4th Joint Meeting of the
Acoustical Society of America
and the
Acoustical Society of Japan
28 November--2 December

University of Parma

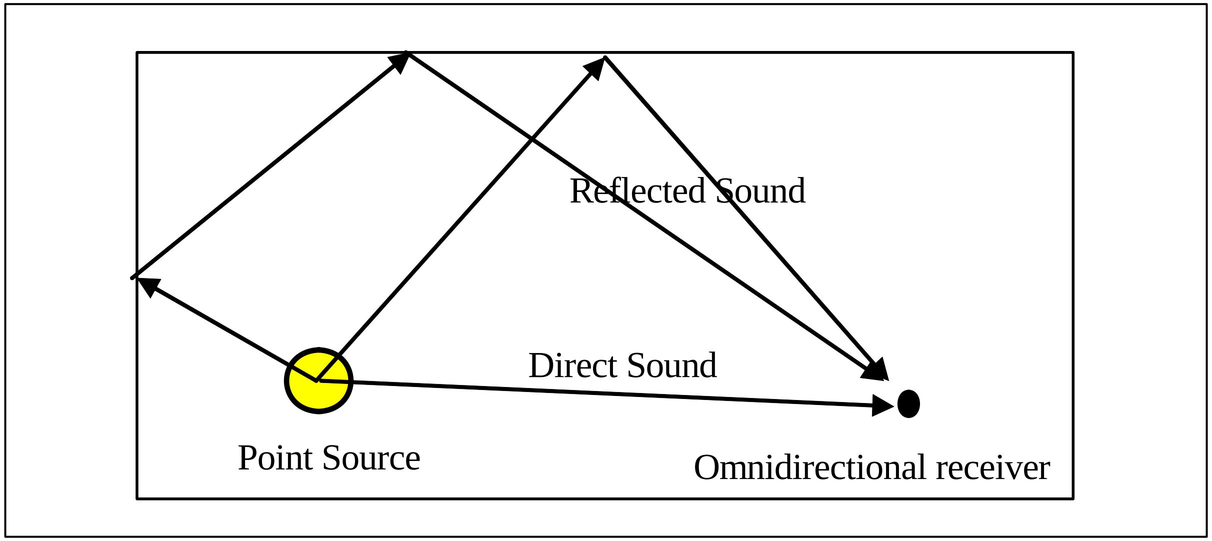
Measuring impulse responses containing complete spatial information

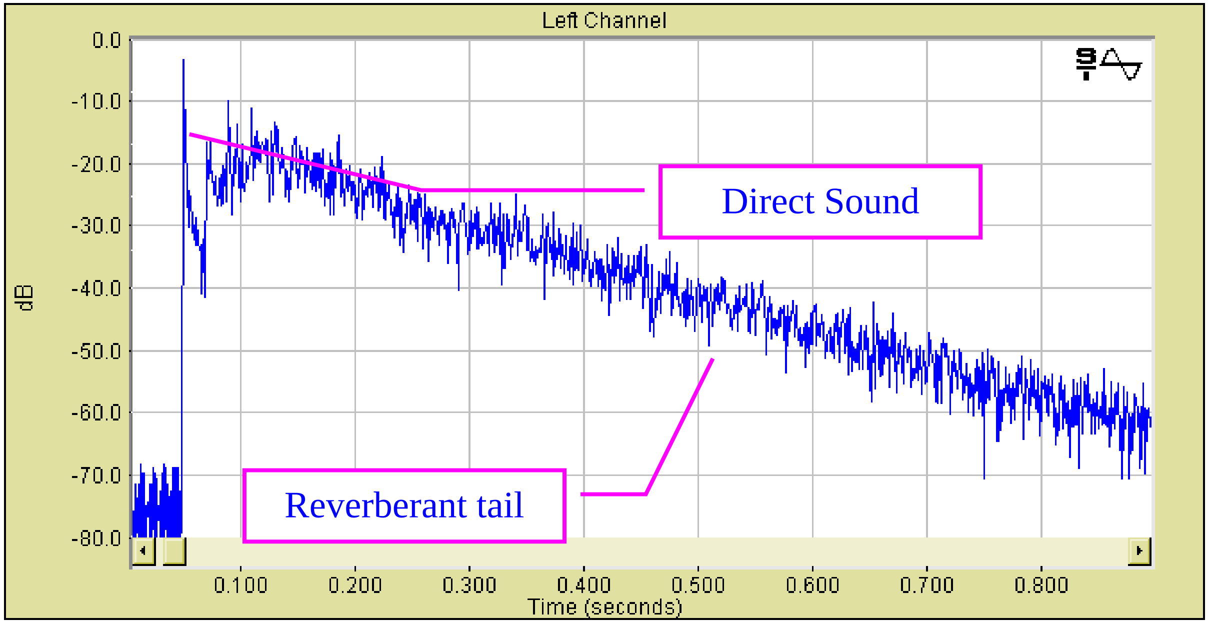
Angelo Farina, Paolo Martignon, Andrea Capra, Simone Fontana
Industrial Engineering Dept.
University of Parma - ITALY
HTTP://www.angelofarina.it

Topics

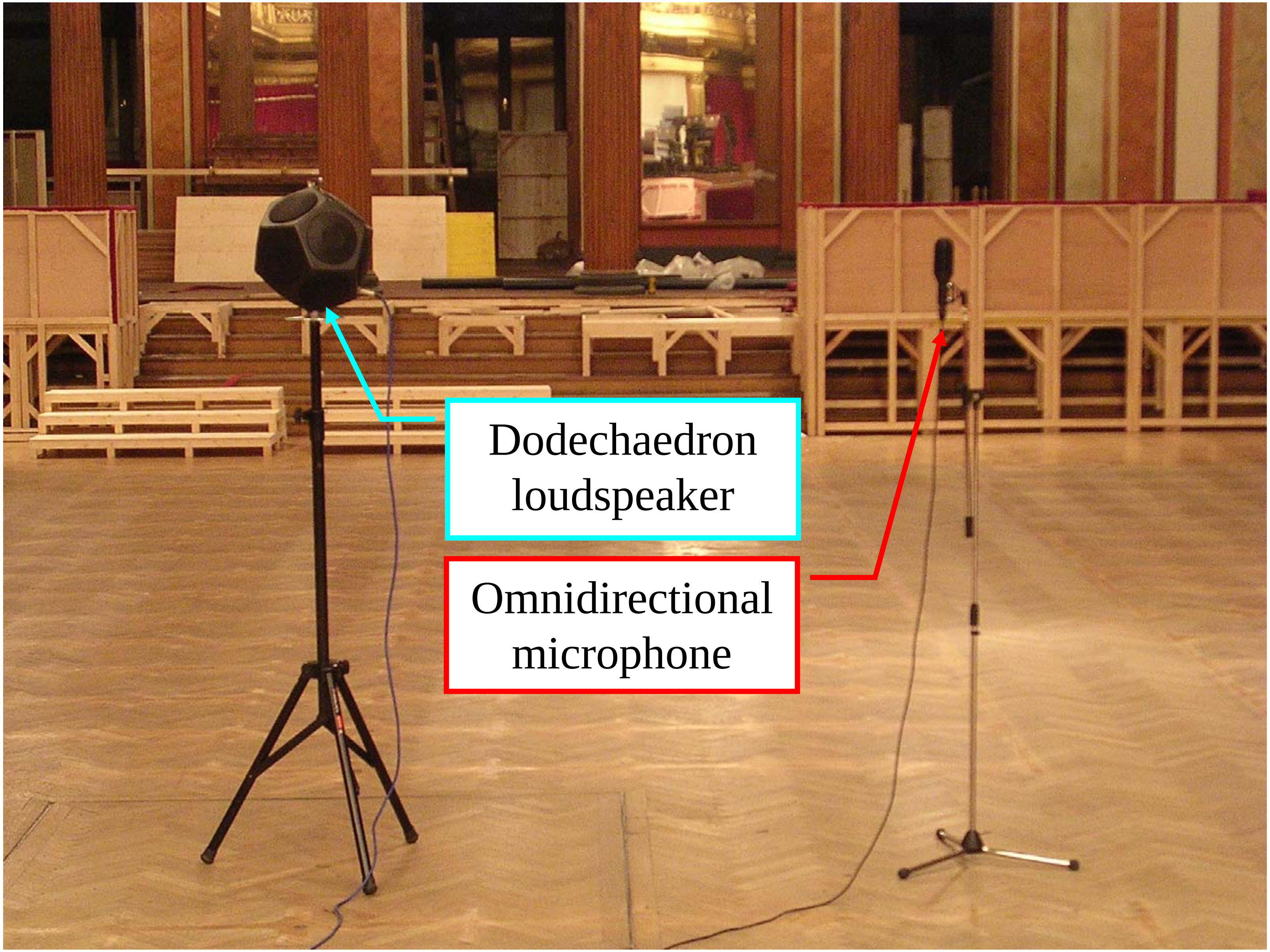
- Directional transducers, the first attempts of spatial analysis
- Orthonormal decomposition of the spatial properties in spherical harmonics: the Ambisonics method
- The reciprocity principle: directive microphones and directive sources
- Generalization of higher-order spherical harmonics representation of both source and receiver directivity
- Joining time and space: from Einstein's view to a comprehensive data structure representing the acoustical transfer function of a room
- Practical usages of measured (or numerically simulated) temporal-spatial impulse responses

Basic sound propagation scheme

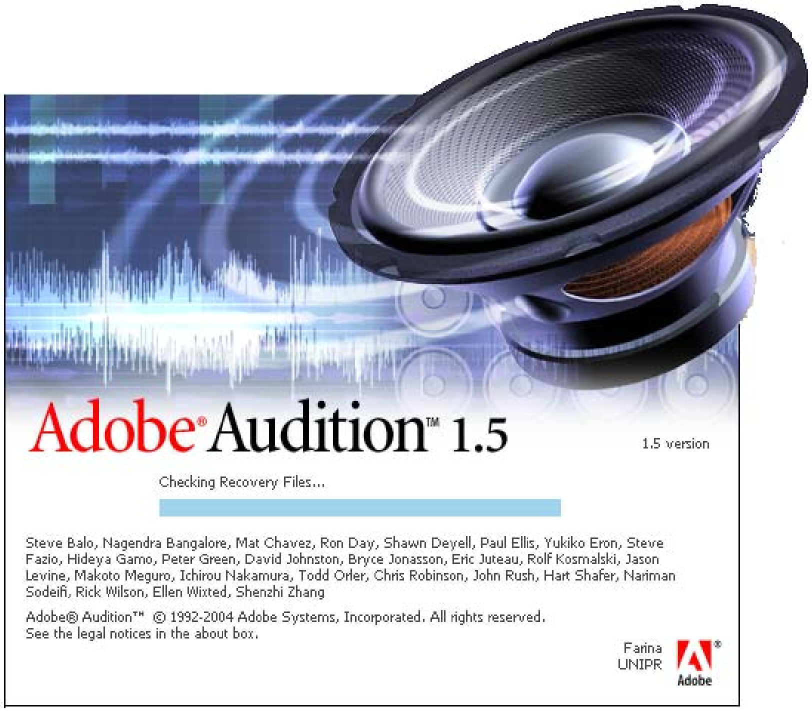




Hardware: loudspeaker & microphone



Software



+

Aurora Plugins

Generate MLS	
Deconvolve MLS	
Generate Sweep	
Deconvolve Sweep	
Convolution	
Kirkeby Inverse Filter	
Speech Transm. Index	

Spatial analysis by directive microphones

- **The initial approach was to use directive microphones for gathering some information about the spatial properties of the sound field "as perceived by the listener"**
- **Two apparently different approaches emerged: binaural dummy heads and pressure-velocity microphones:**

