Workshop
Casa della Musica
Parma, 18 October 2008

istituzione casadellamusica parma

IMPULSE RESPONSE MEASUREMENTS BY EXPONENTIAL SINE SWEEPS

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

University of Parma

Time Line

The Past

- Traditional time-domain measurements with pulsive sounds and omnidirectional transducers
- MLS and TDS methods for electroacoustical masurements

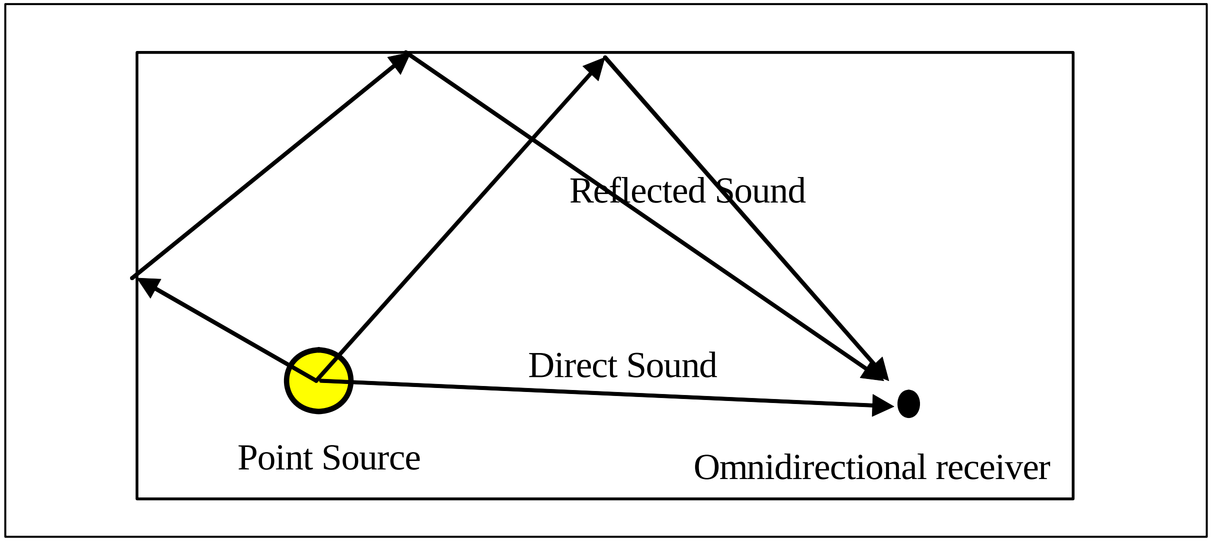
The Present

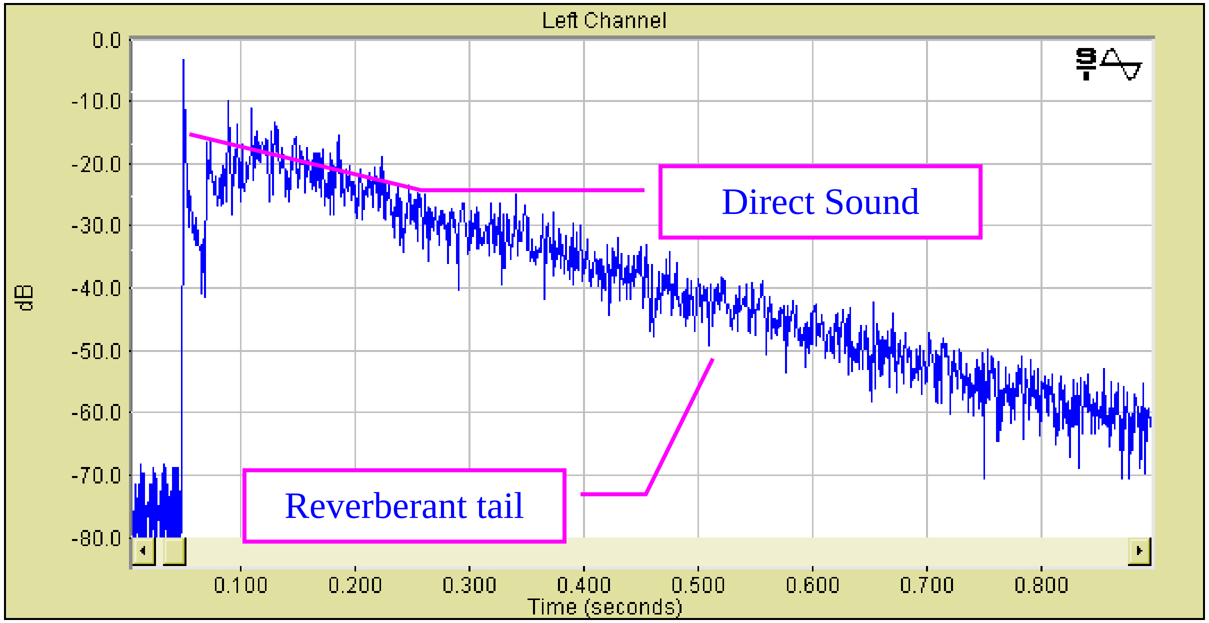
- Electroacoustical measurements employing the Exponential Sine Sweep method (ESS)

The Future

- Capturing the complete spatial information by means of arrays of transducers
- Employment of not-linear impulse responses in the auralization process

Starting point: room impulse response



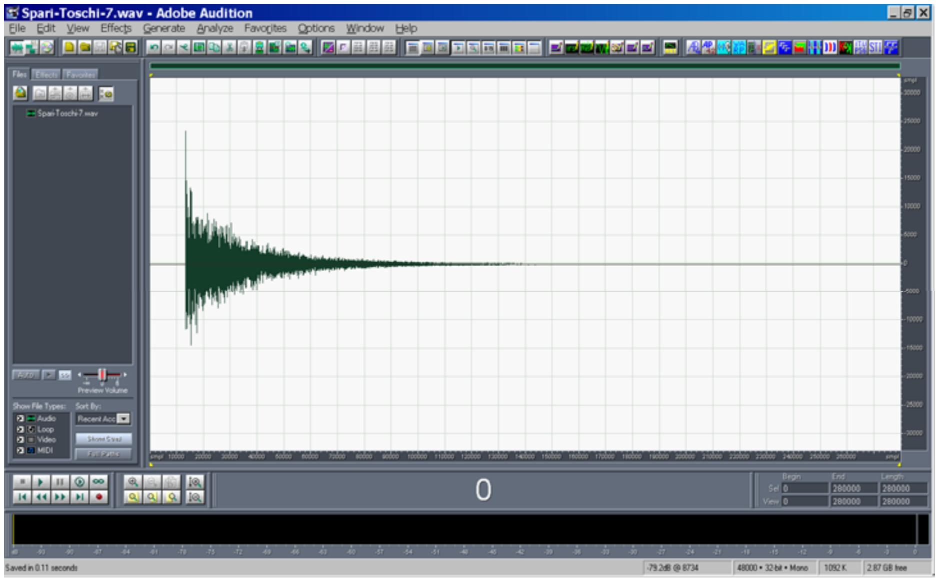


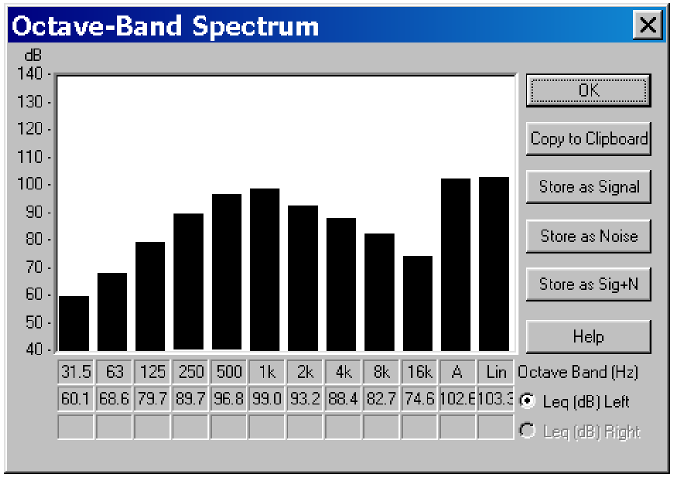
The Past

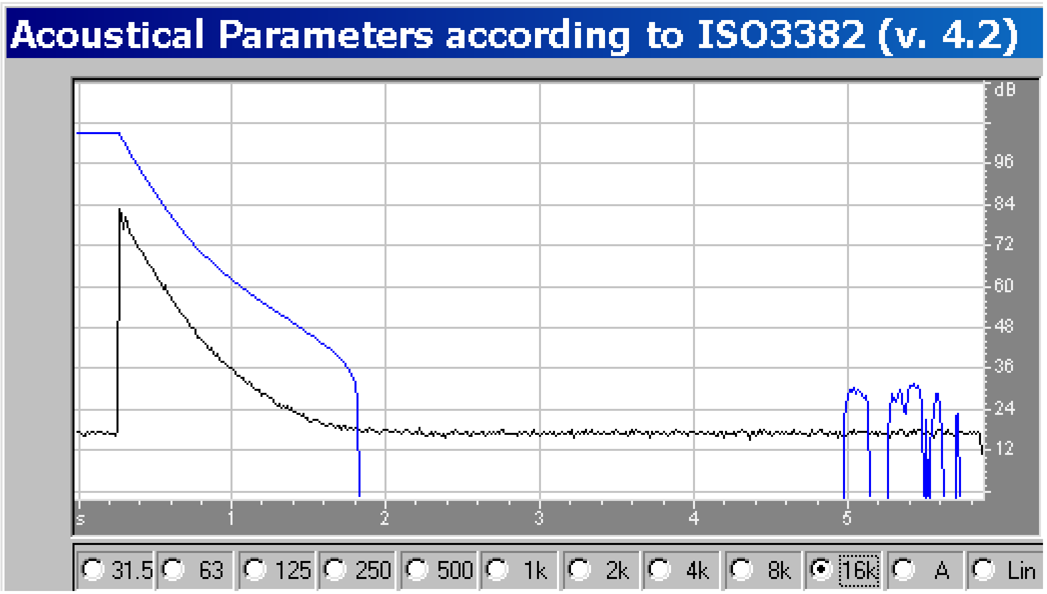
Traditional measurement methods

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

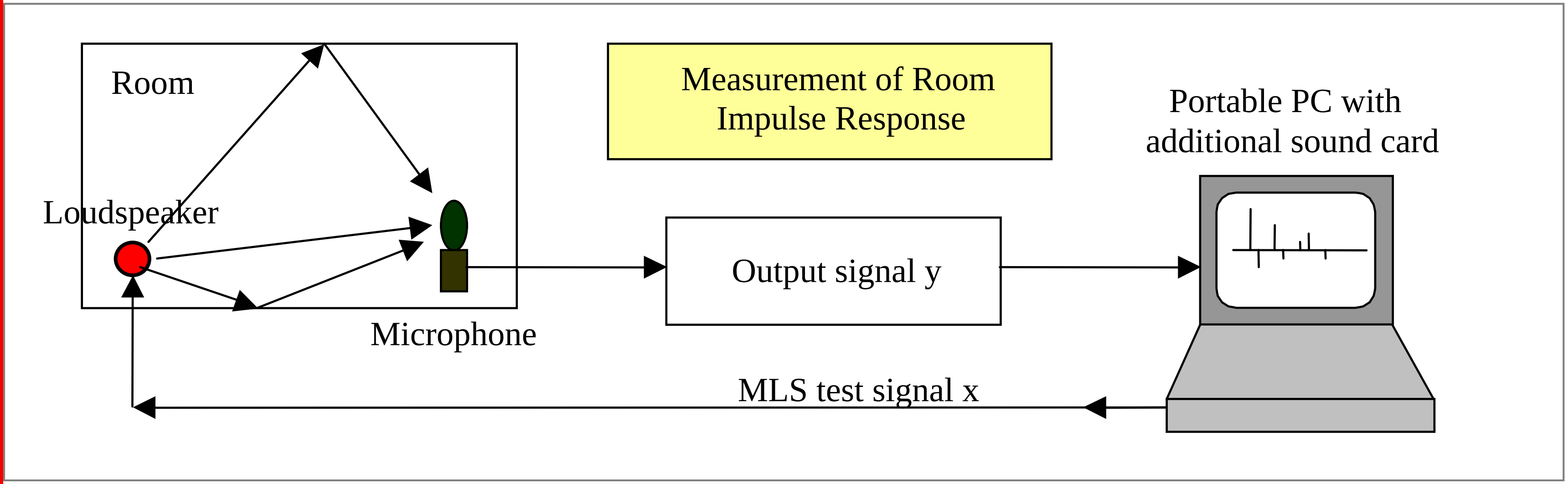
Example of a pulsive impulse response





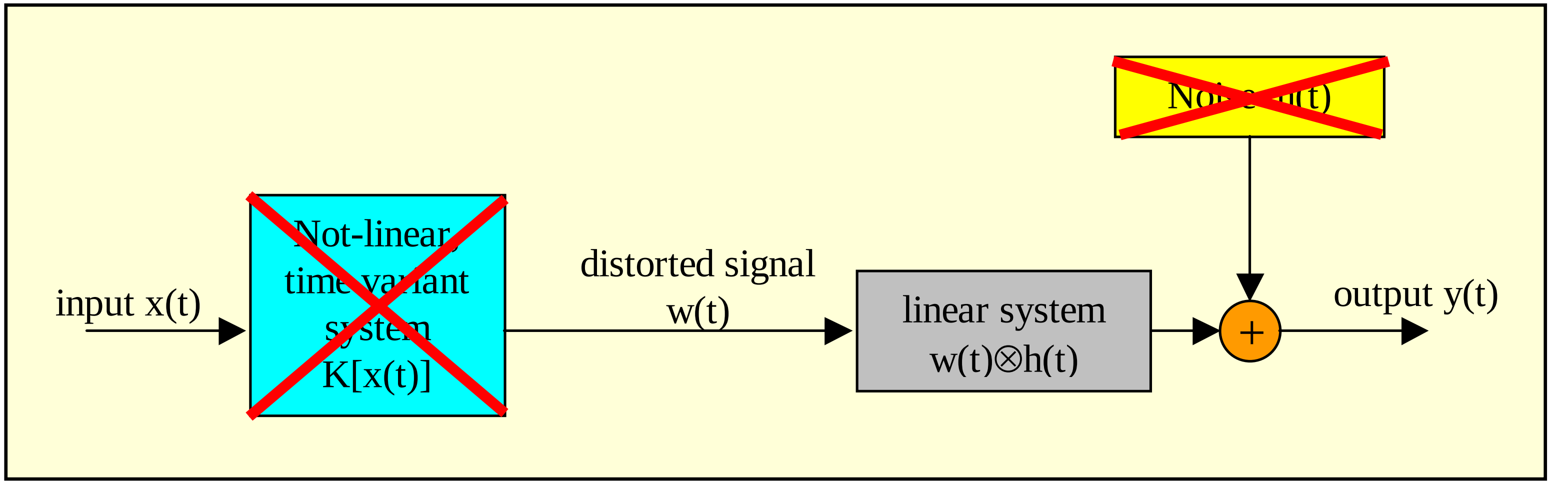


Loudspeaker as sound source



- A loudspeaker is fed with a special test signal x(t), while a microphone records the room response
- A proper deconvolution technique is required for retrieving the impulse response h(t) from the recorded signal y(t)

