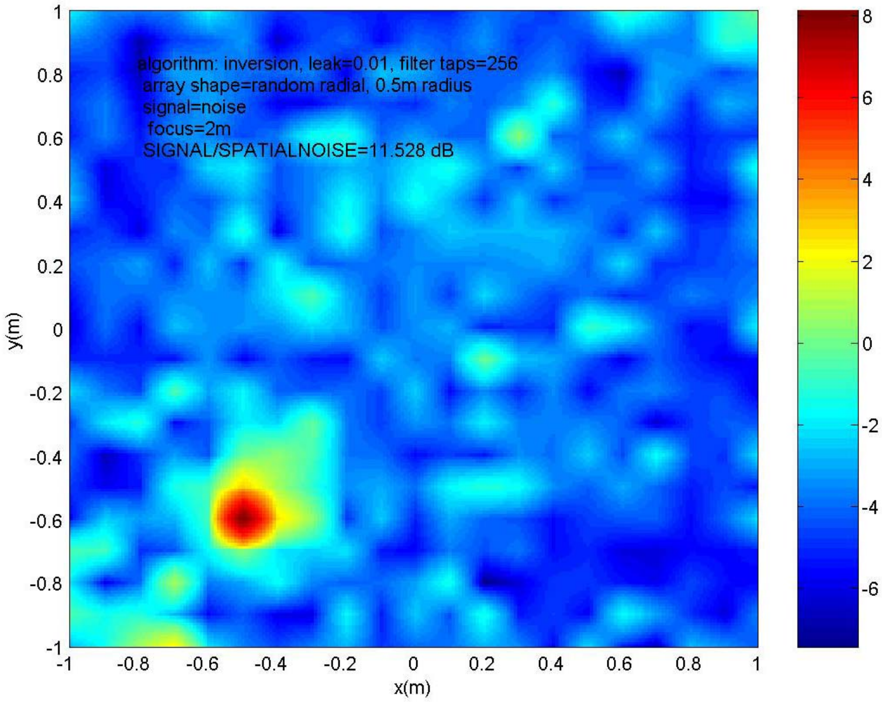
Beamforming vs. Inverse filters

Random Array - Matrix inversion - 1 kHz

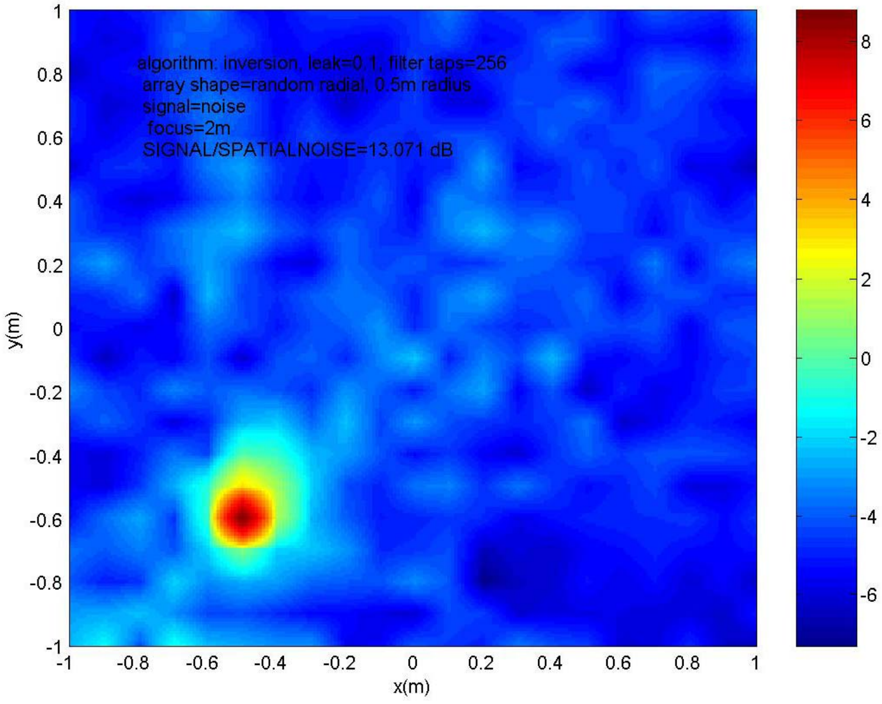


Effect of the regularization parameter

Random Array - Matrix Inversion - 5 kHz



ε = 0.01

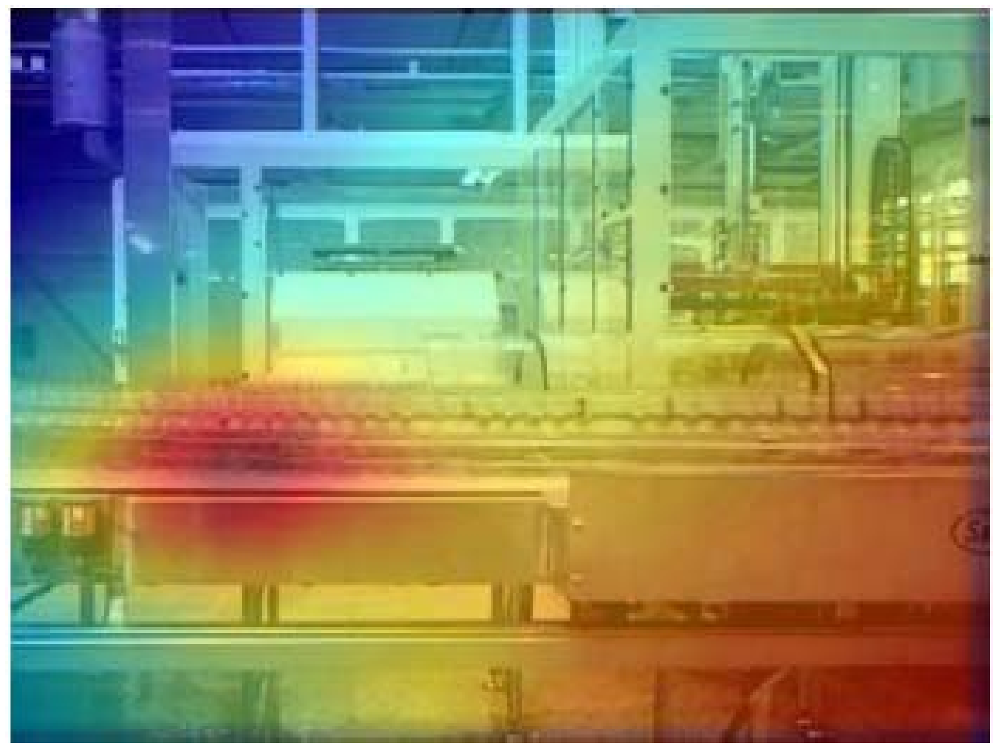


ε = 0.1

Indoor application (source localisation)

- The array is equipped with a webcam

- The filtering process discriminates each spherical wave (*Tr'$_k$(t)*) coming by the pixels of a virtual screen placed on the plane where machine to be tested lyes