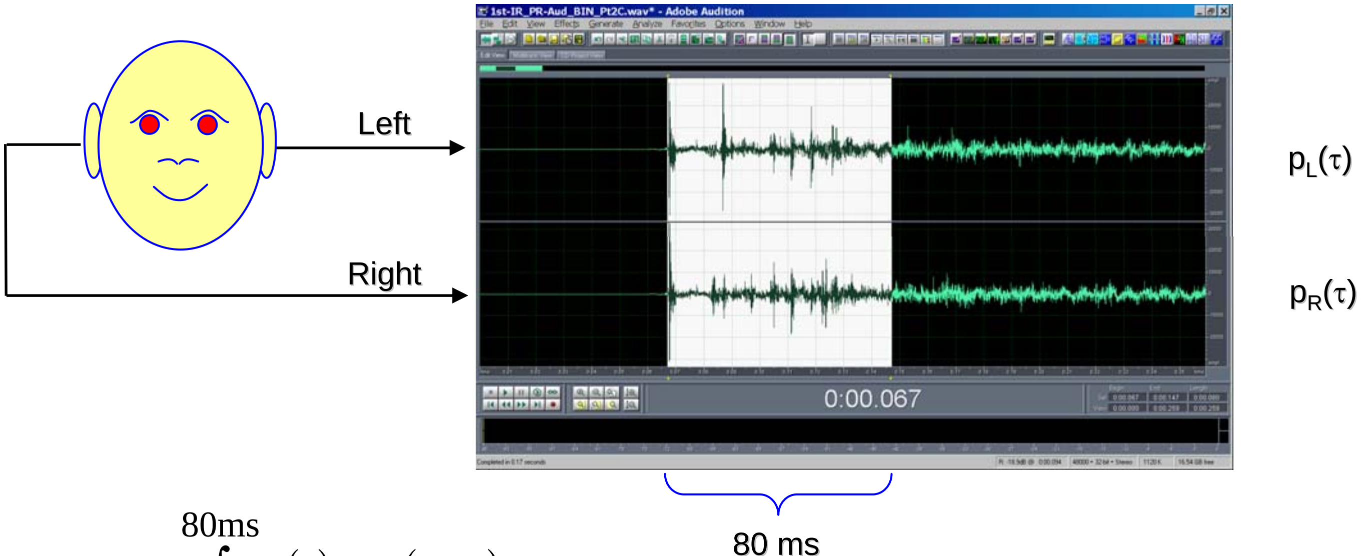
Binaural microphone (left)

and

variable-directivity microphone (right)

“objective” spatial parameters

- **It was attempted to “quantify” the “spatiality” of a room by means of “objective” parameters, based on 2-channels impulse responses measured with directive microphones**
- **The most famous “spatial” parameter is IACC (Inter Aural Cross Correlation), based on binaural IR measurements**

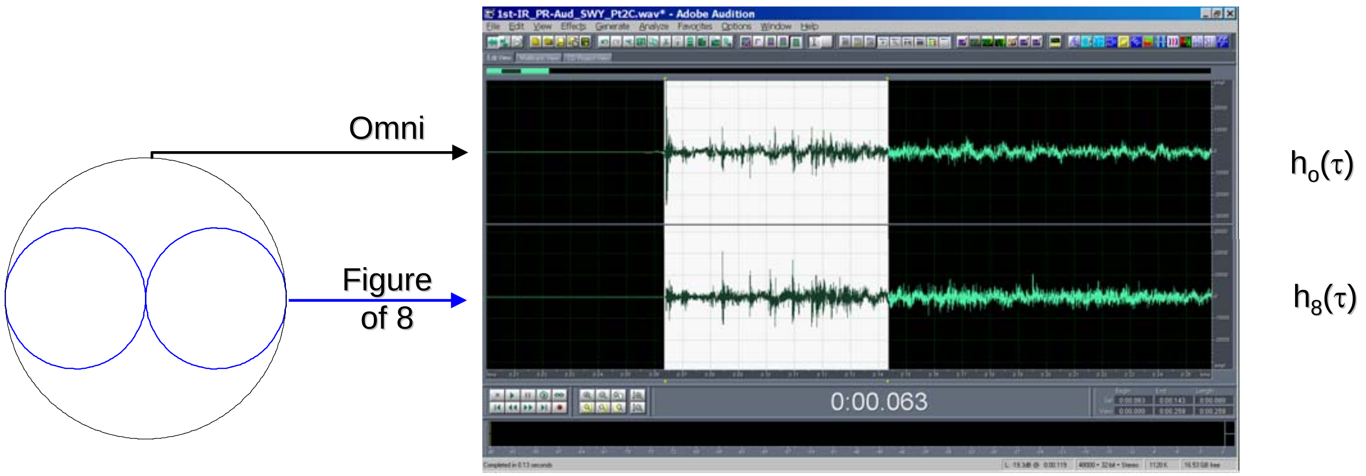


$$\rho(t)=\frac{\int_0^{80ms} p_L(\tau)\cdot p_R(\tau+t)\cdot d\tau}{\sqrt{\int_0^{80ms} p_L^2(\tau)\cdot d\tau\cdot\int_0^{80ms} p_R^2(\tau+t)\cdot d\tau}}$$

$$IACC_E = Max[\rho(t)] \qquad t\in[-1ms\ldots+1ms]$$

"objective" spatial parameters

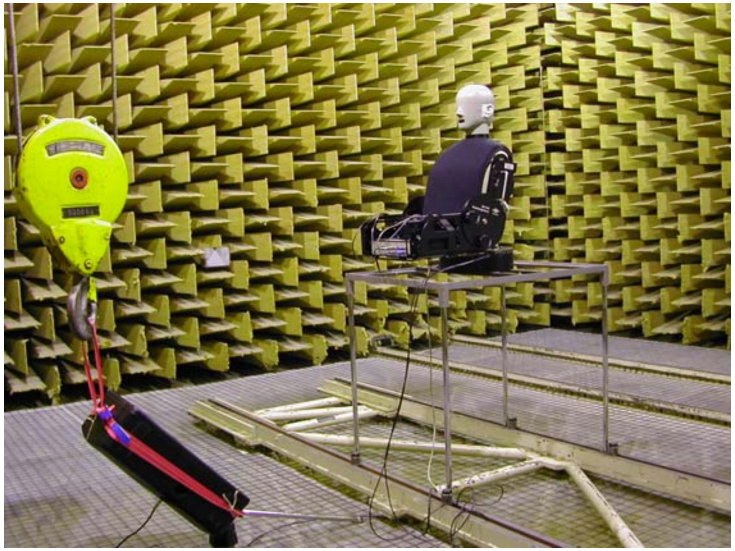
- **Other "spatial" parameters are the Lateral Energy ratios: LE, LF, LFC**
- **These are defined from a 2-channels impulse response, the first channel is a standard omni microphone, the second channel is a "figure-of-eight" microphone:**

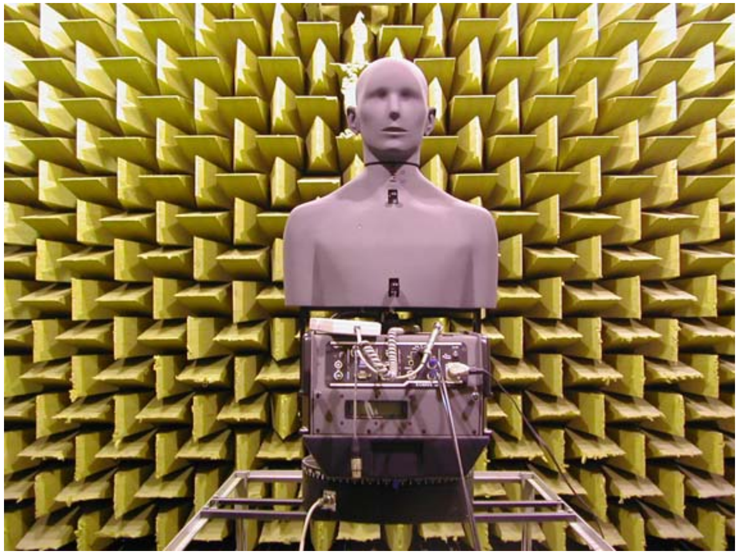


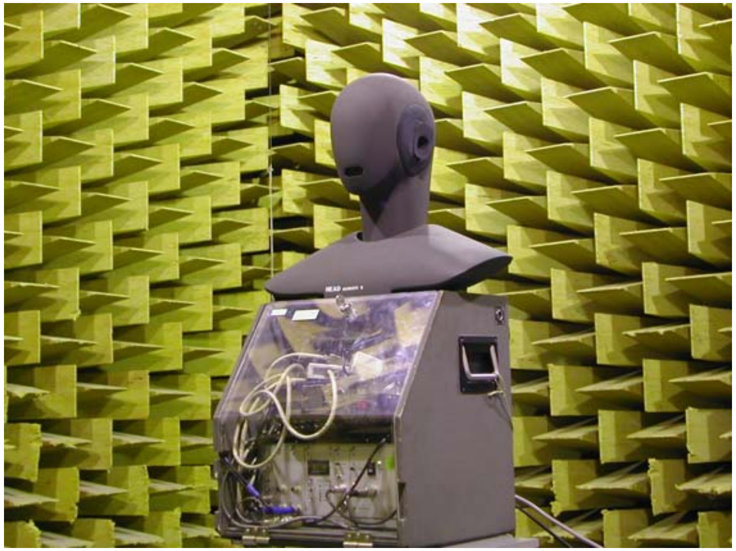
$$LE = \frac{\int_{25ms}^{80ms} h_8^2(\tau)\cdot d\tau}{\int_{0ms}^{80ms} h_o^2(\tau)\cdot d\tau} \qquad LF = \frac{\int_{5ms}^{80ms} h_8^2(\tau)\cdot d\tau}{\int_{0ms}^{80ms} h_o^2(\tau)\cdot d\tau} \qquad LFC = \frac{\int_{5ms}^{80ms} h_8(\tau)\cdot h_o(\tau)\cdot d\tau}{\int_{0ms}^{80ms} h_o^2(\tau)\cdot d\tau}$$

Are binaural measurents reproducible?

- **Experiment performed in anechoic room - same loudspeaker, same source and receiver positions, 5 binaural dummy heads**

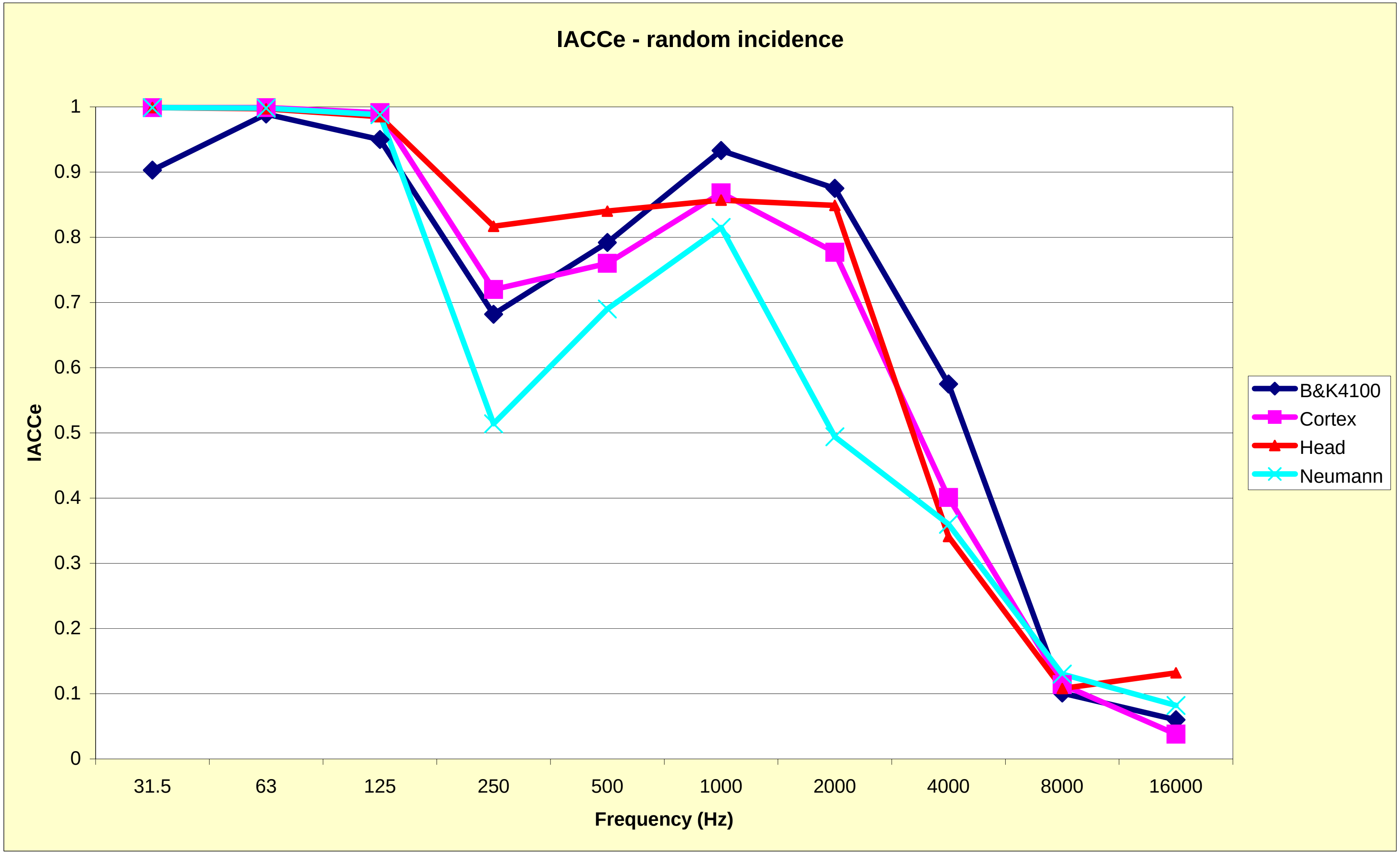






Are binaural measurents reproducible?