Measurement process

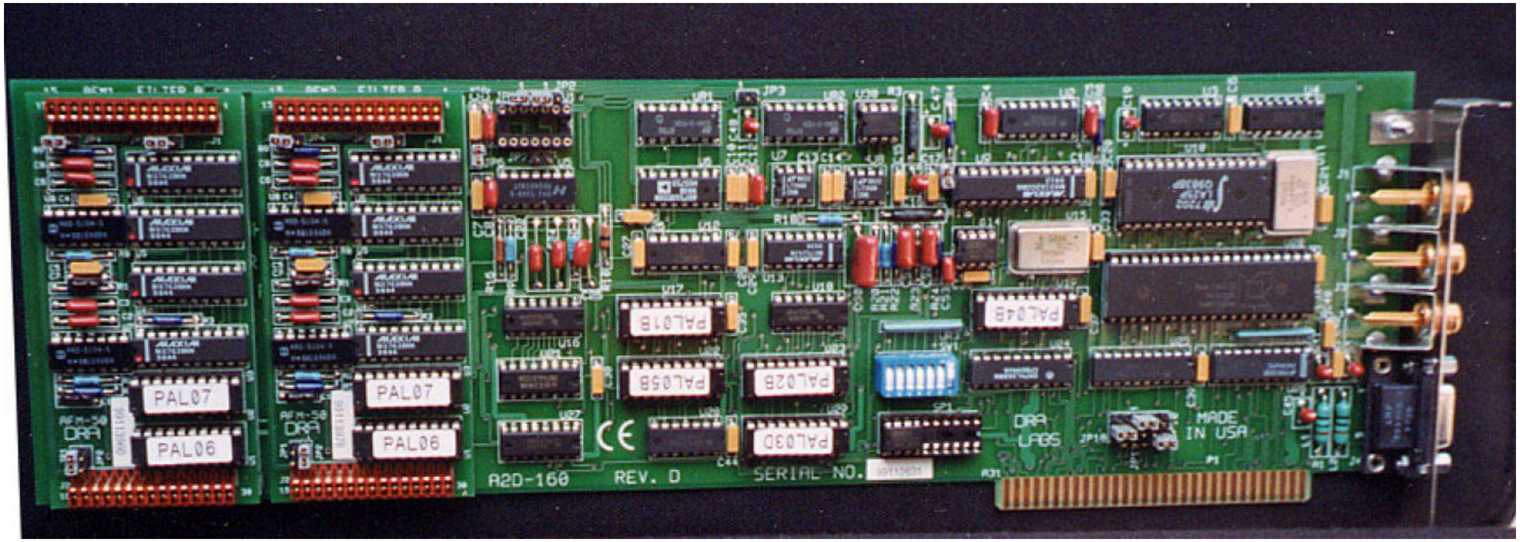


- The desidered result is the linear impulse response of the acoustic propagation h(t). It can be recovered by knowing the test signal x(t) and the measured system output y(t).
- It is necessary to exclude the effect of the not-linear part K and of the background noise n(t).

Electroacoustical methods

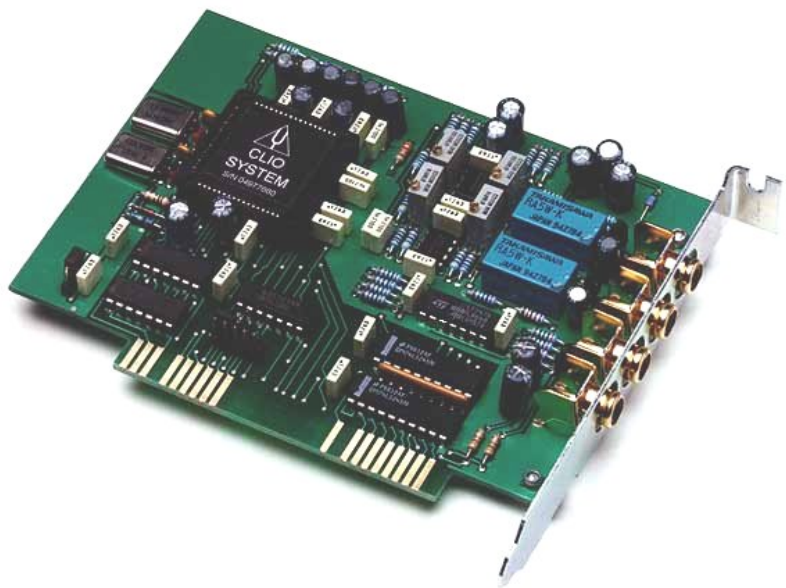
- **Different types of test signals have been developed, providing good immunity to background noise and easy deconvolution of the impulse response:**
  - MLS (Maximum Lenght Sequence, pseudo-random white noise)
  - TDS (Time Delay Spectrometry, which basically is simply a linear sine sweep, also known in Japan as “stretched pulse” and in Europe as “chirp”)
  - 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**

The first MLS apparatus - MLSSA



- **MLSSA was the first apparatus for measuring impulse responses with MLS**

More recently - the CLIO system



PB-4281

SC-01

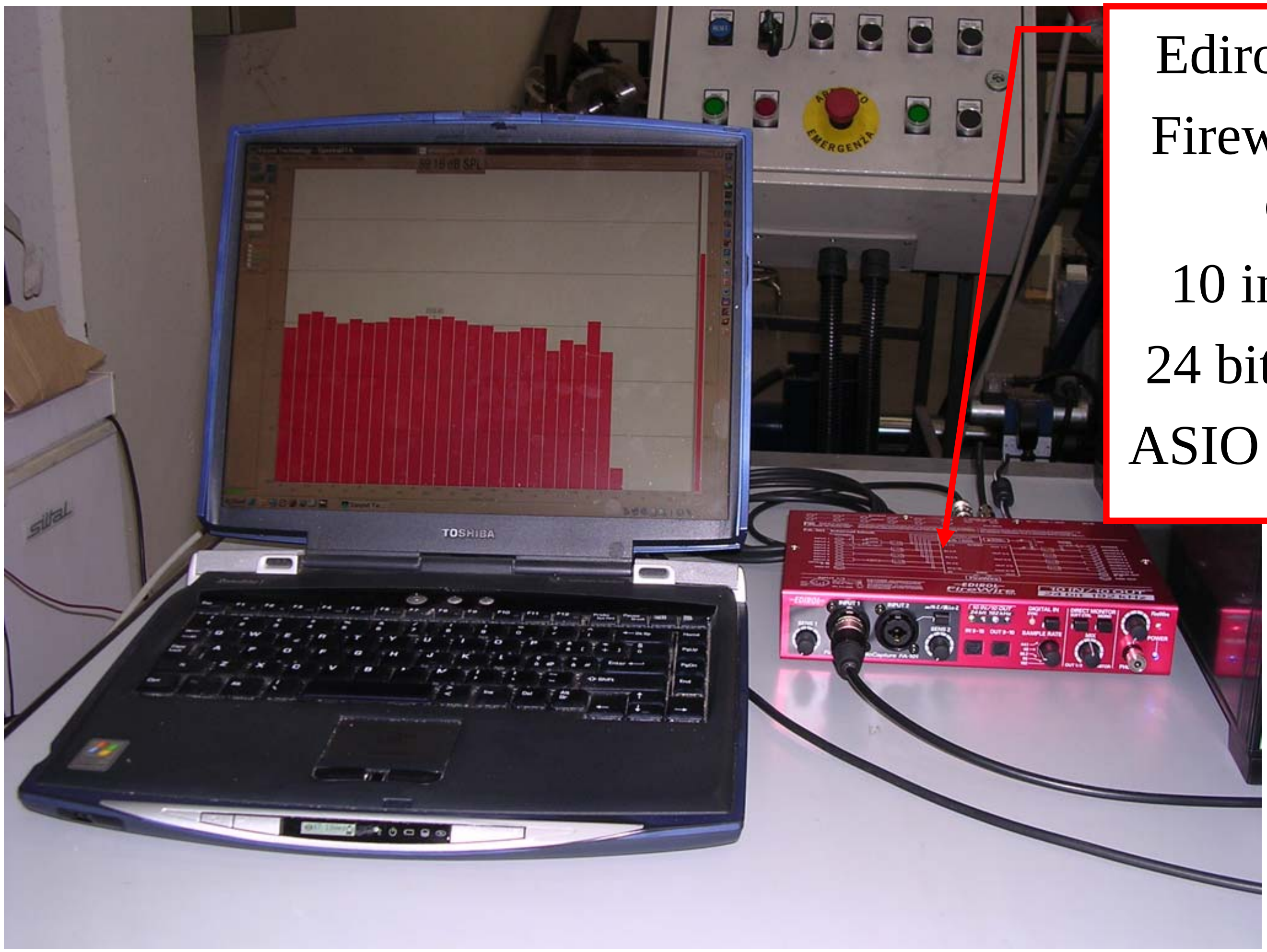
- **The Italian-made CLIO system has superseded MLSSA for most low-cost electroacoustics applications (measurement of loudspeakers, quality control)**

The first TDS apparatus - TEF

- Techron TEF 10 was the first apparatus for measuring impulse responses with TDS
- Subsequent versions (TEF 20, TEF 25) also support MLS

The Present

Today's Hardware: PC and audio interface



Edirol FA-101

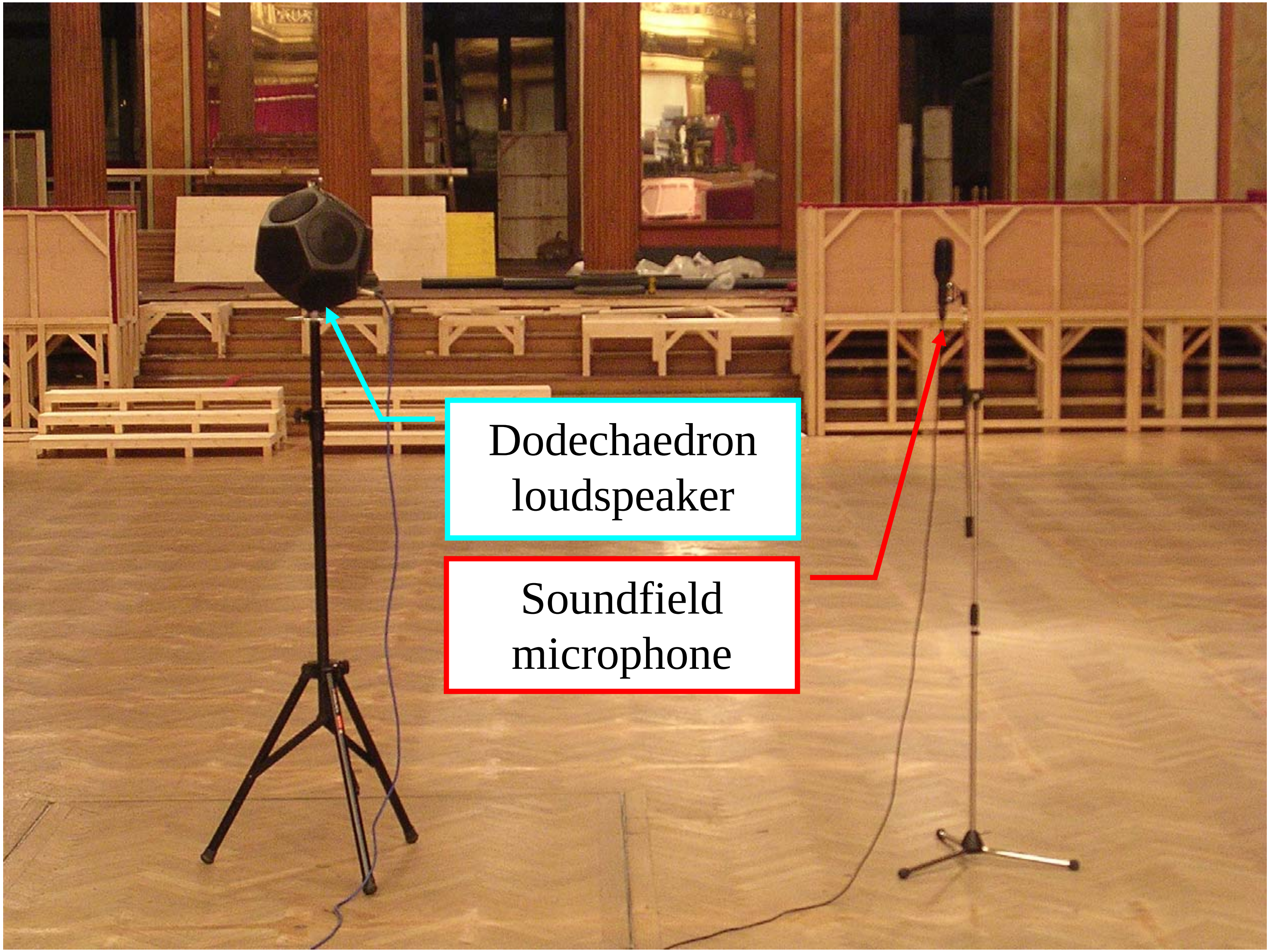
Firewire sound card:

10 in / 10 out

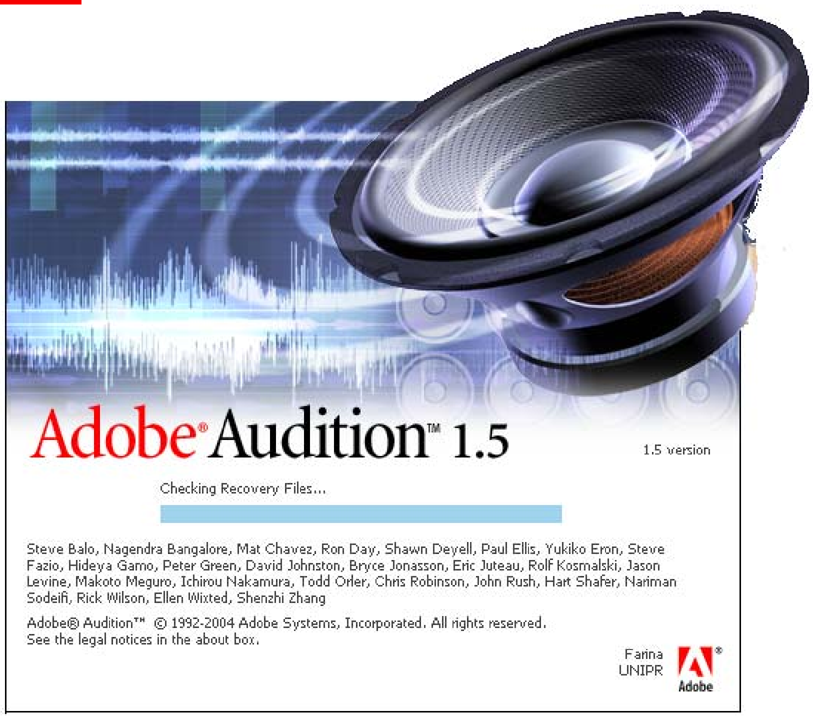
24 bit, 192 kHz

ASIO and WDM

Hardware: loudspeaker & microphone

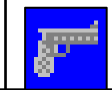
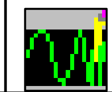
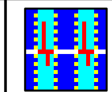
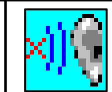
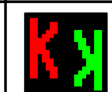


The first ESS system - AURORA



+

Aurora Plugins

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

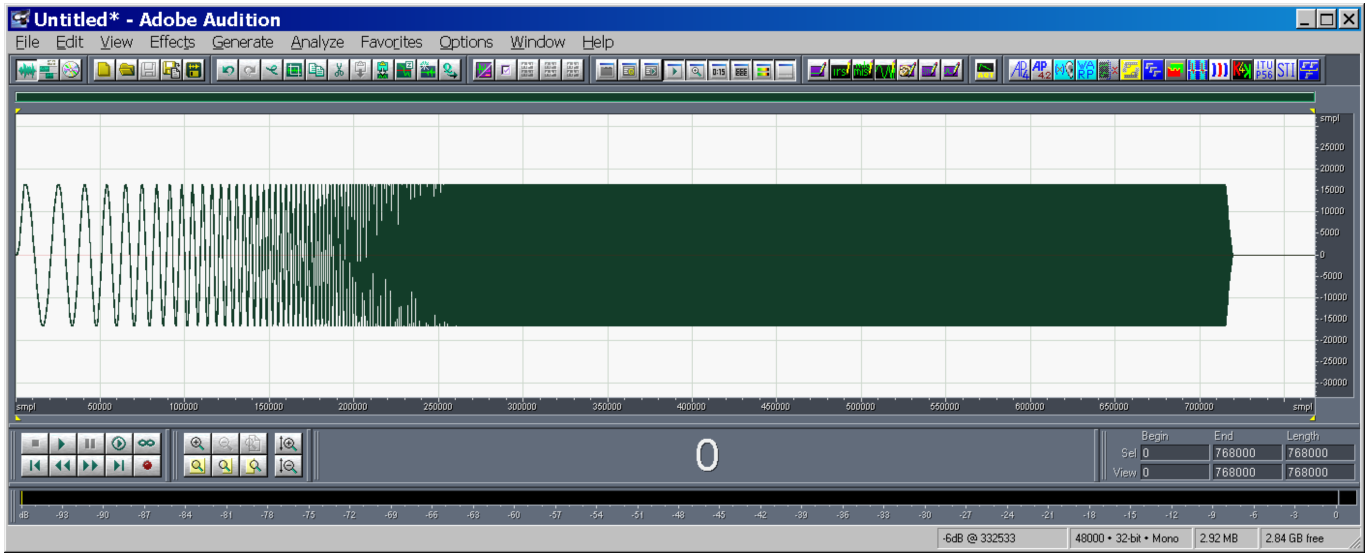
- **Aurora was the first measurement system based on standard sound cards and employing the Exponential Sine Sweep method**
- **It also works with traditional TDS and MLS methods, so the comparison can be made employing exactly the same hardware**

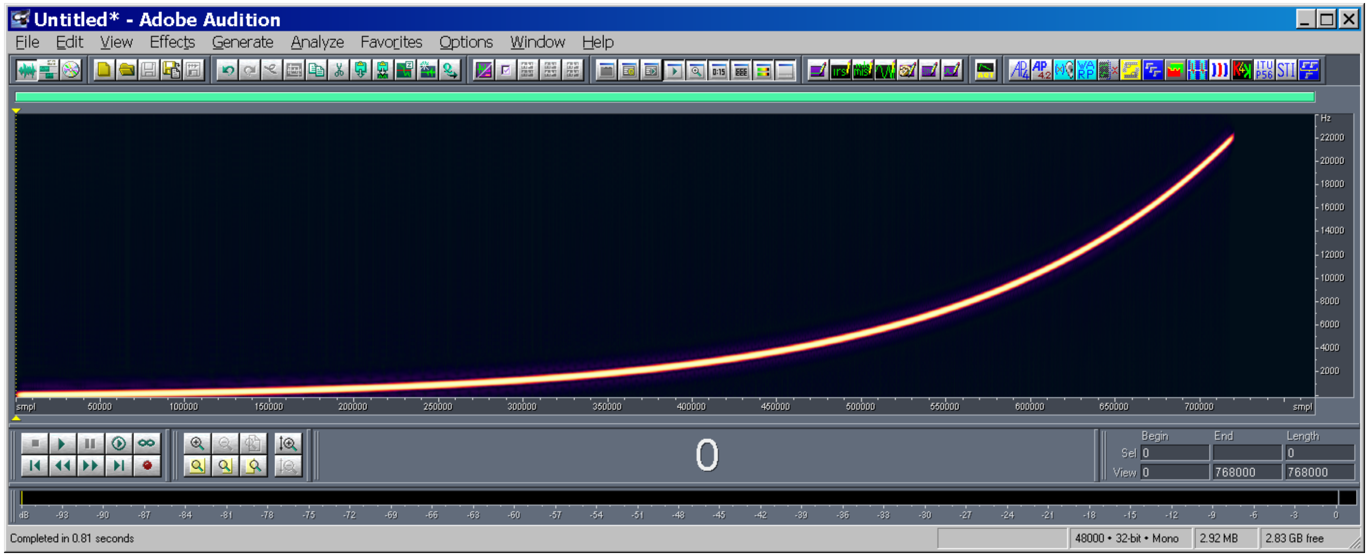
Exponential Sine Sweep method

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

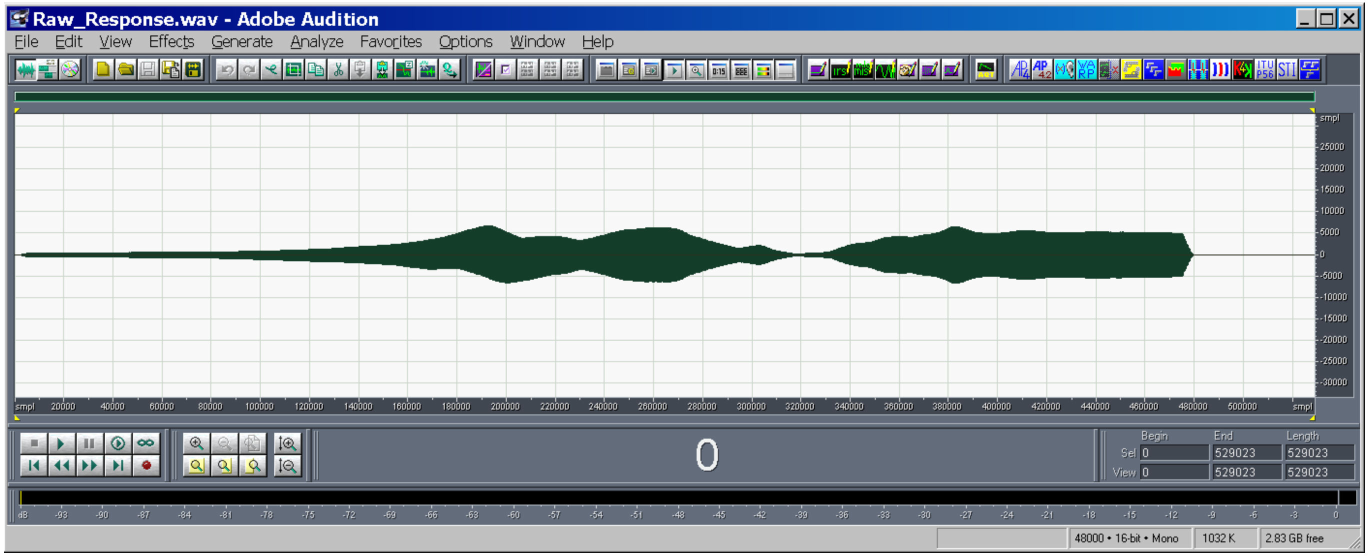
$$x(t) = \sin\left[\frac{2 \cdot \pi \cdot f_1 \cdot T}{\ln\left(\frac{f_2}{f_1}\right)} \cdot \left(e^{\frac{t}{T} \cdot \ln\left(\frac{f_2}{f_1}\right)} - 1\right)\right]$$

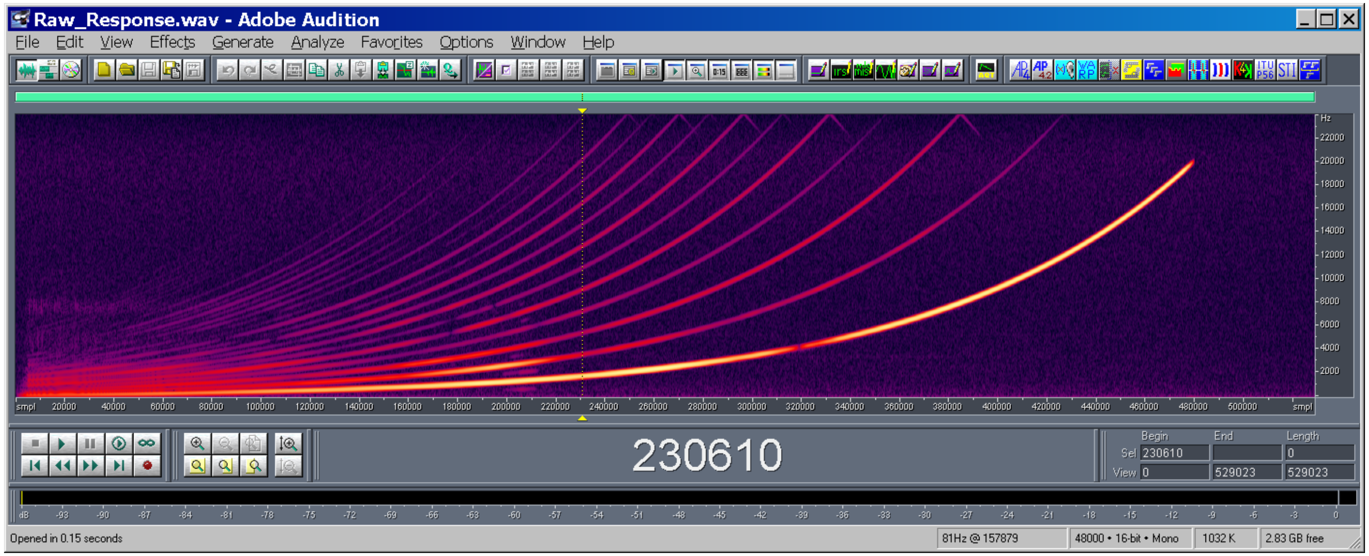
Test Signal – x(t)





