WESPAC IX 2006
The 9th Western Pacific Acoustics Conference
Seoul, Korea, June 26-28, 2006

ASK
Automotive Industries Group
zendan
SIPE

University of Parma

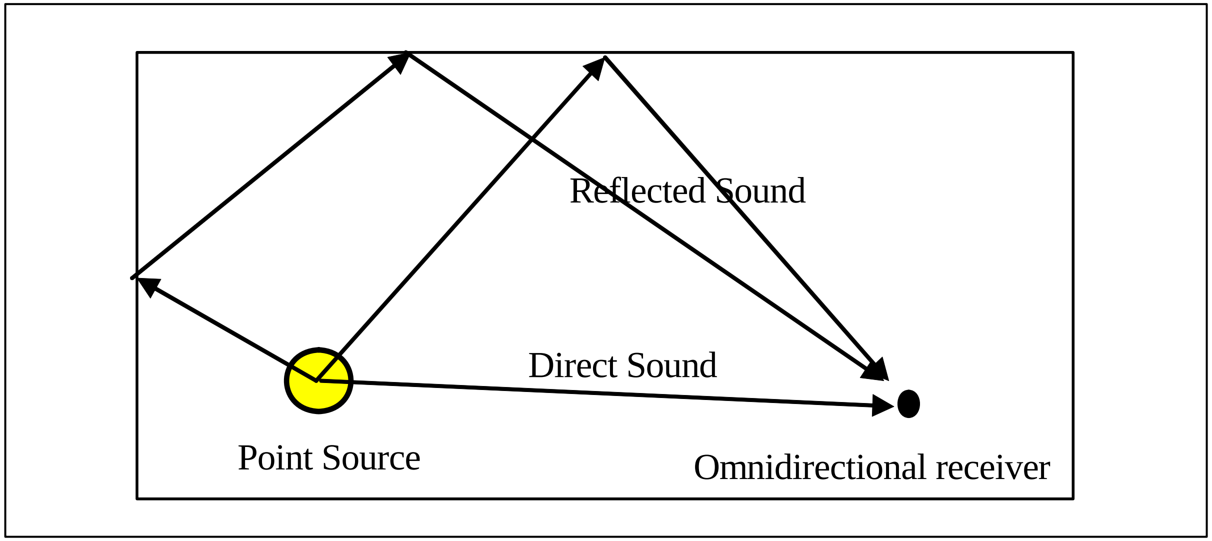
ROOM IMPULSE RESPONSES AS TEMPORAL AND SPATIAL FILTERS

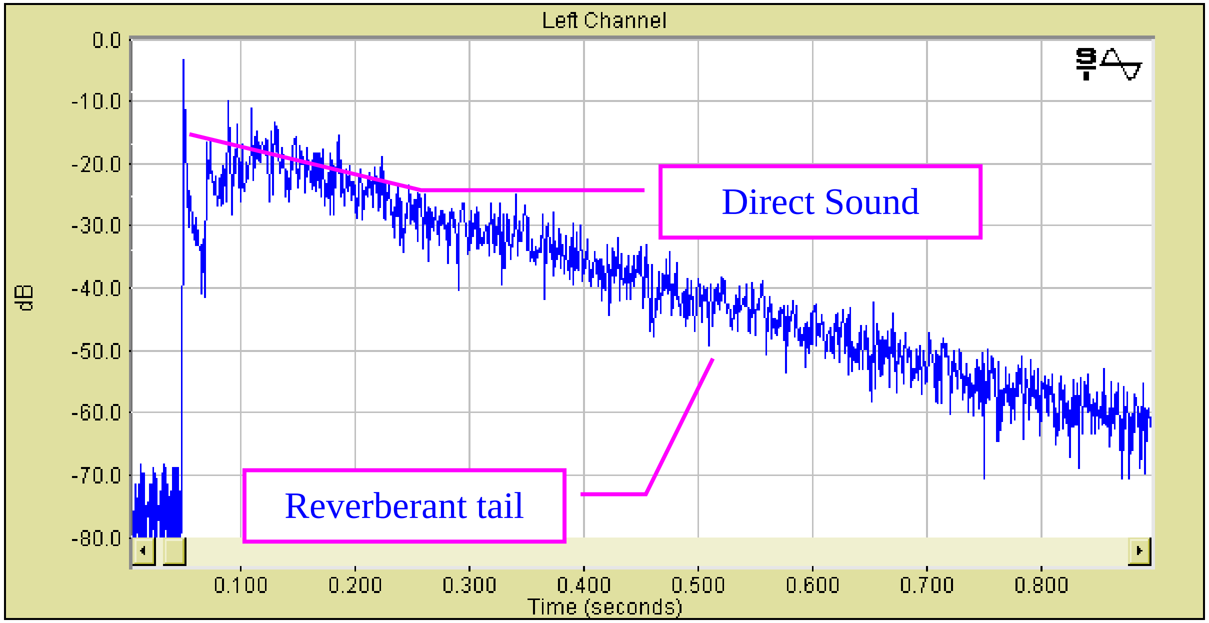
Angelo Farina
Industrial Engineering Dept.
University of Parma - ITALY
HTTP://www.angelofarina.it

Topics

- Traditional time-domain measurements with omnidirectional transducers
- Advanced impulse response measurement methods
- 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 response

Basic sound propagation scheme





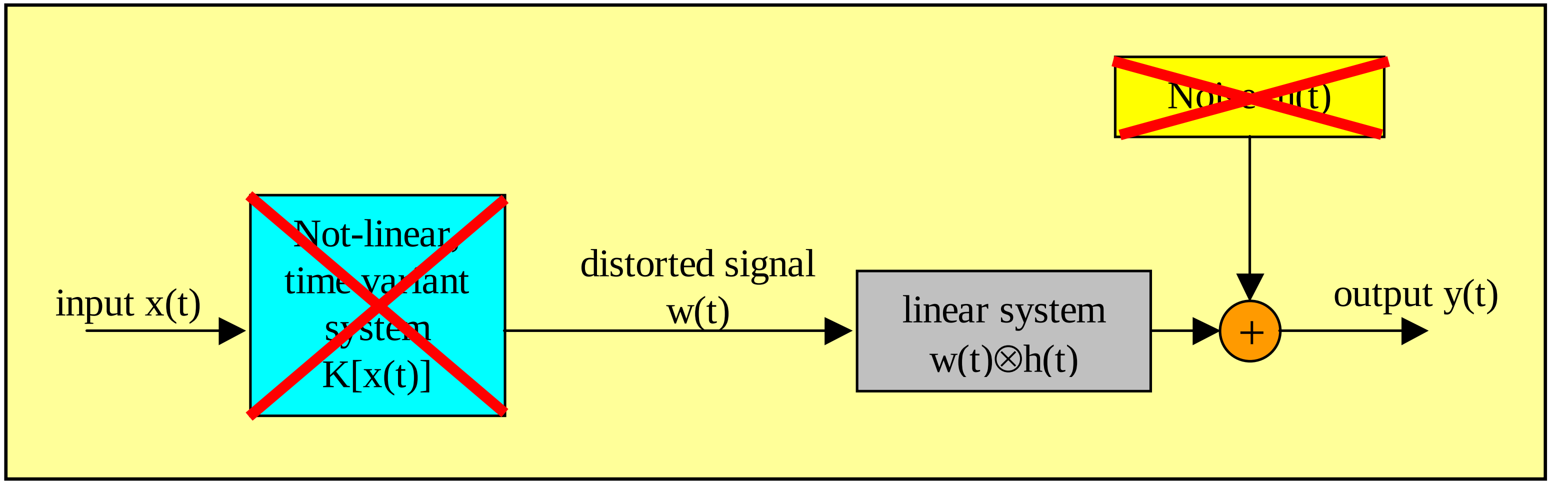
Traditional measurement methods

- **Pulsive sources: ballons, blank pistol**

Modern electroacoustical methods

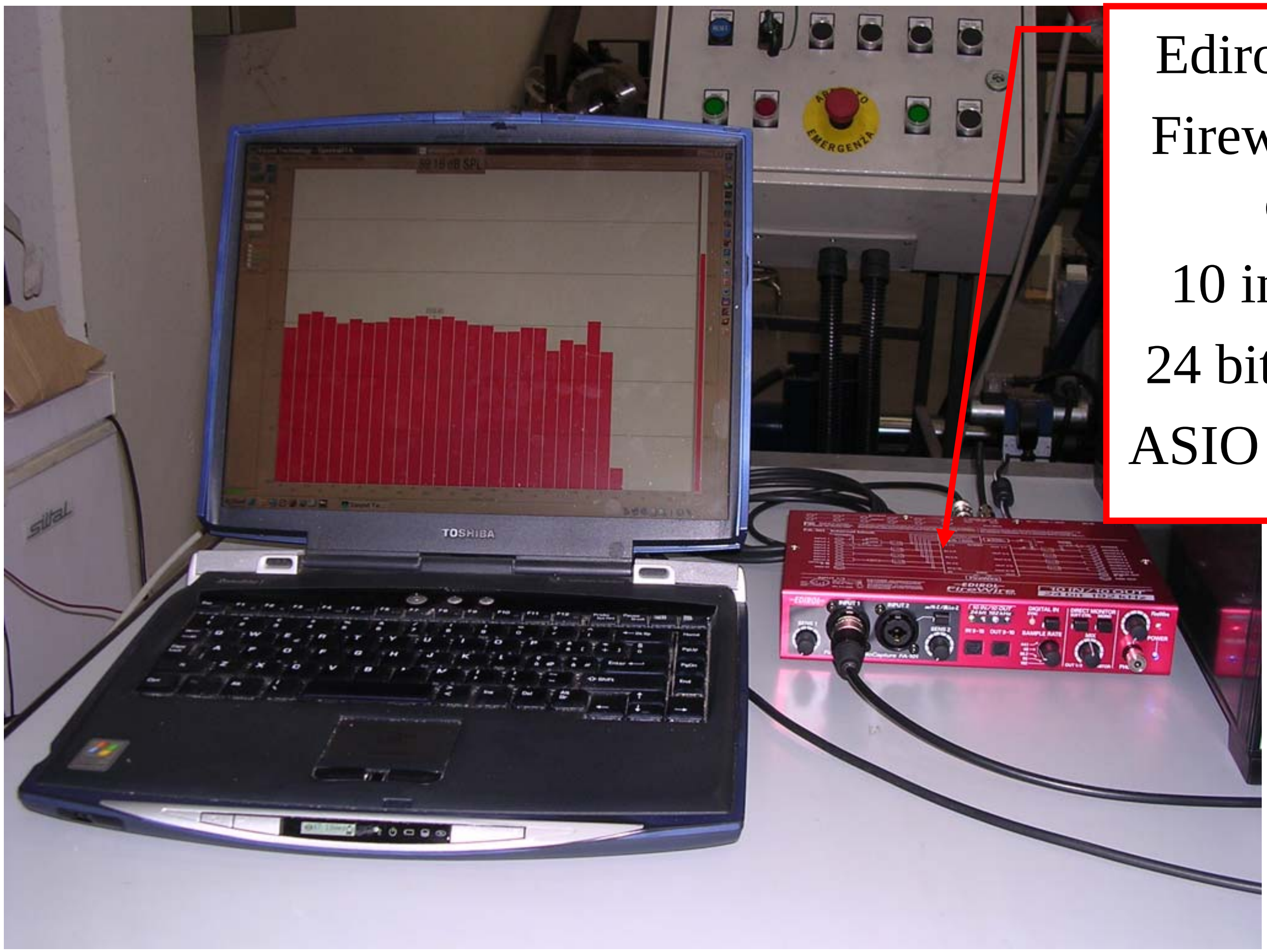
- **The sound is generated by means of an omnidirectional loudspeaker**
- **The signal is computer-generated**
- **The same computer is also employed for recording the room'response by means of one or more omnidirectional microphones**
- **Also directive microphones can be used: binaural, figure-of-eight**
- **Different types of test signals have been developed, providing good immunity to background noise and easy deconvolution of the impulse response:**
  - MLS (Maximum Lenght Sequence, pseudo-random white noise)
  - TDS (Time Delay Spectrometry, which basically is simply a linear sine sweep, also known in Japan as "stretched pulse")
  - ESS (Exponential Sine Sweep)
- **Each of these test signals can be employed with different deconvolution techniques, resulting in a number of "different" measurement methods**
- **Due to theoretical and practical considerations, the preference is nowadays generally oriented for the usage of ESS with not-circular deconvolution**

Measurement process



- The desidered result is the linear impulse response of the acoustic propagation h(t). It can be recovered by knowing the test signal x(t) and the measured system output y(t).
- It is necessary to exclude the effect of the not-linear part K and of the background noise n(t).

Hardware: PC and audio interface



Edirol FA-101

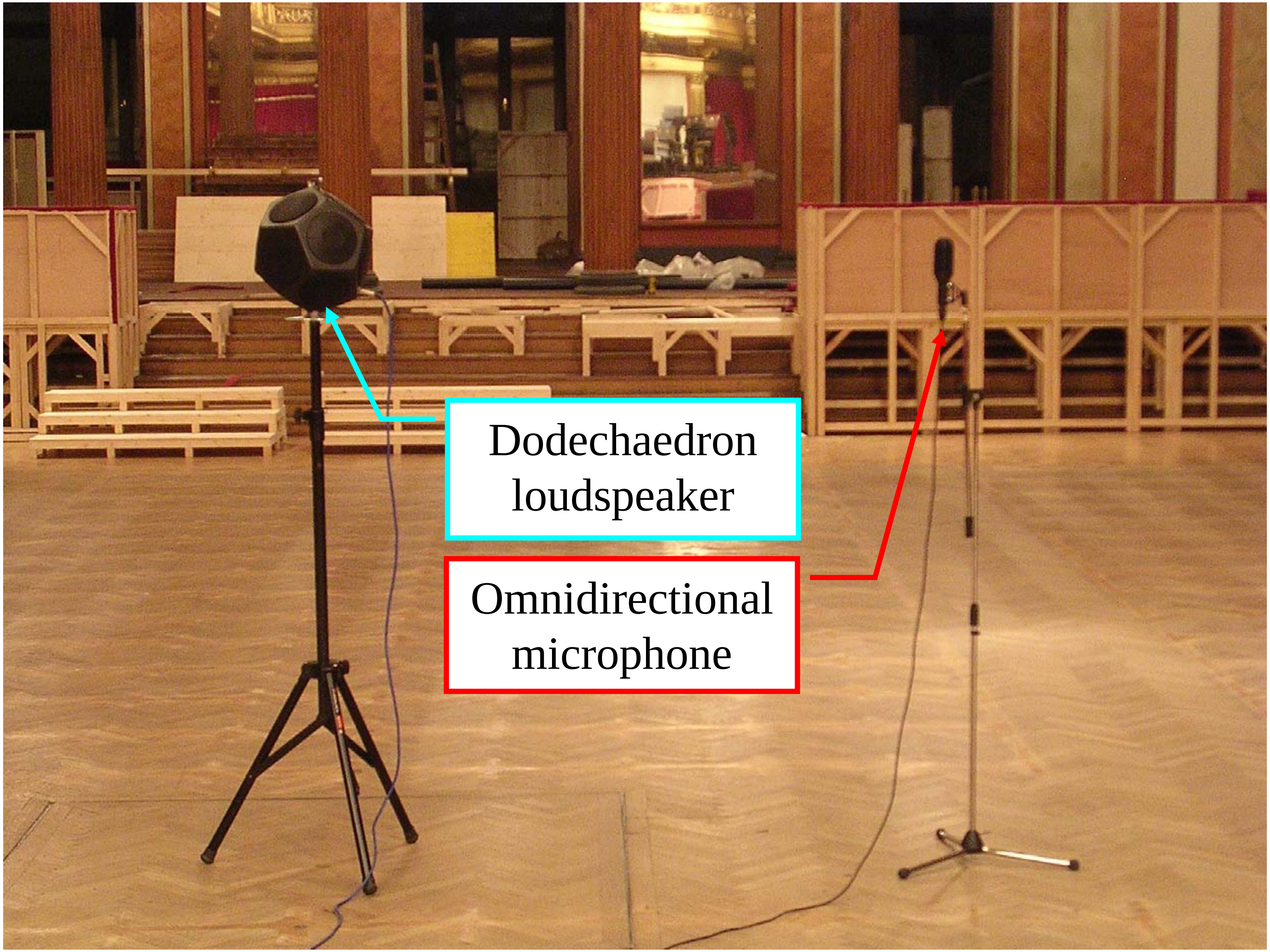
Firewire sound card:

10 in / 10 out

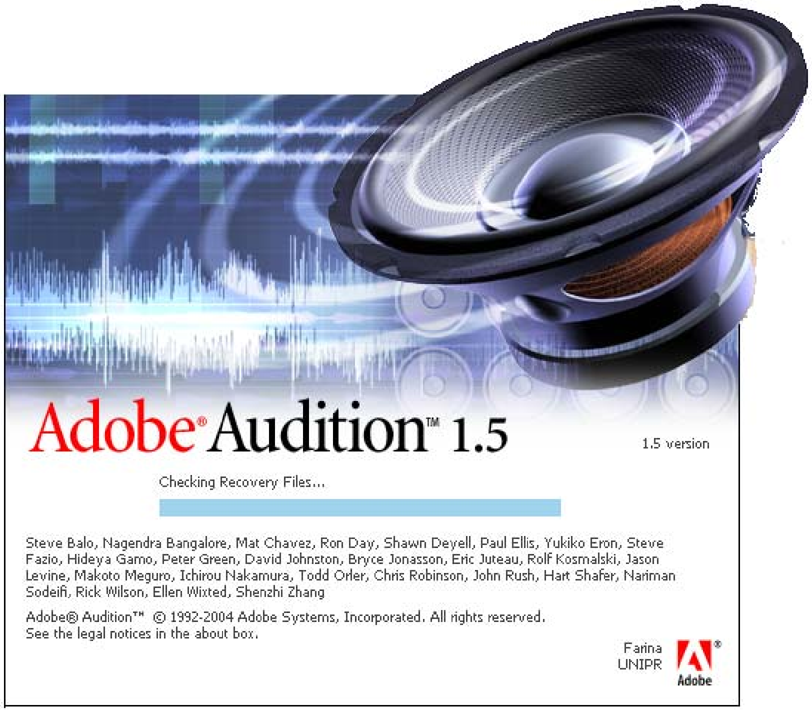
24 bit, 192 kHz

ASIO and WDM

Hardware: loudspeaker & microphone



Software

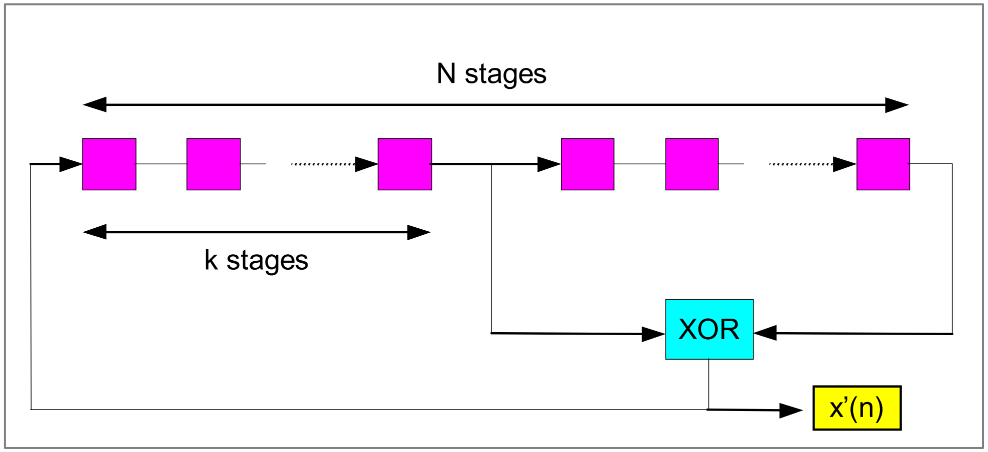


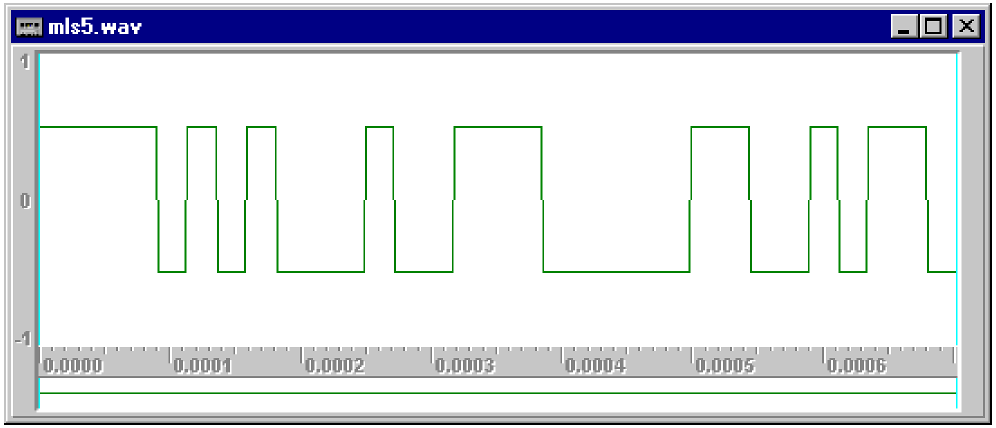
+

Aurora Plugins

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

MLS method





- X(t) is a periodic binary signal obtained with a suitable shift-register, configured for maximum lenght of the period.