- **Diffuse field - the difference between the heads is now dramatic**

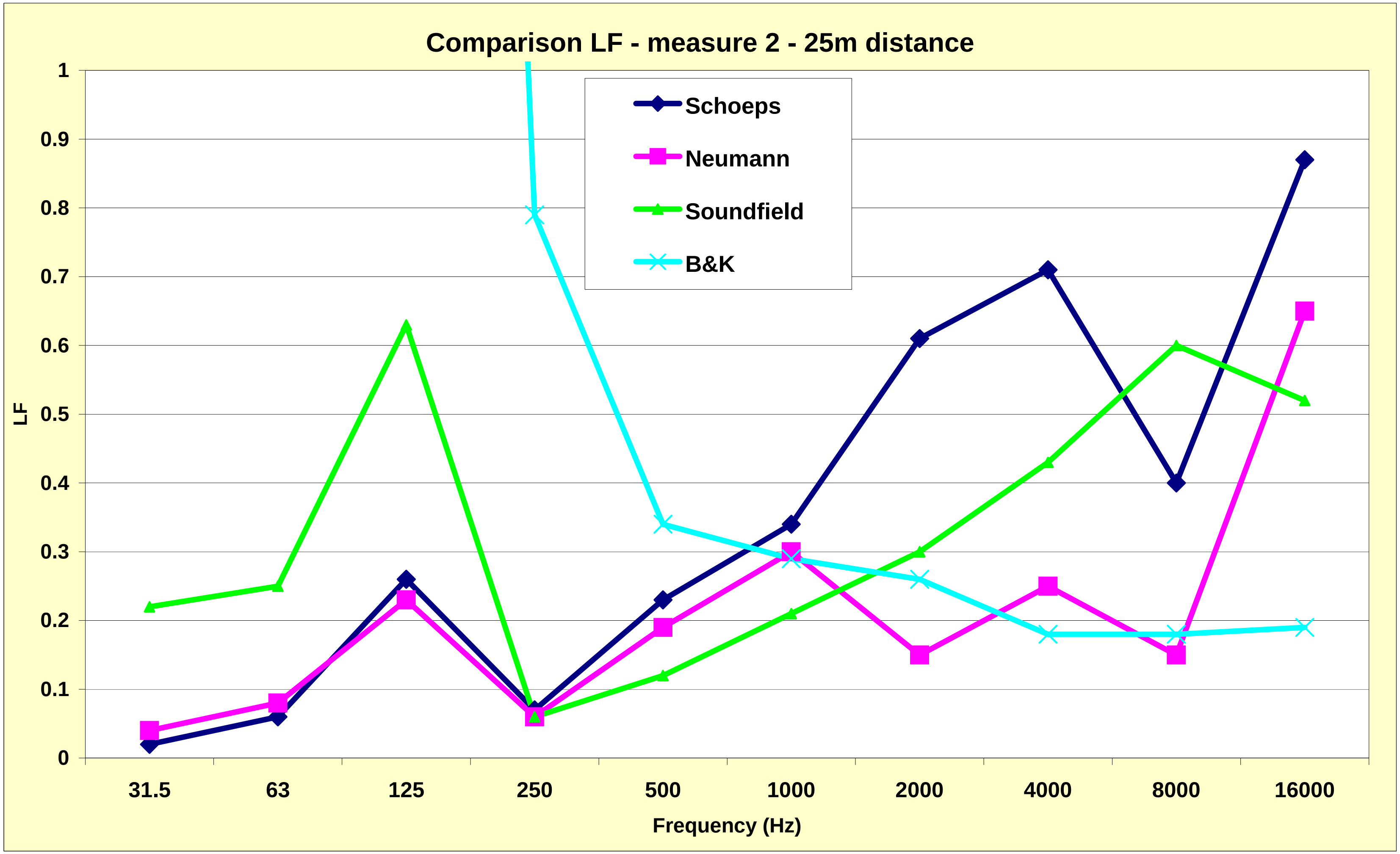


Are LF measurents reproducible?

- **Experiment performed in the Auditorium of Parma - same loudspeaker, same source and receiver positions, 5 pressure-velocity microphones**

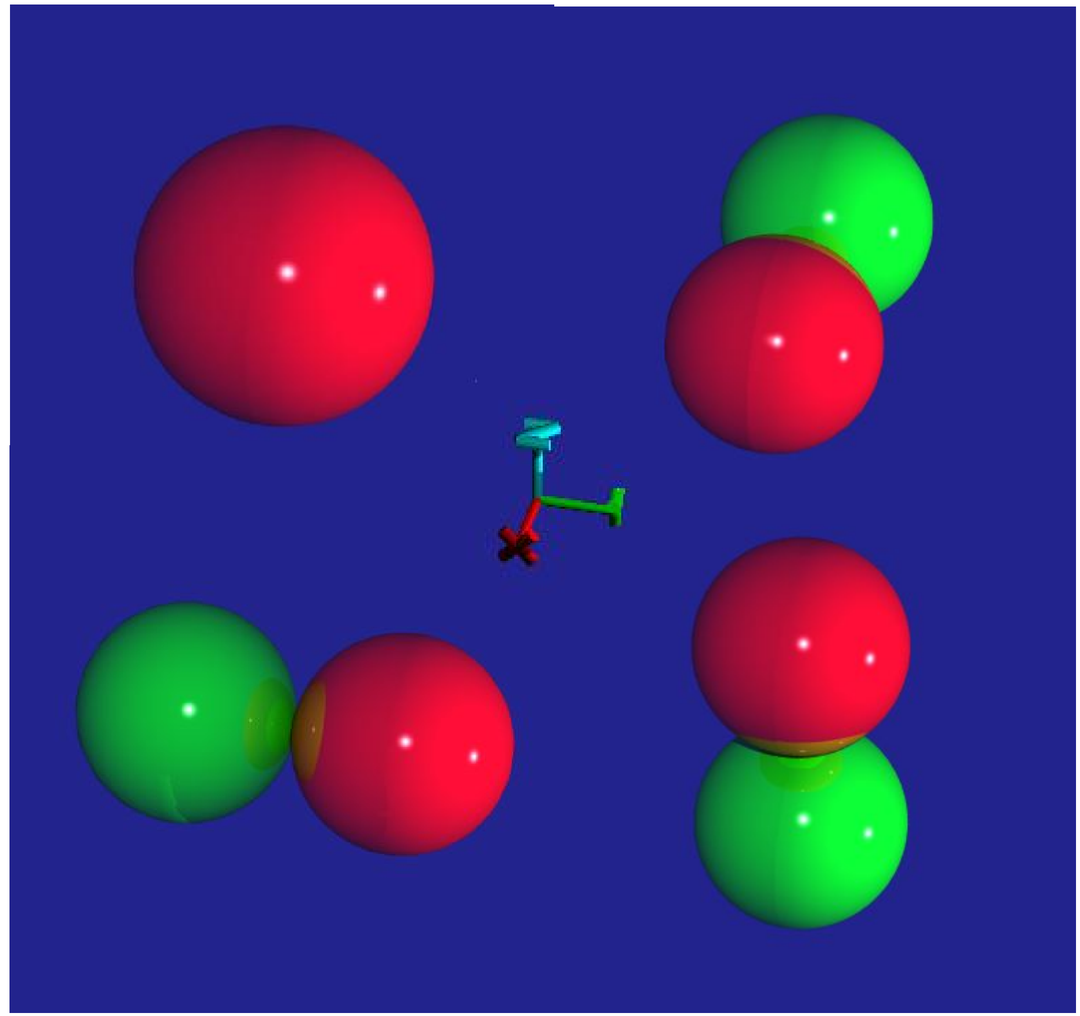
Are LF measurents reproducible?

- **At 25 m distance, the scatter is even larger....**

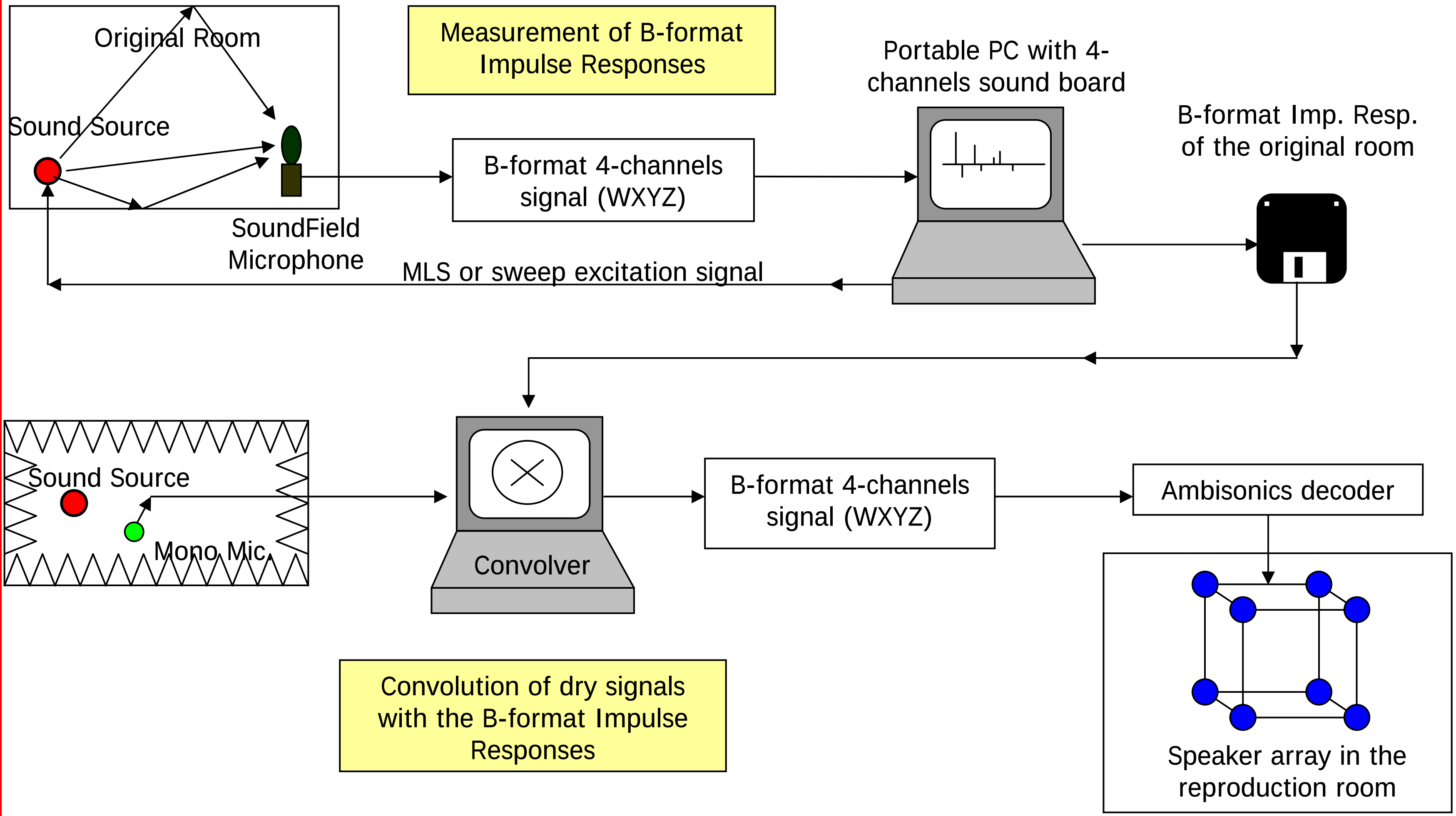


3D extension of the pressure-velocity measurements

- **The Soundfield microphone allows for simultaneous measurements of the omnidirectional pressure and of the three cartesian components of particle velocity (figure-of-8 patterns)**