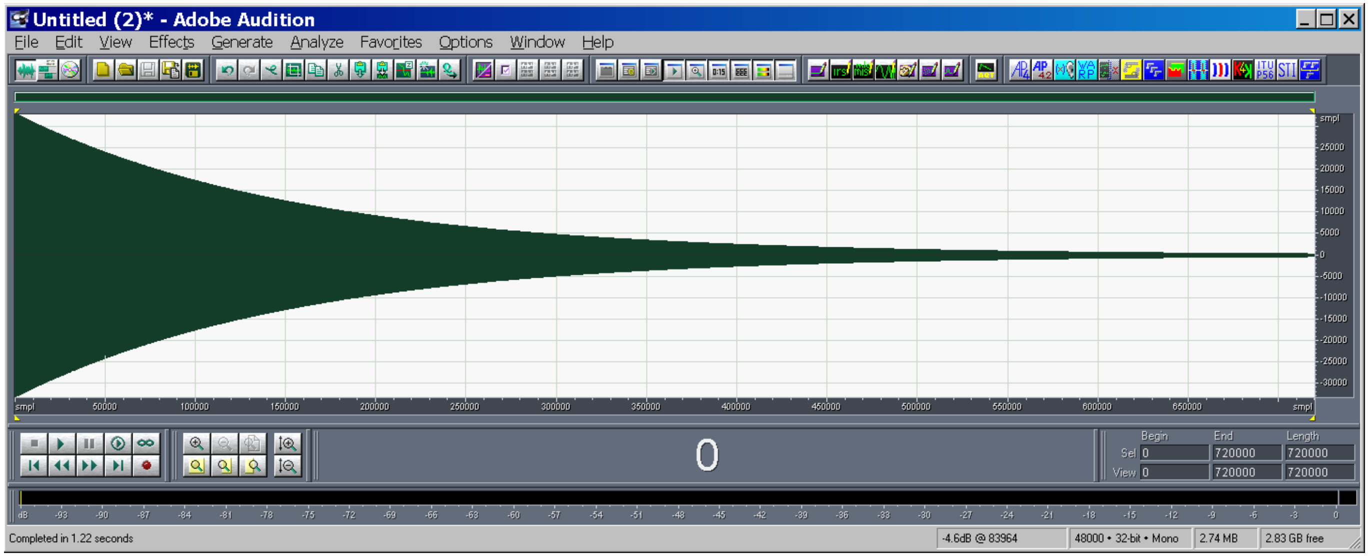
Measured signal - y(t)

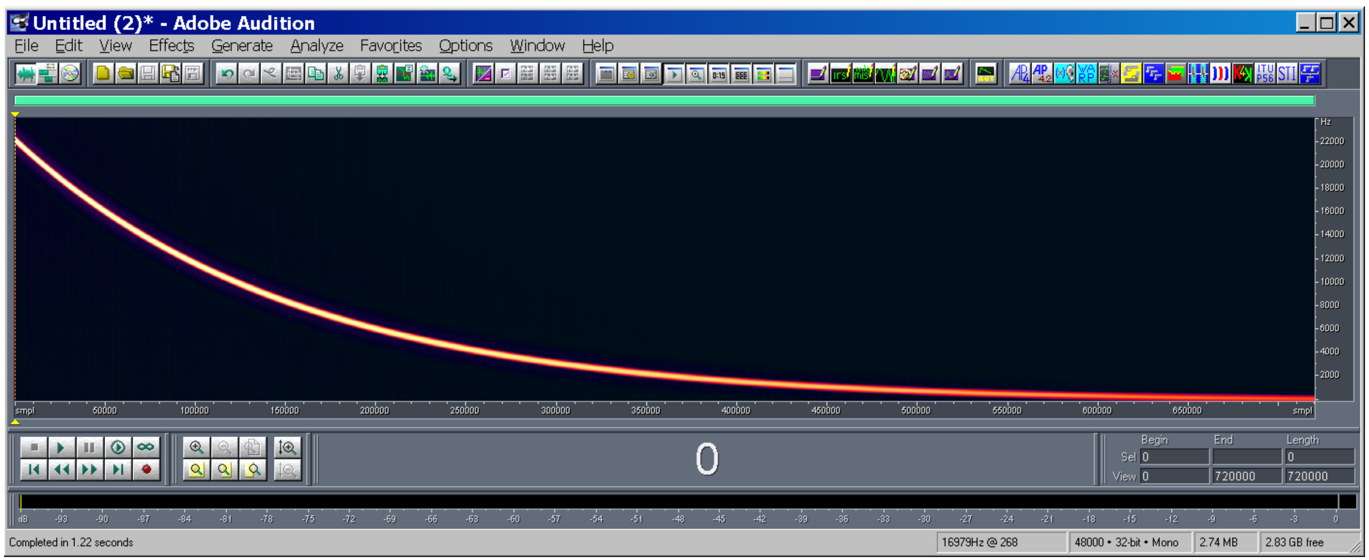




- **The not-linear behaviour of the loudspeaker causes many harmonics to appear**

Inverse Filter – z(t)

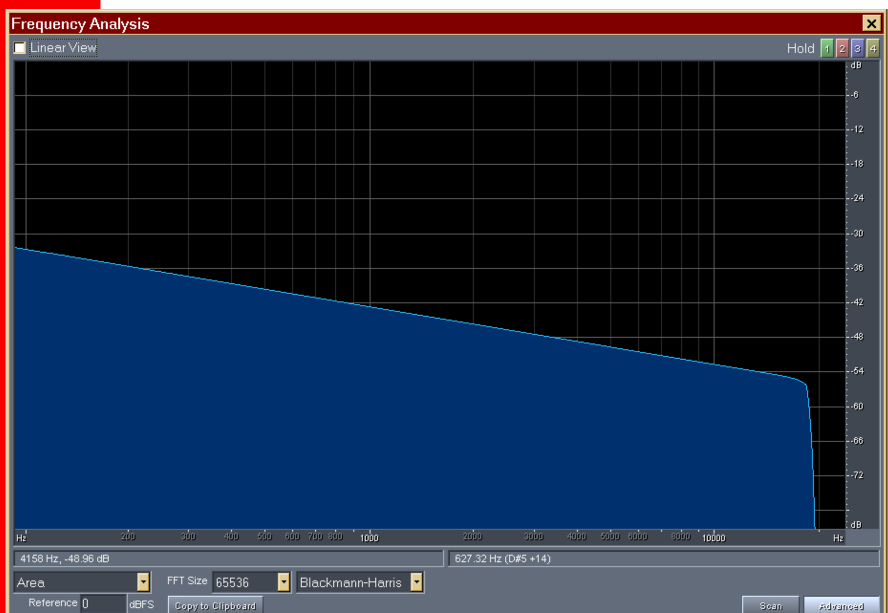




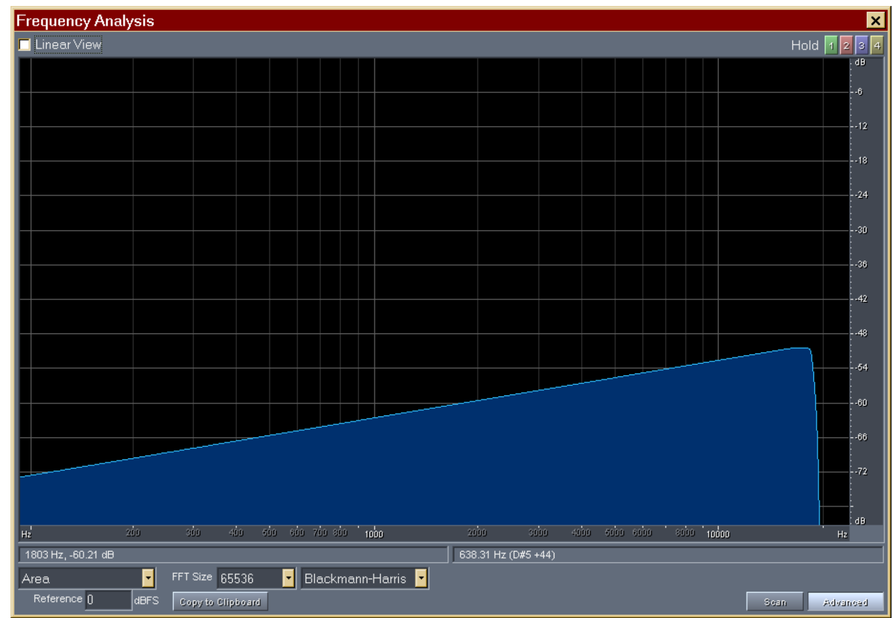
The deconvolution of the IR is obtained convolving the measured signal y(t) with the inverse filter z(t) [equalized, time-reversed x(t)]

Deconvolution of Exponential 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

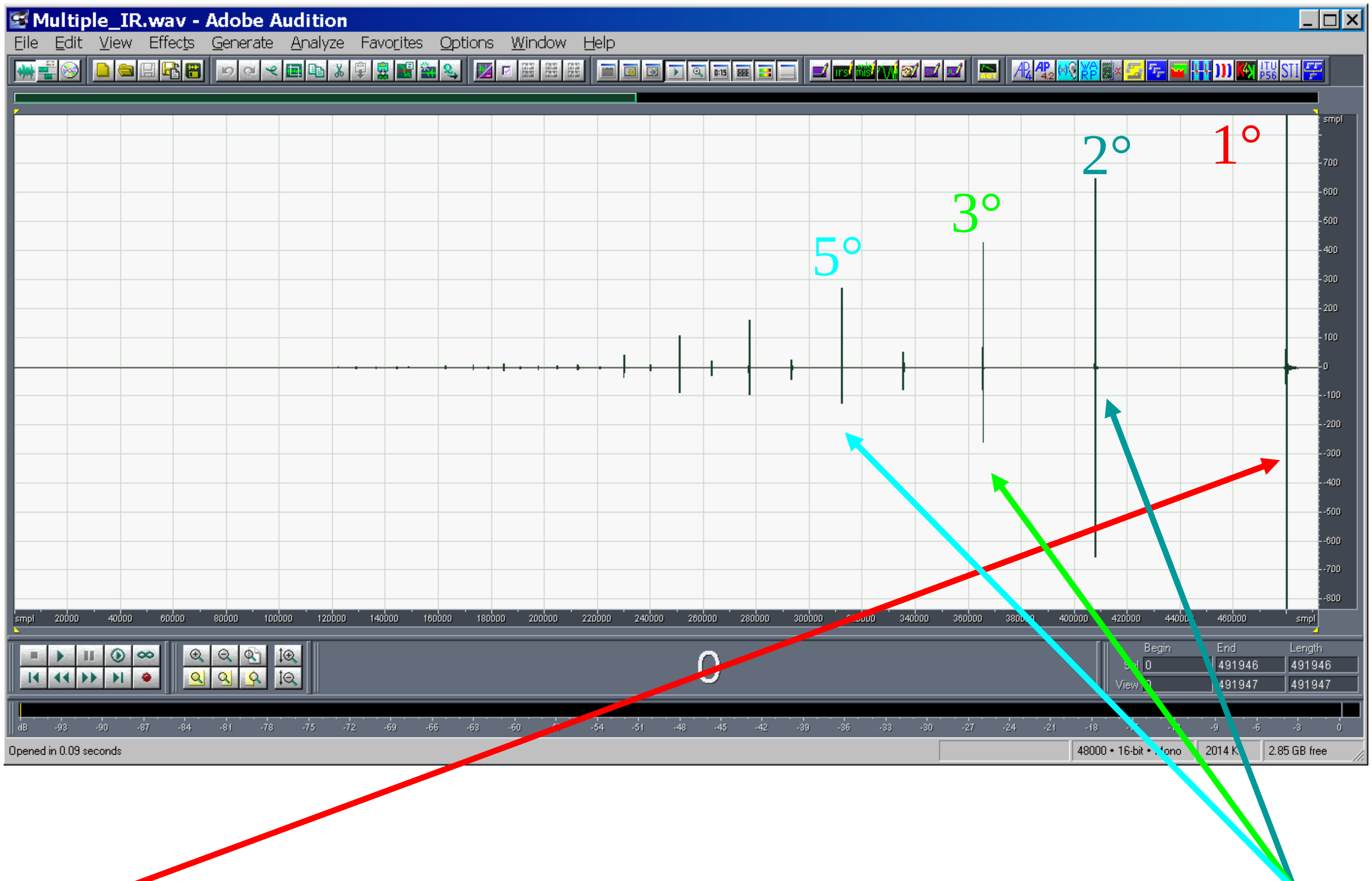


Test Signal x(t)



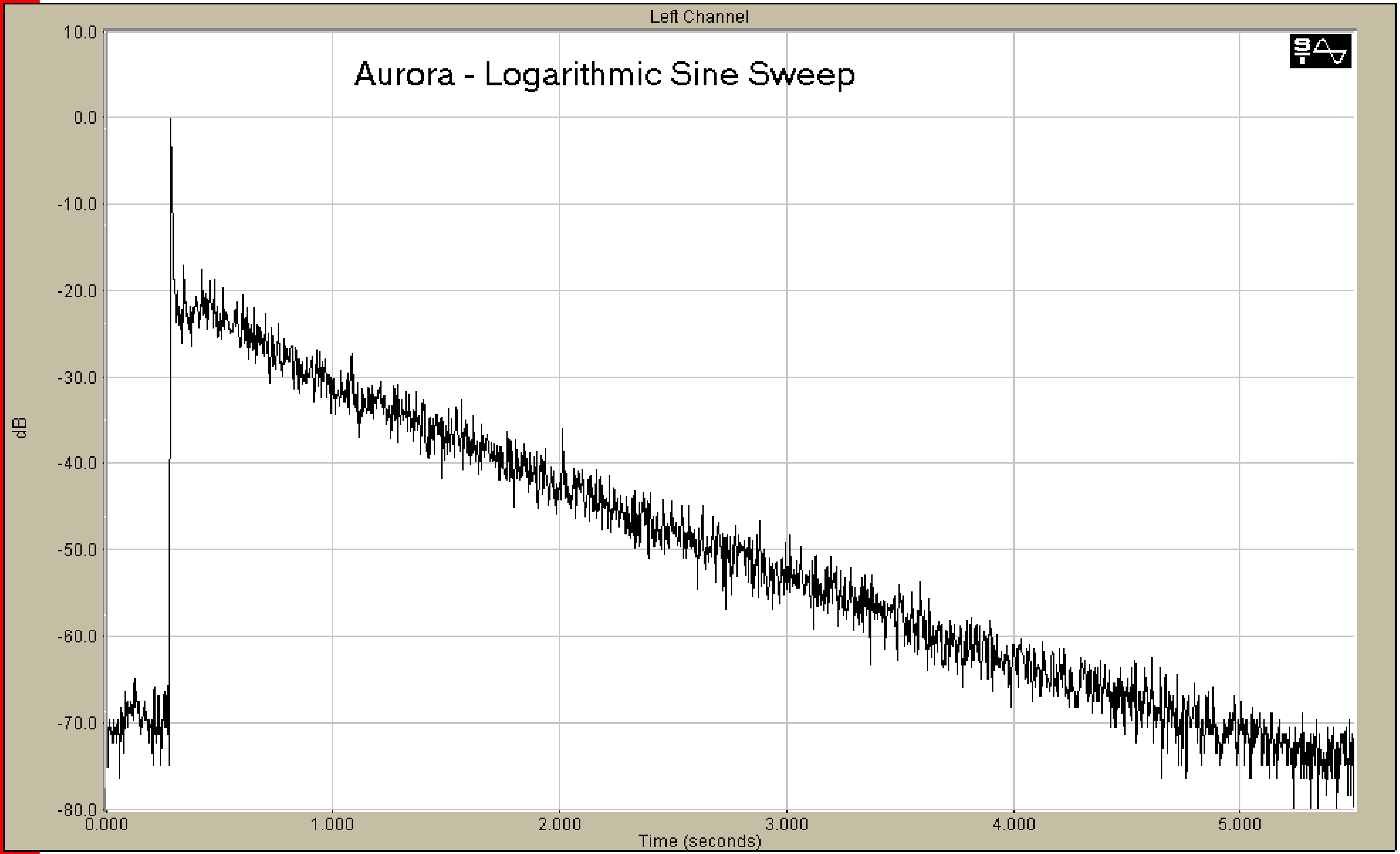
Inverse Filter z(t)