$$L = 2^N - 1$$

MLS deconvolution

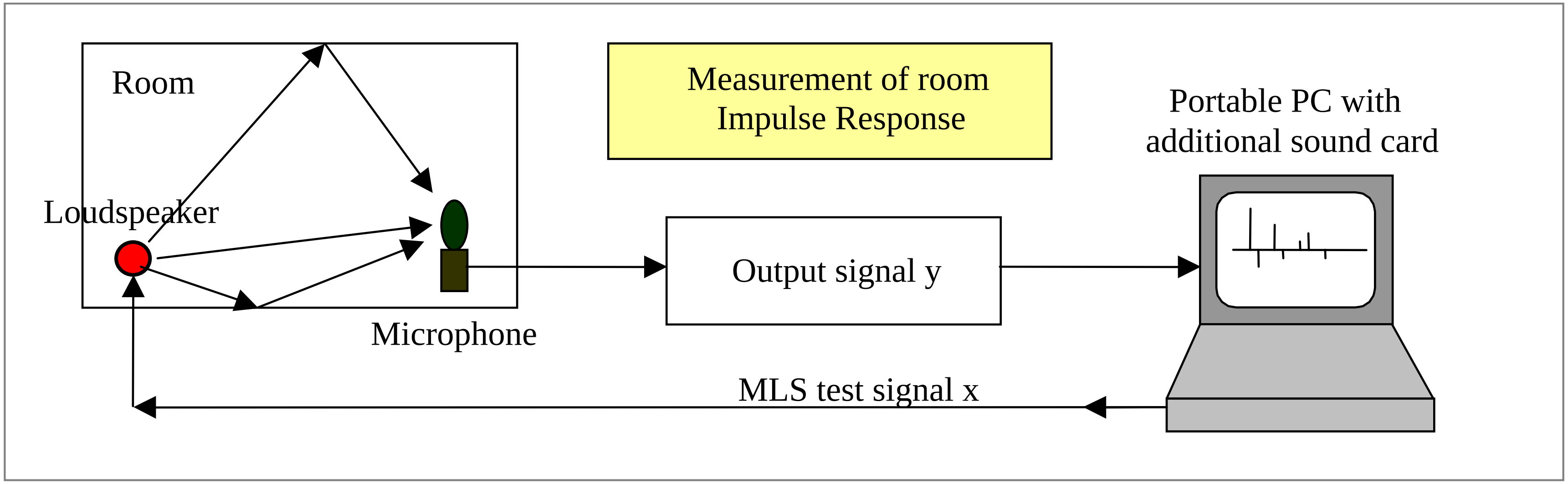
- **The re-recorded signal y(i) is cross-correlated with the excitation signal thanks to a fast Hadamard transform. The result is the required impulse response h(i), if the system was linear and time-invariant**

$$h = \frac{1}{L+1} \cdot \tilde{M} \cdot y$$

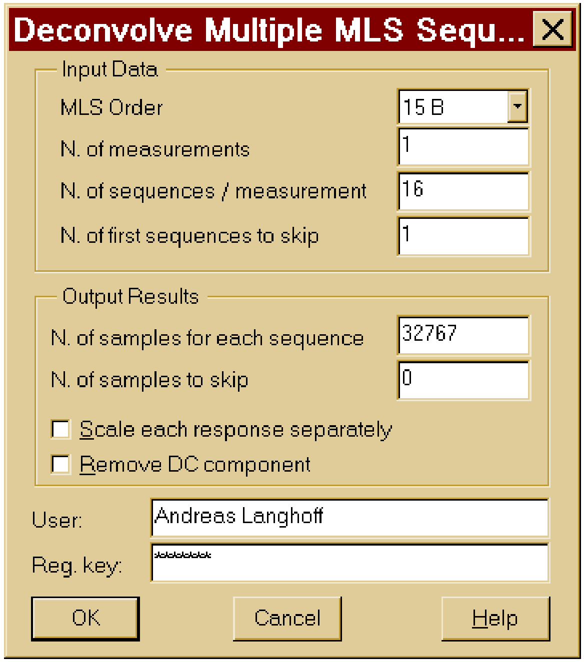
- **Where M is the Hadamard matrix, obtained by permutation of the original MLS sequence m(i)**

$$\tilde{M}(i, j) = m[(i + j - 2) \bmod L] - 1$$

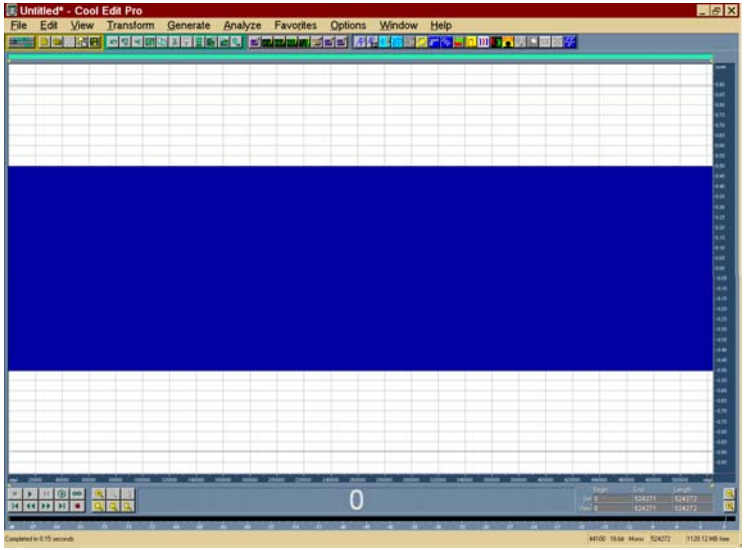
MLS example



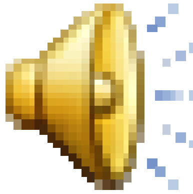
Room
Loudspeaker
Microphone
Measurement of room Impulse Response
Output signal y
Portable PC with additional sound card
MLS test signal x



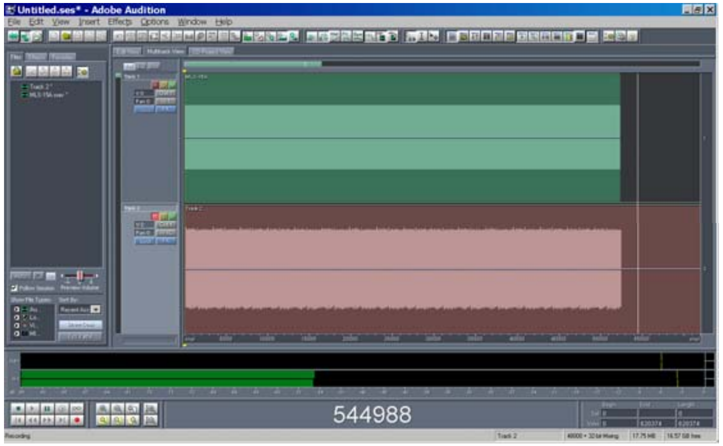
Deconvolve Multiple MLS Sequ...
Input Data
MLS Order
15 B
N. of measurements
1
N. of sequences / measurement
16
N. of first sequences to skip
1
Output Results
N. of samples for each sequence
32767
N. of samples to skip
0
Scale each response separately
Remove DC component
User:
Andreas Langhoff
Reg. key:
OK
Cancel
Help

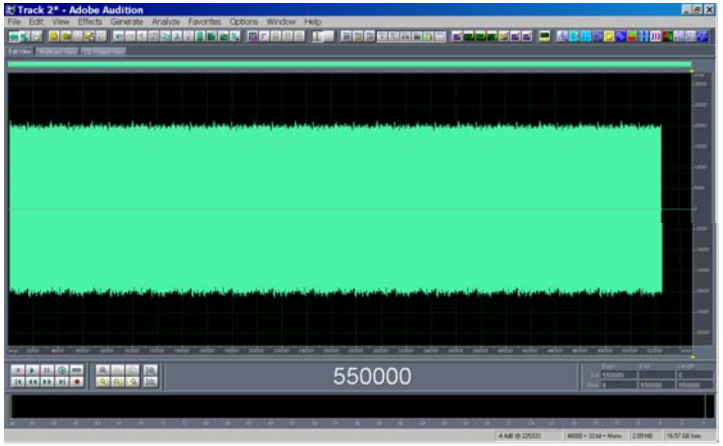


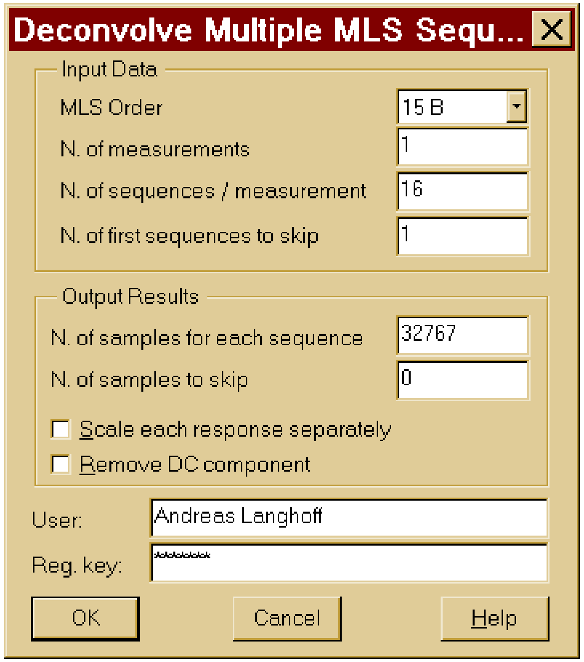
Untitled* - Cool Edit Pro
File Edit View Transform Generate Analyze Favorites Options Window Help
0

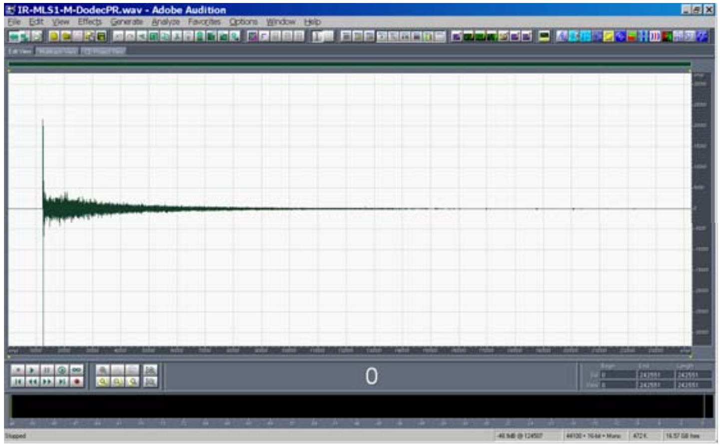


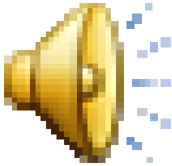
MLS example


Untitled.ses* - Adobe Audition
544988


Track 2* - Adobe Audition
550000


Deconvolve Multiple MLS Sequ...
Input Data
MLS Order
15 B
N. of measurements
1
N. of sequences / measurement
16
N. of first sequences to skip
1
Output Results
N. of samples for each sequence
32767
N. of samples to skip
0
Scale each response separately
Remove DC component
User:
Andreas Langhoff
Reg. key:
OK
Cancel
Help


IR-MLS1-M-DodecPR.wav - Adobe Audition
0

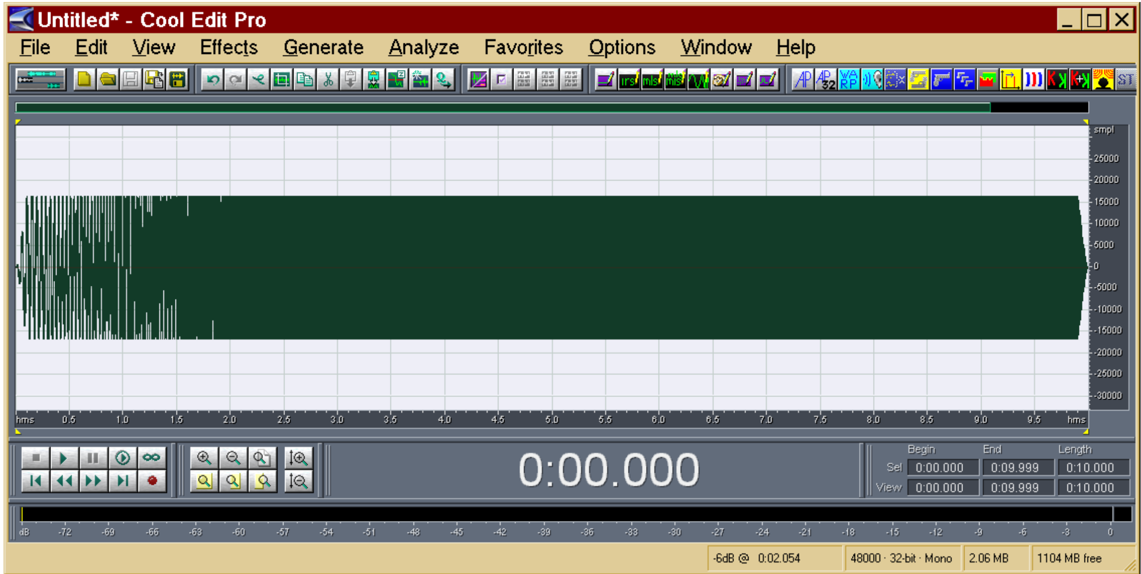


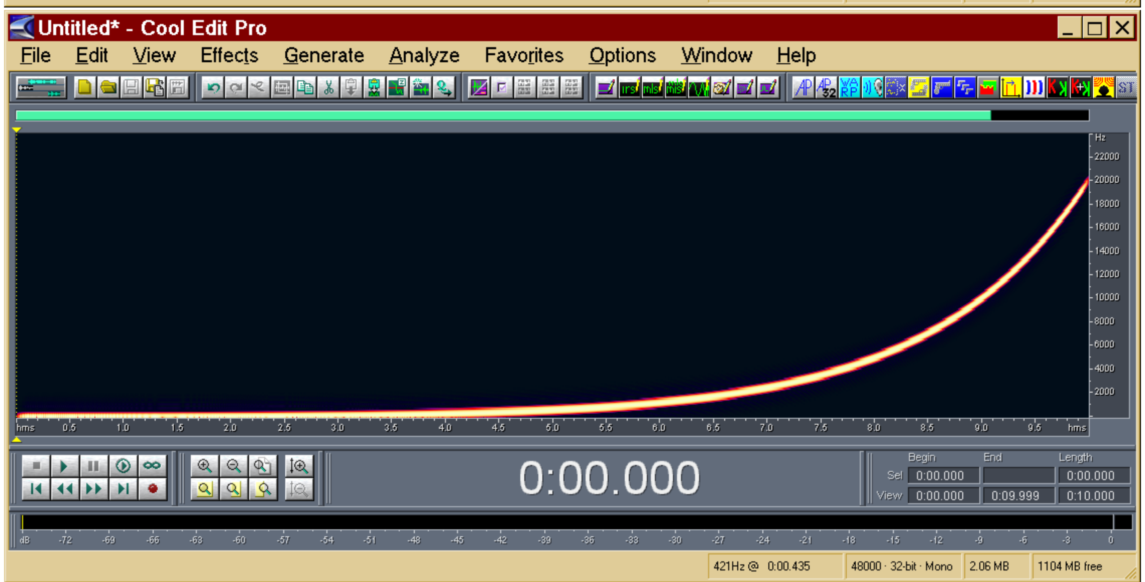
Exponential Sine Sweep method

- **x(t) is a sine signal, which frequency is varied exponentially with time, starting at f_1 and ending at f_2.**

$$x(t) = \sin\left[\frac{2\cdot\pi\cdot f_1\cdot T}{\ln\left(\frac{f_2}{f_1}\right)}\cdot\left(e^{\frac{t}{T}\cdot\ln\left(\frac{f_2}{f_1}\right)} - 1\right)\right]$$

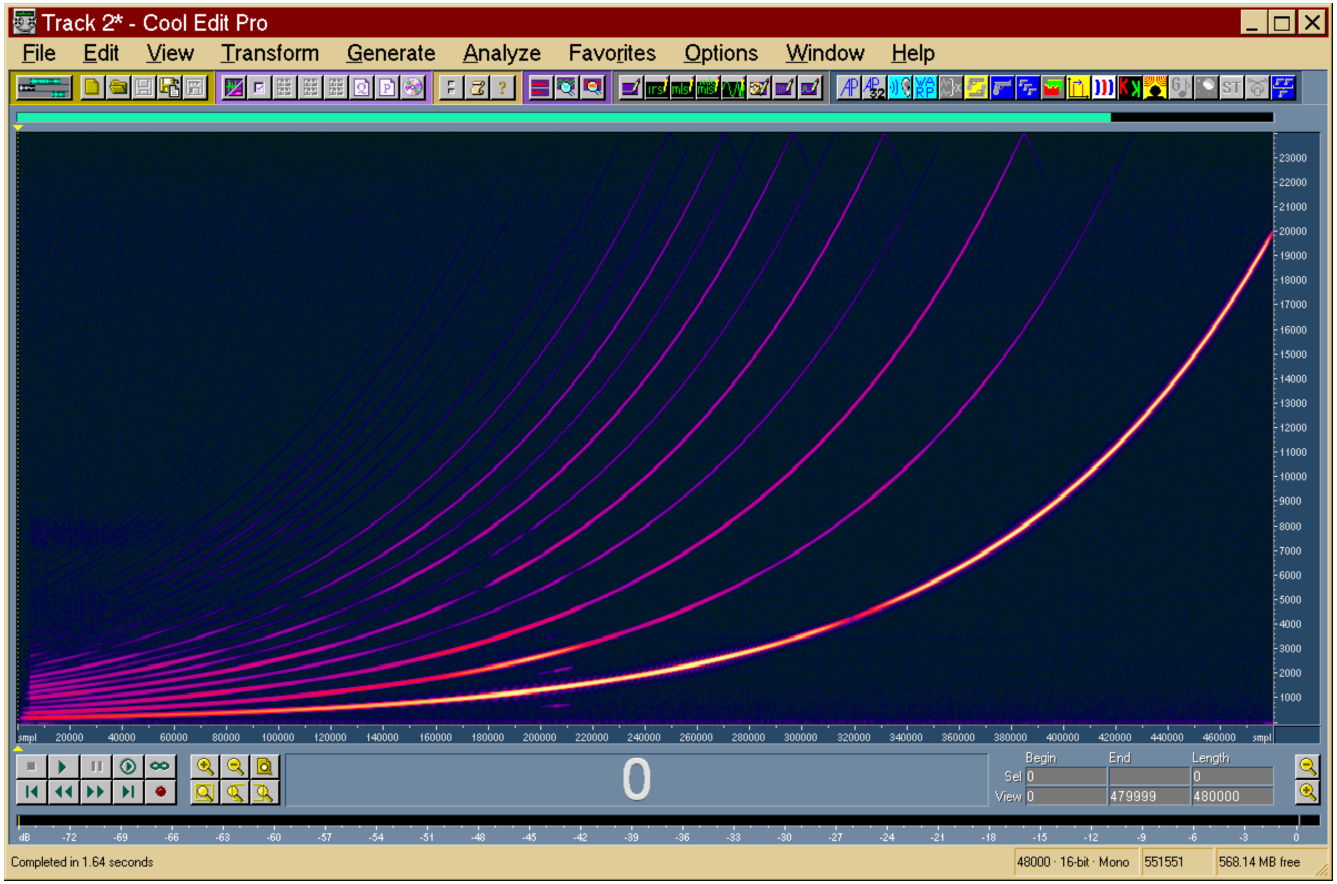
Test Signal – x(t)





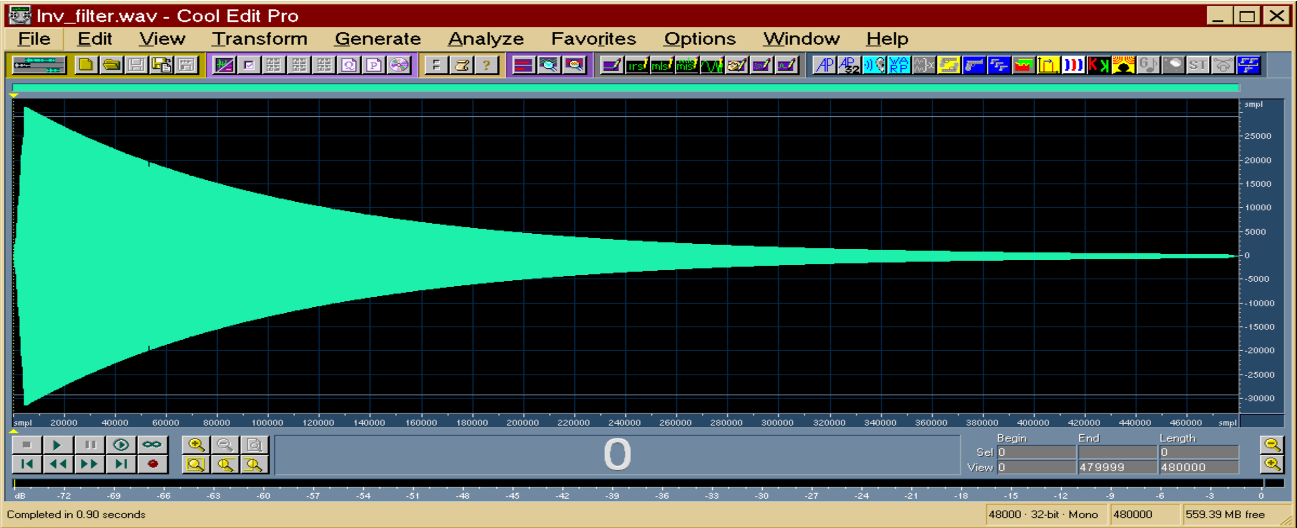
Measured signal - y(t)

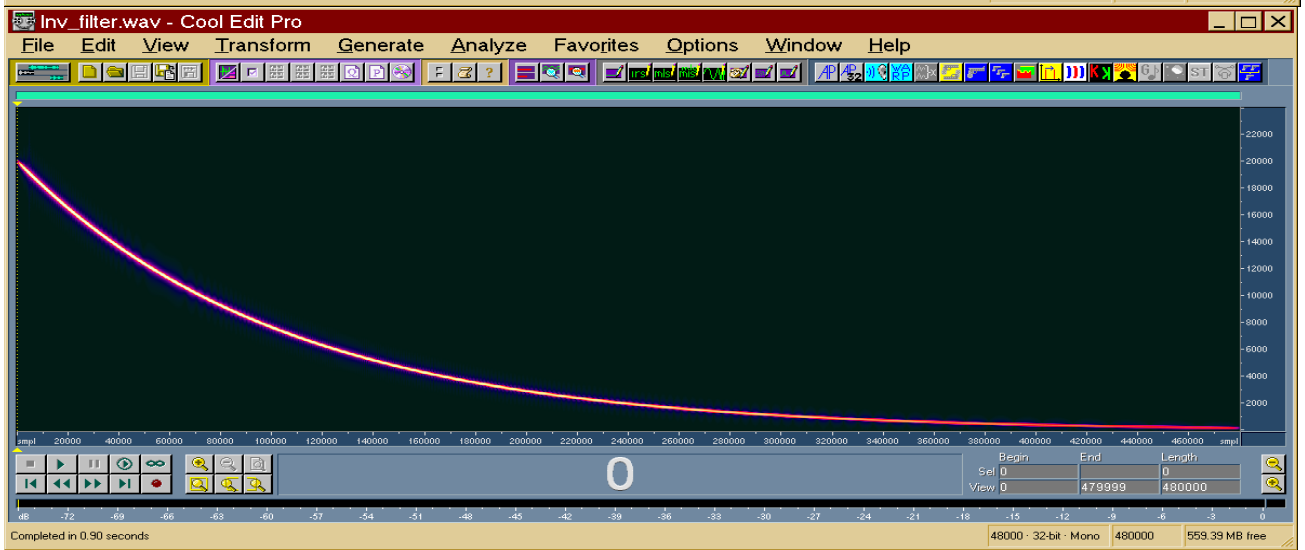
- **The not-linear behaviour of the loudspeaker causes many harmonics to appear**



Inverse Filter – z(t)