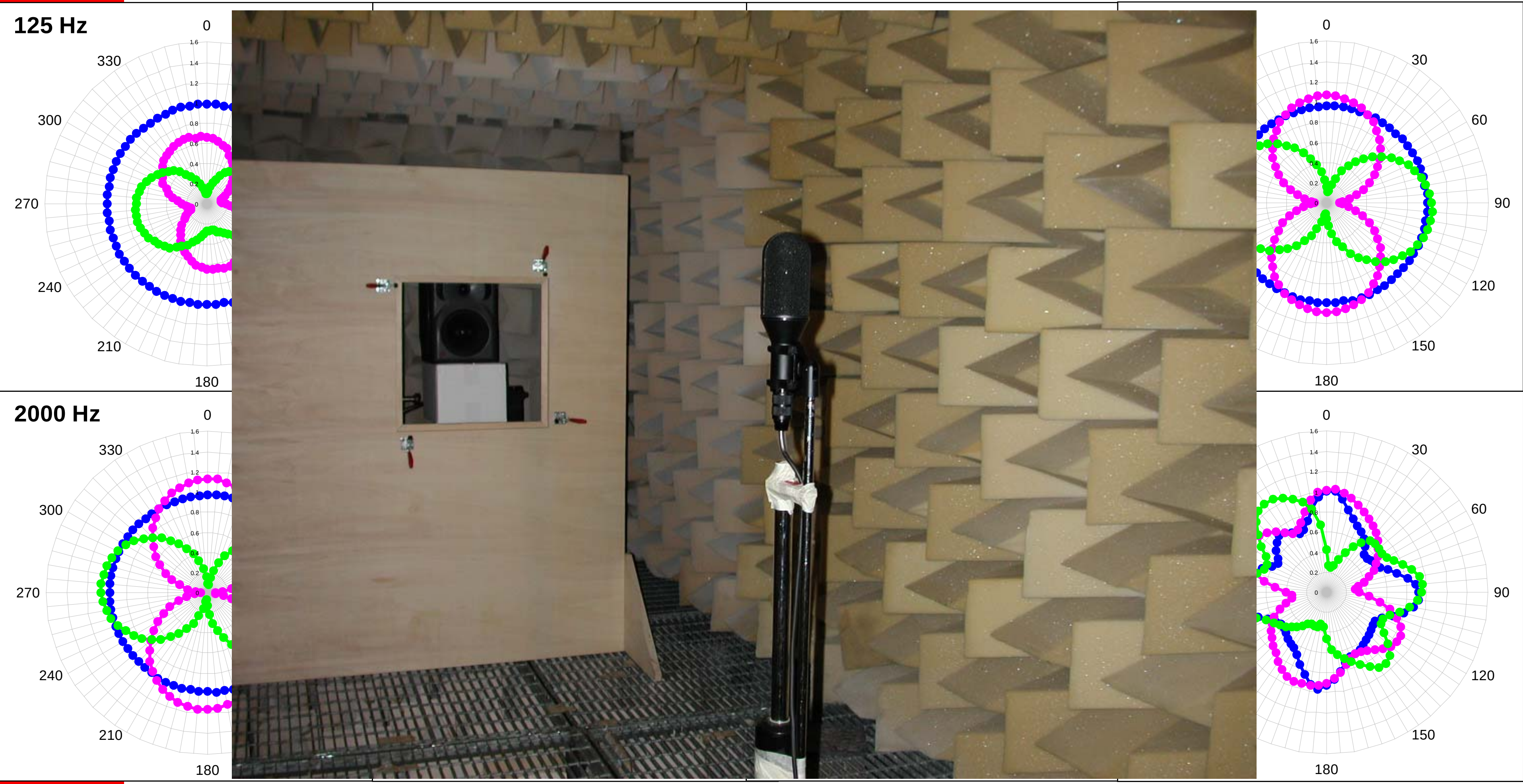


3D Impulse Response (Gerzon, 1975)



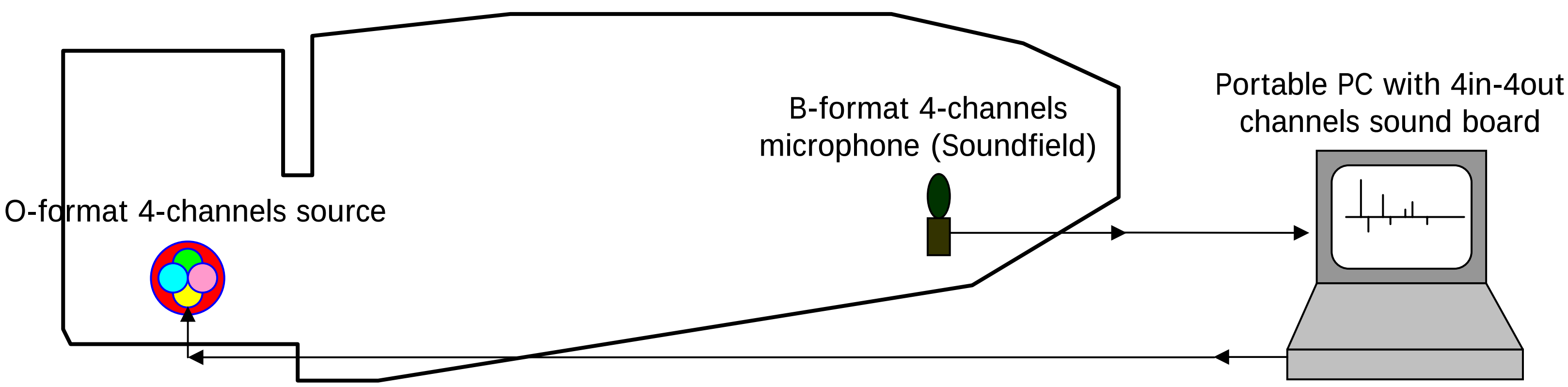
Directivity of transducers

Soundfield ST-250 microphone



What about source directivity ?

- Current 3D IR sampling is still based on the usage of an "omnidirectional" source
- The knowledge of the 3D IR measured in this way provide no information about the soundfield generated inside the room from a directive source (i.e., a musical instrument, a singer, etc.)
- Dave Malham suggested to represent also the source directivity with a set of spherical harmonics, called O-format - this is perfectly reciprocal to the representation of the microphone directivity with the B-format signals (Soundfield microphone).
- Consequently, a complete and reciprocal spatial transfer function can be defined, employing a 4-channels O-format source and a 4-channels B-format receiver:



1st order MIMO impulse response

- **If only spherical harmonics of order 0 and 1 are taken into account, a complete spatial transfer function measurement requires 16 impulse responses:**

$$\begin{Bmatrix} y_w \\ y_x \\ y_y \\ y_z \end{Bmatrix} = \begin{bmatrix} h_{ww} & h_{wx} & h_{wy} & h_{wz} \\ h_{xw} & h_{xx} & h_{xy} & h_{xz} \\ h_{yw} & h_{yx} & h_{yy} & h_{yz} \\ h_{zw} & h_{zx} & h_{zy} & h_{zz} \end{bmatrix} \otimes \begin{Bmatrix} x_w \\ x_x \\ x_y \\ x_z \end{Bmatrix} \quad \text{or} \quad \{y\} = [h] \otimes \{x\}$$

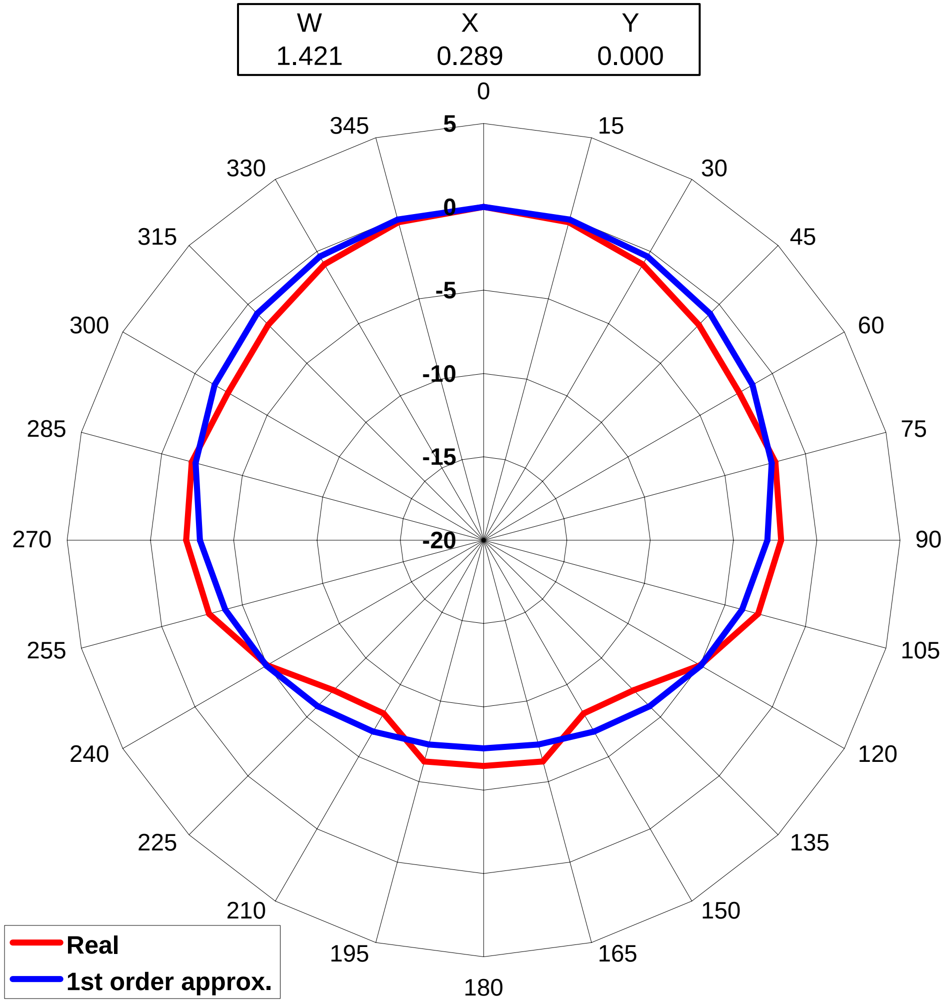
- **Once these 16 IRs have been measured, it is possible to compute the response of the room with a source and a receiver having arbitrary directivity patterns, given by the O-format source functions $\{r_w, r_x, r_y, r_z\}$, and the B-format receiver functions $\{r_w, r_x, r_y, r_z\}$:**

$$t_{sr} = \{r\} \otimes \{y\} = \{r\} \otimes [h] \otimes \{s\}$$

- **In which also each of {s} and {r} are sets of 4 impulse responses, representing the frequency-dependent directivities of the source and of the receiver**

Limits of the 1st-order method

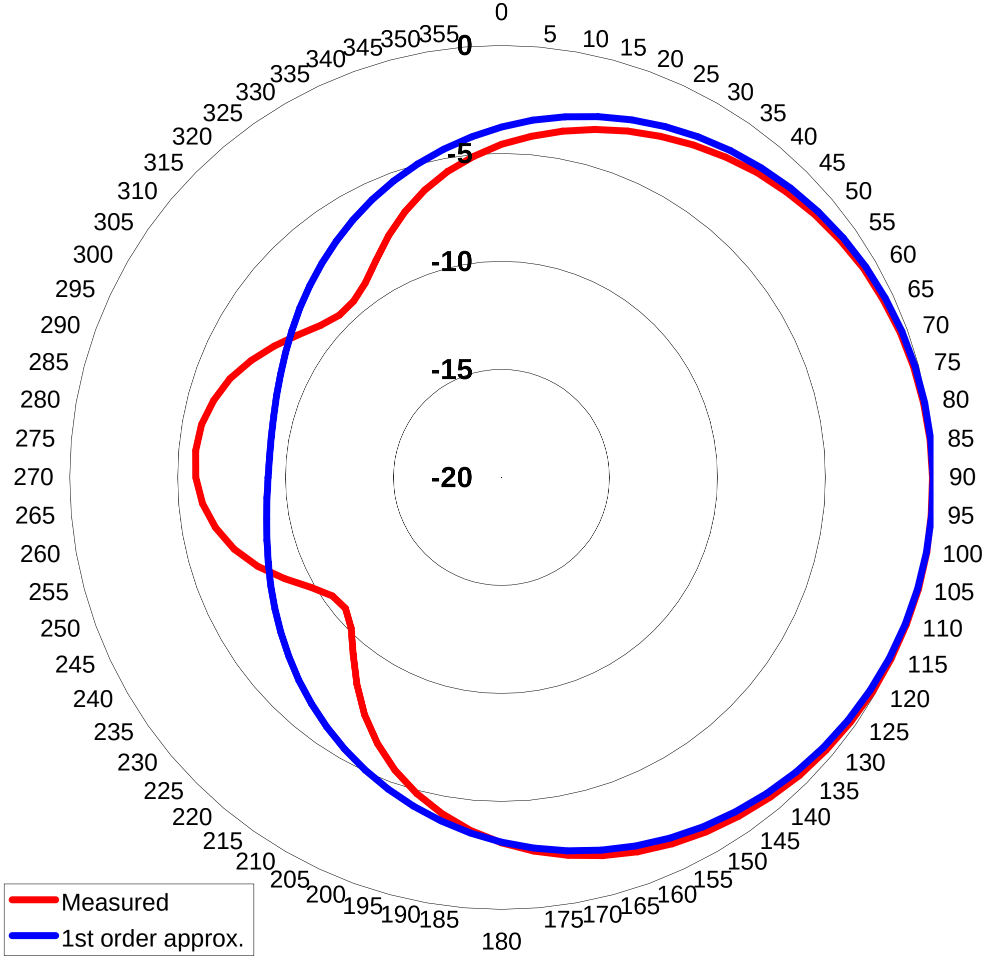
- **Albeit mathematically elegant and easy to implement with currently-existing hardware, the 1st-order method presented here cannot represent faithfully the complex directivity pattern of an human voice or of an human ear:**