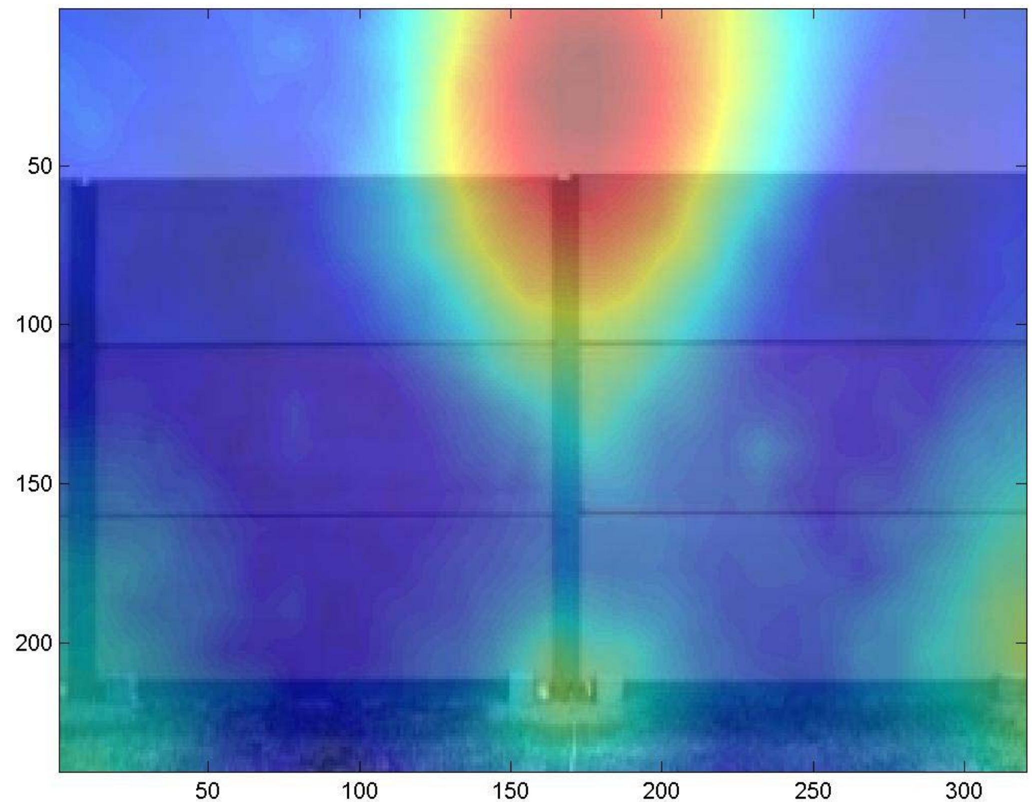


- At each pixel is associated the A-weighted equivalent level (L_{Aeq}, in dB) of each soundtrack extracted

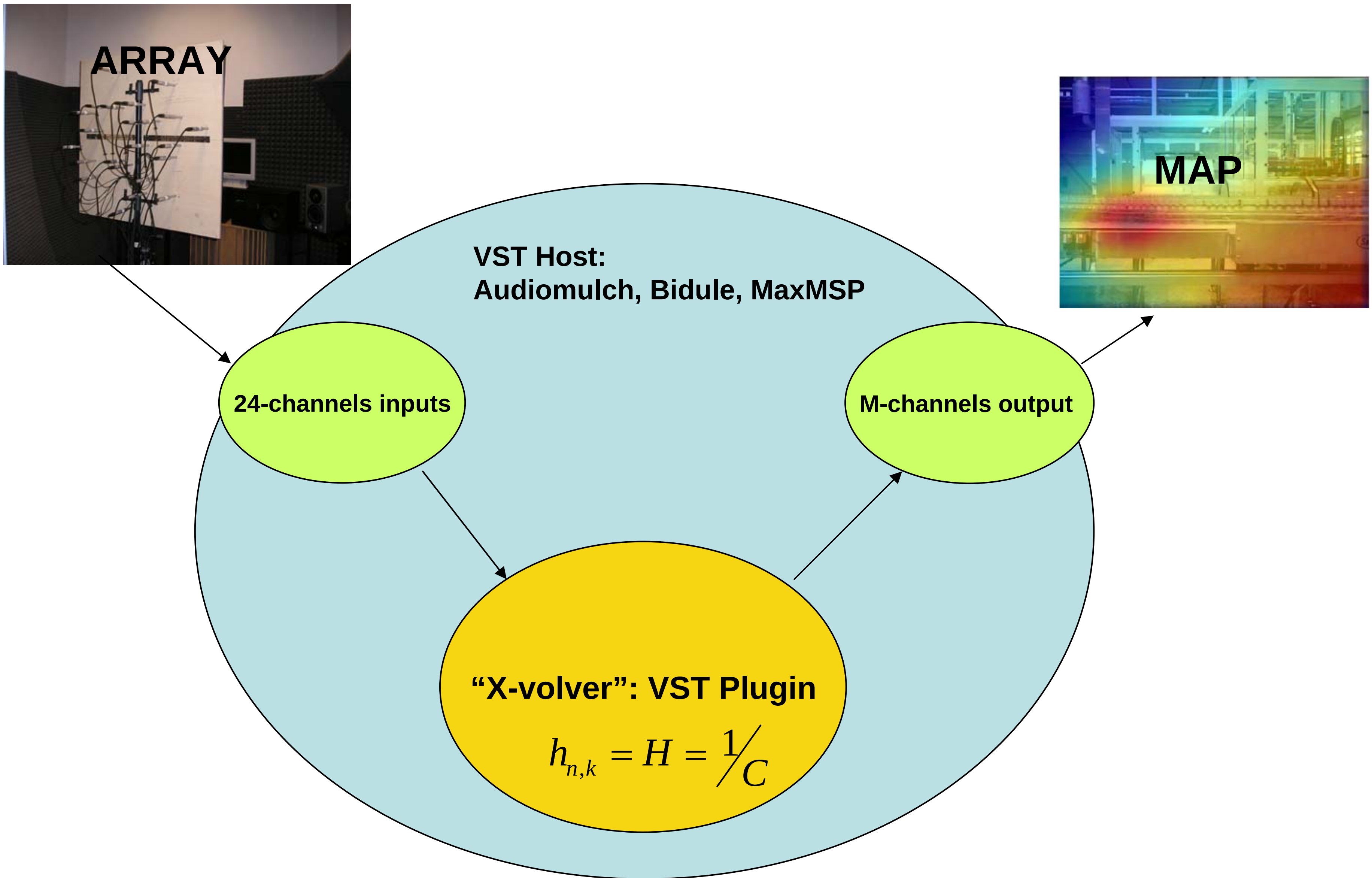
- The dB(A) color map is matched and superimposed to the webcam image