Result of the deconvolution

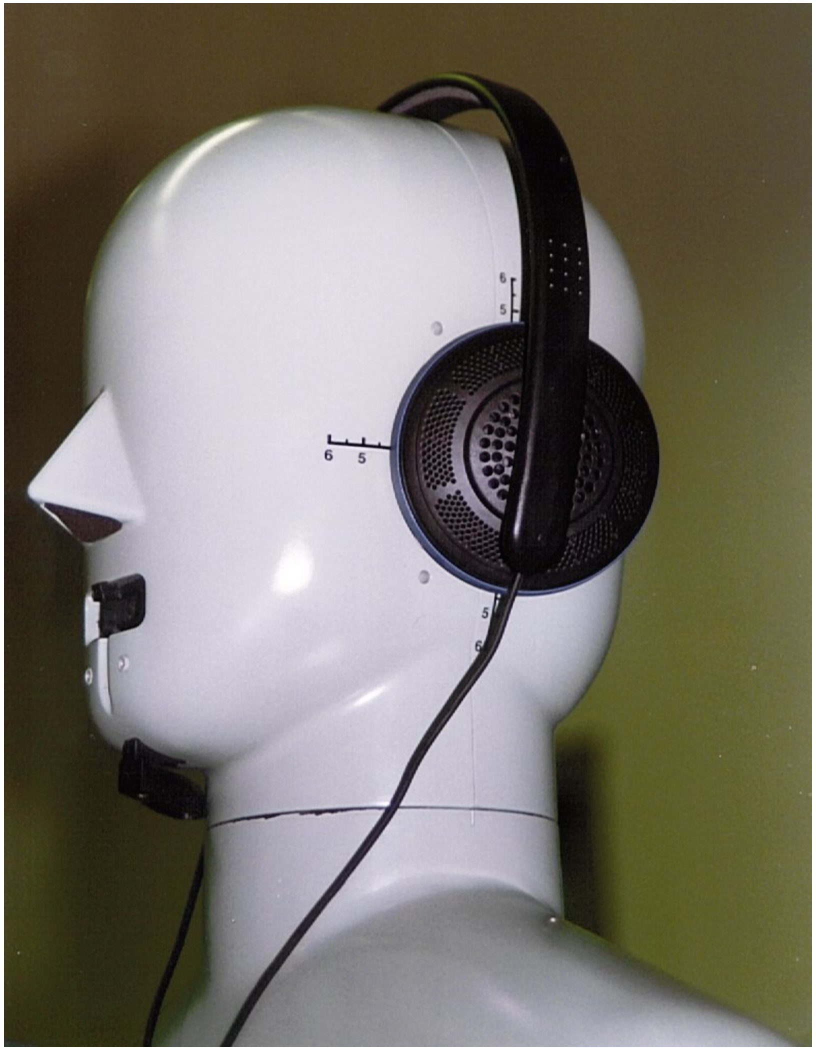


The last impulse response is the linear one, the preceding are the harmonics distortion products of various orders

Maximum Length Sequence vs. Exp. Sine Sweep

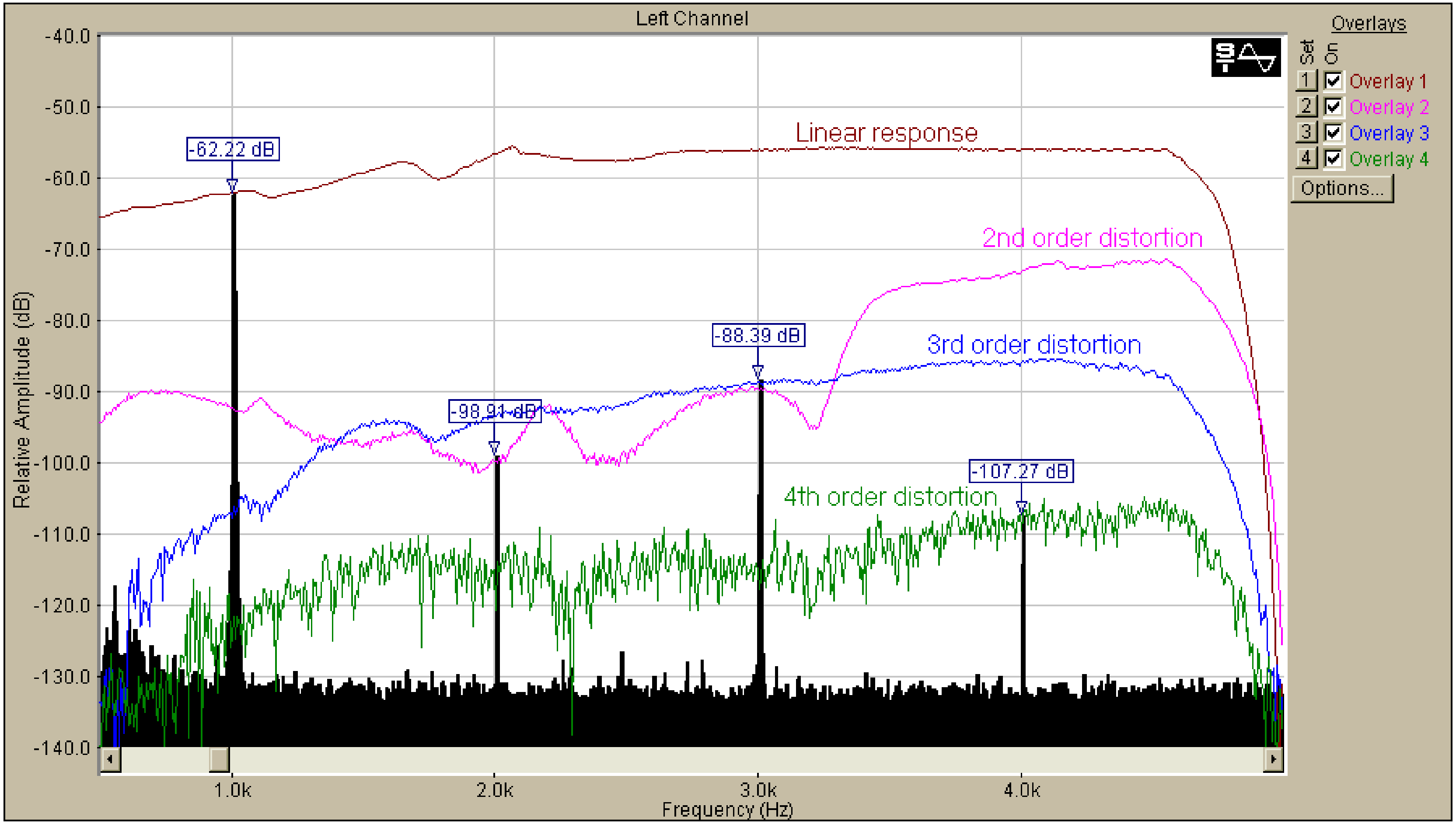


Distortion measurements



- A headphone was driven with a 1 V RMS signal, causing severe distortion in the small loudspeaker.
- The measurement was made placing the headphone on a dummy head.
- Measurements: ESS and traditional sine at 1 kHz

Distortion measurements



- **Comparison between:**
- **traditional distortion measurement with fixed-frequency sine (the black histogram)**
- **the new exponential sweep (the 4 narrow, coloured lines)**

Spatial analysis by directive impulse responses

- **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:**

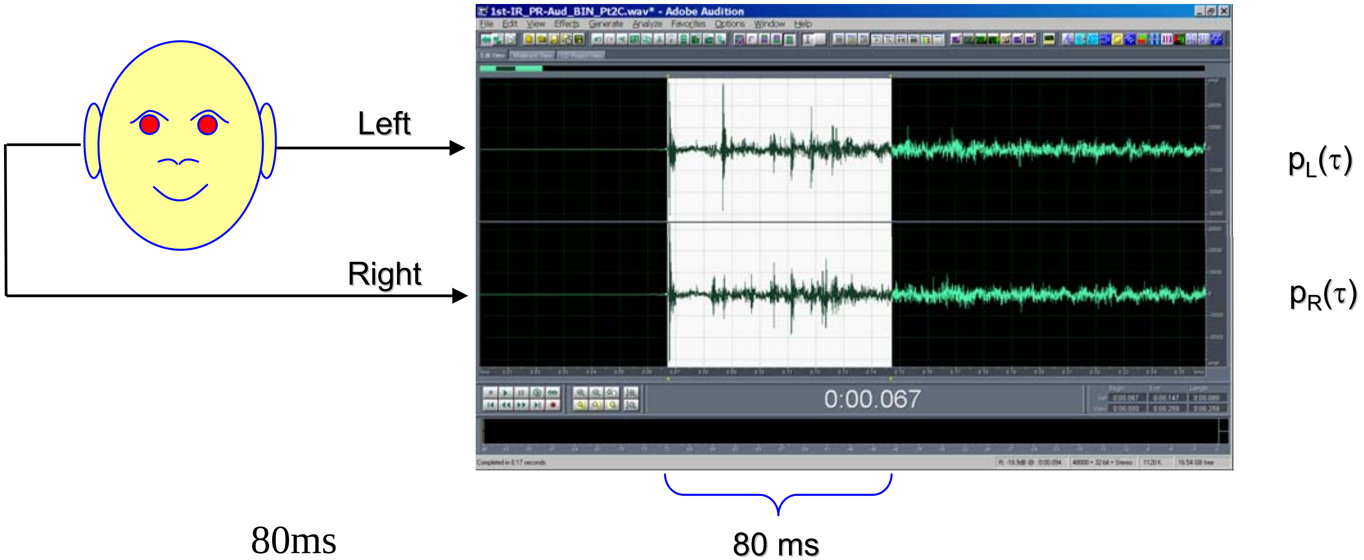
Binaural microphone (left)

and

Pressure-velocity microphone (right)

IACC "objective" spatial parameter

- 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

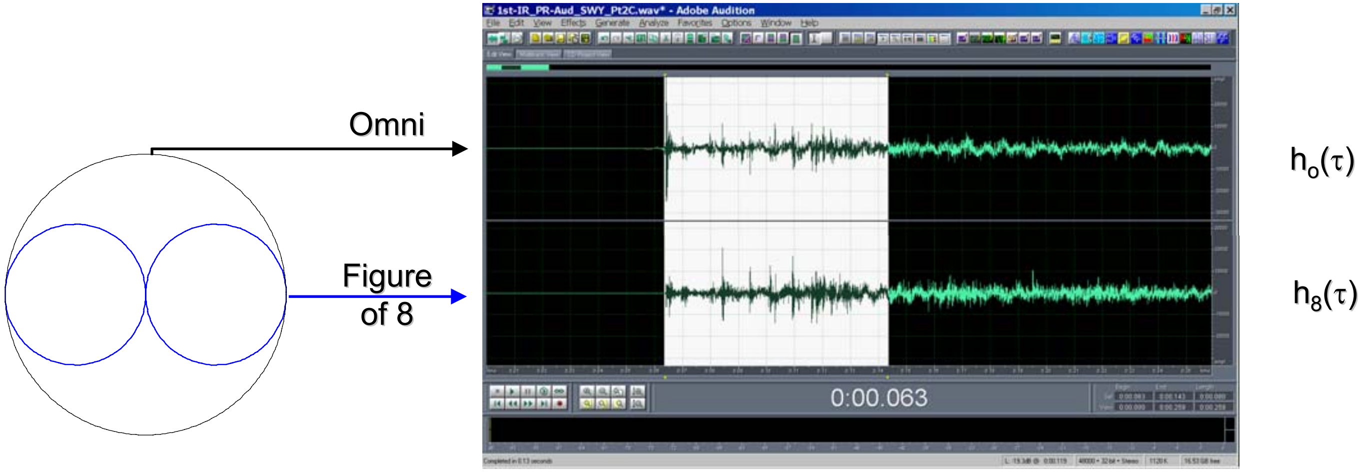


$$\rho(t) = \frac{\int_0^{80ms} p_L(\tau)\cdot p_R(\tau + t)\cdot d\tau}{\sqrt{\int_0^{80ms} p_L^2(\tau)\cdot d\tau \cdot \int_0^{80ms} p_R^2(\tau + t)\cdot d\tau}}$$