Limits of the 1st-order method

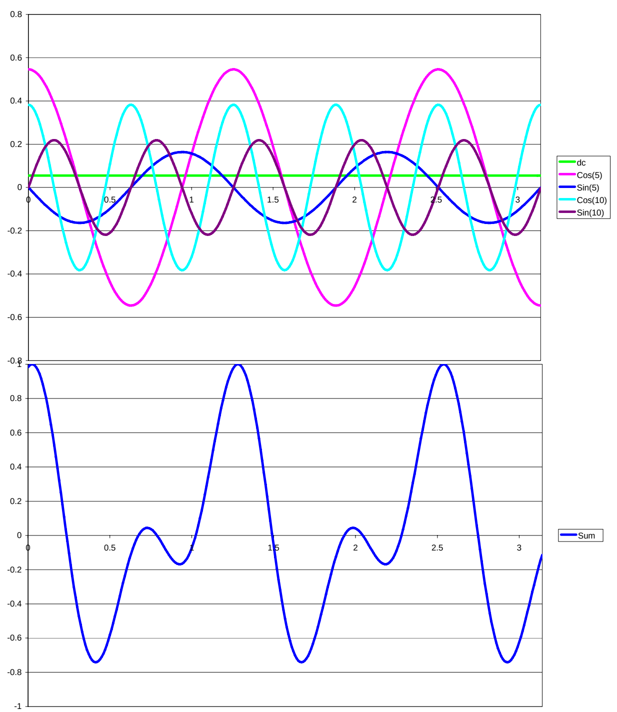
- **The polar pattern of a binaural dummy head is even more complex, as shown here (1 kHz, right ear):**

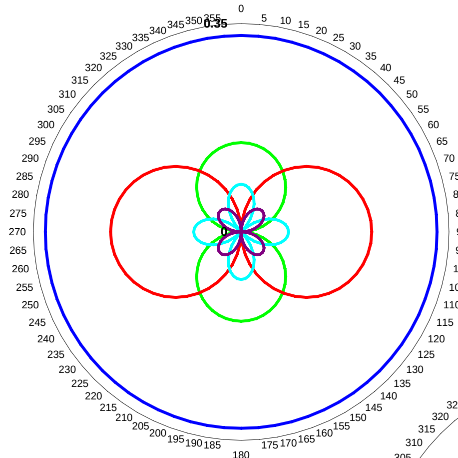
W	X	Y
0.395	-0.015	0.191

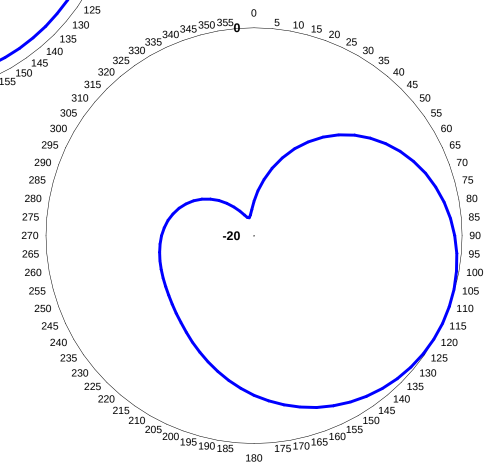


How to get better spatial resolution?

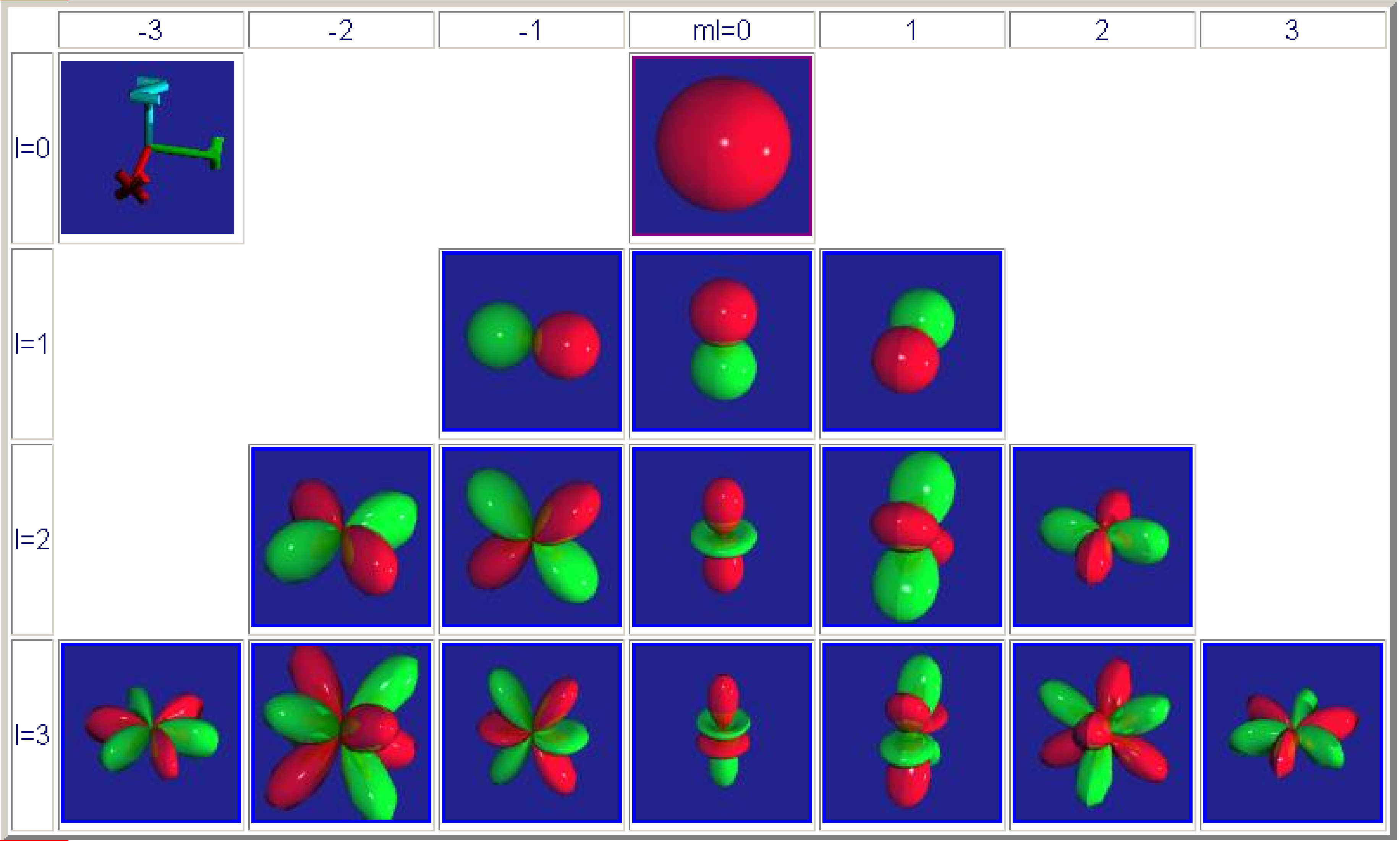
- The answer is simple: analyze the spatial distribution of both source and receiver by means of higher-order spherical harmonics expansion
- Spherical harmonics analysis is the equivalent, in space domain, of the Fourier analysis in time domain
- As a complex time-domain waveform can be thought as the sum of a number of sinusoidal and cosinusoidal functions, so a complex spatial distribution around a given notional point can be expressed as the sum of a number of spherical harmonic functions