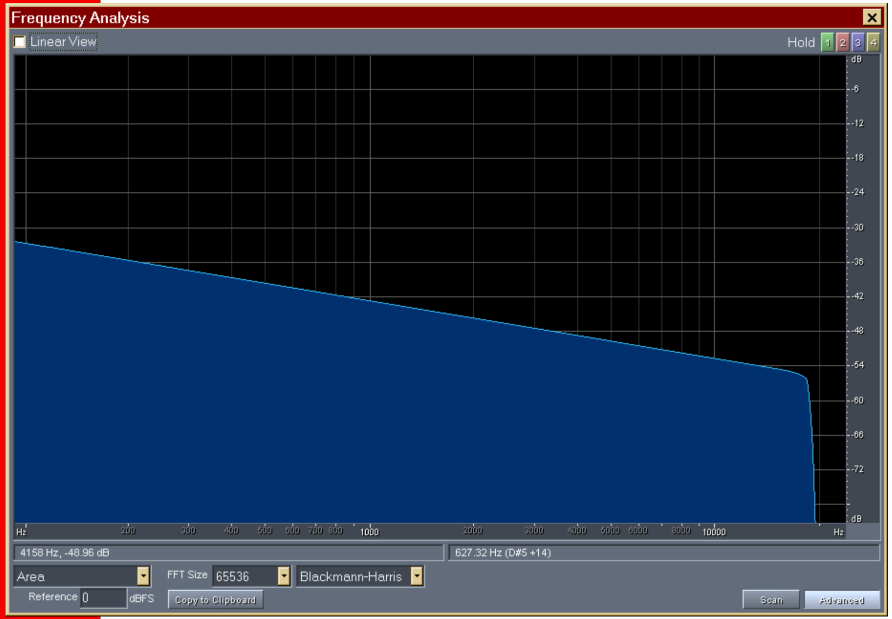
The deconvolution of the IR is obtained convolving the measured signal y(t) with the inverse filter z(t) [equalized, time-reversed x(t)]

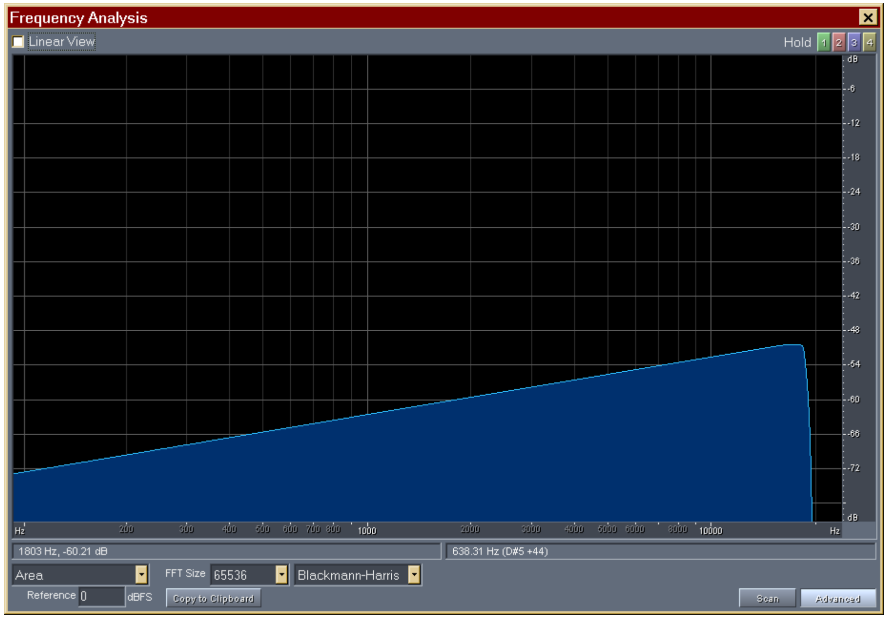




Deconvolution of Log Sine Sweep

The "time reversal mirror" technique is employed: the system's impulse response is obtained by convolving the measured signal y(t) with the time-reversal of the test signal x(-t). As the log sine sweep does not have a "white" spectrum, proper equalization is required

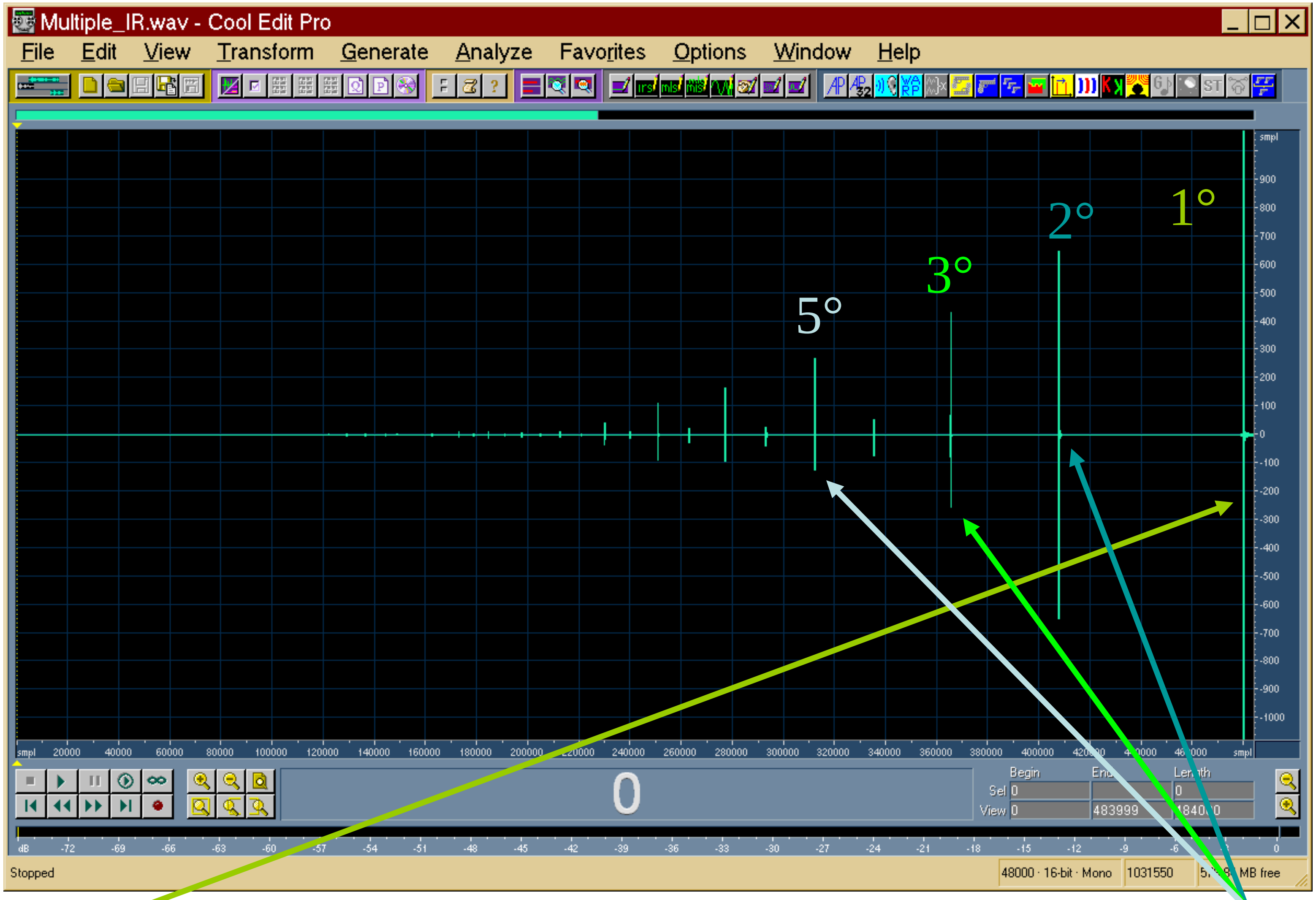




Test Signal x(t)

Inverse Filter z(t)

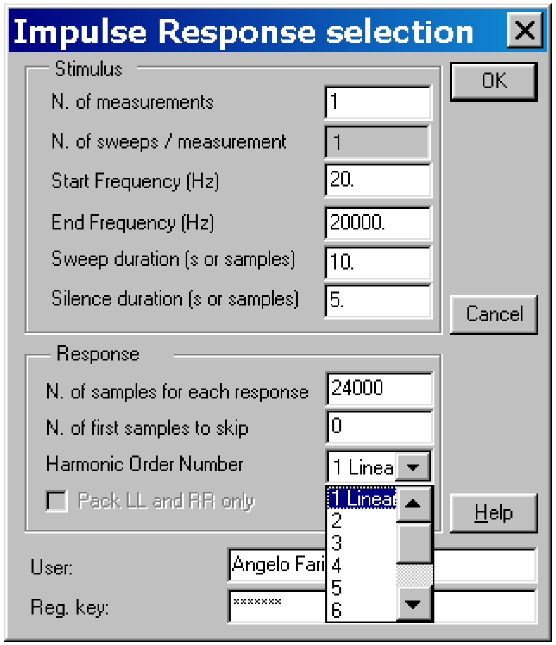
Result of the deconvolution

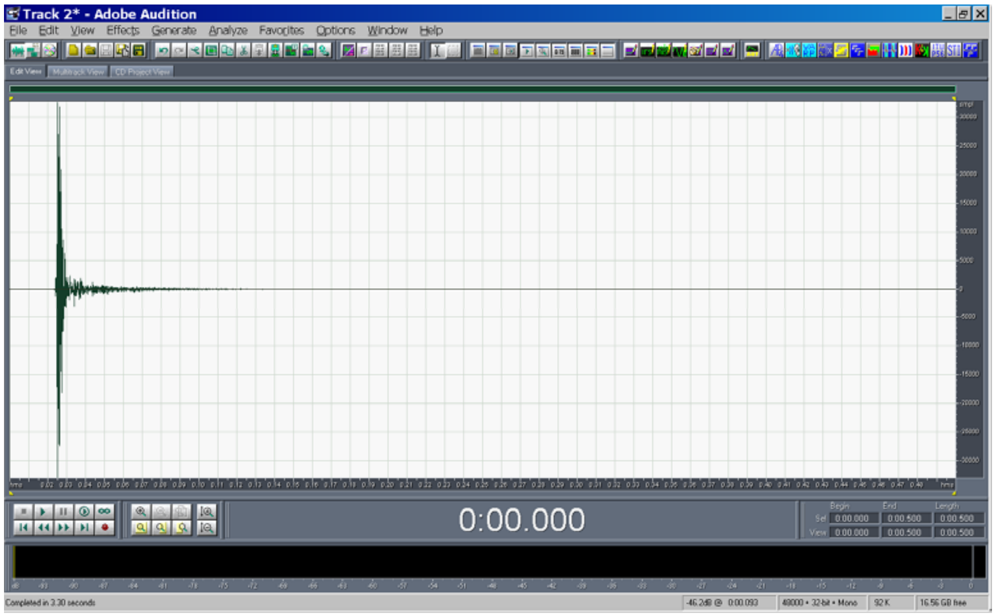


The last impulse response is the linear one, the preceding are the harmonics distortion products of various orders

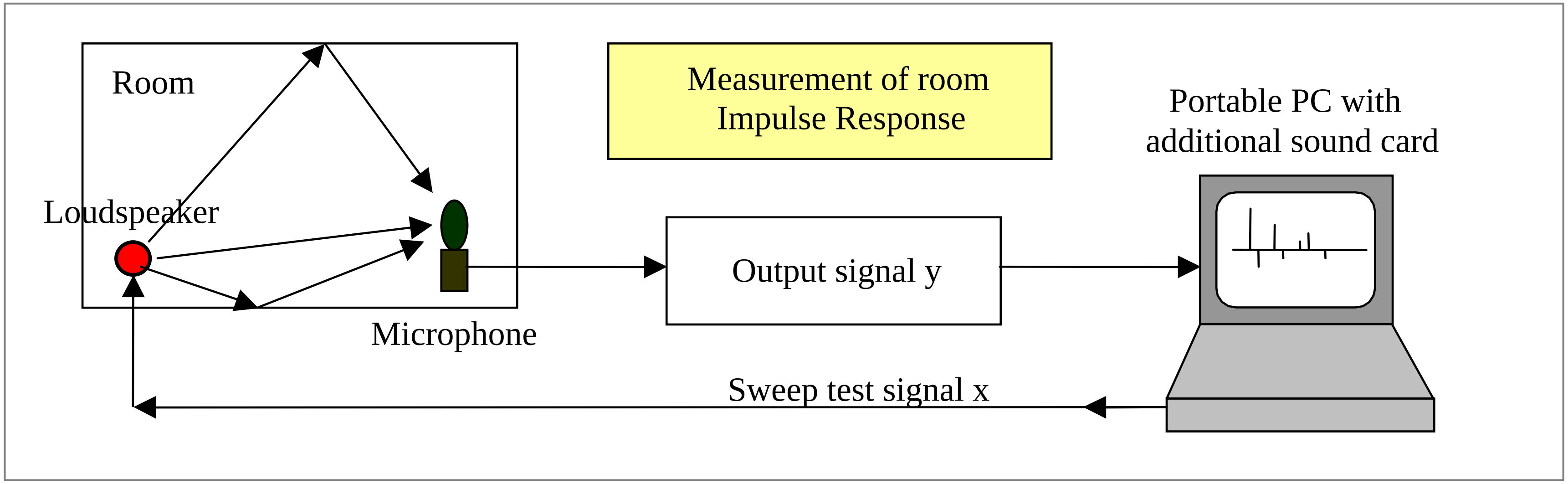
IR Selection

- **After the sequence of impulse responses has been obtained, it is possible to select and insulate just one of them:**

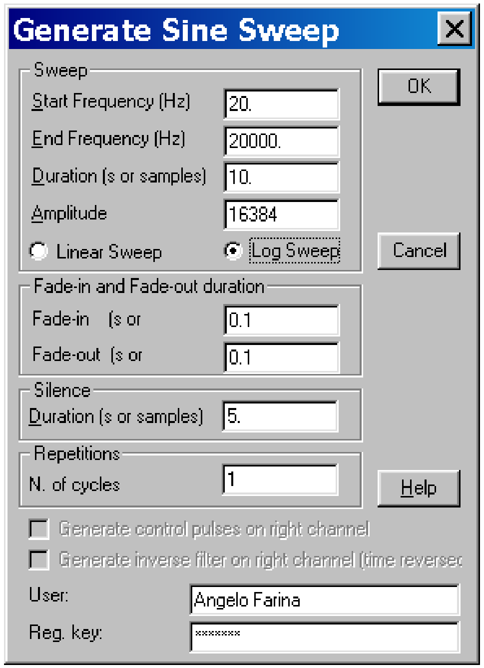




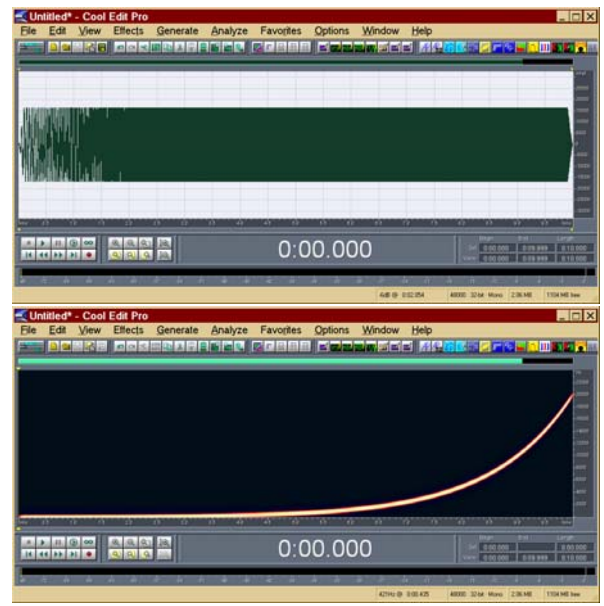
ESS example



Room
Loudspeaker
Microphone
Measurement of room Impulse Response
Output signal y
Portable PC with additional sound card
Sweep test signal x

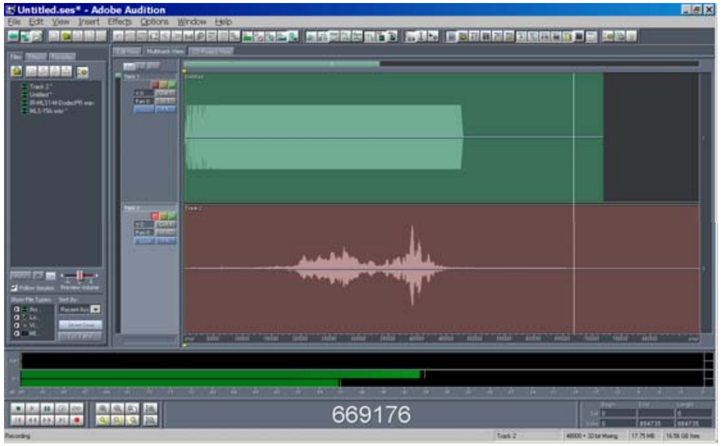


Generate Sine Sweep
Sweep
Start Frequency (Hz) 20.
End Frequency (Hz) 20000.
Duration (s or samples) 10.
Amplitude 16384
Linear Sweep
Log Sweep
OK
Cancel
Fade-in and Fade-out duration
Fade-in (s or 0.1
Fade-out (s or 0.1
Silence
Duration (s or samples) 5.
Repetitions
N. of cycles 1
Help
Generate control pulses on right channel
Generate inverse filter on right channel (time reversec
User: Angelo Farina
Reg. key: *******

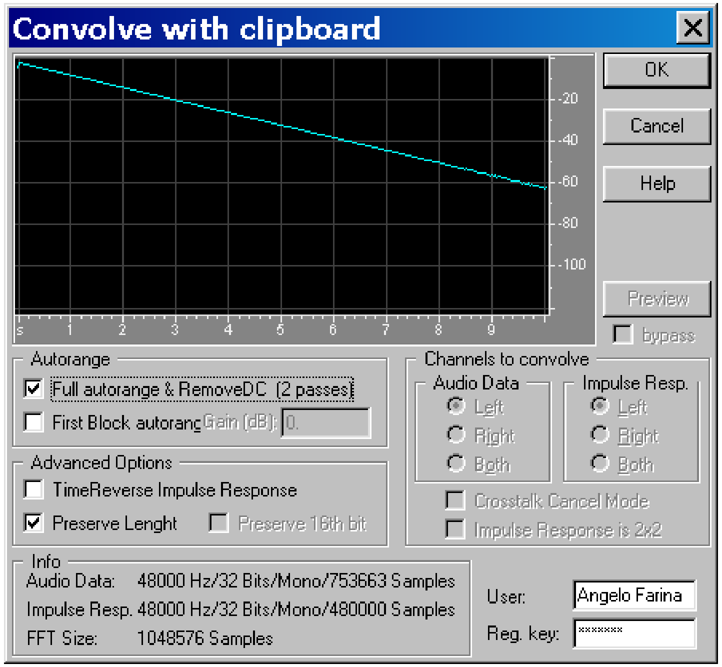


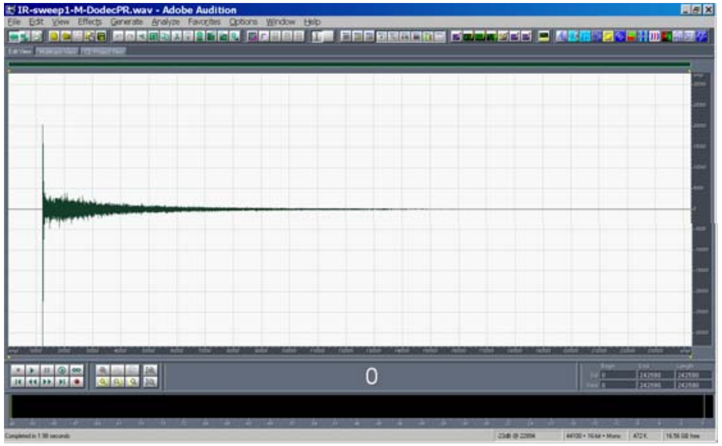
Untitled* - Cool Edit Pro
File Edit View Effects Generate Analyze Favorites Options Window Help
0:00.000
Untitled* - Cool Edit Pro
File Edit View Effects Generate Analyze Favorites Options Window Help
0:00.000

ESS example

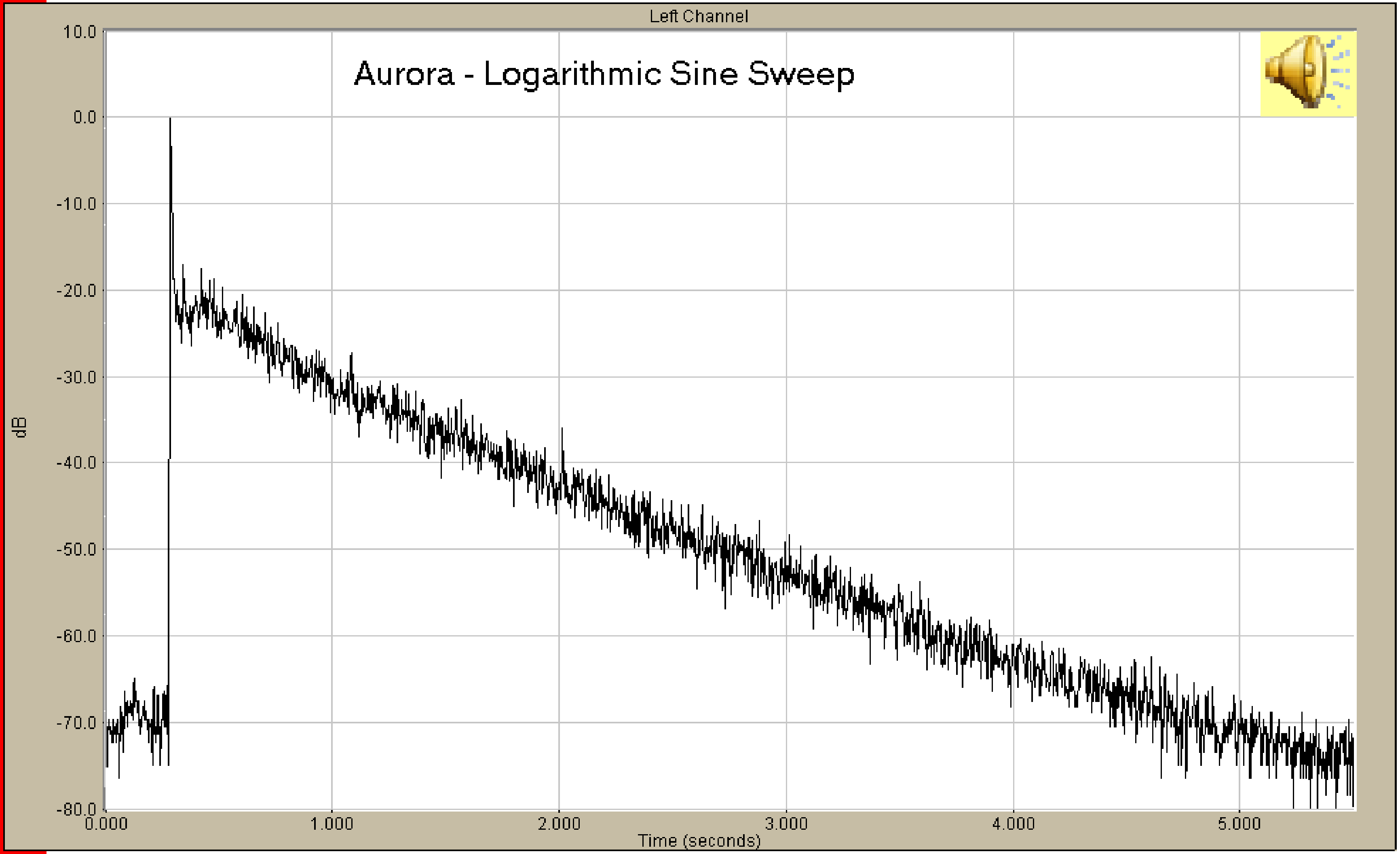

Untitled.ses* - Adobe Audition
669176


Track 2* - Adobe Audition
0


Convolve with clipboard
-20
-40
-60
-80
-100
s 1 2 3 4 5 6 7 8 9
OK
Cancel
Help
Preview
bypass
Autorange
Full autorange & RemoveDC (2 passes)
First Block autorang Gain (dB): 0.
Advanced Options
TimeReverse Impulse Response
Preserve Lenght
Preserve 16th bit
Channels to convolve
Audio Data
Left
Right
Both
Impulse Resp.
Left
Right
Both
Crosstalk Cancel Mode
Impulse Response is 2x2
Info
Audio Data: 48000 Hz/32 Bits/Mono/753663 Samples
Impulse Resp. 48000 Hz/32 Bits/Mono/480000 Samples
FFT Size: 1048576 Samples
User: Angelo Farina
Reg. key: *******