Outdoor application

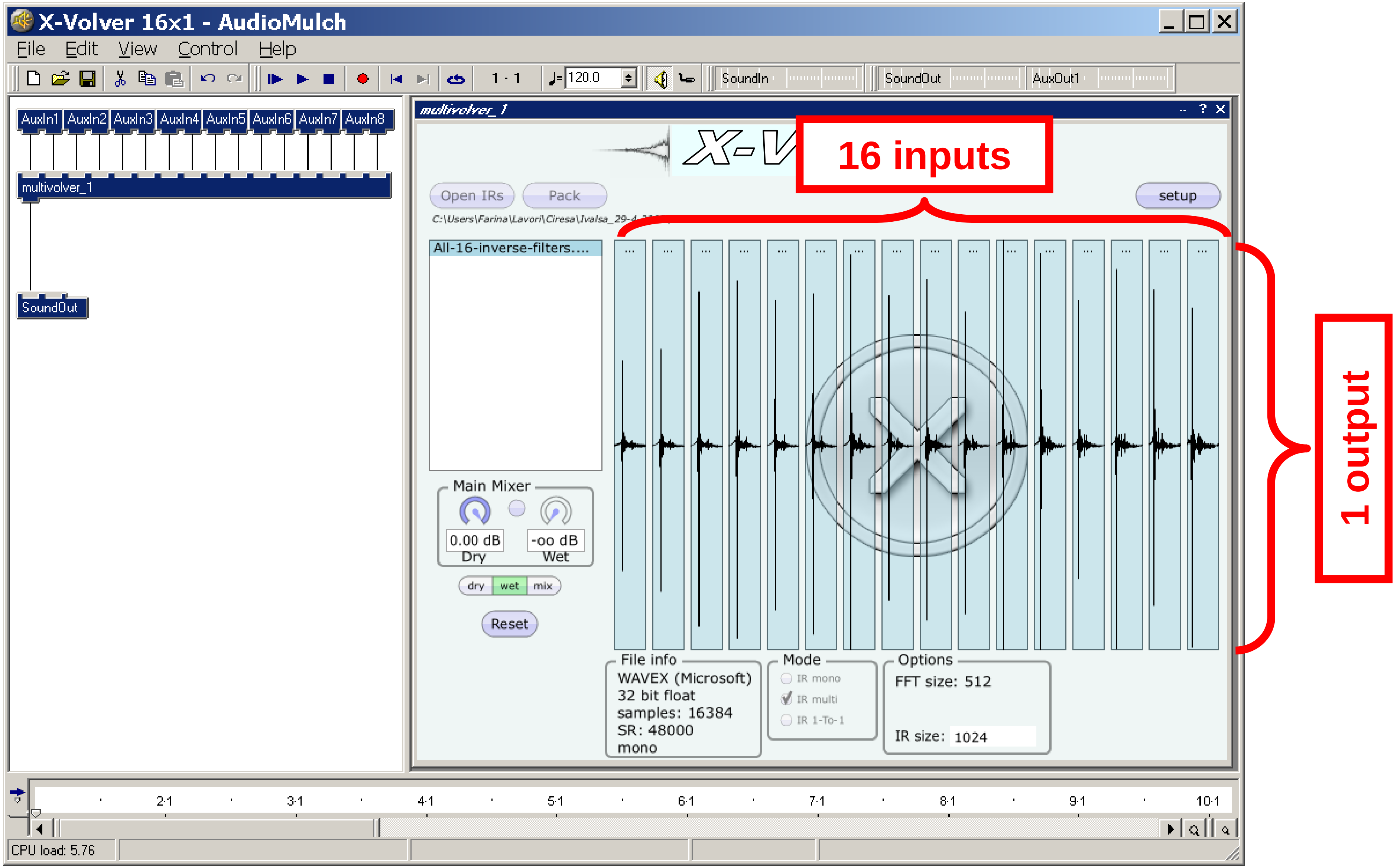
- The array is placed behind a noise barrier
- The colour map of the SPL shows the leakage under the barrier