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

Lateral Fraction (LF) spatial parameter

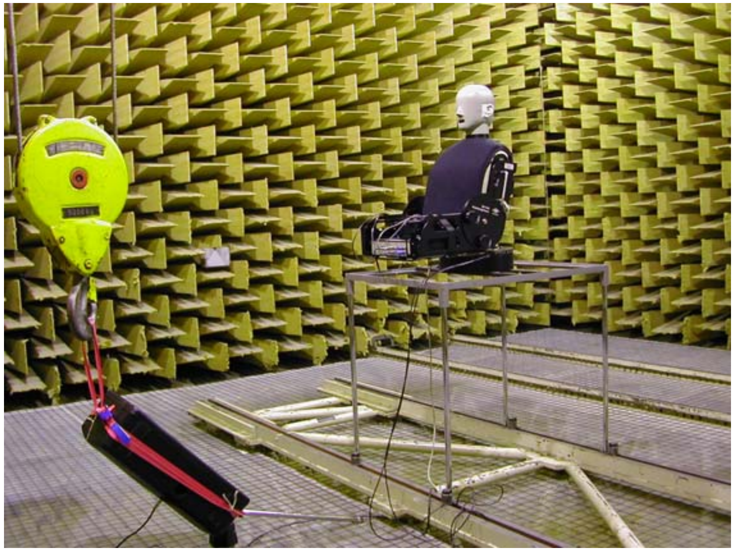
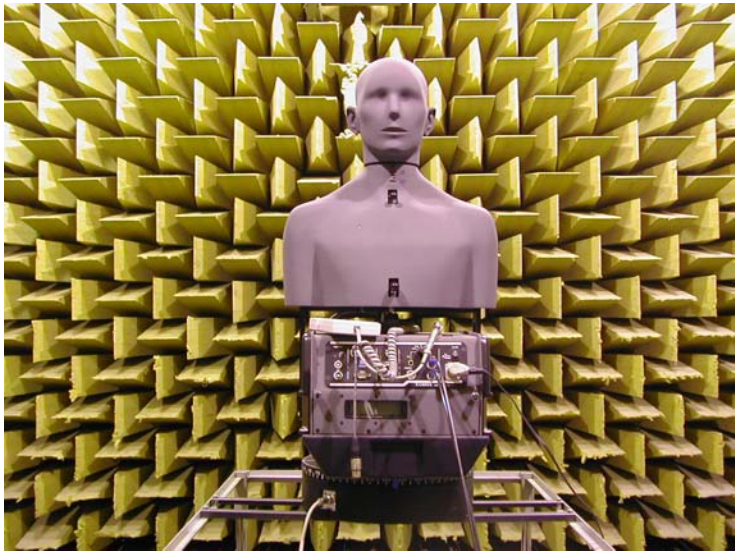
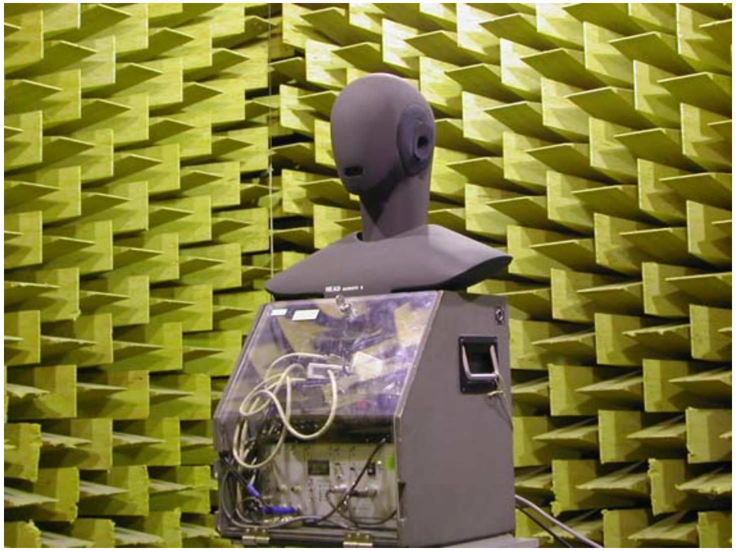
- **Another "spatial" parameter is the Lateral Fraction LF**
- **This is defined from a 2-channels impulse response, the first channel is a standard omni microphone, the second channel is a "figure-of-eight" microphone:**



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

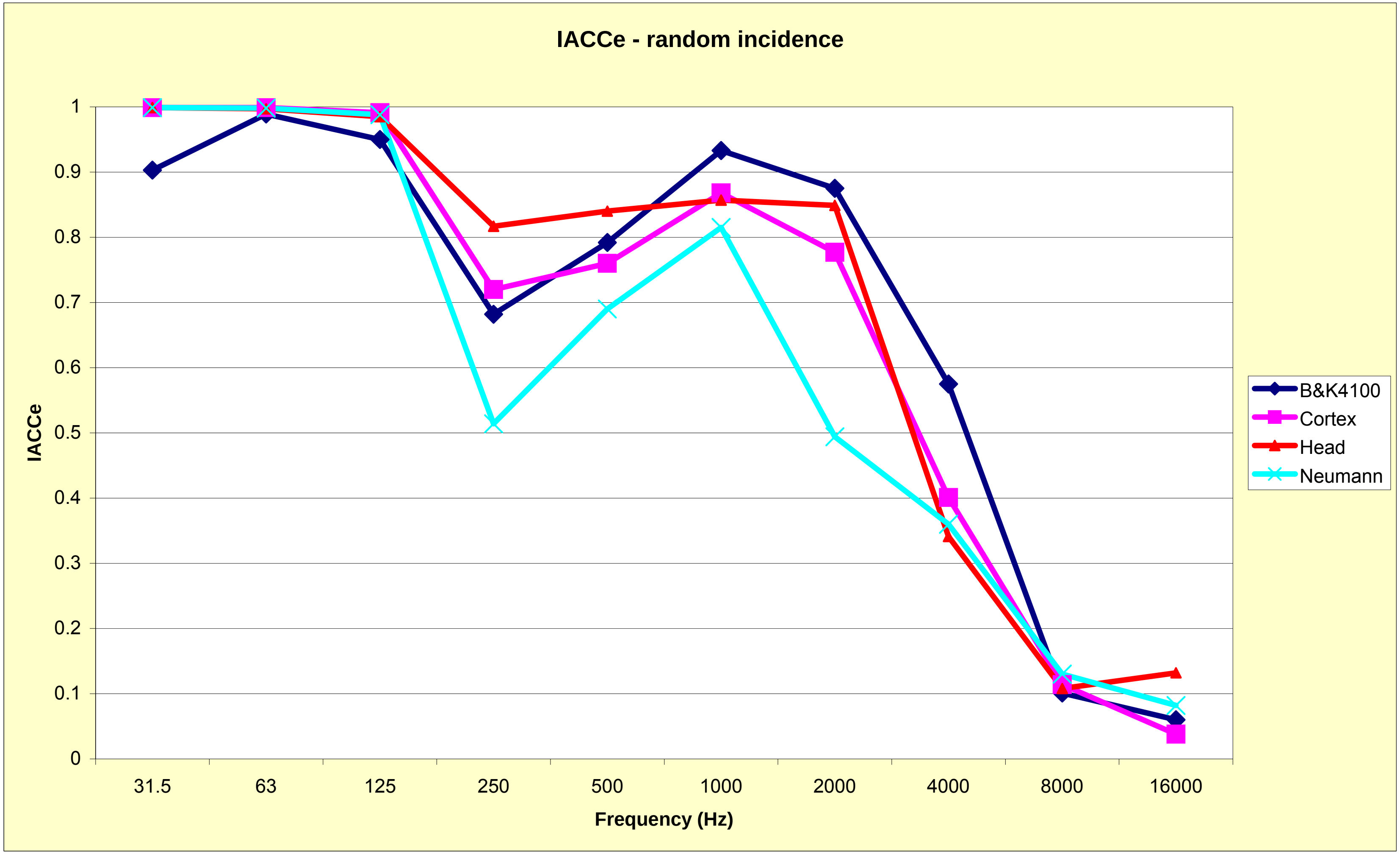
Are binaural measurents reproducible?

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


Are IACC measurents reproducible?

- **Diffuse field - huge difference among the 4 dummy heads**

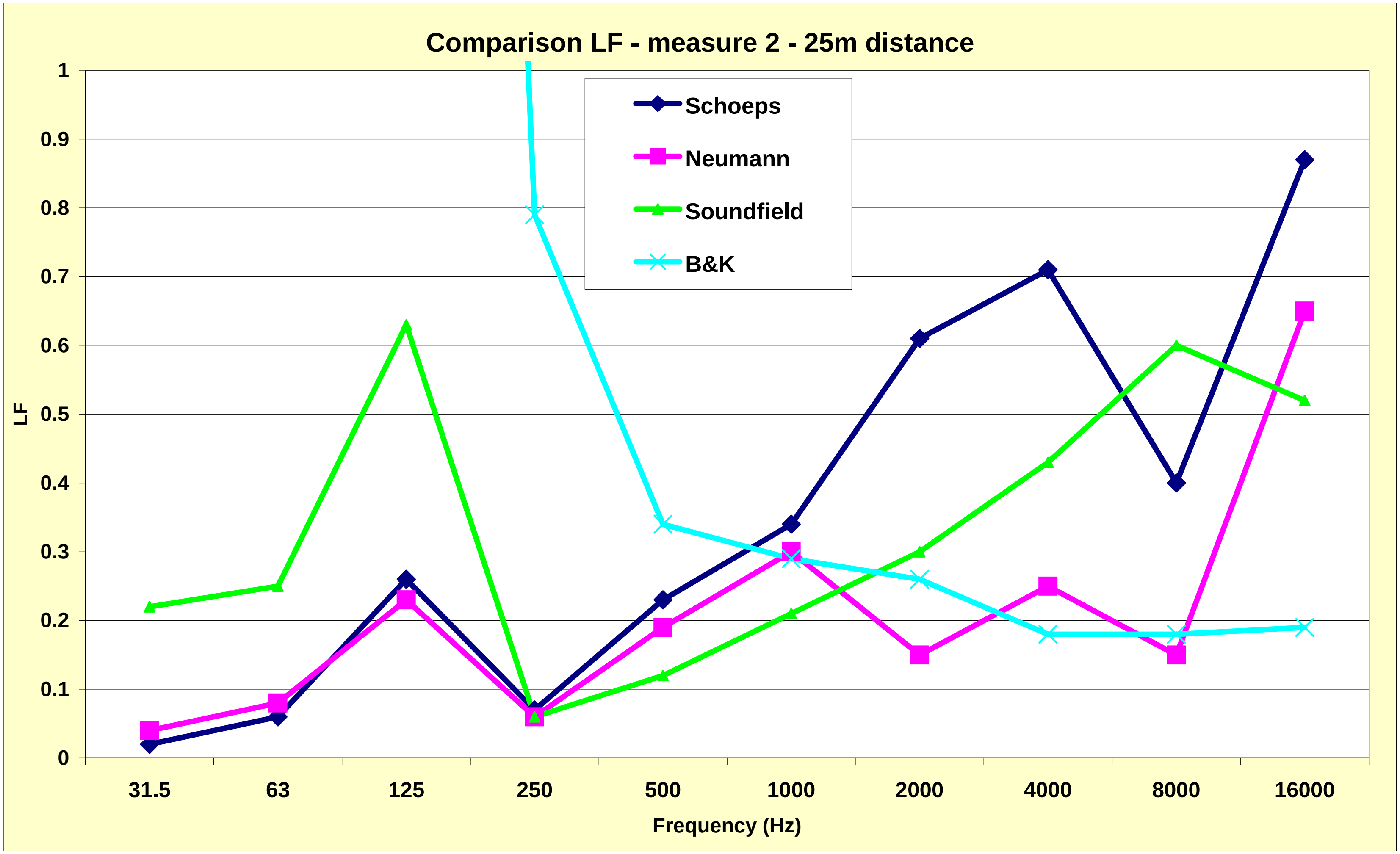


Are LF measurents reproducible?

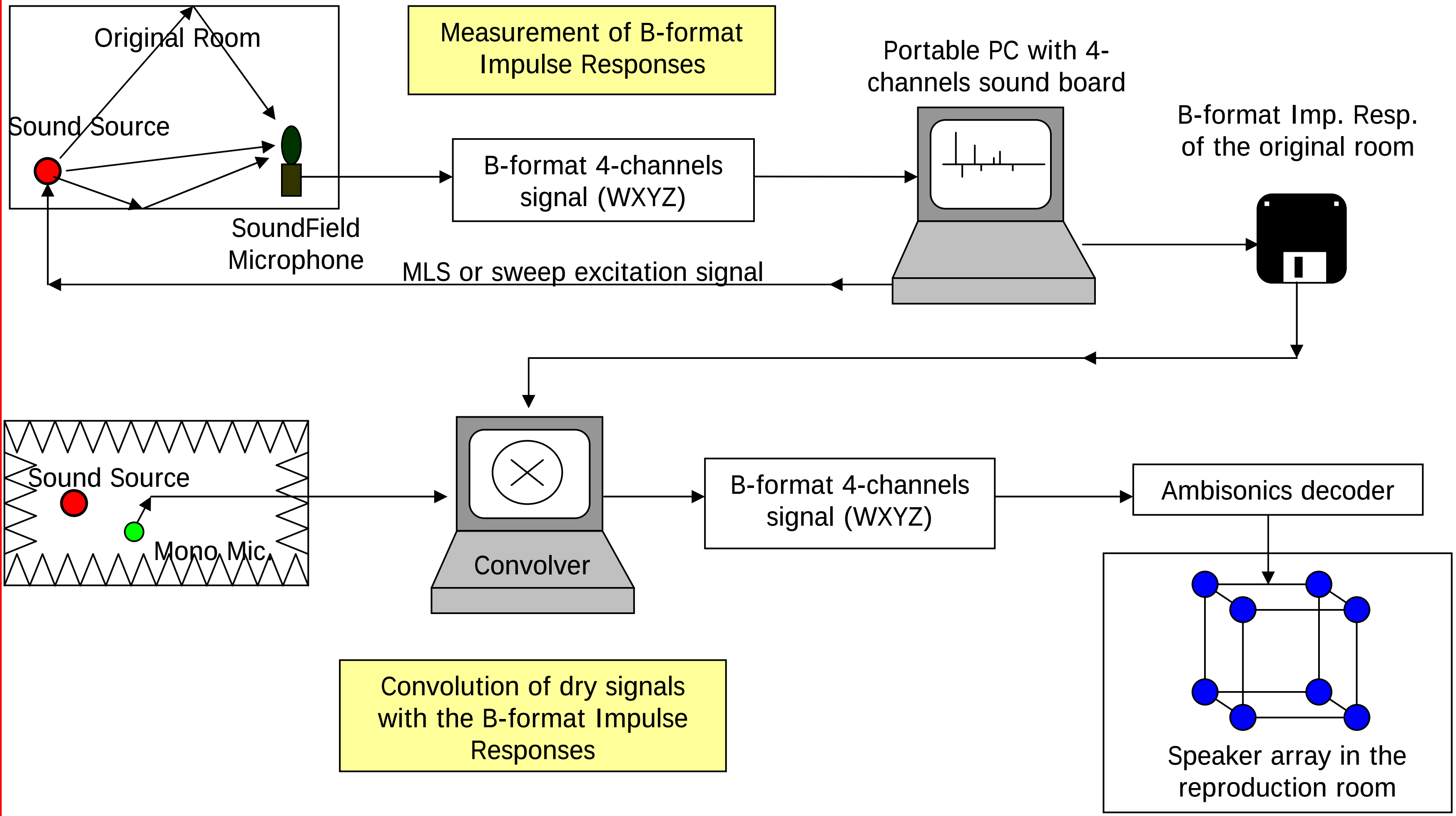
- **Experiment performed in the Auditorium of Parma - same loudspeaker, same source and receiver positions, 4 pressure-velocity microphones**

Are LF measurents reproducible?