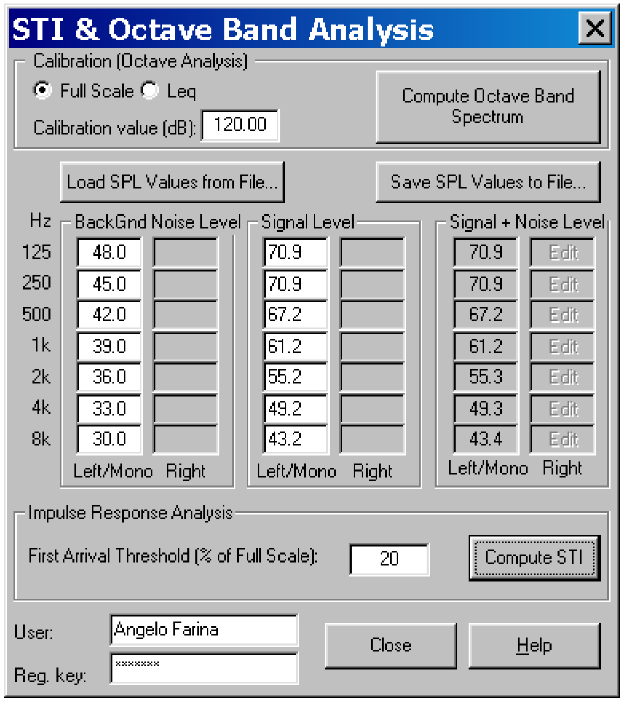

IR-sweep1-M-DodecPR.wav - Adobe Audition
0

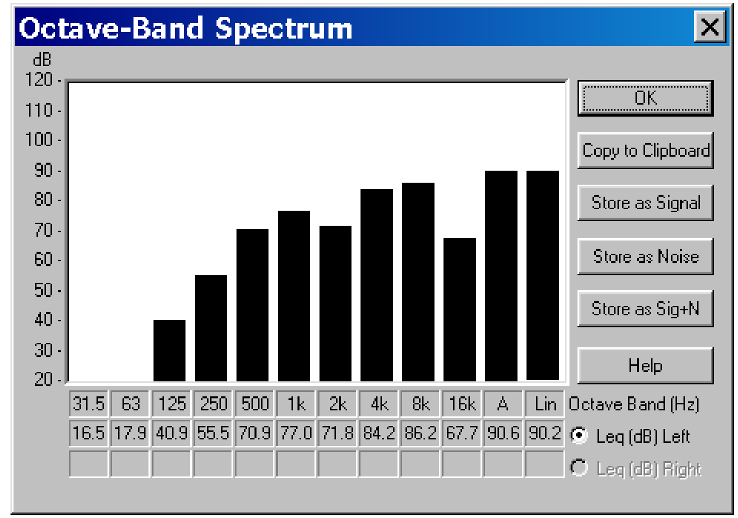
Maximum Length Sequence vs. Sweep

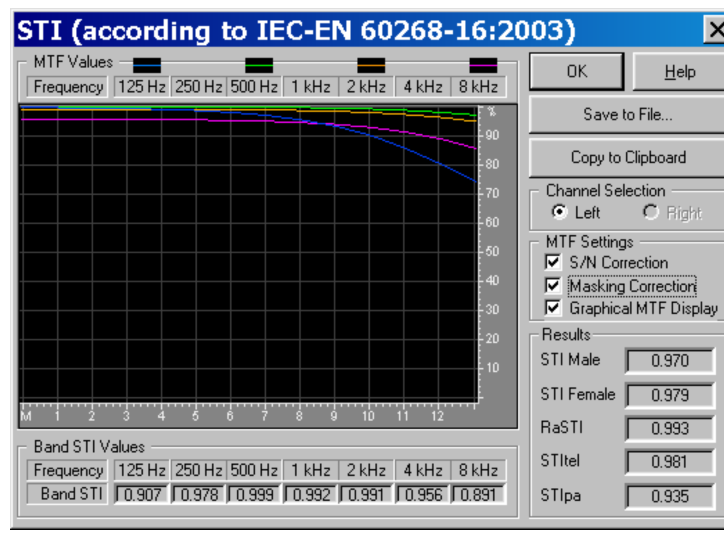


Post processing of impulse responses

- **A special plugin has been developed for the computation of STI according to IEC-EN 60268-16:2003**

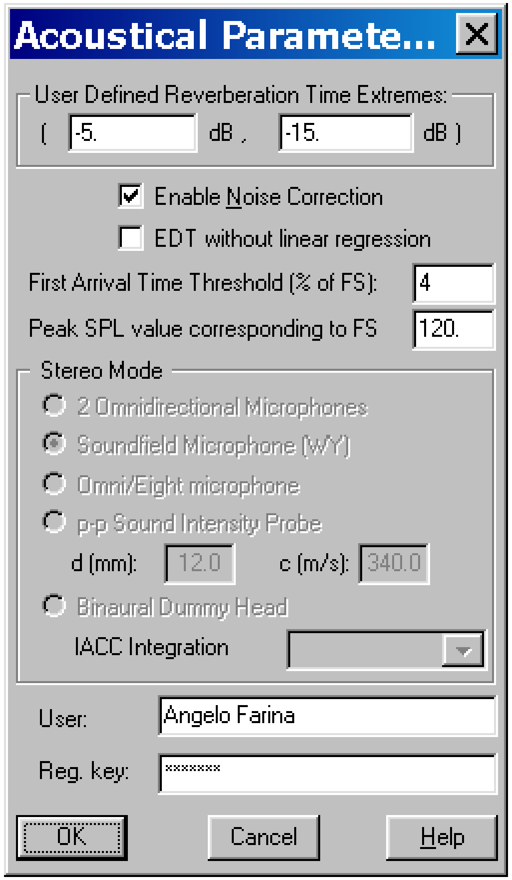


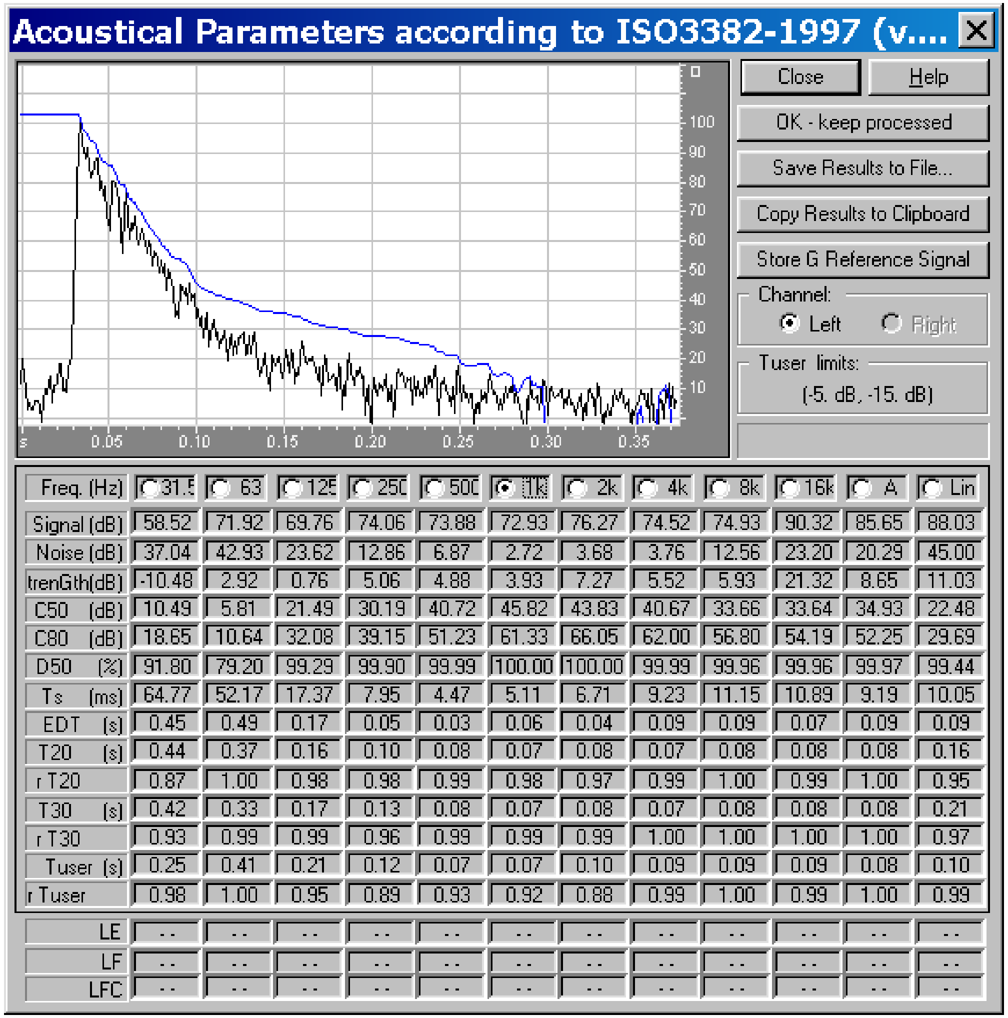




Post processing of impulse responses

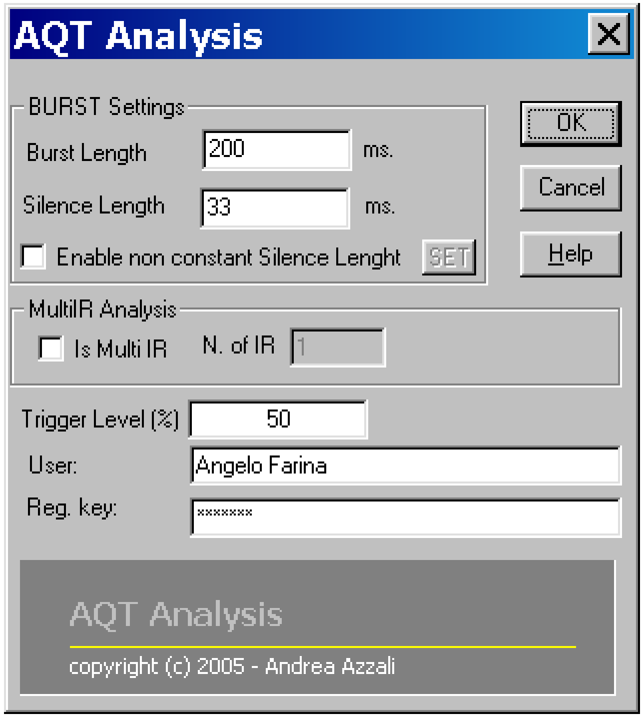
- **A special plugin has been developed for performing analysis of acoustical parameters according to ISO-3382**

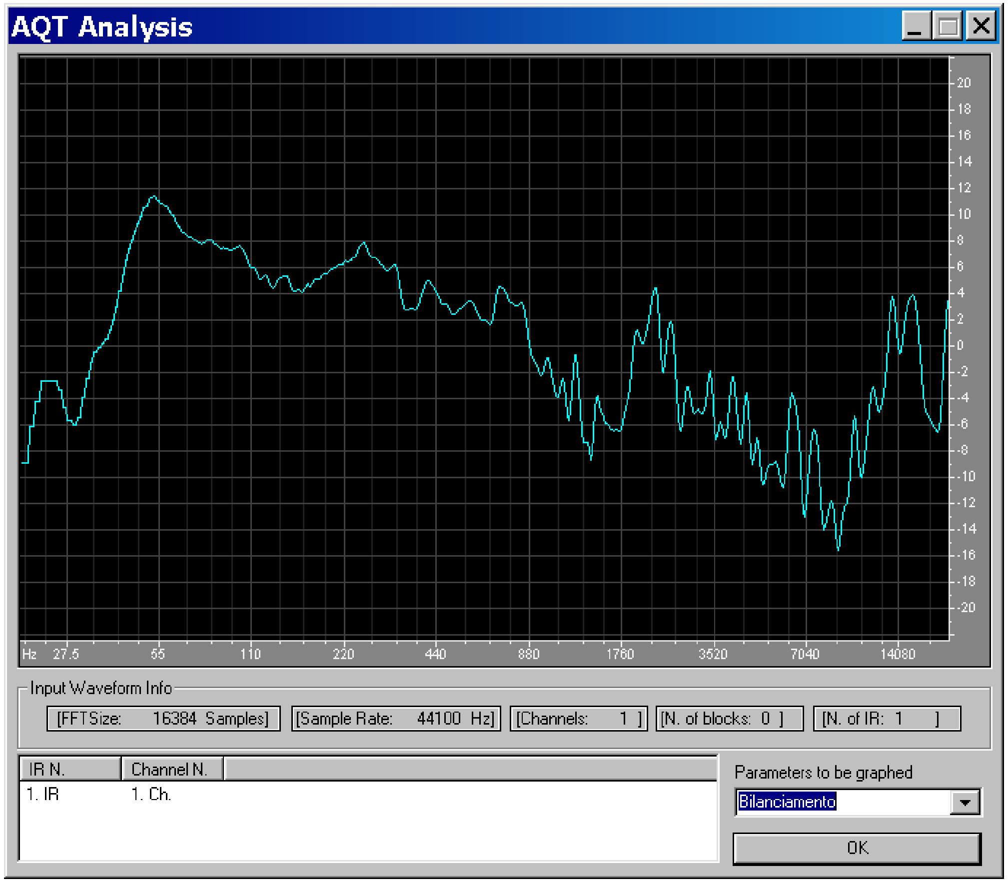




The new AQT plugin for Audition

- **The new module is still under development and will allow for very fast computation of the AQT (Dynamic Frequency Response) curve from within Adobe Audition**





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:**

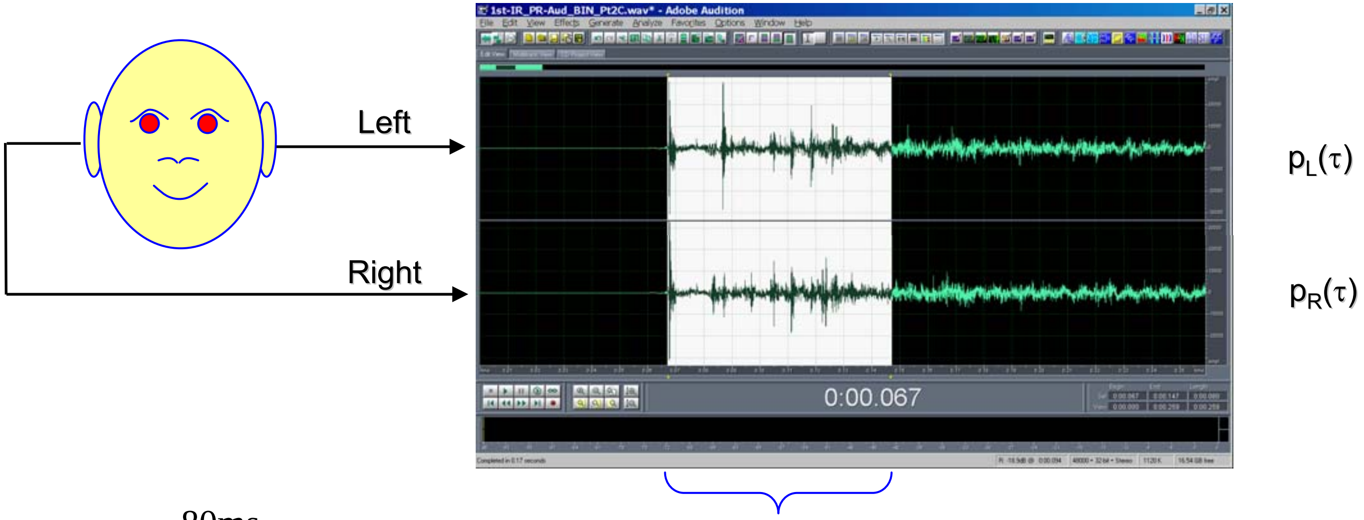
Binaural microphone (left)

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

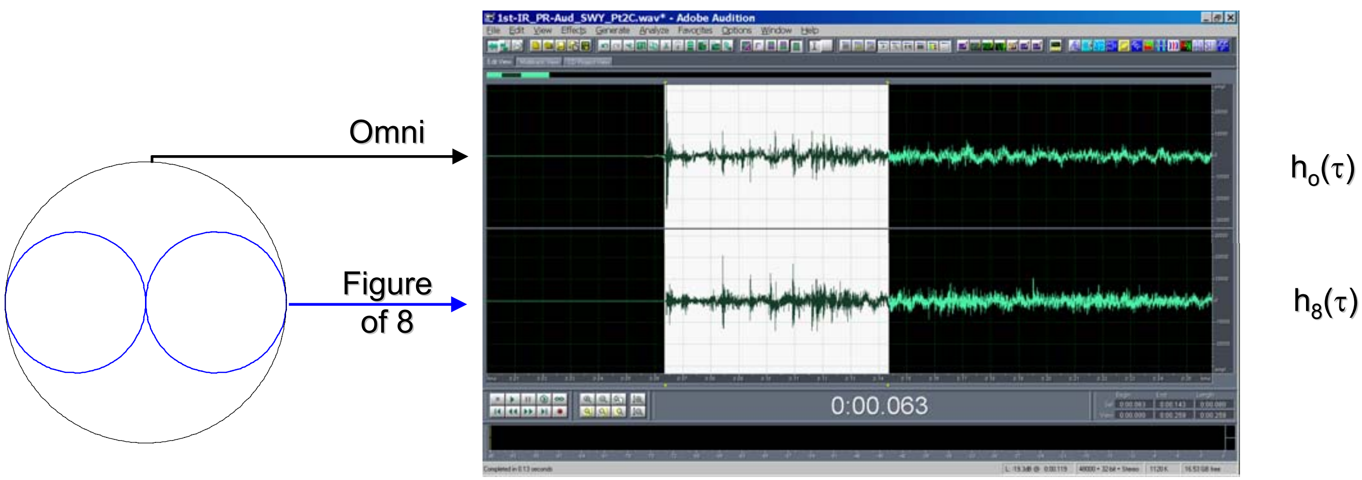


$$\rho(t) = \frac{\int_0^{80ms} p_L(\tau) \cdot p_R(\tau + t) \cdot d\tau}{\sqrt{\int_0^{80ms} p_L^2(\tau) \cdot d\tau \cdot \int_0^{80ms} p_R^2(\tau + t) \cdot d\tau}}$$

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

“objective” spatial parameters

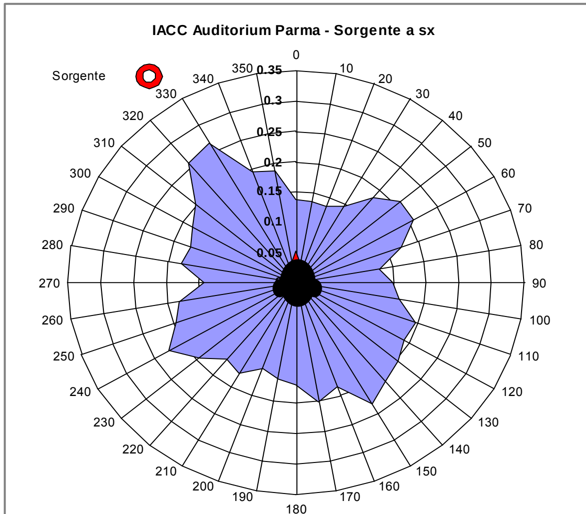
- **Other “spatial” parameters are the Lateral Energy ratios: LE, LF, LFC**
- **These are defined from a 2-channels impulse response, the first channel is a standard omni microphone, the second channel is a “figure-of-eight” microphone:**

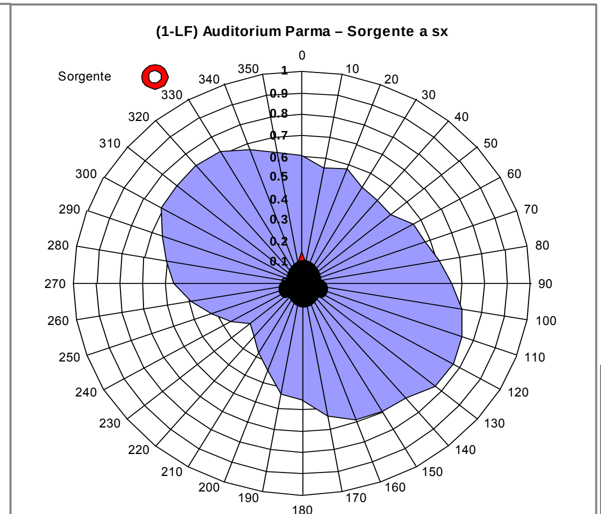


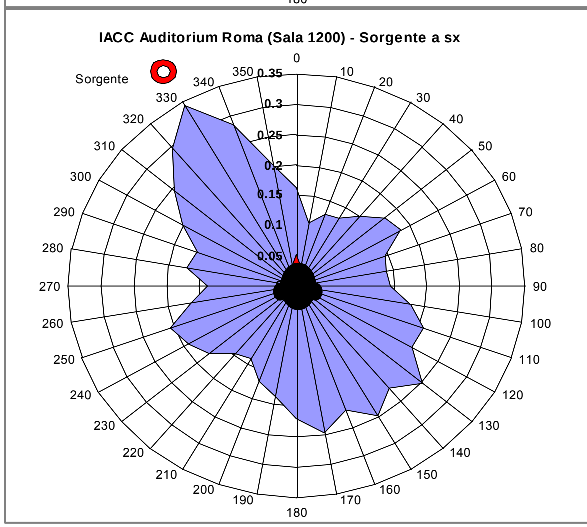
$$LE = \frac{\int_{25ms}^{80ms} h_8^{2}(\tau)\cdot d\tau}{\int_{0ms}^{80ms} h_o^{2}(\tau)\cdot d\tau} \qquad LF = \frac{\int_{5ms}^{80ms} h_8^{2}(\tau)\cdot d\tau}{\int_{0ms}^{80ms} h_o^{2}(\tau)\cdot d\tau} \qquad LFC = \frac{\int_{5ms}^{80ms} h_8(\tau)\cdot h_o(\tau)\cdot d\tau}{\int_{0ms}^{80ms} h_o^{2}(\tau)\cdot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