Real time processing: work on progress



“X-volver” VST plugin



3D arrays

- DPA-4 A-format microphone
- 4 closely-spaced cardioids
- A set of 4x4 filters is required for getting B-format signals
- Global approach for minimizing errors over the whole sphere

IR measurements on the DPA-4

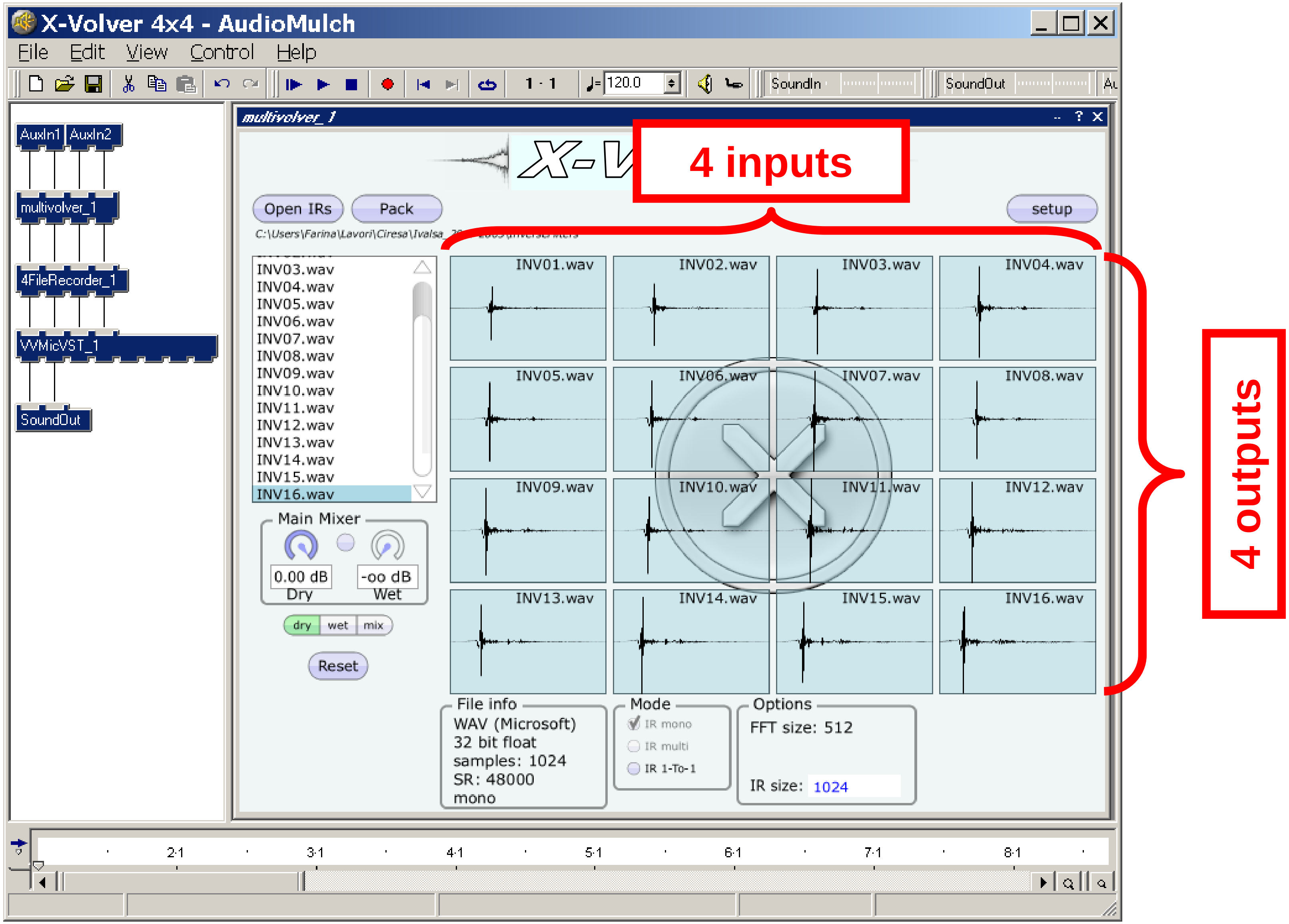
84 IRs were measured, uniformly scattered around a sphere