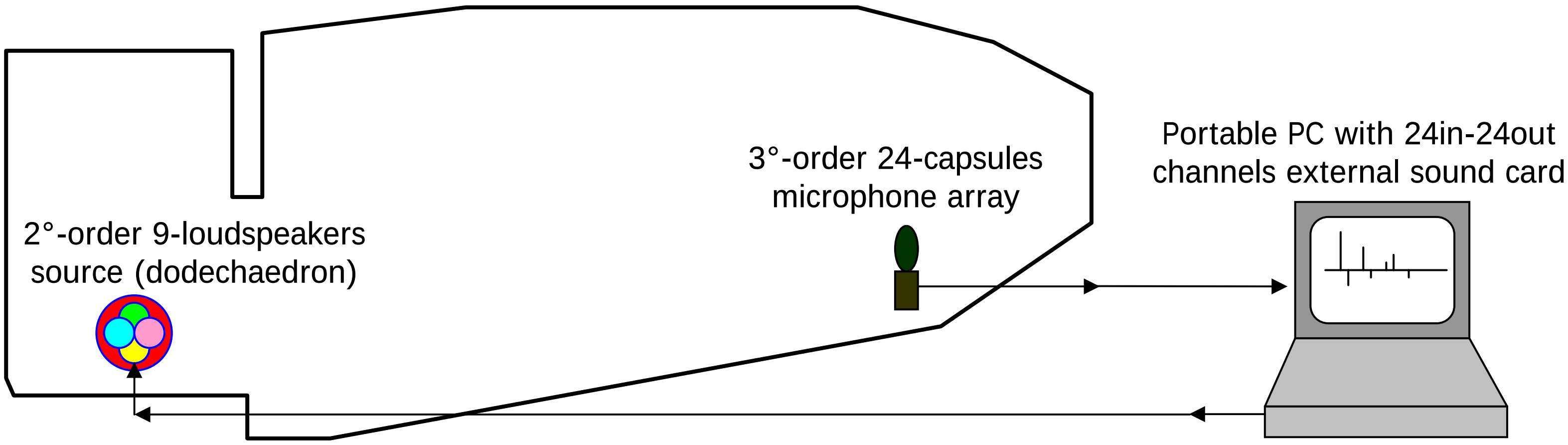


Higher-order spherical harmonics expansion



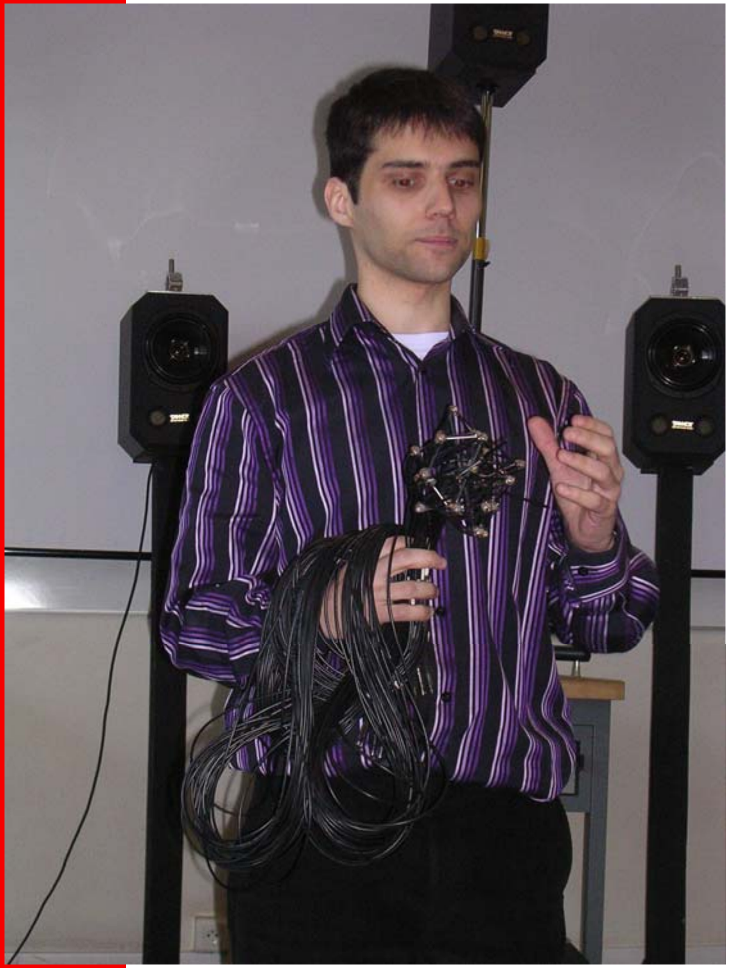
Complete high-order MIMO method

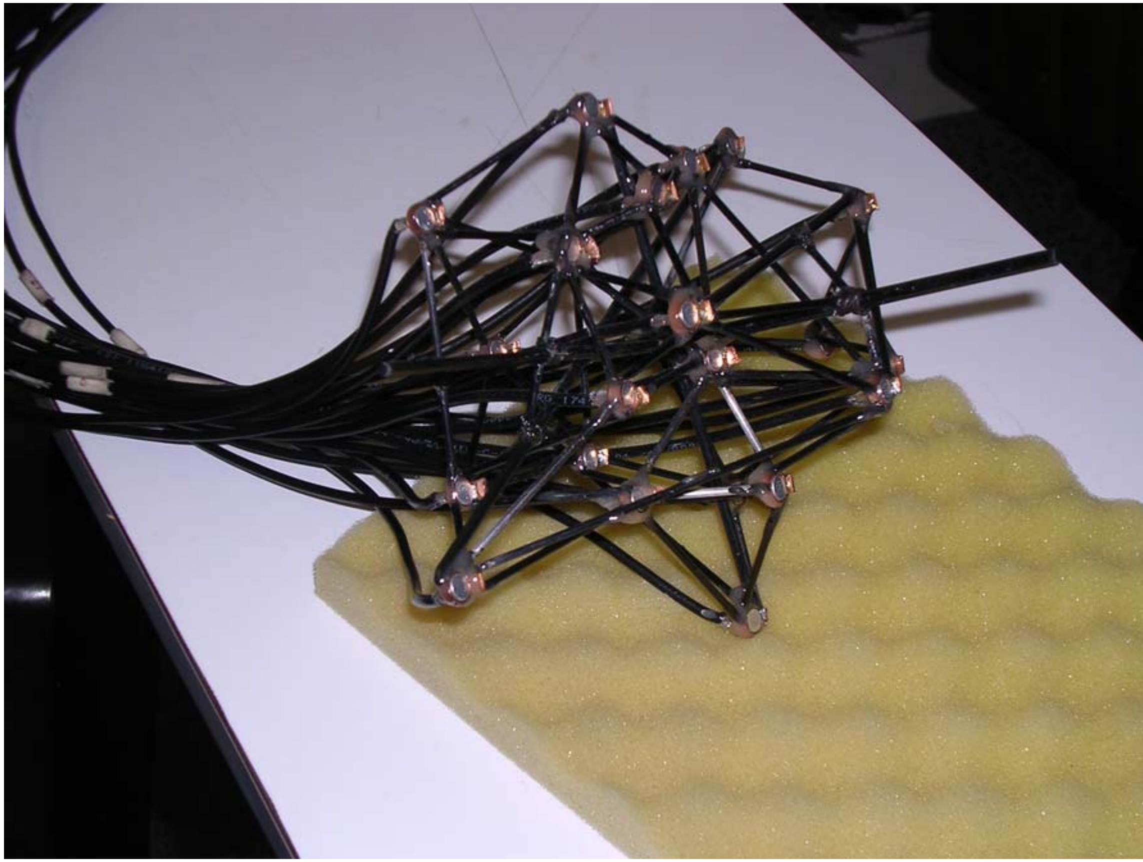
- **Employing massive arrays of transducers, it is nowaday feasible to sample the acoustical temporal-spatial transfer function of a room**
- **Currently available hardware and software tools make this practical up to 4° order, which means 25 inputs and 25 outputs**
- **A complete measurement for a given source-receiver position pair takes approximately 10 minutes (25 sine sweeps of 15s each are generated one after the other, while all the microphone signals are sampled simultaneously)**
- **However, it has been seen that real-world sources can be already approximated quite well with 2°-order functions, and even the human HRTF directivites are reasonally approximated with 3°-order functions.**



3°-order microphone (Trinnov - France)

- **Arnoud Laborie developed a 24-capsule compact microphone array - by means of advanced digital filtering, spherical ahrmonic signals up to 3° order are obtained (16 channels)**



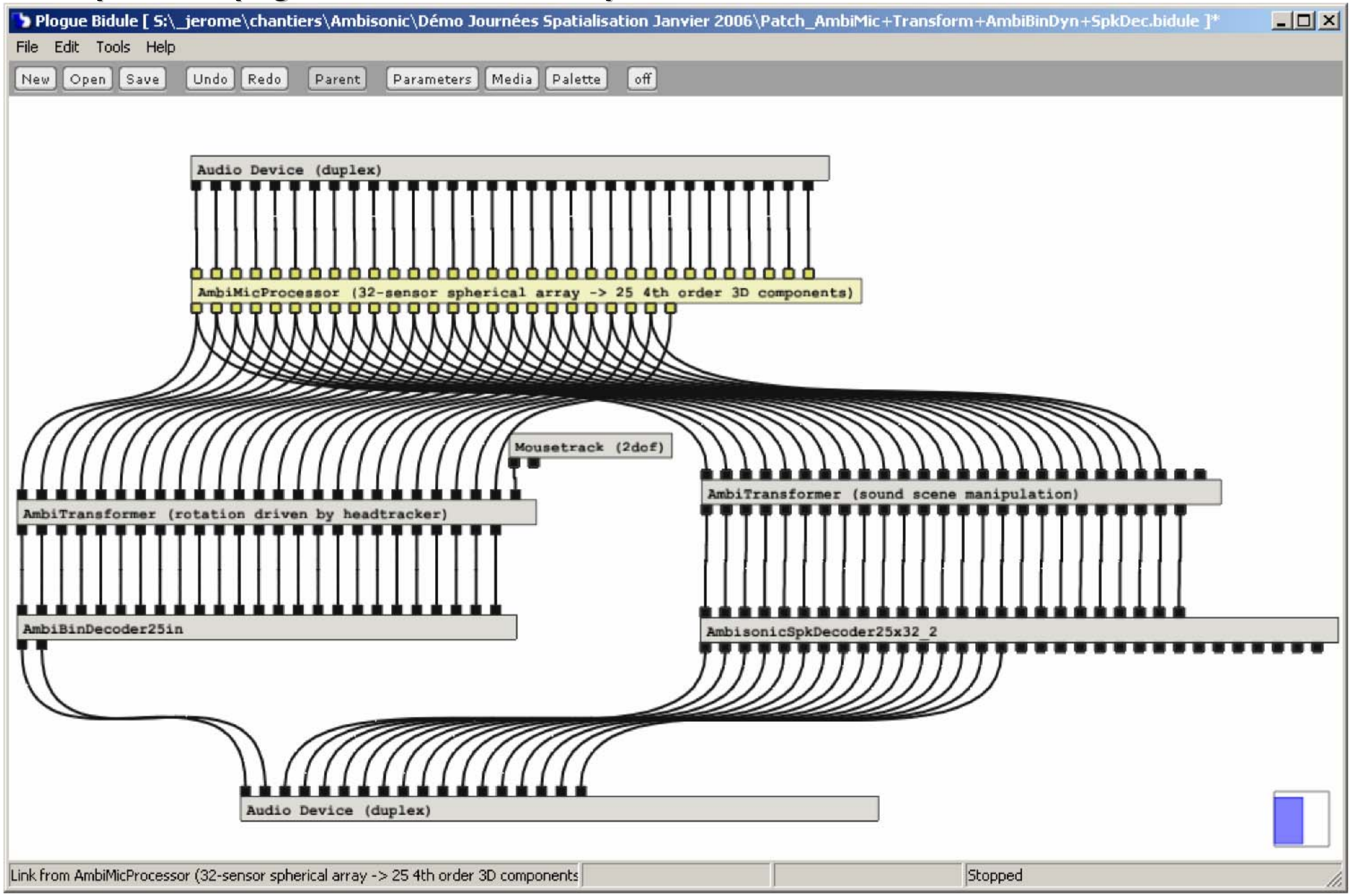


4°-order microphone (France Telecom)

- **Jerome Daniel and Sebastien Moreau built samples of 32-capsules spherical arrays - these allow for extraction of microphone signals up to 4° order (25 channels)**

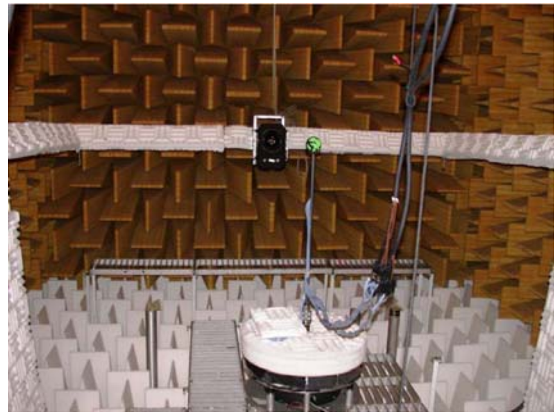
Multichannel software for high-order

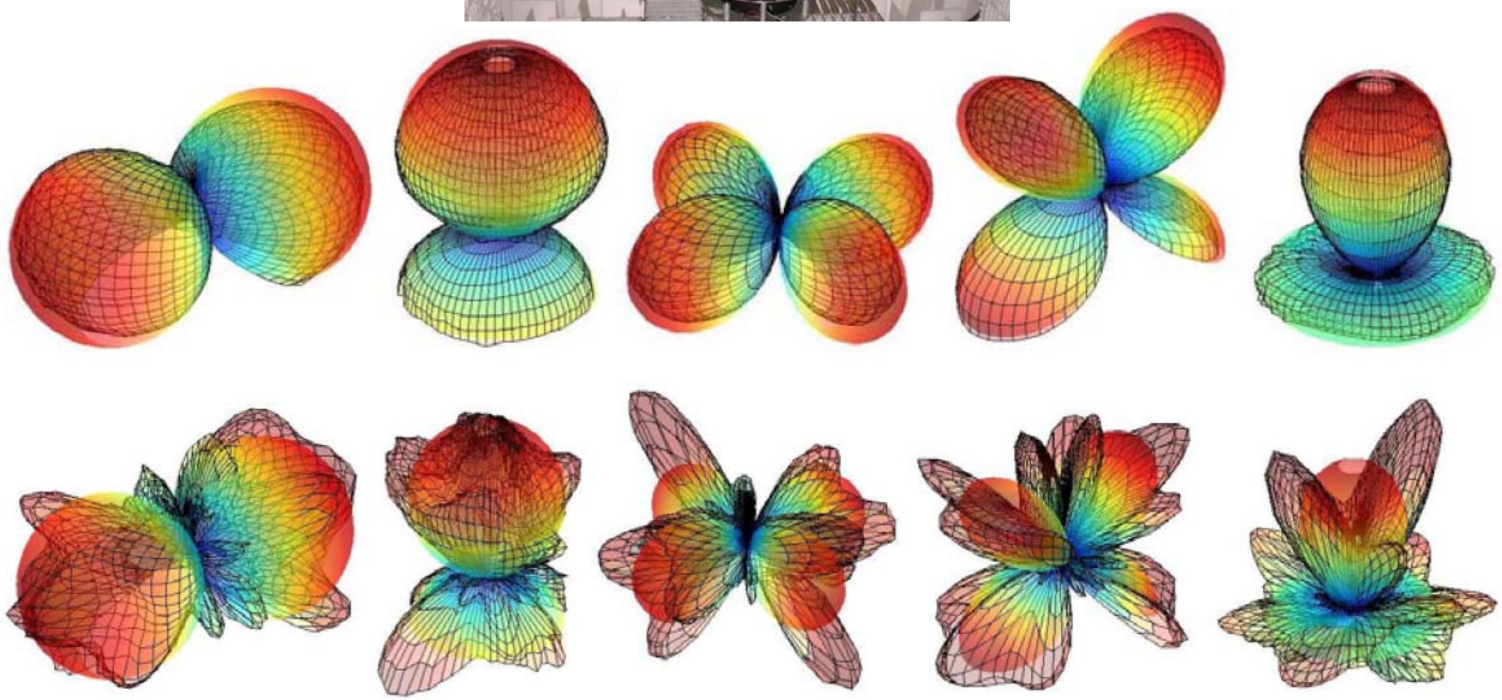
- **Plogue Bidule can be used as multichannel host software, running a number of VST plugins developed by France Telecom - these include spherical harmonics extraction from the spherical microphone arrays, rotation and manipulation of the multichannel B-format signals, and final rendering either on head-.tracked headphones or on a static array of loudspeakers (high-order Ambisonics)**



Verification of high-order patterns

- **Sebastien Moreau and Olivier Warusfel verified the directivity patterns of the 4°-order microphone array in the anechoic room of IRCAM (Paris)**