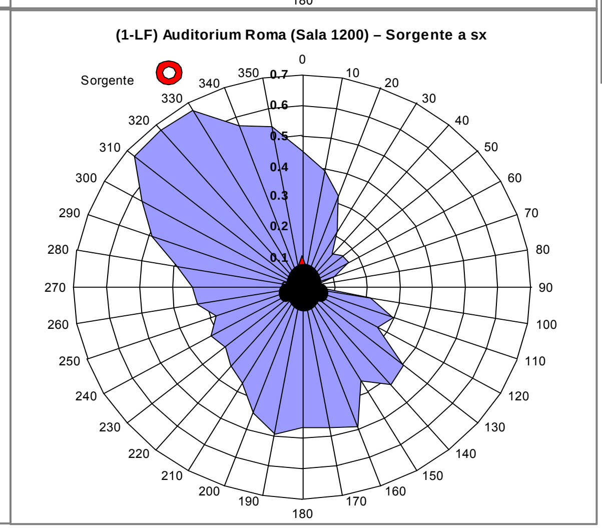
Robustness of spatial parameters

- **Both IACC and LF depend strongly on the orientation of the microphones**
- **Binaural and pressure-velocity measurements were performed in 2 theatres employing a rotating table for turning the microphones**





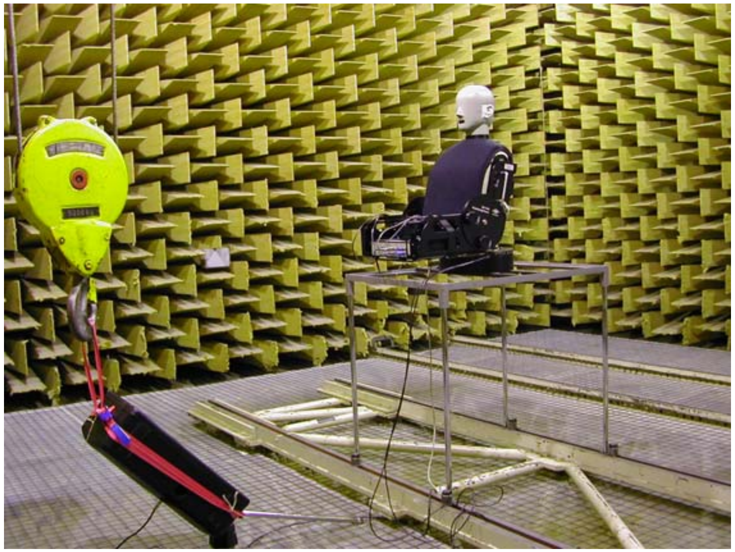


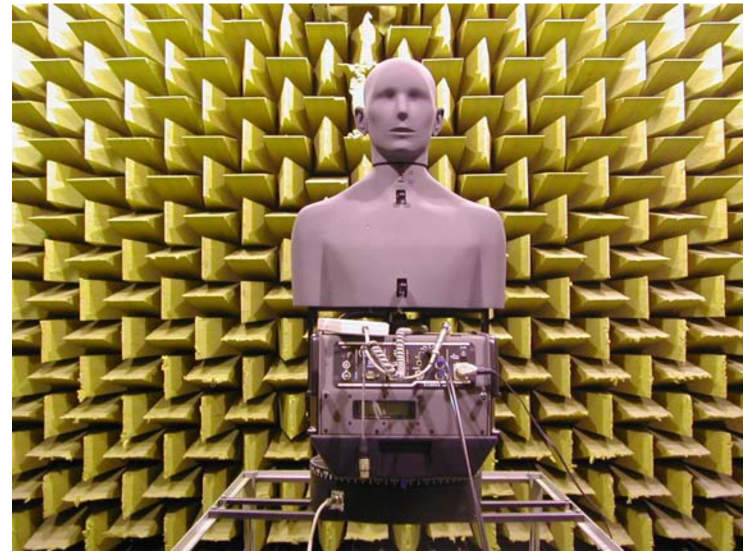


Theatre	1-LF	IACC
Parma	0.725	0.266
Roma	0.676	0.344

Are binaural measurents reproducible?

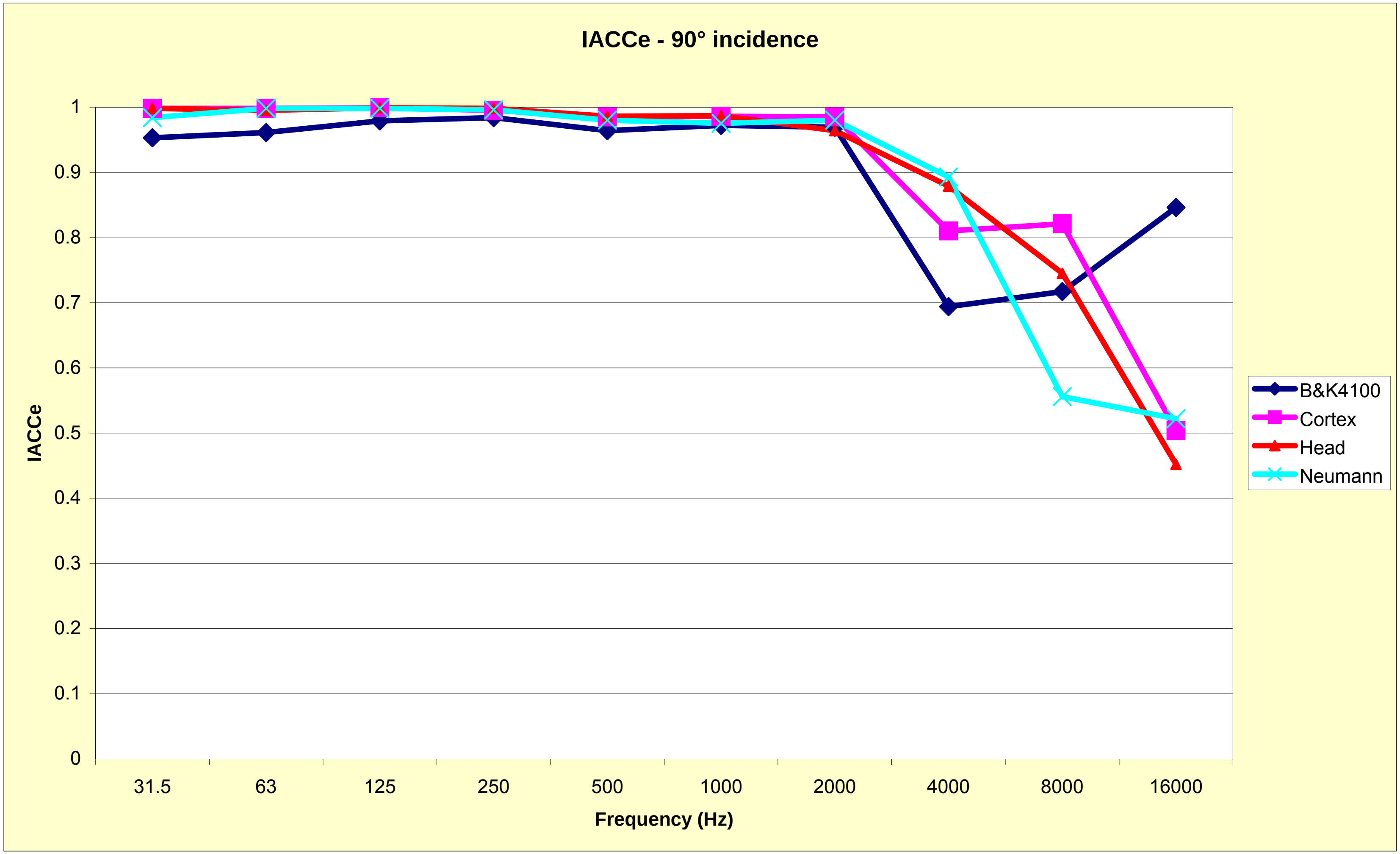
- **Experiment performed in anechoic room - same loudspeaker, same source and receiver positions, 5 binaural dummy heads**





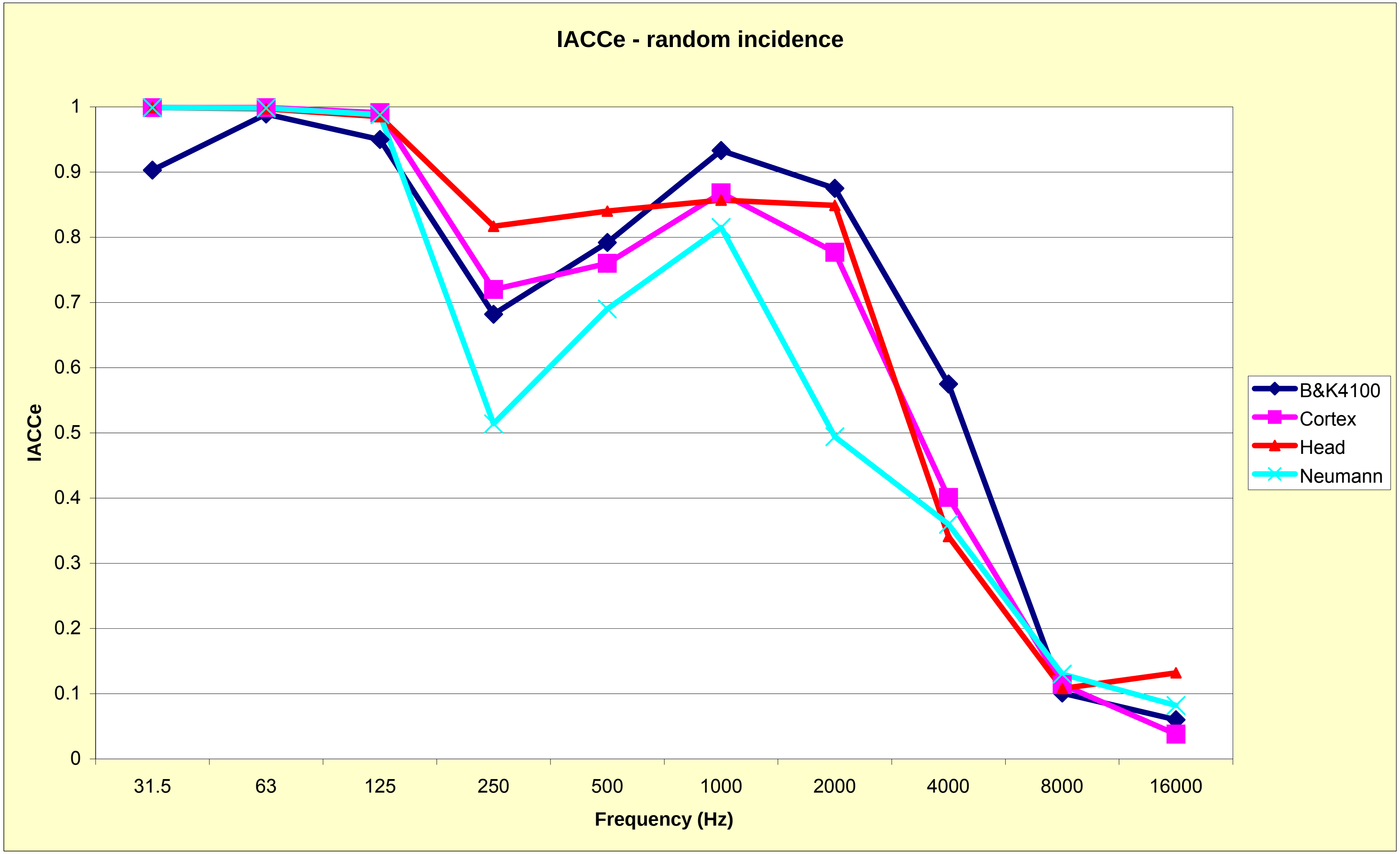
Are binaural measurents reproducible?

- **90° incidence - at low frequency IACC is almost 1, at high frequency the difference between the heads becomes evident**



Are binaural measurents reproducible?

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

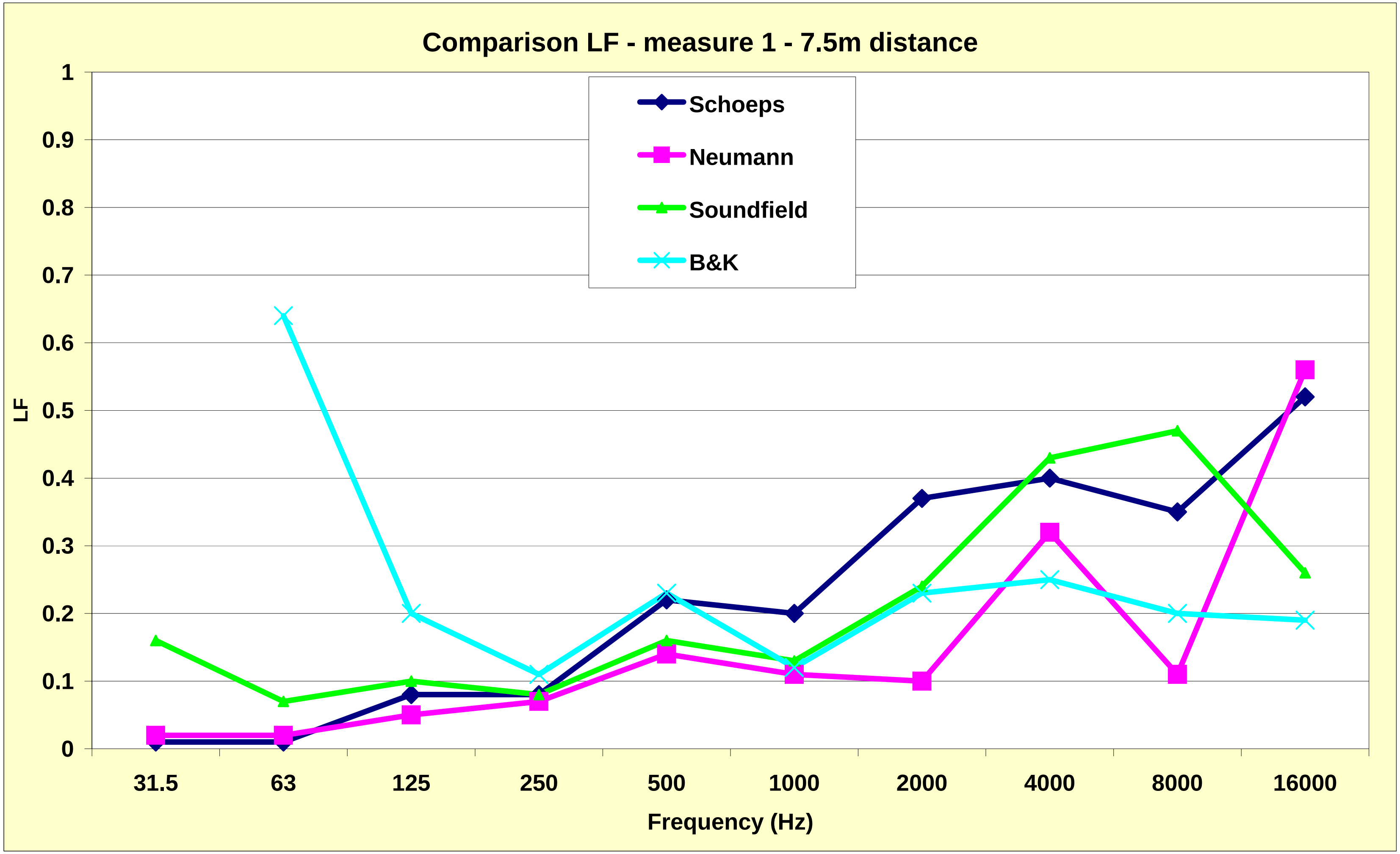


Are LF measurents reproducible?

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

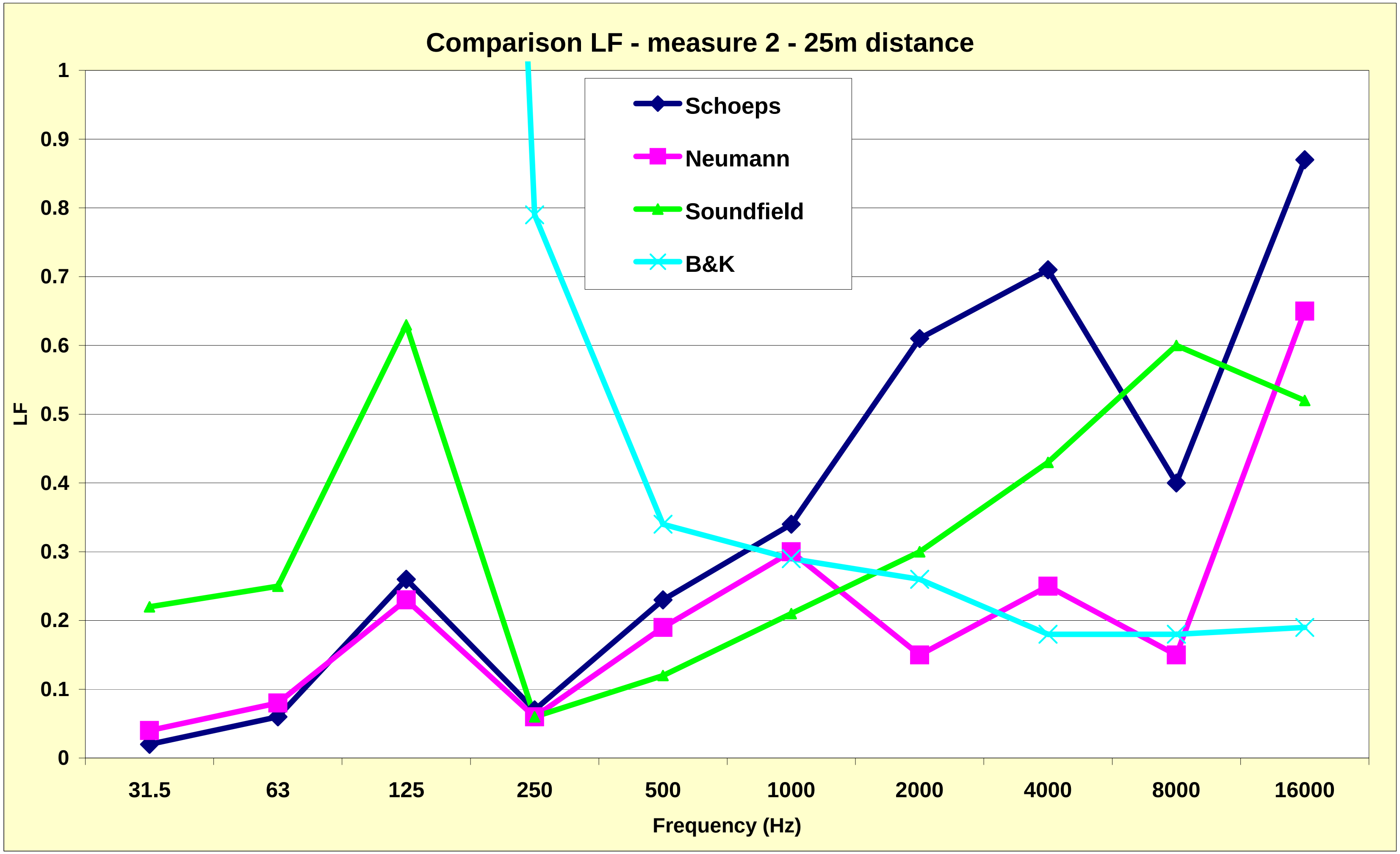
Are LF measurents reproducible?

- **At 7.5 m distance, the results already exhibit significant scatter**

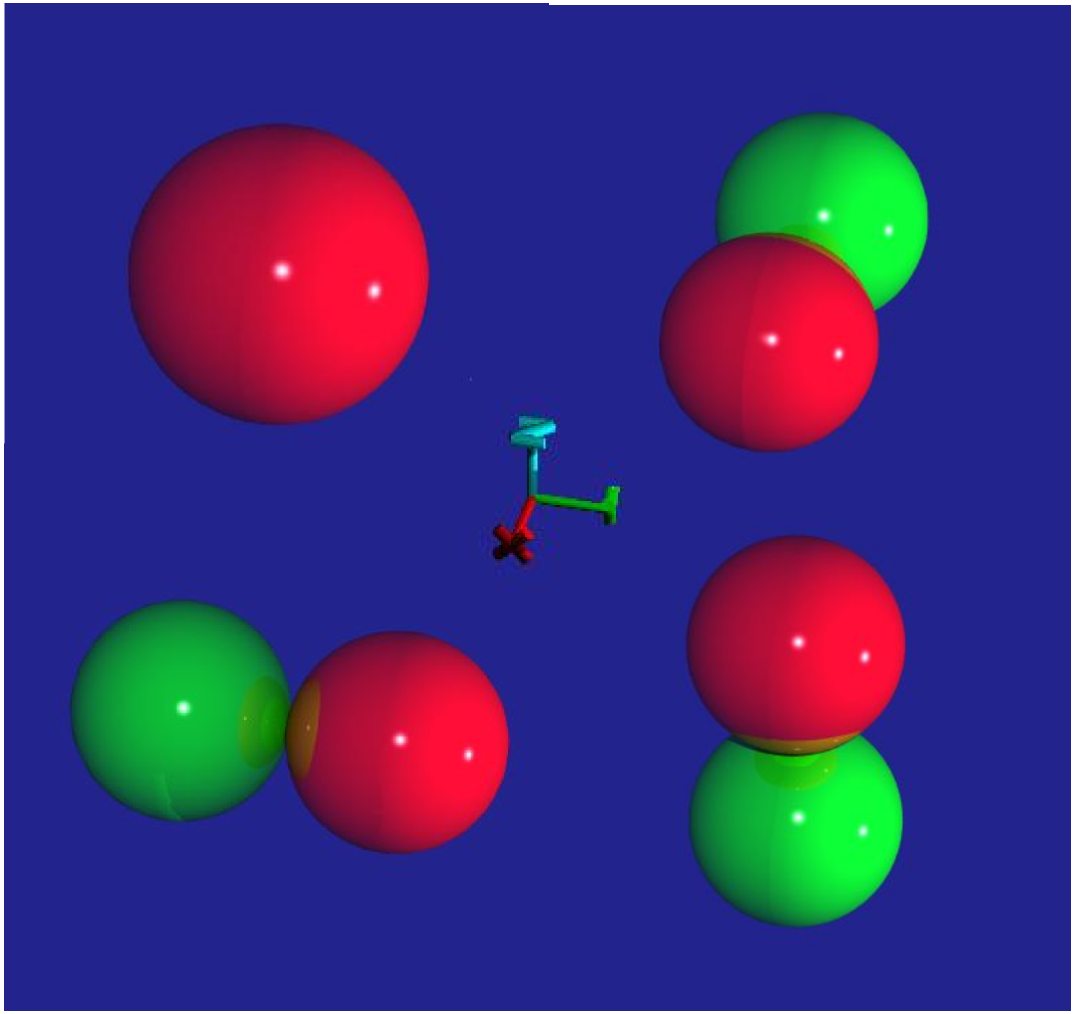


Are LF measurents reproducible?