Computation of the inverse filters

- A set of 16 inverse filters is required (4 inputs, 4 outputs)
- For any of the 84 measured directions, a theoretical response can be computed for each of the 4 output channels (W,X,Y,Z)
- So 84x4=336 conditions can be set:

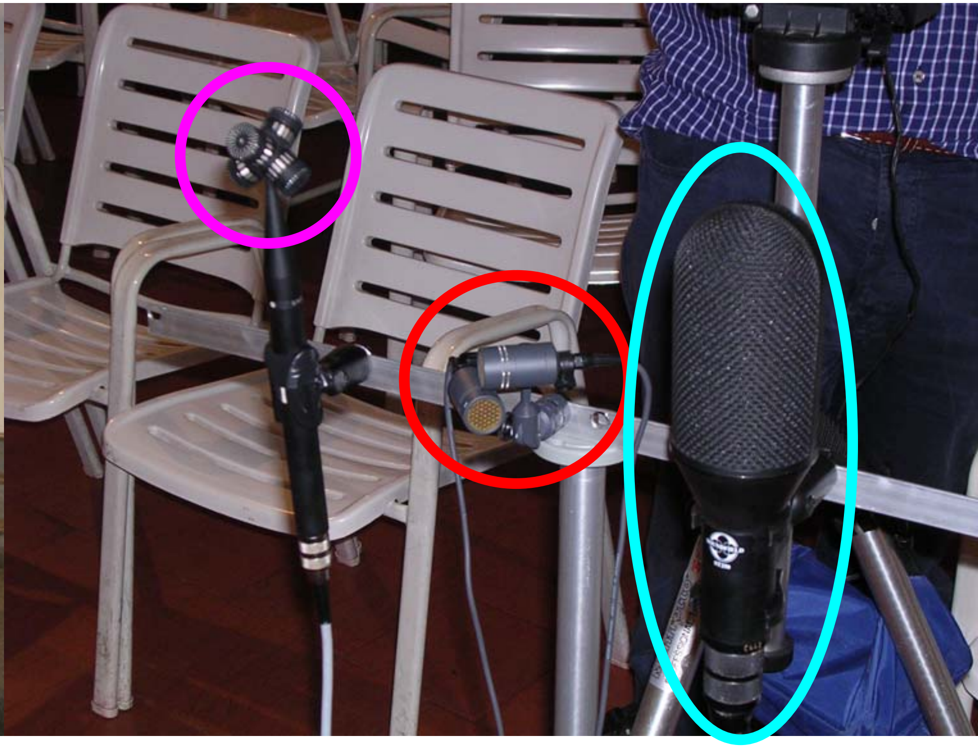
$$\left.\begin{array}{l}
c_1 \otimes h_{1,W} + c_2 \otimes h_{2,W} + c_3 \otimes h_{3,W} + c_4 \otimes h_{4,W} = out_{k,W} \\
c_1 \otimes h_{1,X} + c_2 \otimes h_{2,X} + c_3 \otimes h_{3,X} + c_4 \otimes h_{4,X} = out_{k,X} \\
c_1 \otimes h_{1,Y} + c_2 \otimes h_{2,Y} + c_3 \otimes h_{3,Y} + c_4 \otimes h_{4,Y} = out_{k,Y} \\
c_1 \otimes h_{1,Z} + c_2 \otimes h_{2,Z} + c_3 \otimes h_{3,Z} + c_4 \otimes h_{4,Z} = out_{k,W}
\end{array}\right\} \quad \mathbf{k = 1...84}$$