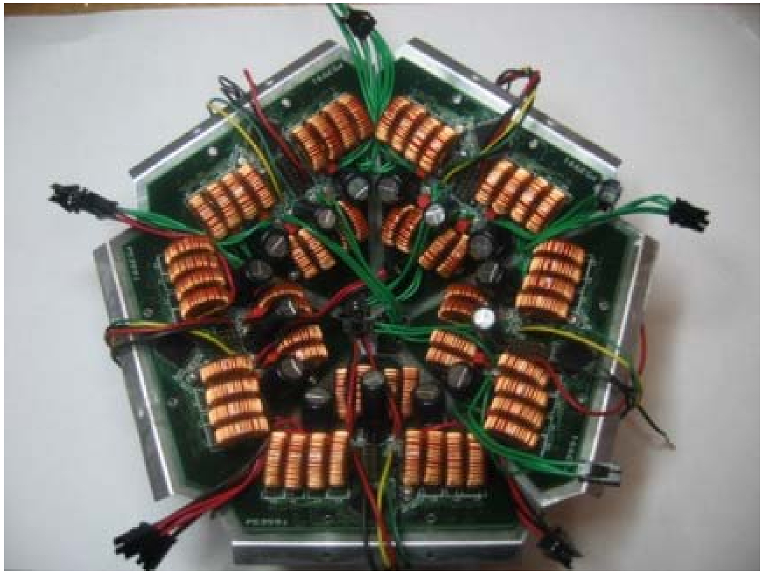
5 kHz

10 kHz

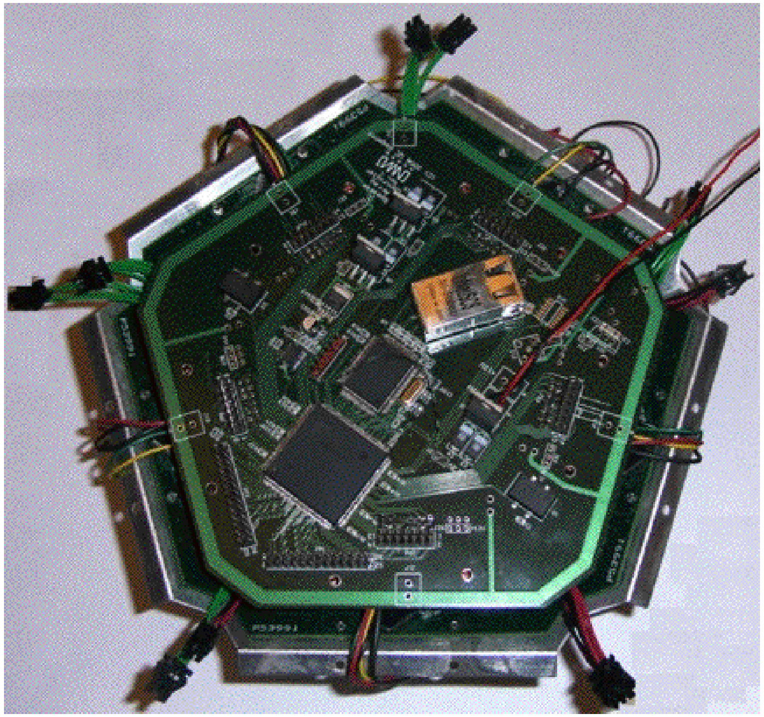
High-order sound source

- **University of California Berkeley's Center for New Music and Audio Technologies (CNMAT) developed a new 120-loudspeakers, digitally controlled sound source, capable of synthesizing sound emission according to spherical harmonics patterns up to 5° order.**

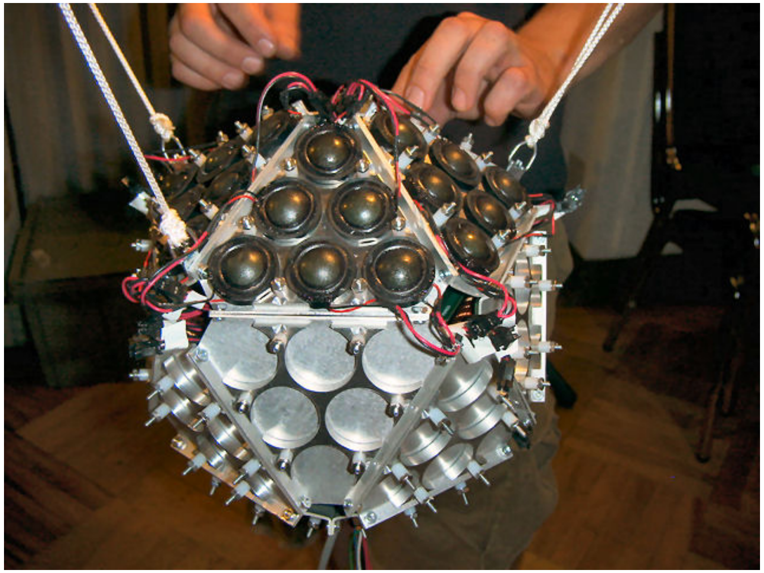
Technical details of high-order source



- Class-D embedded amplifiers



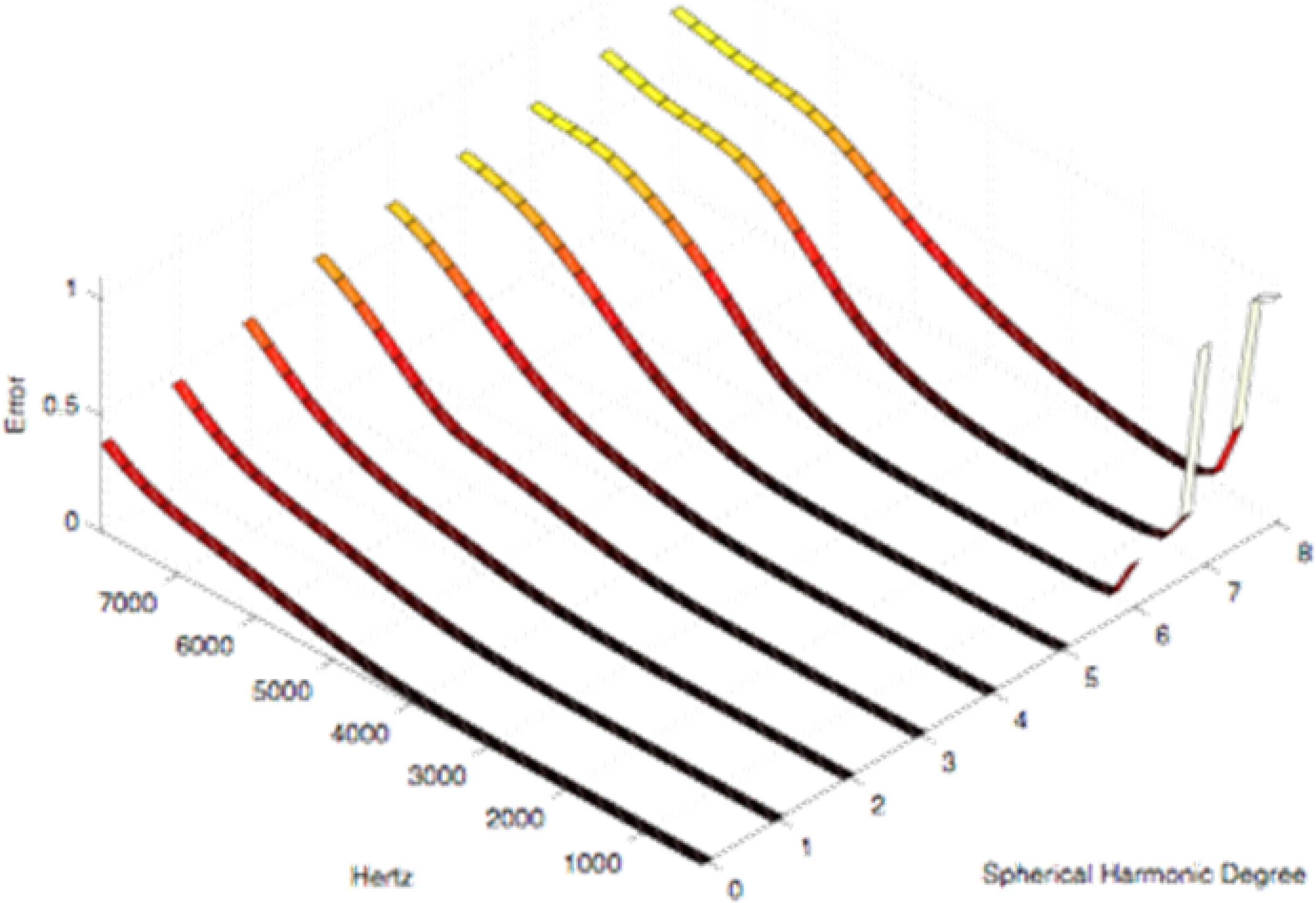
- Embedded ethernet interface and DSP processing



- Long-excursion special Meyer Sound drivers

Accuracy of spatial synthesis

- **The spatial reconstruction error of a 120-loudspeakers array is frequency dependant, as shown here:**

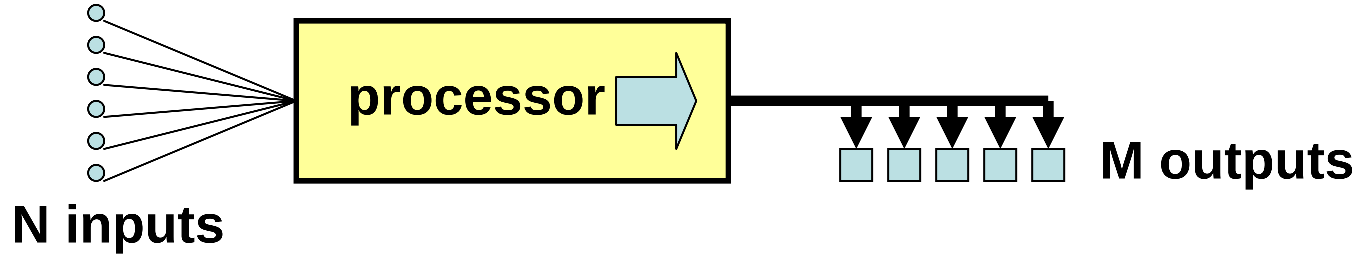


- **The error is acceptably low over an extended frequency range up to 5°-order**

Computer and sound card

- A complete 24-inputs, 24-outputs system can now assembled for less than 2000 USD
- Low-noise PC case, RME Hammerfall sound card, 3 Behringer Ultragain Pro-8 digital converters slaved to the same master clock

Advanced digital filtering techniques

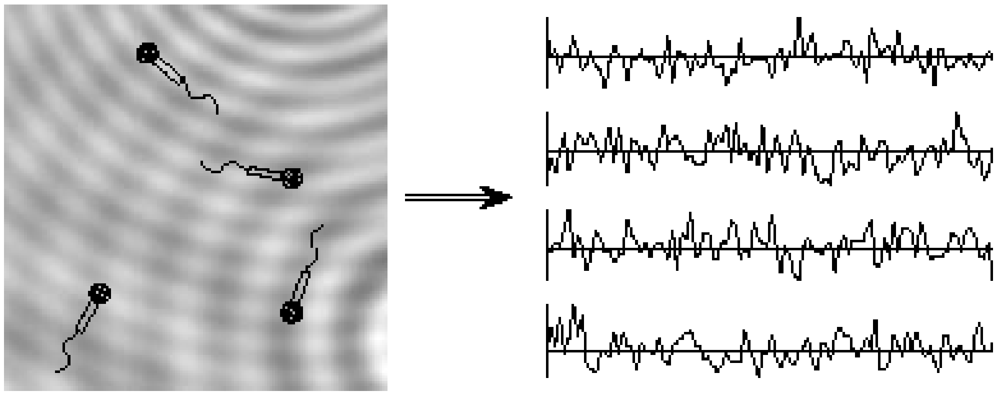


- A set of digital filters can be employed for sinthesizing the required spatial pattern (spherical harmonics), either when dealing with a microphone array or when dealing with a loudspeaker array
- Whatever theory or method is chosen, we always start with N input signals x_i, and we derive from them M output signals y_j
- And, in any case, each of these M outputs can be expressed by:

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

Example with a microphone array

- **The sound field is sampled in N points by means of a microphone array**



$x_1(t)$

$x_2(t)$

$x_3(t)$

$x_4(t)$

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 design of digital filters

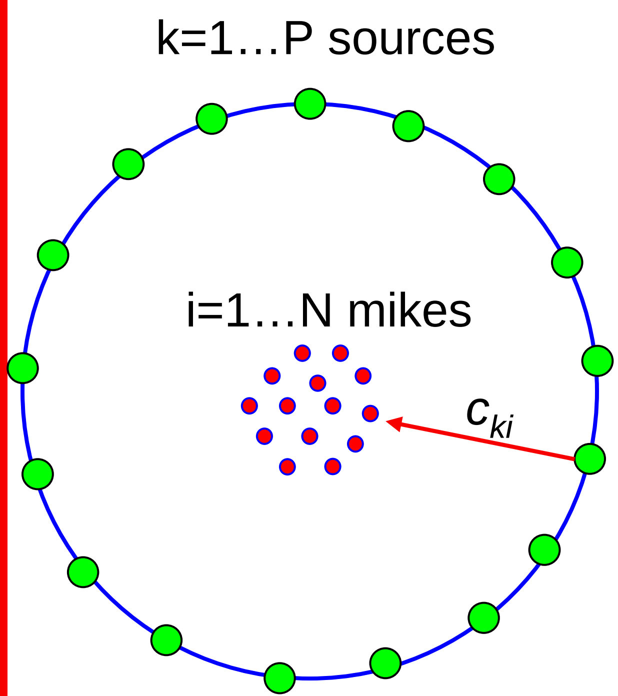
- The processing filters h_{ij} are usually computed following one of several, complex mathematical theories, based on the solution of the wave equation (often under certaing simplifications), and assuming that the microphones are ideal and identical
- In some implementations, the signal of each microphone 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: synthesizing 0-order shape

The microphone array impulse responses $c_{k,i}$, are measured for a number of *P* incoming directions.



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

…we get the **filters** to be applied to the microphonic signals from processing the matrix of measured impulse responses

Example: synthesizing 0-order shape

We design the filters in such a way that the response of the system is the prescribed theoretical function v_k for the k-th source (an unit-amplitude Dirac's Delta function in the case of the example, as the 0th-order function is omnidirectional).
So we set up a linear equation system of P equations, imposing that:

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

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

....