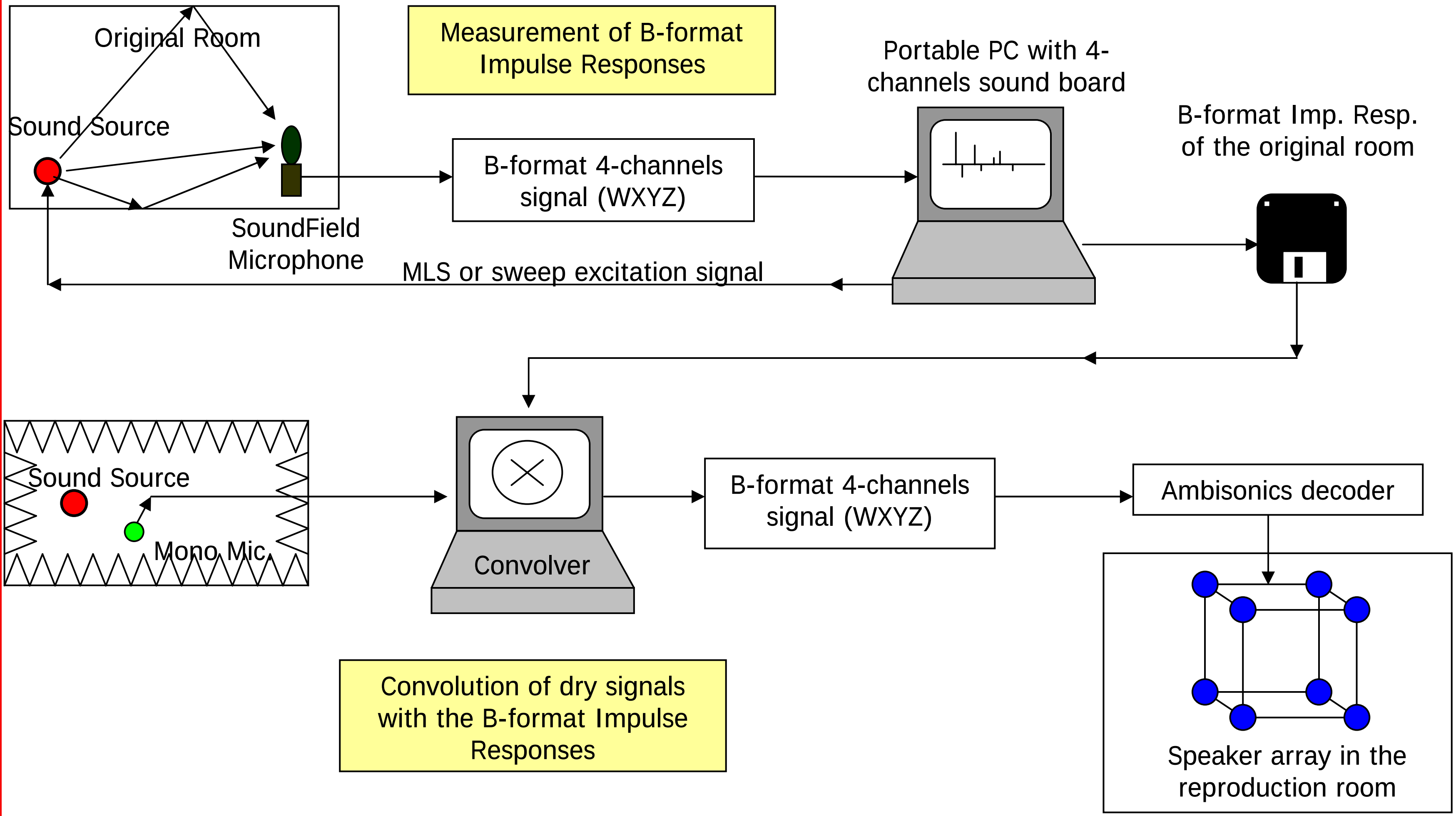
- At 25 m distance, the scatter is even larger....



- **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)

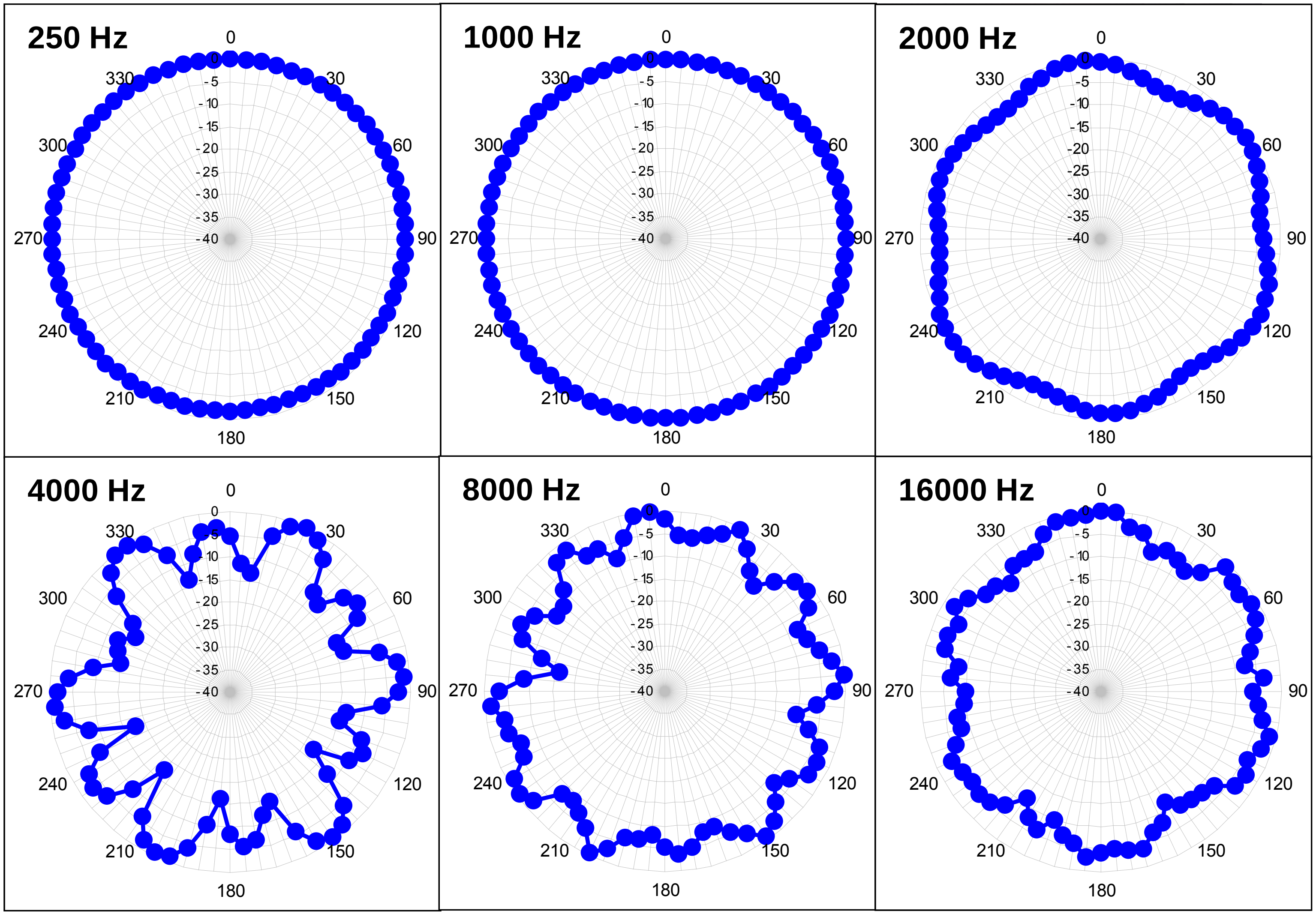


The Waves project (2003)

- The original idea of Michael Gerzon was finally put in practice in 2003, thanks to the Israeli-based company WAVES
- More than 50 theatres all around the world were measured, capturing 3D IRs (4-channels B-format with a Soundfield microphone)
- The measurments did also include binaural impulse responses, and a circular-array of microphone positions
- More details on WWW.ACOUSTICS.NET

Directivity of transducers

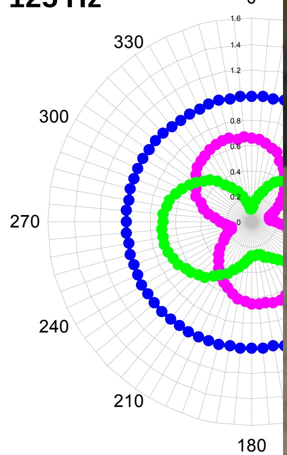
LookLine D200 dodechaedron



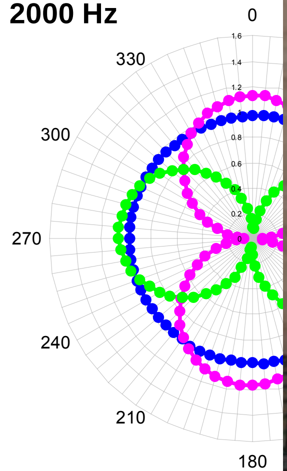
Directivity of transducers

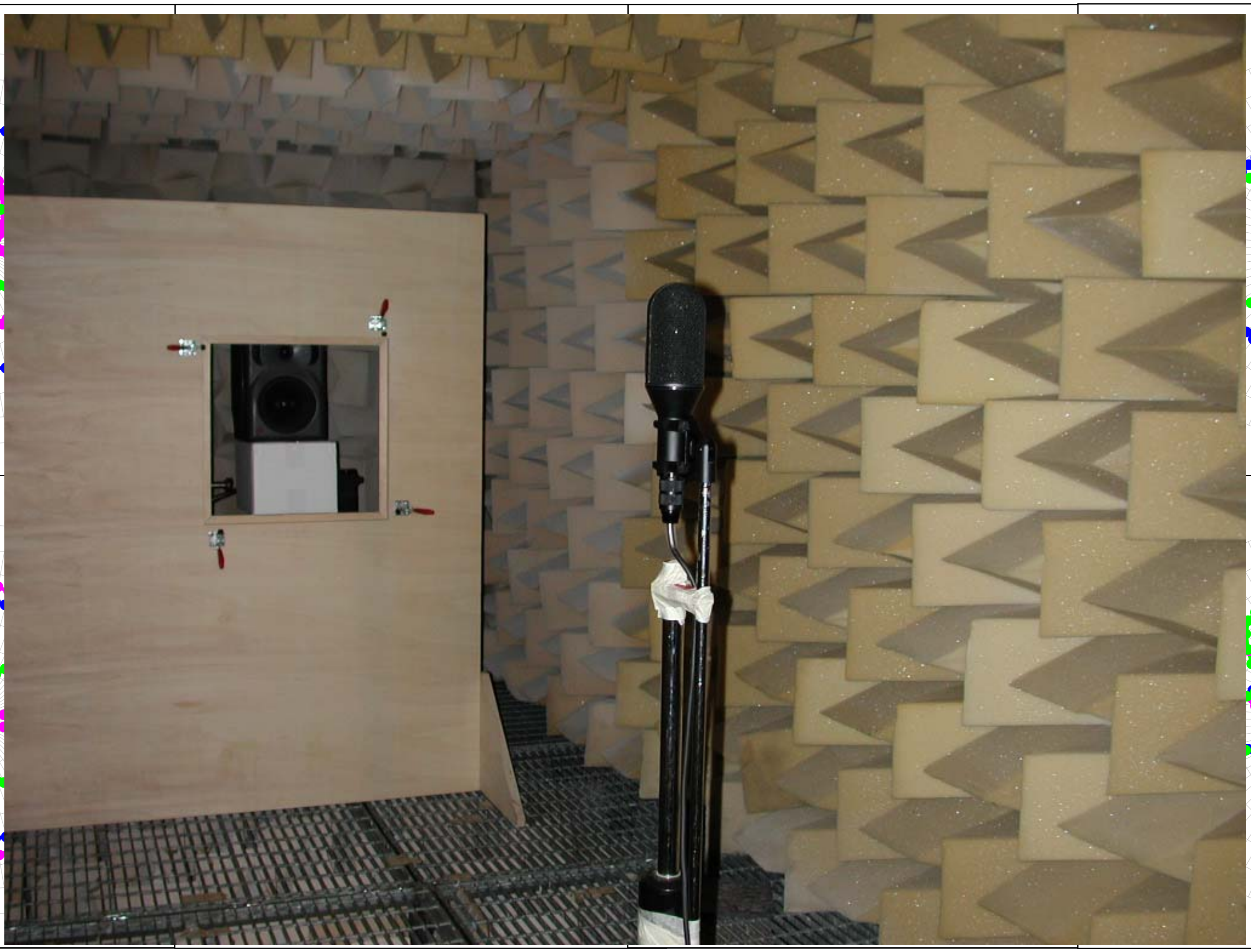
Soundfield ST-250 microphone

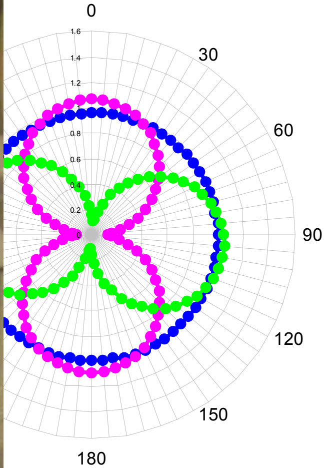
125 Hz

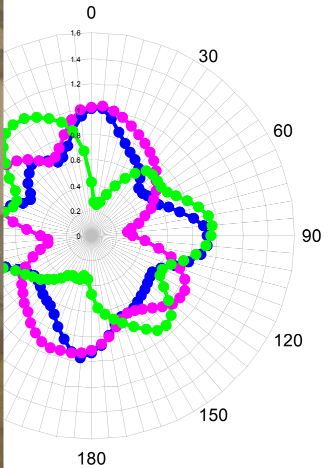


2000 Hz



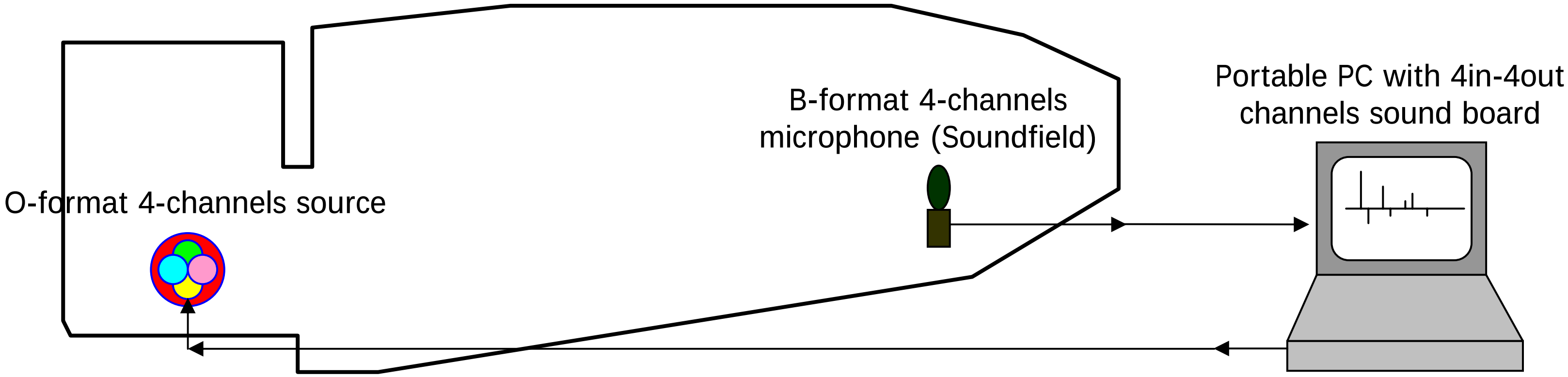






What about source directivity ?

- 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 $\{r_w, r_x, r_y, r_z\}$:

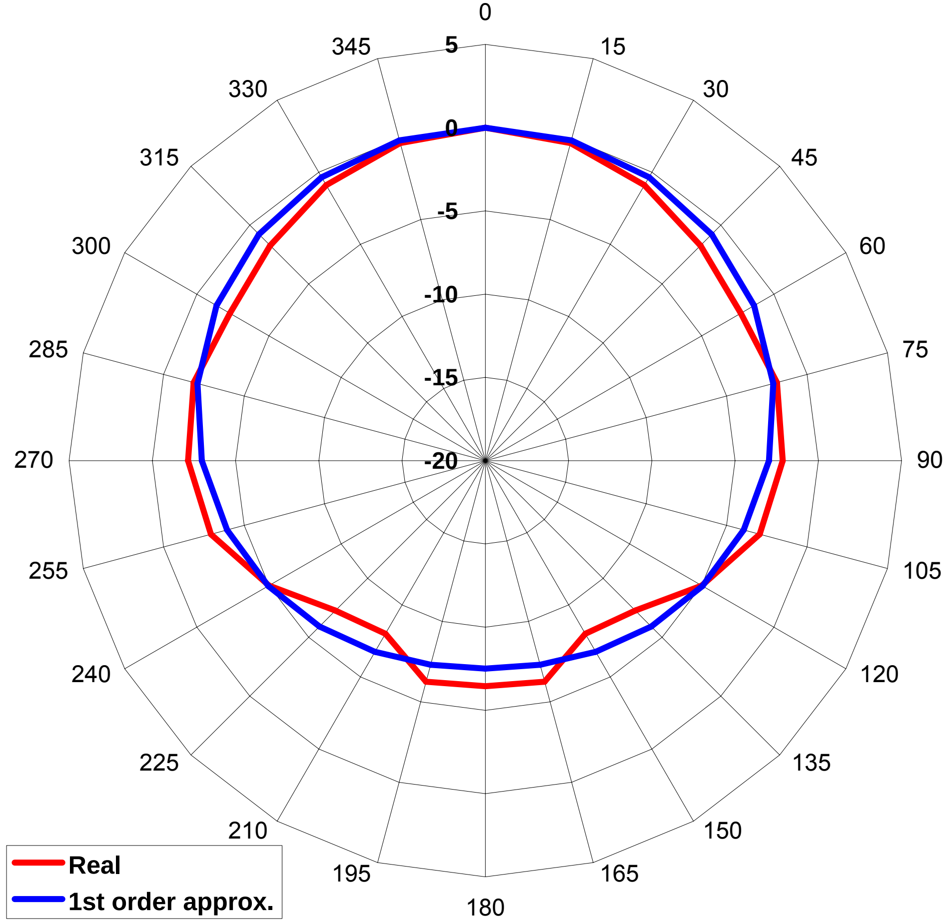
$$t_{sr} = \{r\} \otimes \{y\} = \{r\} \otimes [h] \otimes \{s\}$$

- 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:**

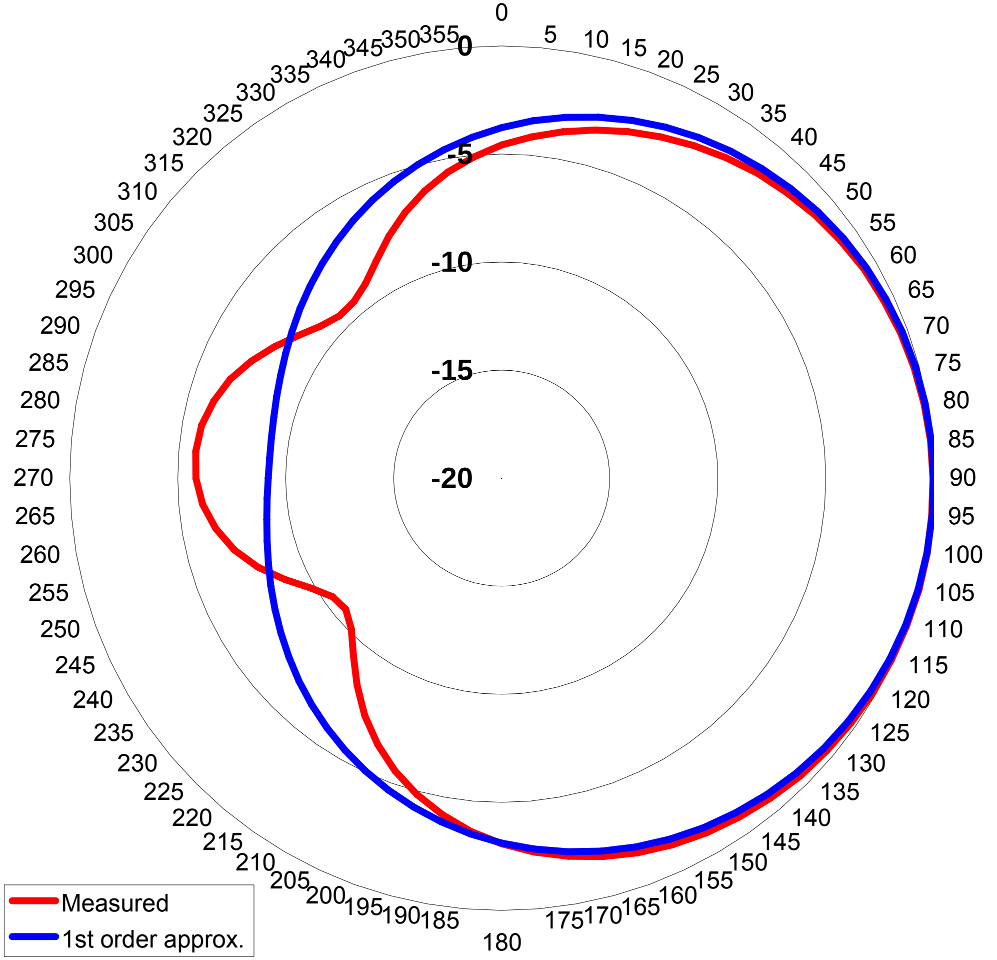
W	X	Y
1.421	0.289	0.000



Limits of the 1st-order method

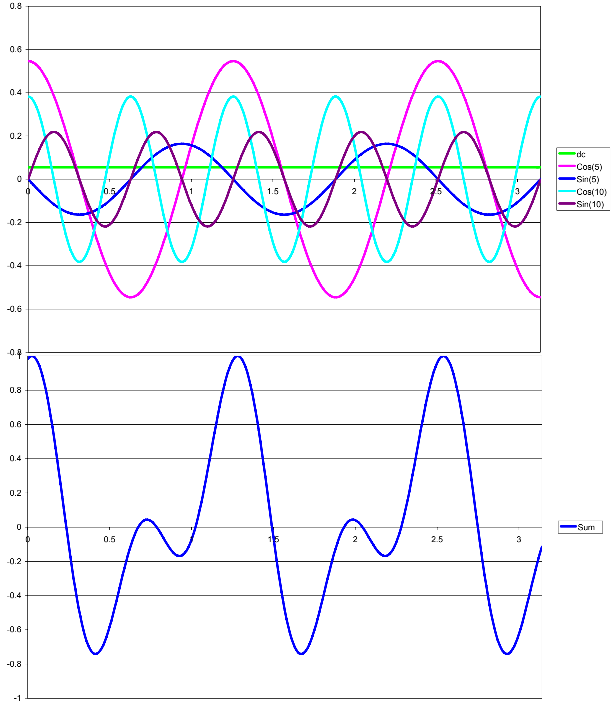
- **The polar pattern of a binaural dummy head is even more complex, as shown here (1 kHz, right ear):**

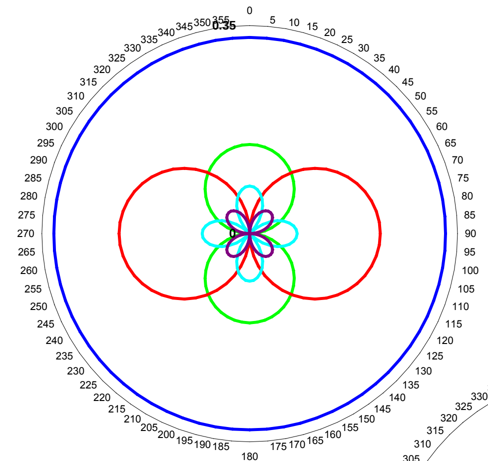
W	X	Y
0.395	-0.015	0.191

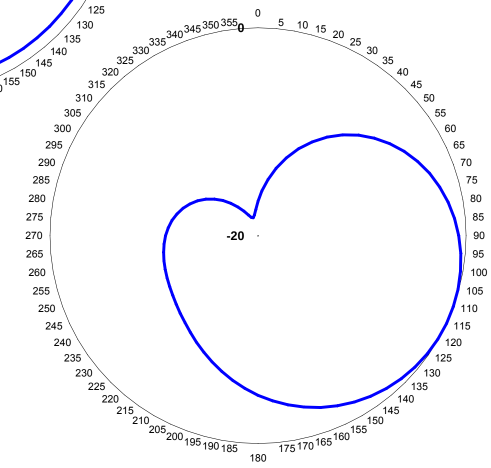


How to get better spatial resolution?