Real-time implementation

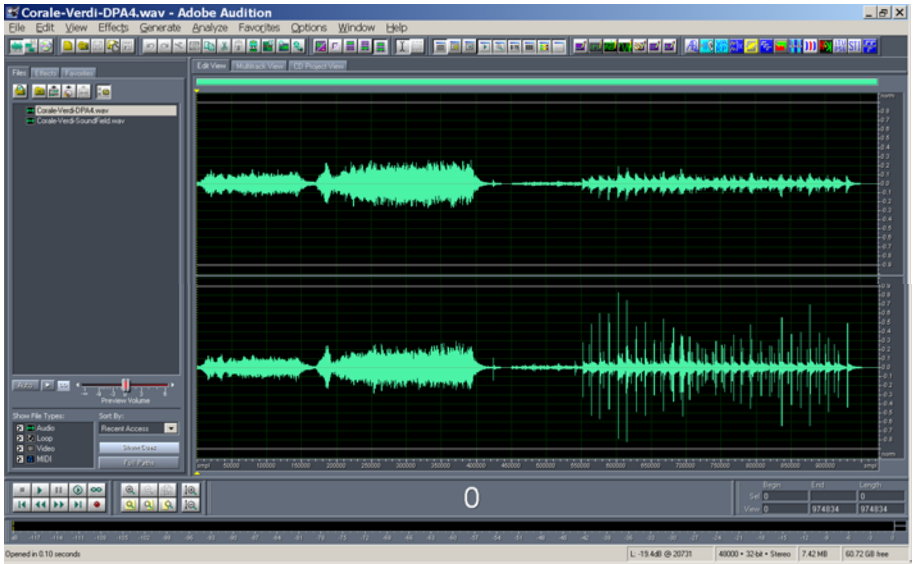


Microphone comparison

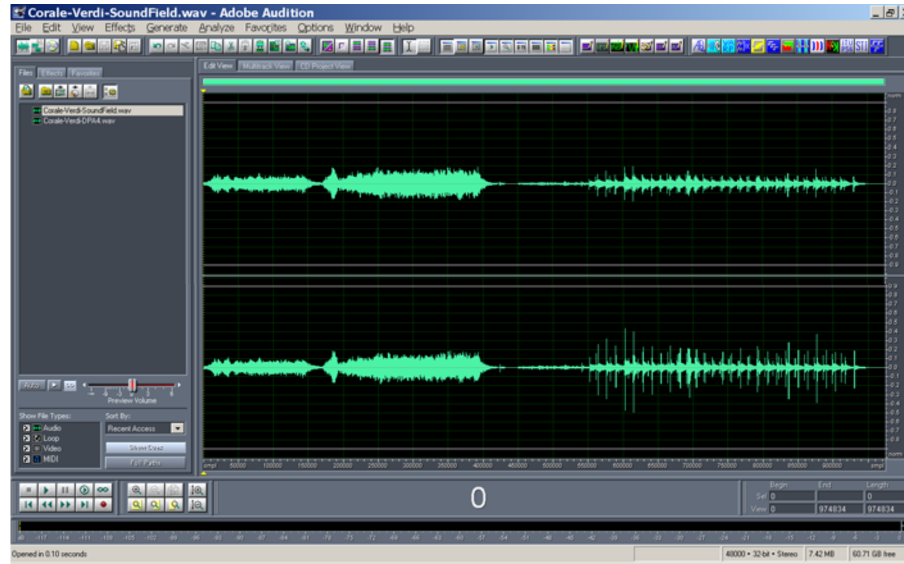
- 2 crossed Neumann K-140 were compared with a pair of virtual cardioids derived from B-format signals, recorded either with a Soundfield ST-250 and with the new DPA-4



Sound samples



DPA-4



Soundfield

- The new DPA-4 outperforms the Soundfield in terms of stereo separation and frequency response, and is indistinguishable from the “reference” Neumann cardioids

Conclusions

- The numerical approach to array processing does not require complex mathematical theories
- The quality of the processing FIR filters depends strongly on the quality of the impulse response measurements
- The method allows for the usage of imperfect arrays, with low-quality transducers and irregular geometry
- A new fast convolver has been developed for real-time applications

Future developements

- A new 24-microphones array is being assembled, employing 24 high quality B&K 4188 microphones

The goal is to record wide-band, high-quality 3°-order Ambisonics signals, which requires a set of 24x16 filters

Future developements

- The Multivolver VST plugin will be improved (Intel IPP 5.0 FFT subroutines, multithread, rebuffering for employing larger FFT blocks even when the host block is limited)
- Fast switching of the set of impulse responses will be added, with MIDI control of the running set (for head-tracking, or realtime spatialisation simulating movement of sources or receivers)
- A new standalone program will be developed for speeding up the computation of the sets of inverse filters (the actual Matlab implementation is very slow and unfriendly)