- At 25 m distance, the scatter is really big

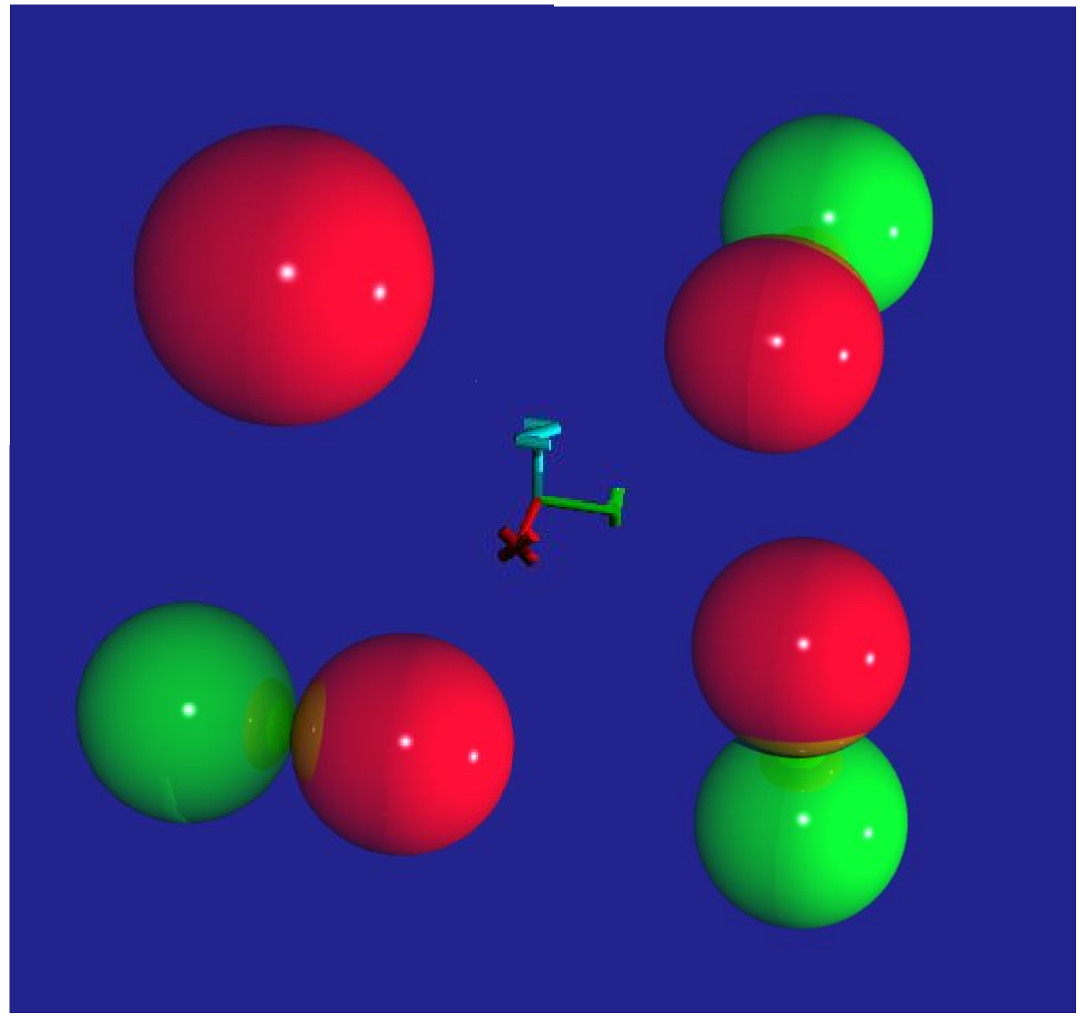


3D Impulse Response (Gerzon, 1975)



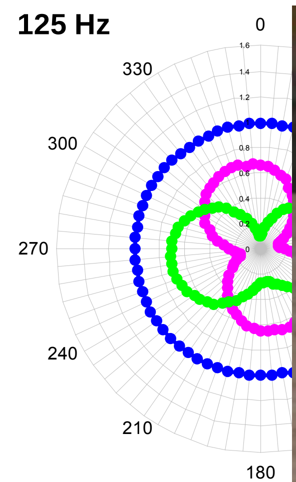
3D extension of the pressure-velocity measurements

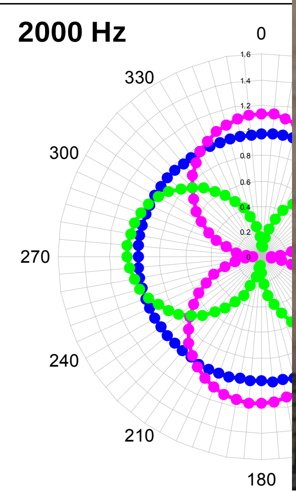
- **The Soundfield microphone allows for simultaneous measurements of the omnidirectional pressure and of the three cartesian components of particle velocity (figure-of-8 patterns)**

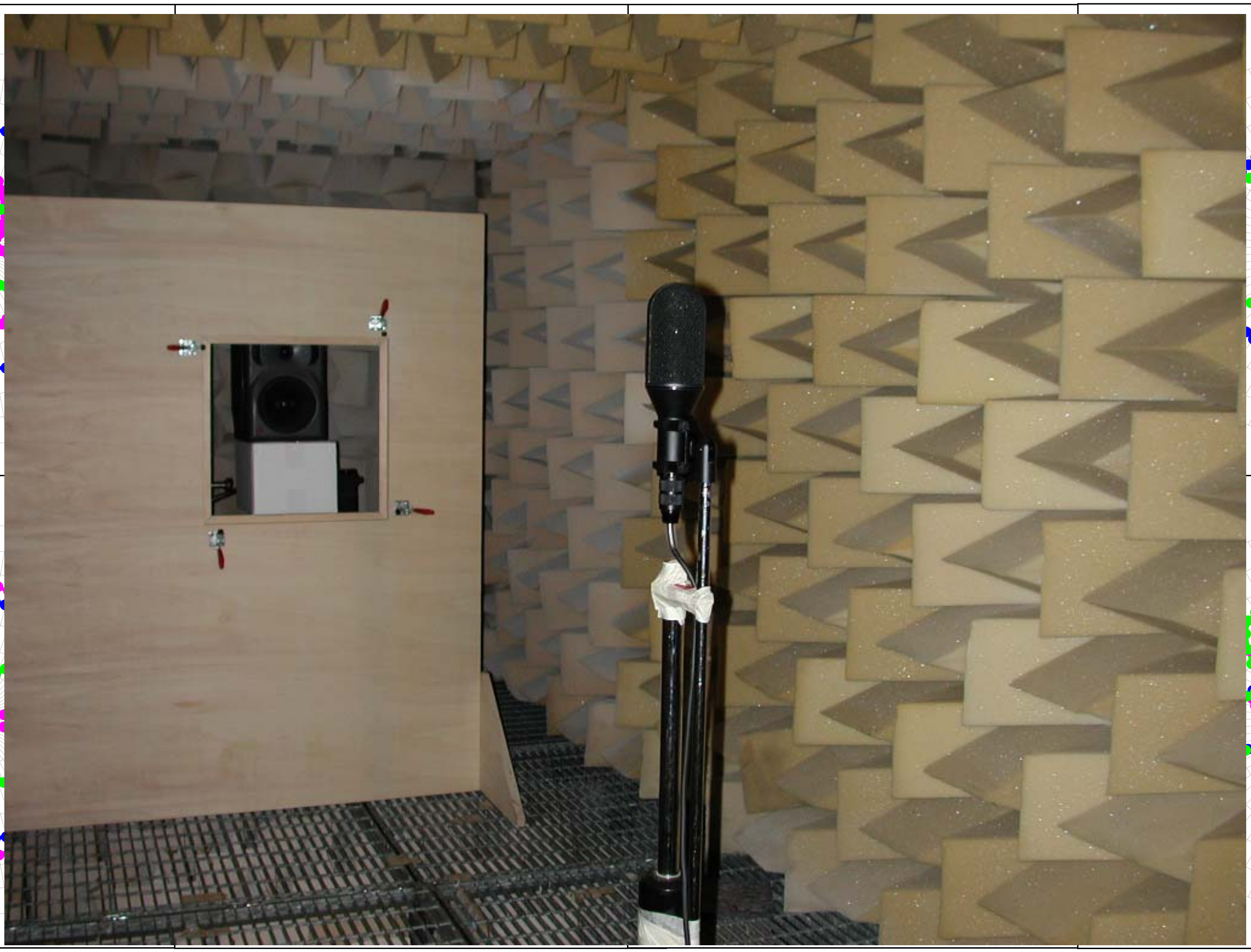


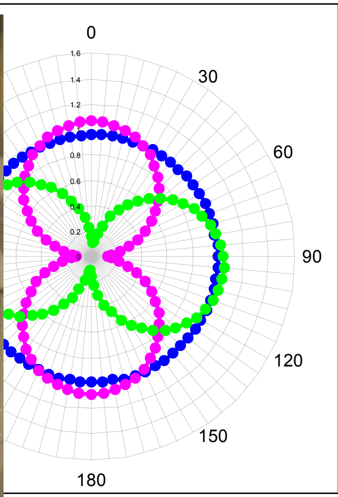
Directivity of transducers

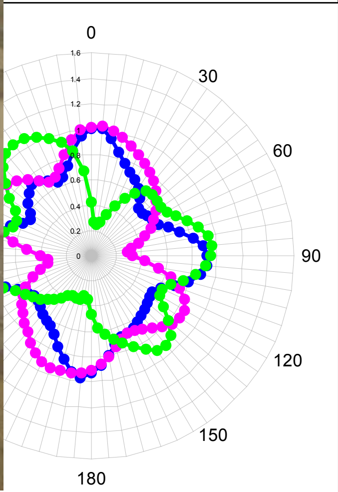
Soundfield ST-250 microphone











A-format microphone arrays

- **Today several alternatives to Soundfield microphones do exists. All of them are providing “raw” signals from the 4 capsules, and the conversion from these signals (A-format) to the standard Ambisonic signals (B-format) is performed digitally by means of software running on the computer**

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

Problems with ESS measurements

- Pre-ringing at high and low frequency before the arrival of the direct sound pulse
- Pre/post equalization of the test signal performed in a way which avoids time-smearing of the impulse response
- Sensitivity to abrupt pulsive noises during the measurement
- Skewing of the measured impulse response when the playback and recording digital clocks are mismatched
- Cancellation of high frequencies in the late part of the tail when performing synchronous averaging

Pre-ringing at high and low frequency

- **Pre-ringing at high frequency due to improper fade-out**

This picture shows the preringing obtained deconvolving directly the test signal, without passing through the system under test

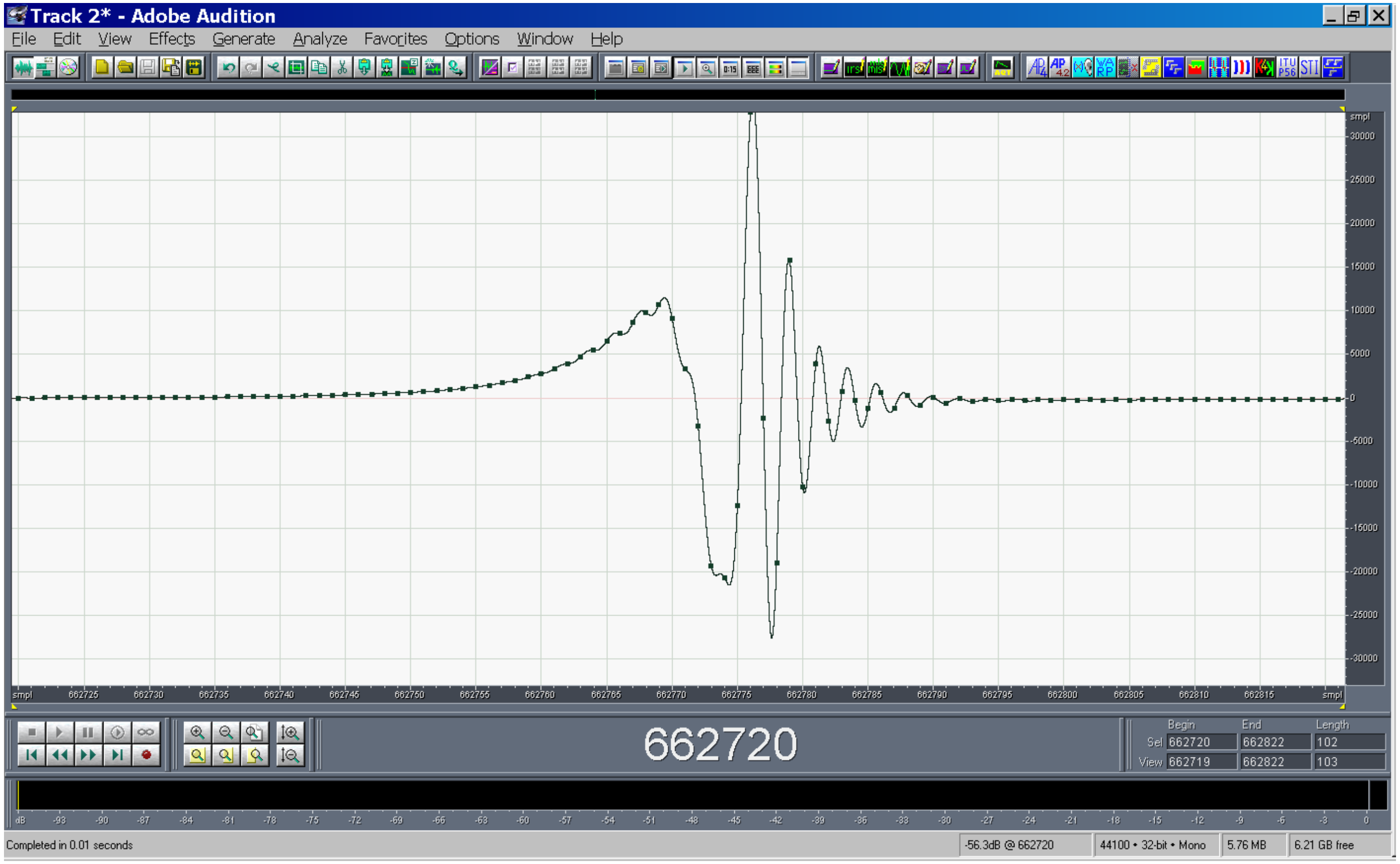
Pre-ringing at high and low frequency

- **Perfect Dirac's delta after removing the fade-out**

This picture shows the result obtained deconvolving directly the test signal, without passing through the system under test, and employing a sine sweep going up to the Nyquist frequency