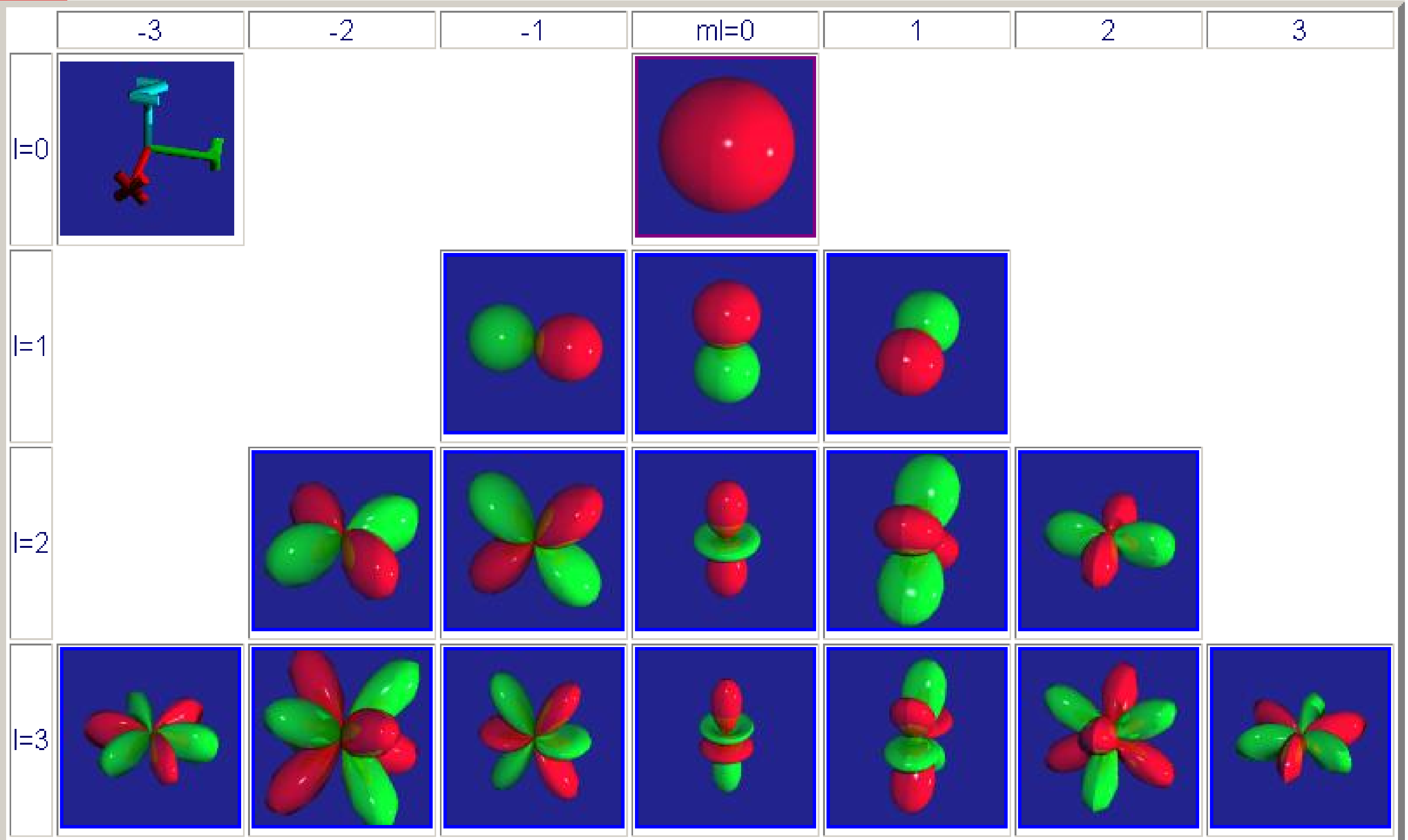
- The answer is simple: analyze the spatial distribution of both source and receiver by means of higher-order spherical harmonics expansion
- Spherical harmonics analysis is the equivalent, in space domain, of the Fourier analysis in time domain
- As a complex time-domain waveform can be though as the 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





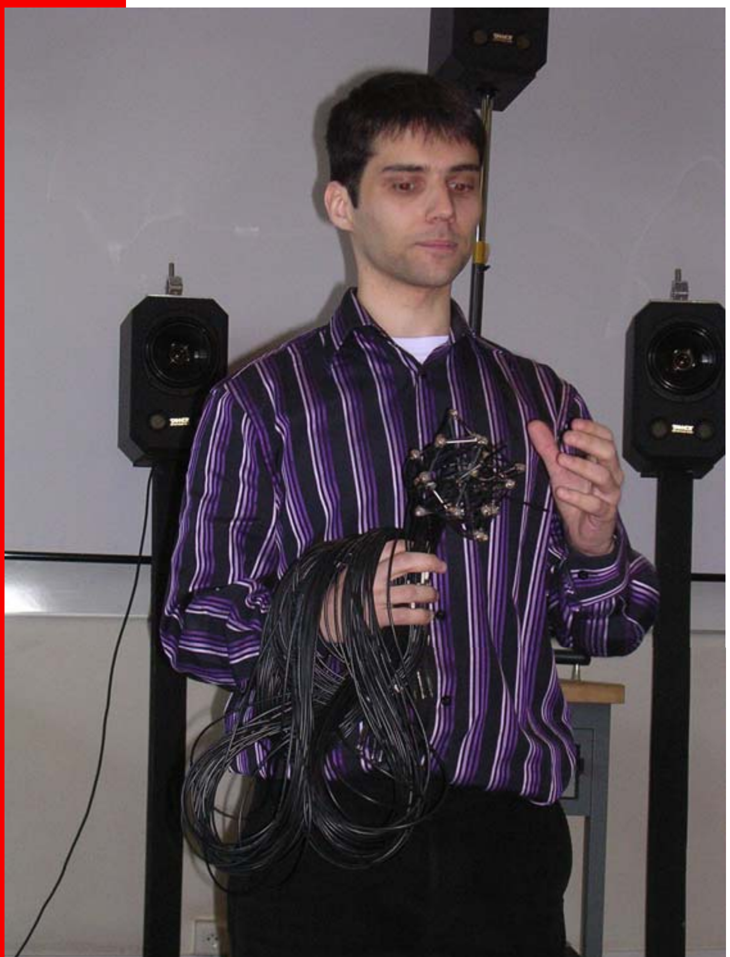


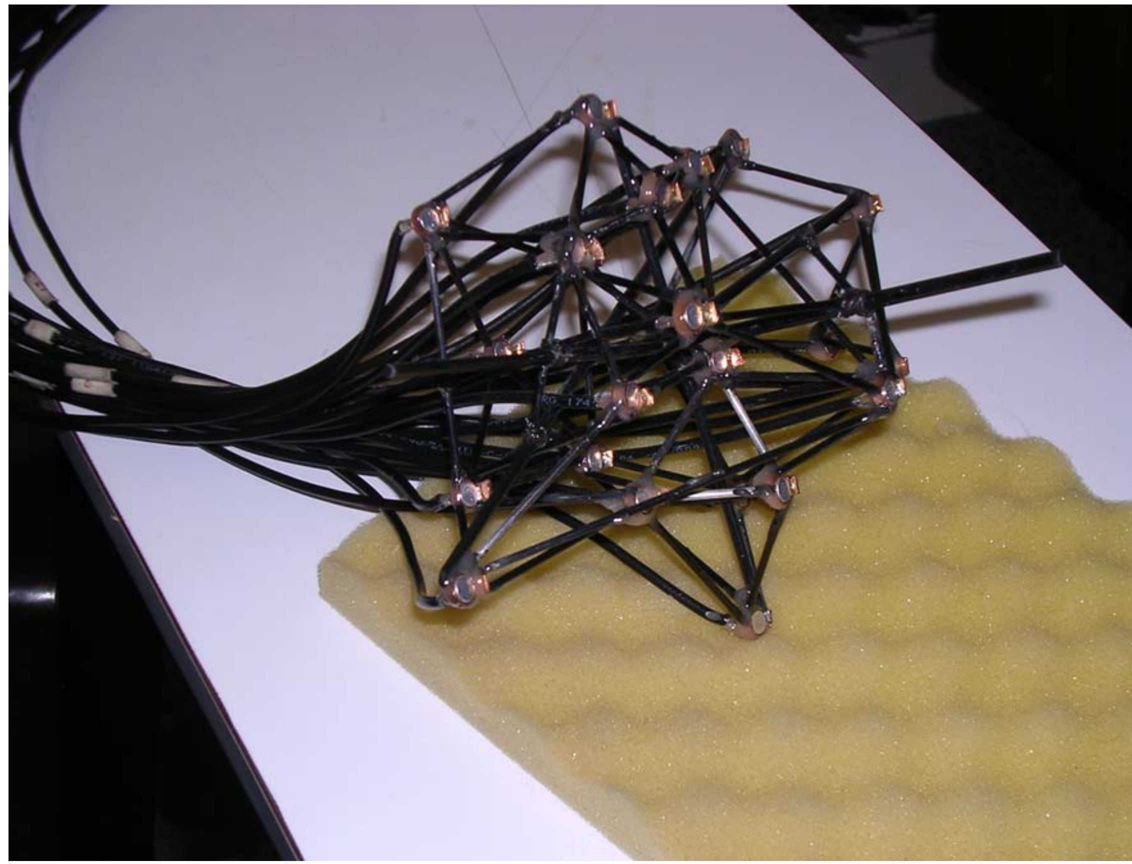
Higher-order spherical harmonics expansion



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)**



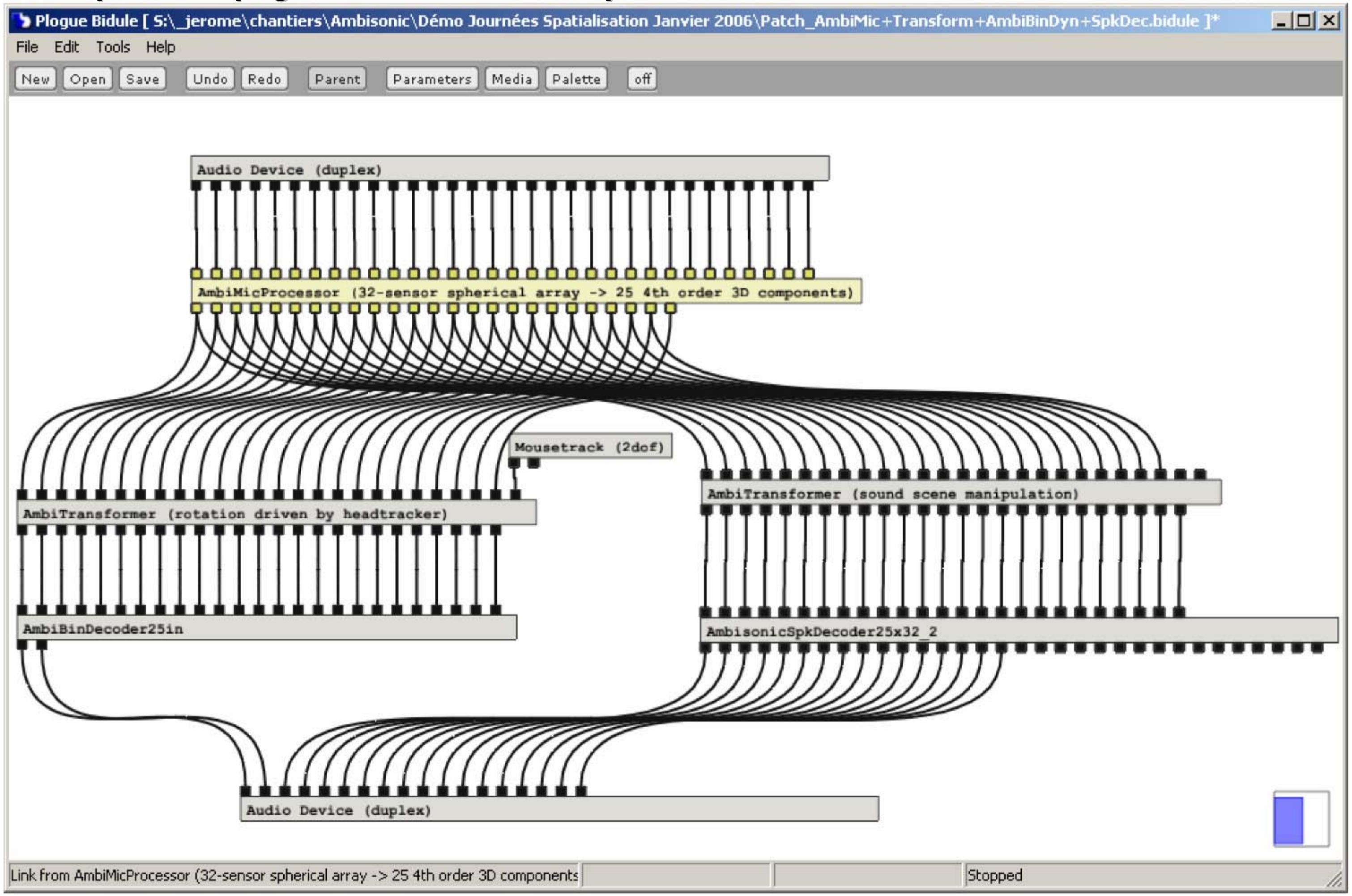


4°-order microphone (France Telecom)

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

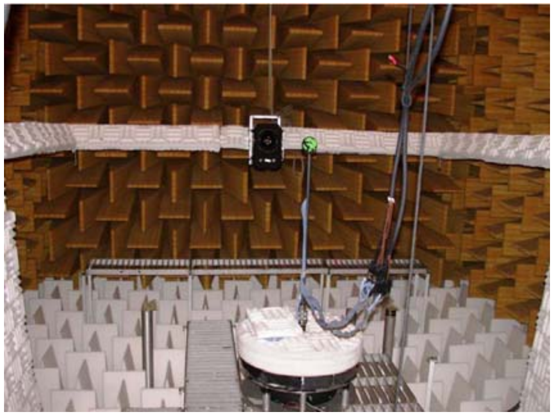
Multichannel software for high-order

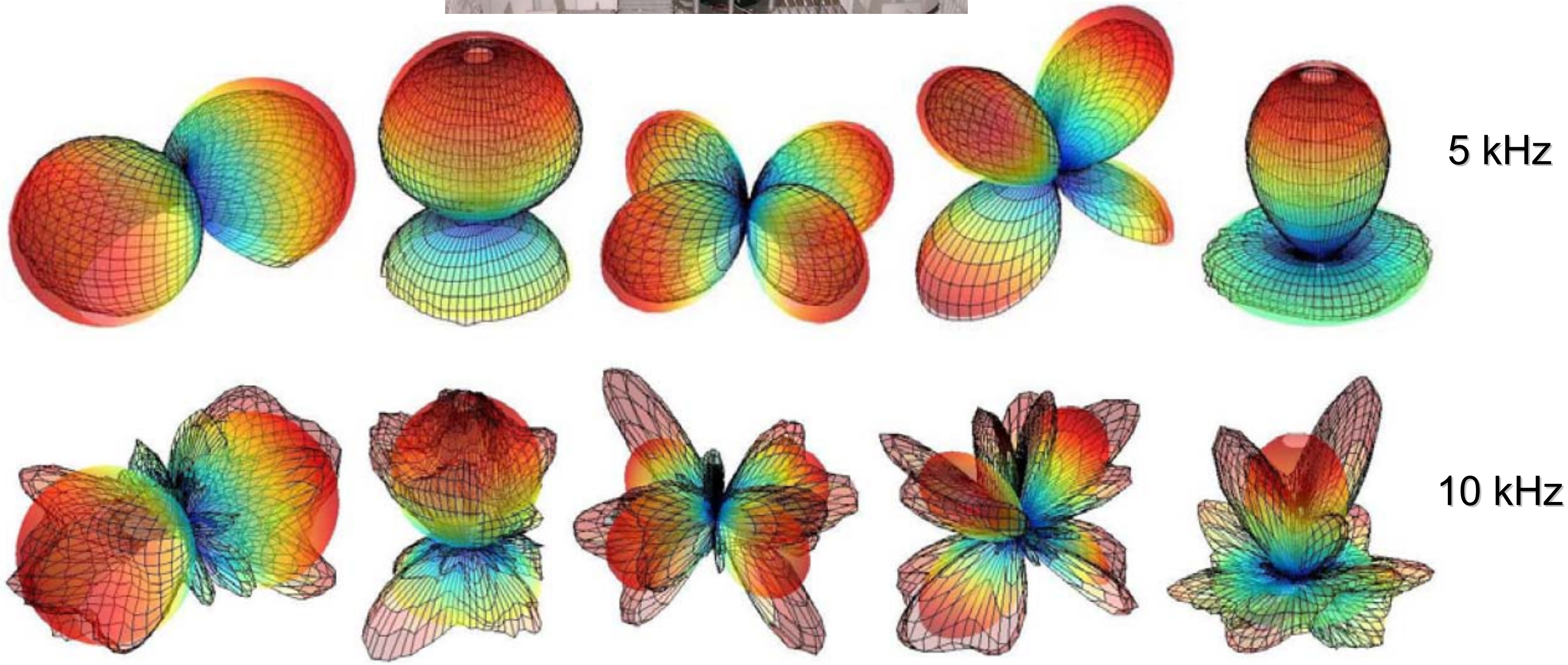
- **Plogue Bidule can be used as multichannel host software, running a number of VST plugins developed by France Telecom - these include spherical harmonics extraction from the spherical microphone arrays, rotation and manipulation of the multichannel B-format signals, and final rendering either on head-.tracked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