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

....

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

Lets call v_k the right-hand vector of known results (they will be different for higher-order shapes).
Once this matrix of N inverse filters are computed (for example, employing the Nelson/Kirkeby method), the output of the microphone array, synthesizing the prescribed 0th-order shape, will again be simply:

$$y_0 = \sum_{i=1}^{N} x_i \otimes h_{i,0}$$

System's least-squares inversion

- For computing the matrix of N filtering coefficients h_{i0}, a least-squares method is employed.
- A "total squared error" ε_{tot} is defined as:

$$\varepsilon_{tot} = \sum_{k=1}^{P} \left[\sum_{i=1}^{N} \left(h_{i0} \otimes c_{ki} \right) - v_k \right]^2$$

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

$$\frac{\partial \varepsilon_{tot}}{\partial h_{i0}} = 0 \qquad (i = 1 \ldots N)$$

Example for a 4-channel mike

- 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

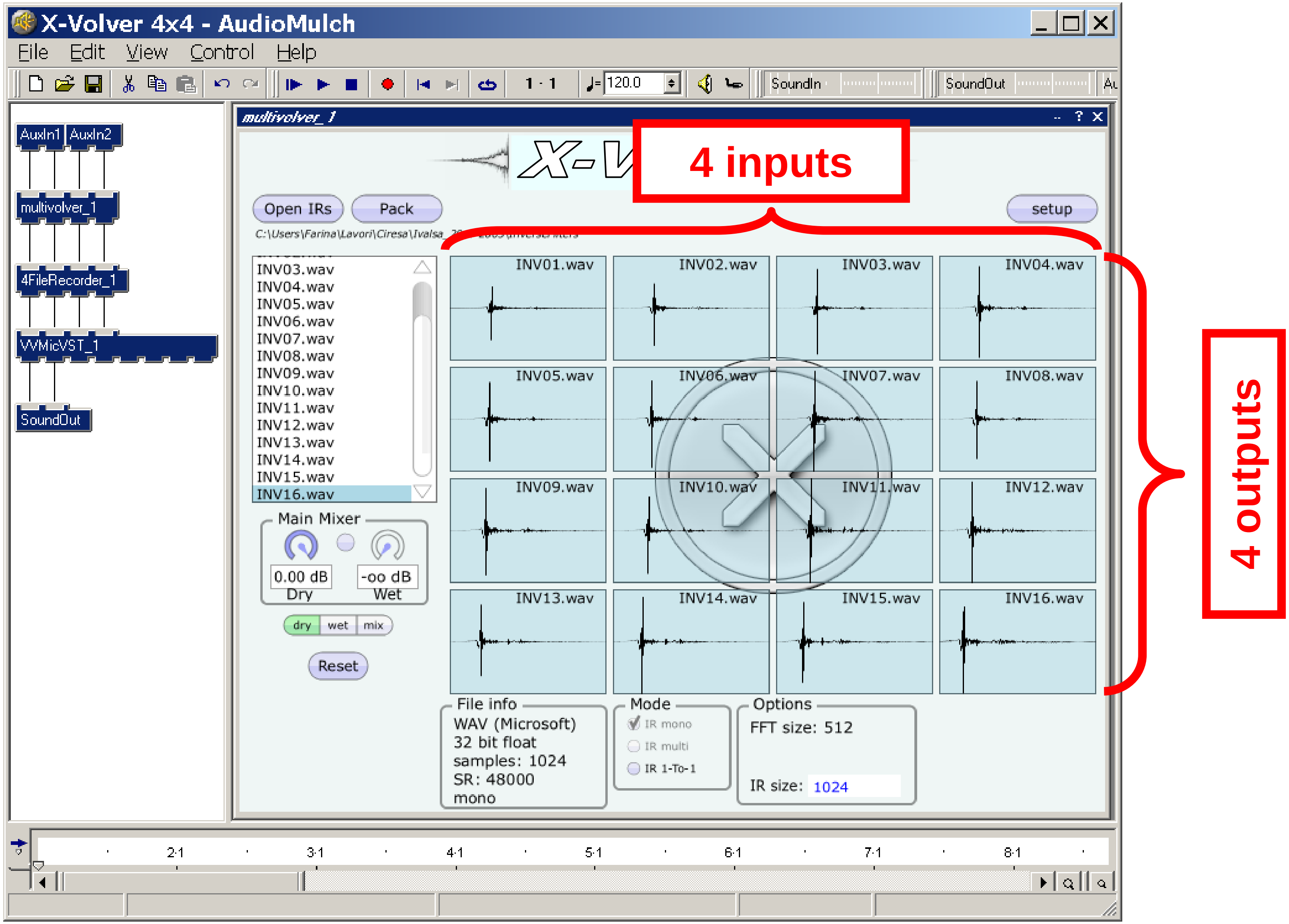
84 IRs were measured, uniformly scattered around a sphere

Computation of the inverse filters

- **A set of 16 inverse filters is required (4 inputs, 4 outputs = 1°-order B-format)**
- **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:**

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

Real-time implementation

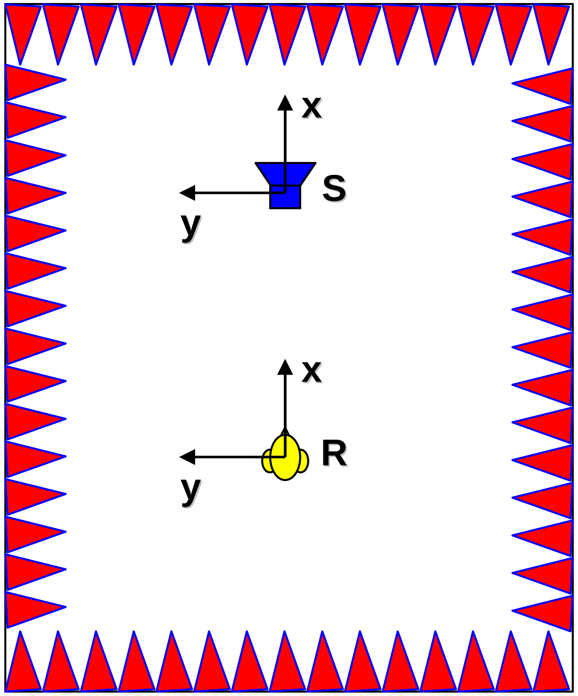


Conclusions

- Traditional methods for measuring “spatial parameters” proved to be unreliable and do not provide complete information
- The 1°-order Ambisonics method can be used for generating and recording sound with a limited amount of spatial information
- For obtaining better spatial resolution, High-Order Ambisonics can be used, limiting the spherical-harmonics expansion to a reasonable order (2°, 3° or 4°).
- Experimental hardware and software tools have been developed (mainly in France, but also in USA), allowing to build an inexpensive complete measurement system
- From the complete matrix of measured impulse responses it is easy to derive any suitable subset, including an highly accurate binaural rendering over head-tracked headphones.

Auralization Example #1

- **Anecoic room, one source, one receiver**

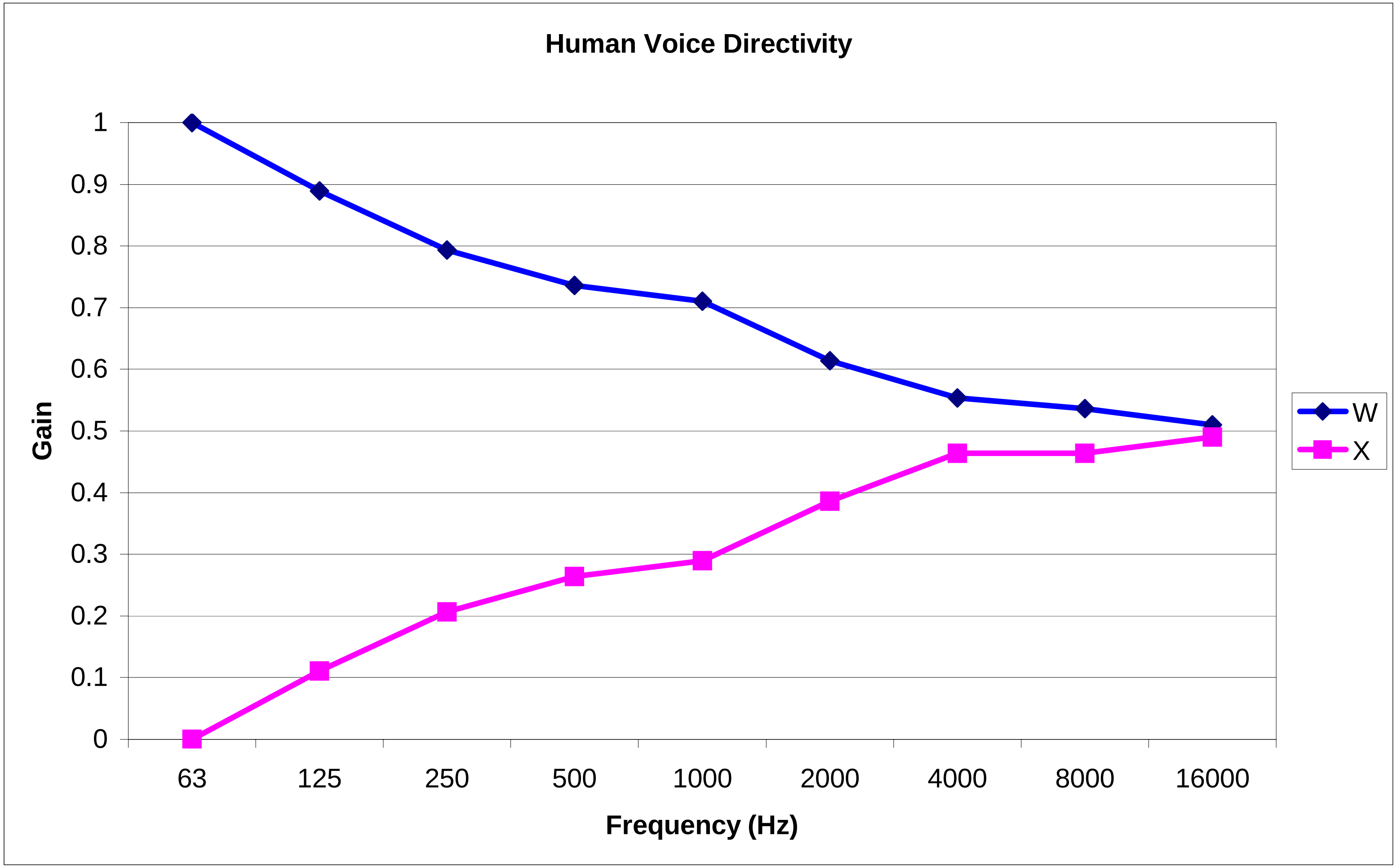


Gains for IR Matrix

	W	X	Y	Z
W	1	-1	0	0
X	1	-1	0	0
Y	0	0	0	0
Z	0	0	0	0

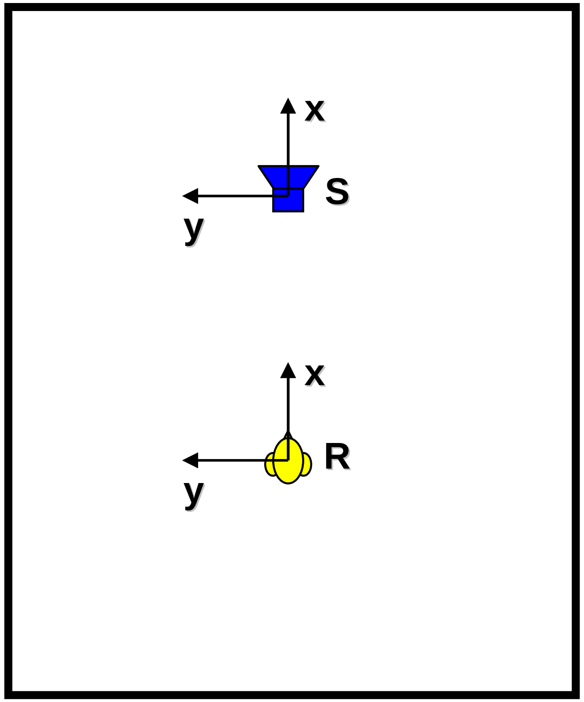
Source Directivity

- **The frequency-dpendent directivity of the human voice has been approximated with first-order components:**



Auralization Example #2

- **Reverbersnt room, one source, one receiver**



Gains for IR Matrix

	W	X	Y	Z
W	h_{WW}	h_{XW}	h_{YW}	h_{ZW}
X	h_{WX}	h_{XX}	h_{YX}	h_{ZX}
Y	h_{WY}	h_{XY}	h_{YY}	h_{ZY}
Z	h_{WZ}	h_{XZ}	h_{YZ}	h_{ZZ}