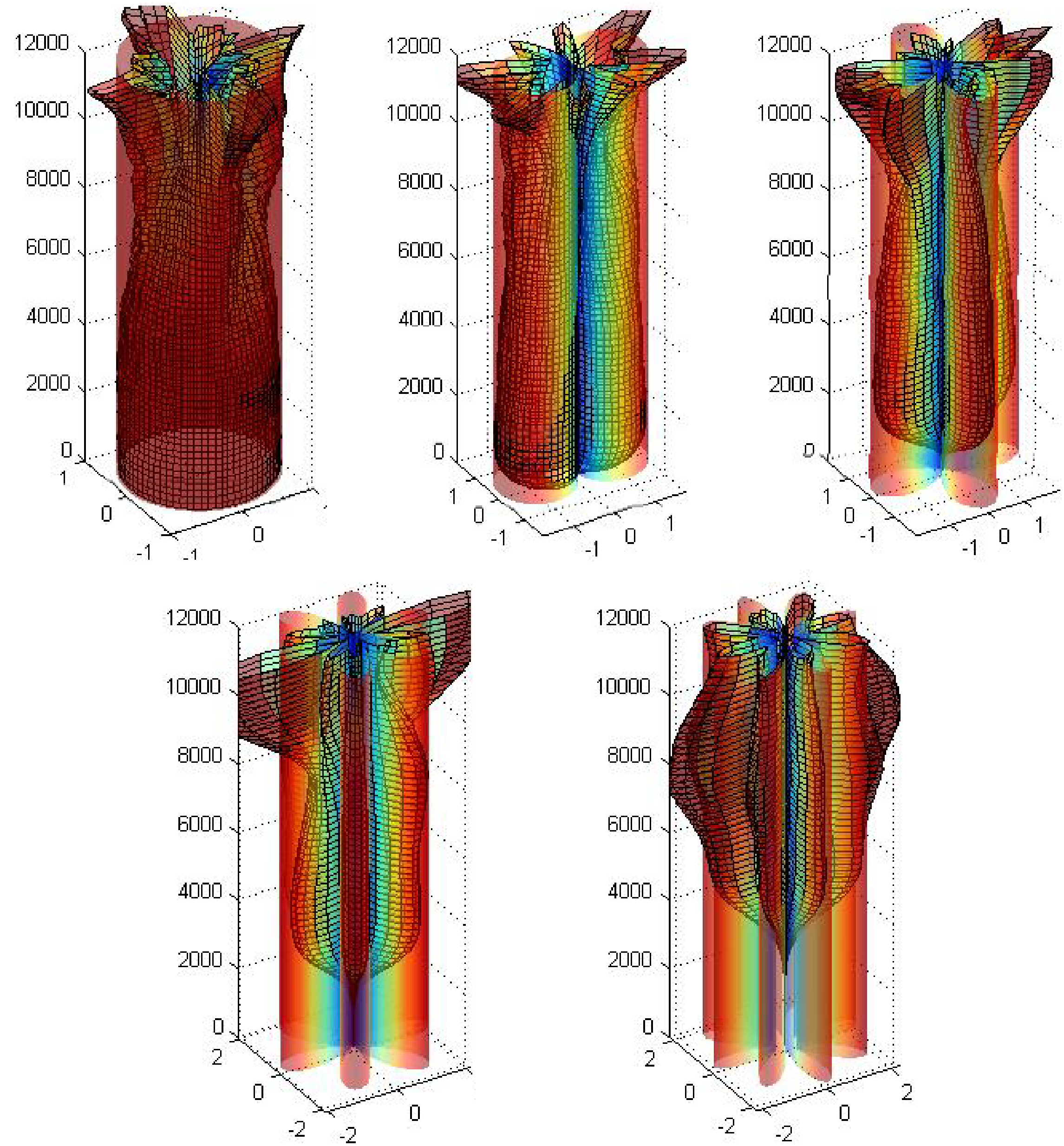
Verification of high-order patterns

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





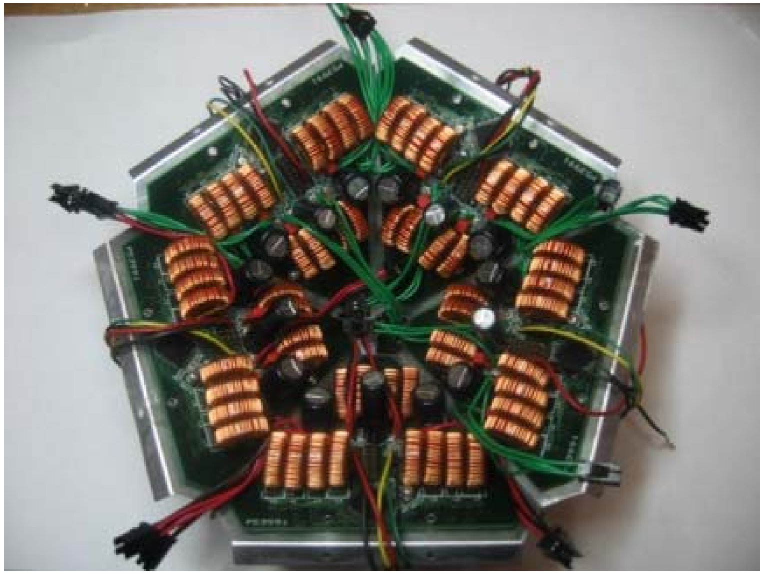
Frequency extension of the patterns



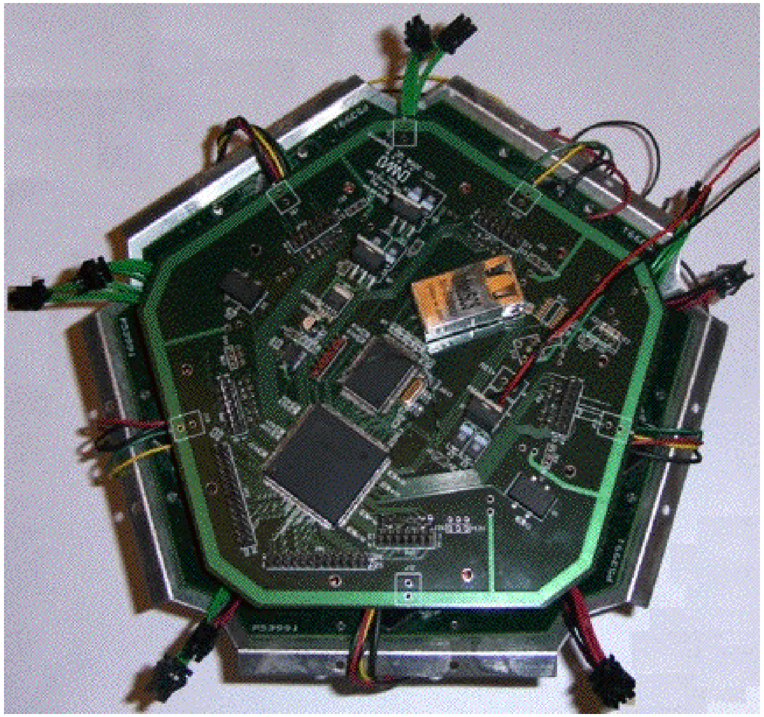
High-order sound source

- **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.**

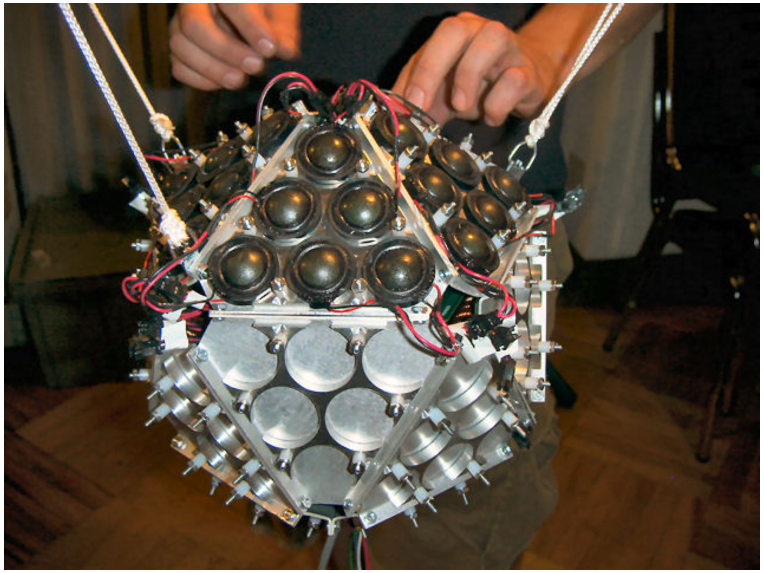
Technical details of high-order source



- Class-D embedded amplifiers



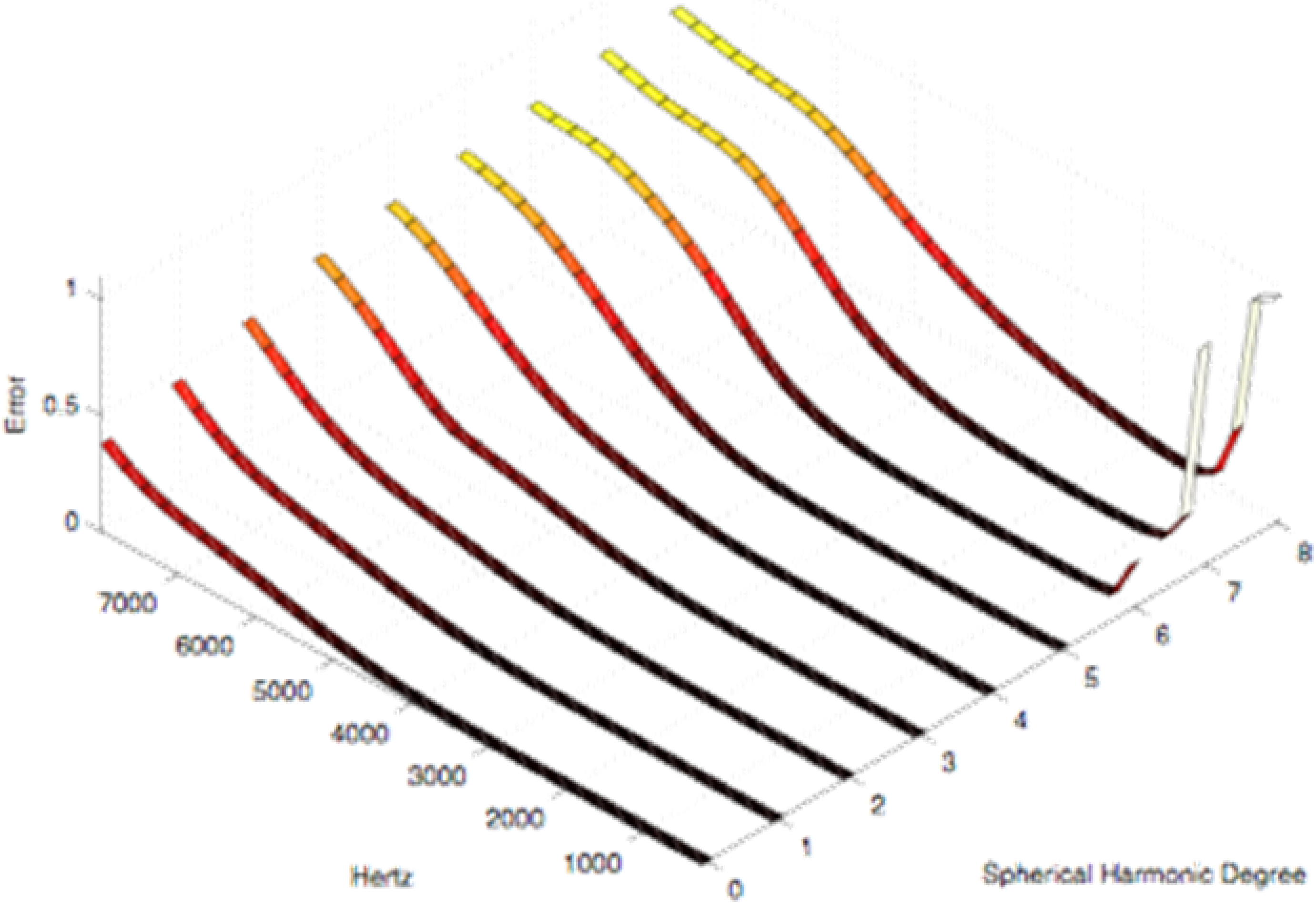
- Embedded ethernet interface and DSP processing



- Long-excursion special Meyer Sound drivers

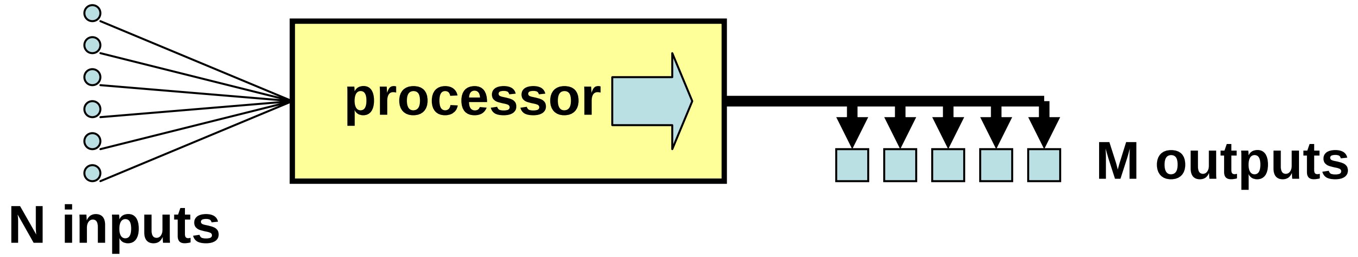
Accuracy of spatial synthesis

- **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**

Advanced digital filtering techniques

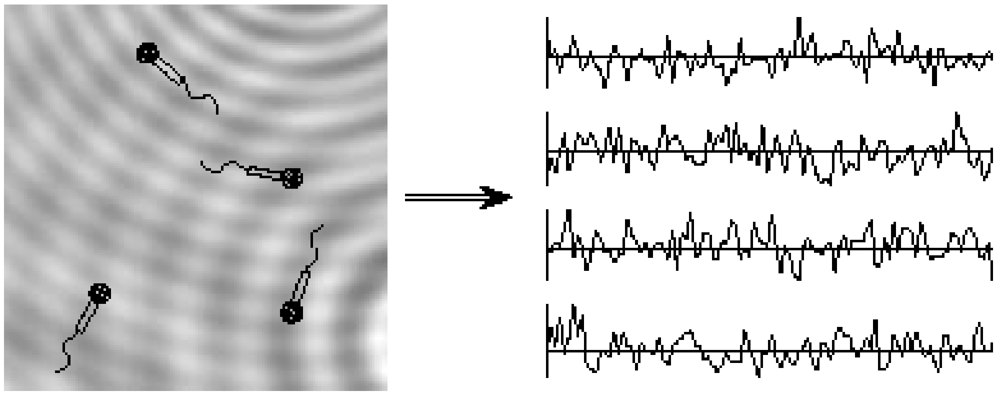


- A set of digital filters can be employed for sinthesizing the required spatial pattern (spherical harmonis), either when dealing with a microphone array or when dealing with a loudspeaker array
- Whatever theory or method is chosen, we always start with N input signals x_i, and we derive from them 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**



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

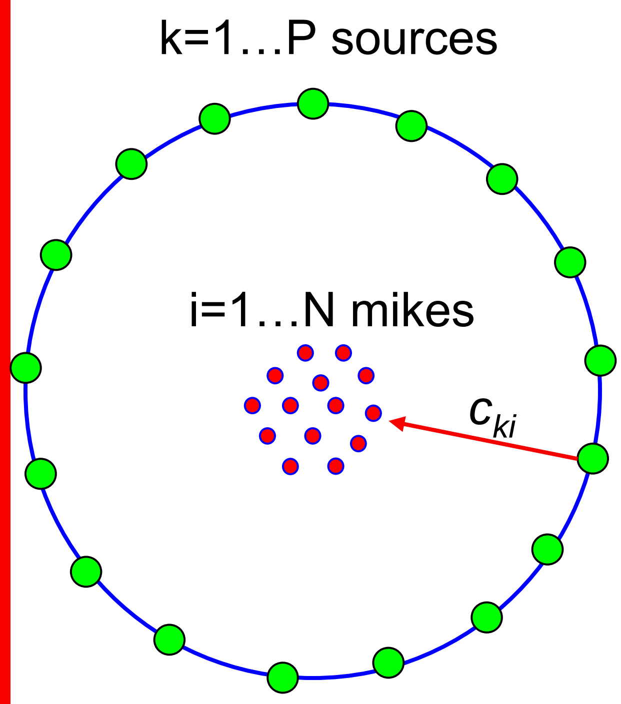
- The processing filters h_{ij} are usually computed following one of several, complex mathematical theories, based on the solution of the wave equation (often under certaing simplifications), and assuming that the microphones are ideal and identical
- In some implementations, the signal of each microphone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

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

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:

$$H=A/D$$

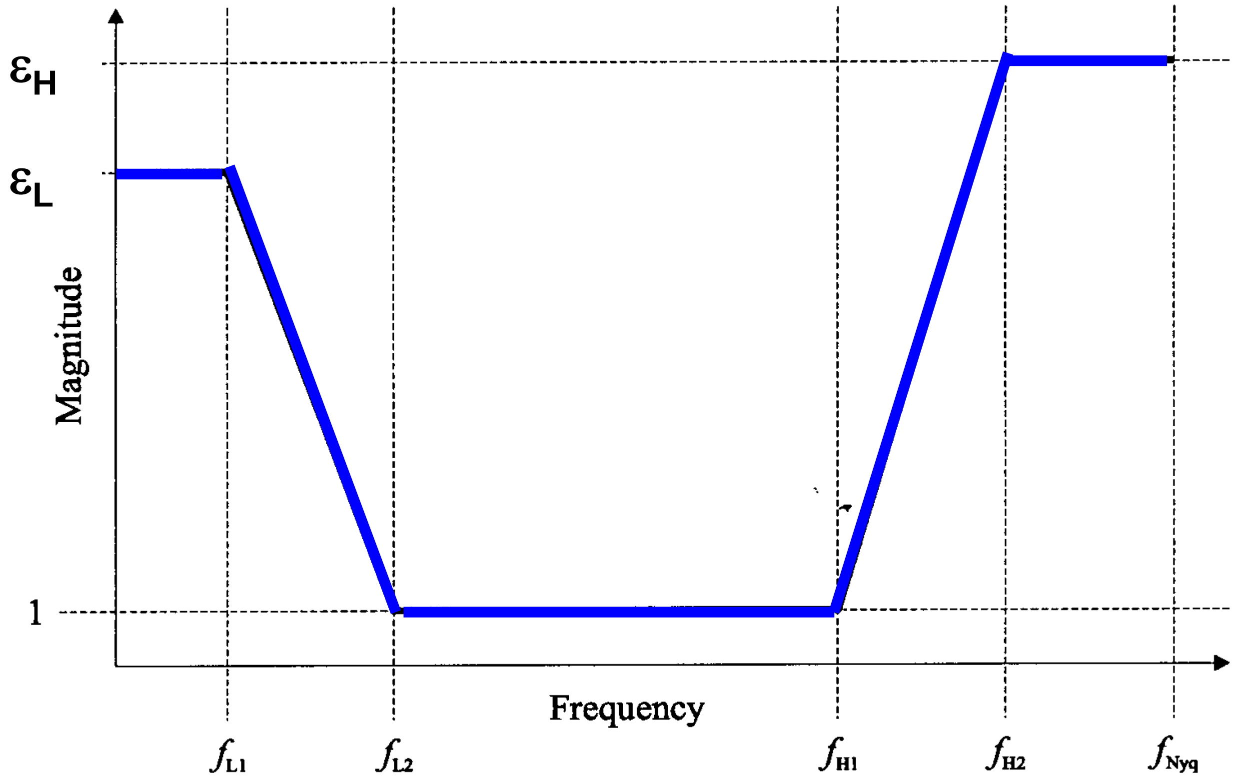
- Computing the reciprocal of the denominator D is generally not trivial, as the inverse of a complex, mixed-phase signal is generally unstable.
- The Nelson/Kirkeby regularization method is usually employed for this task:

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

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

Spectral shape of the regularization parameter ε(ω)

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



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

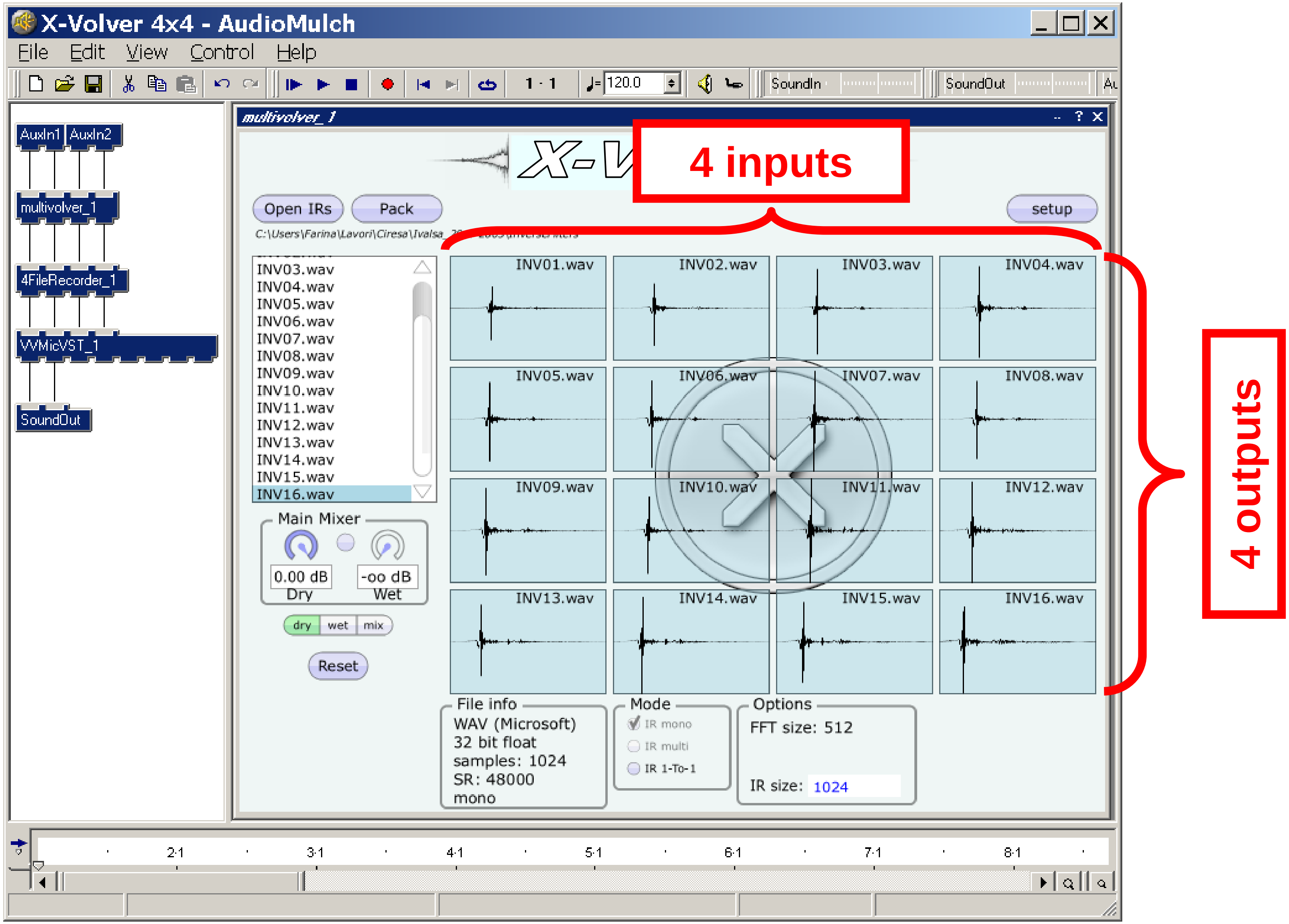
84 IRs were measured, uniformly scattered around a sphere

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:**

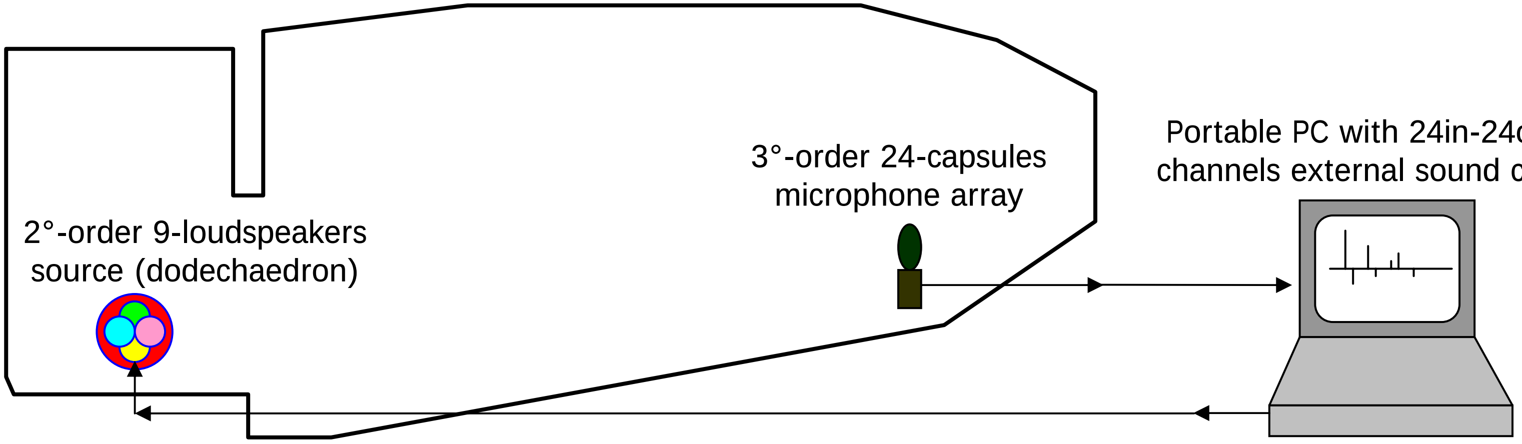
$$\left.\begin{array}{l} c_1 \otimes h_{1,W} + c_2 \otimes h_{2,W} + c_3 \otimes h_{3,W} + c_4 \otimes h_{4,W} = out_{k,W} \\ c_1 \otimes h_{1,X} + c_2 \otimes h_{2,X} + c_3 \otimes h_{3,X} + c_4 \otimes h_{4,X} = out_{k,X} \\ c_1 \otimes h_{1,Y} + c_2 \otimes h_{2,Y} + c_3 \otimes h_{3,Y} + c_4 \otimes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

Real-time implementation



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 only 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.



Conclusions

- The sine sweep method revealed to be systematically superior to the MLS method for measuring electroacoustical impulse responses
- In fact, it is now employed in top-grade measurement systems, including Audio Precision (TM) or Bruel & Kjaer's DIRAC software
- 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 obtained 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.