Pre-ringing at high and low frequency

- **Pre-ringing at low frequency due to a bad sound card featuring frequency-dependent latency**



This artifact can be corrected if the frequency-dependent latency remains the same, by creating a suitable inverse filter with the Kirkeby method

Kirkeby inverse filter

- **The Kirkeby inverse filter is computed inverting the measured IR**

1) The IR to be inverted is FFT transformed to frequency domain:

$$H(f) = FFT\,[h(f)]$$

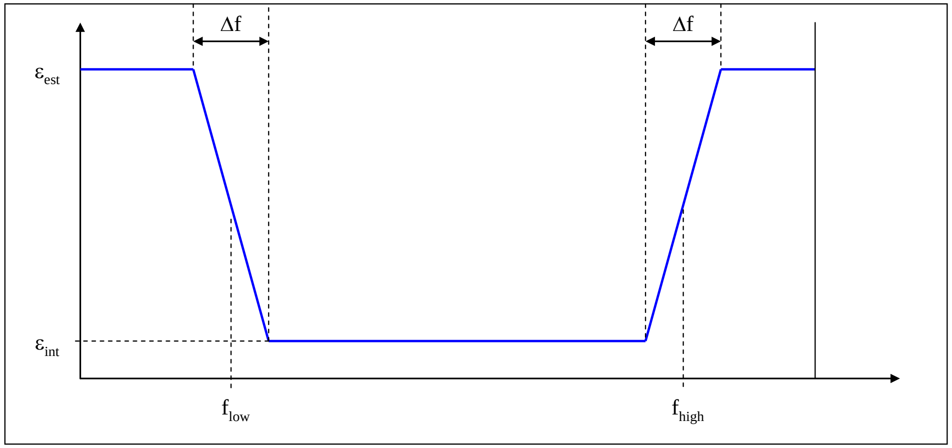
2) The computation of the inverse filter is done in frequency domain:

$$C(f) = \frac{Conj[H(f)]}{Conj[H(f)] \cdot H(f) + \varepsilon(f)}$$

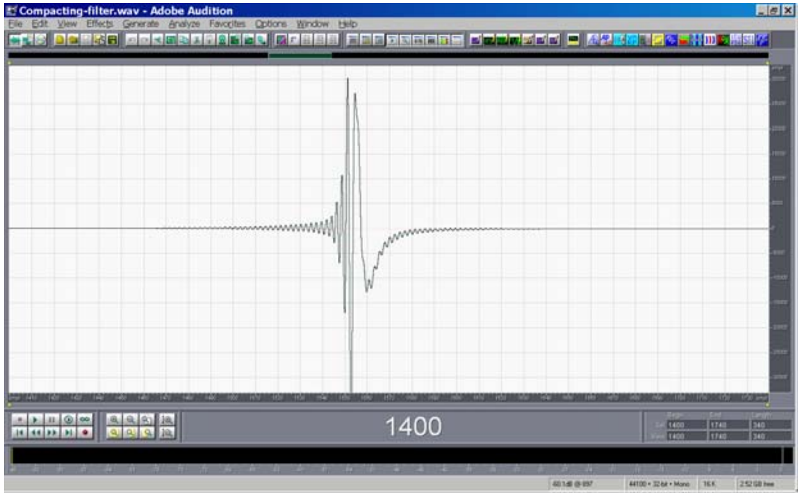
Where $\varepsilon(f)$ is a small, frequency-dependent regularization parameter

3) Finally, an IFFT brings back the inverse filter to time domain:

$$c(t) = IFFT\,[C(f)]$$



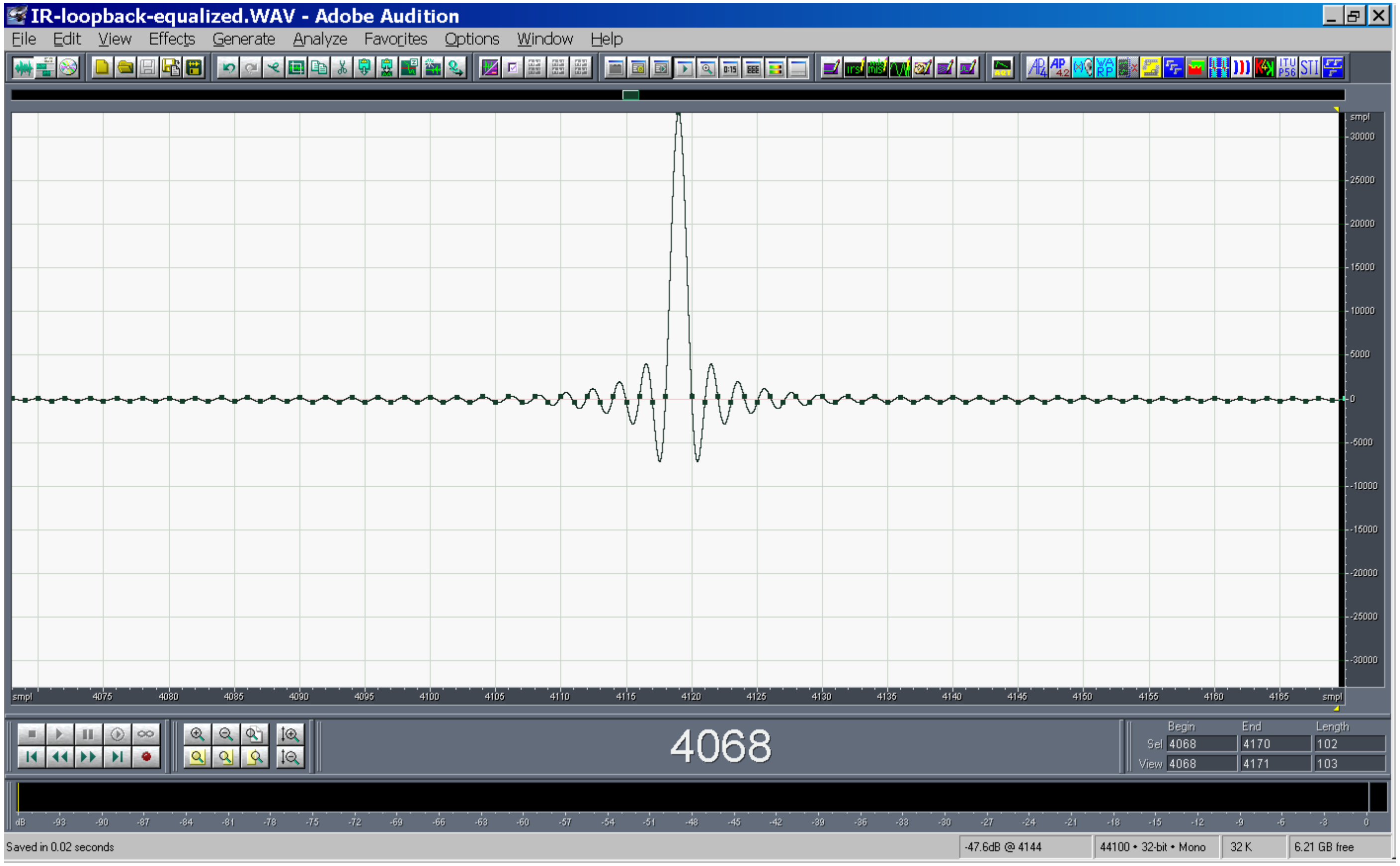
Frequency-dependent regularization parameter



Inverse filter

Pre-ringing at high and low frequency

- **Convolving the time-smeared IR with the Kirkeby compacting filter, a very sharp IR is obtained**



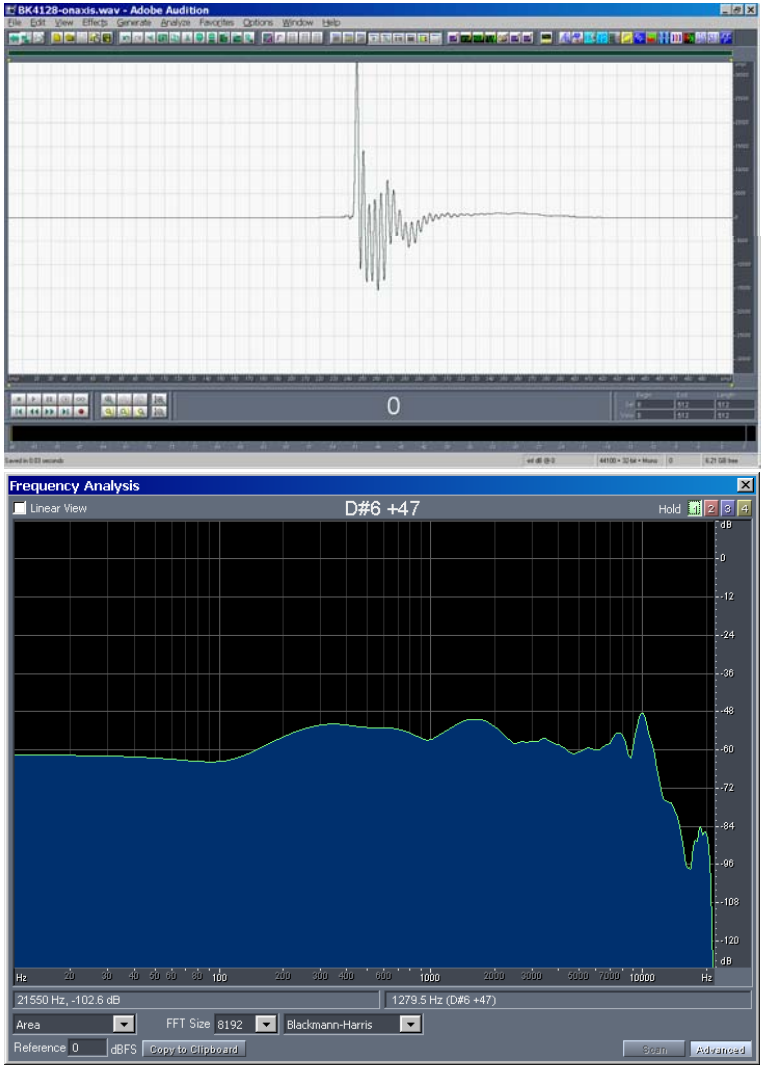
The same method can also be applied for correcting the response of the loudspeaker/microphone system, if an anechoic preliminary test is done

Problems with ESS measurements

- Pre-ringing at high and low frequency before the arrival of the direct sound pulse
- Pre/post equalization of the test signal performed in a way which avoids time-smearing of the impulse response
- Sensitivity to abrupt pulsive noises during the measurement
- Skewing of the measured impulse response when the playback and recording digital clocks are mismatched
- Cancellation of high frequencies in the late part of the tail when performing synchronous averaging

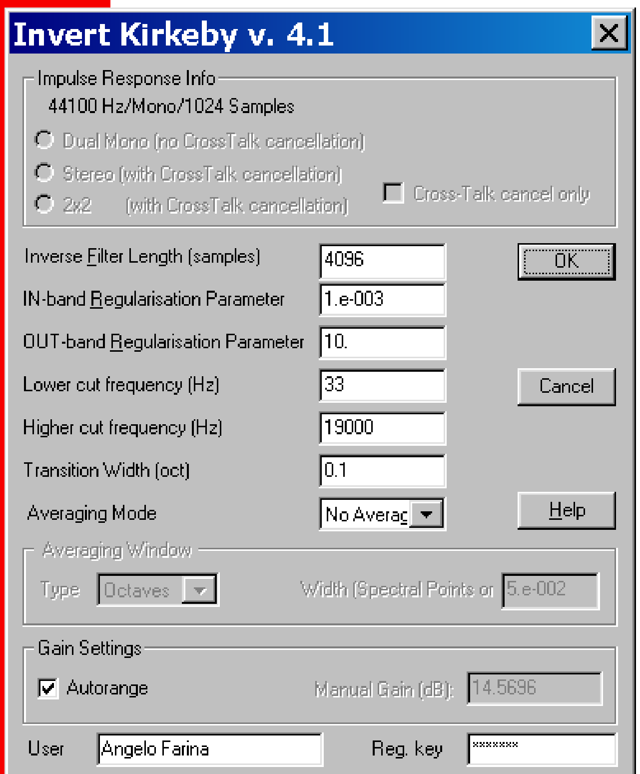
Equalization of the whole system

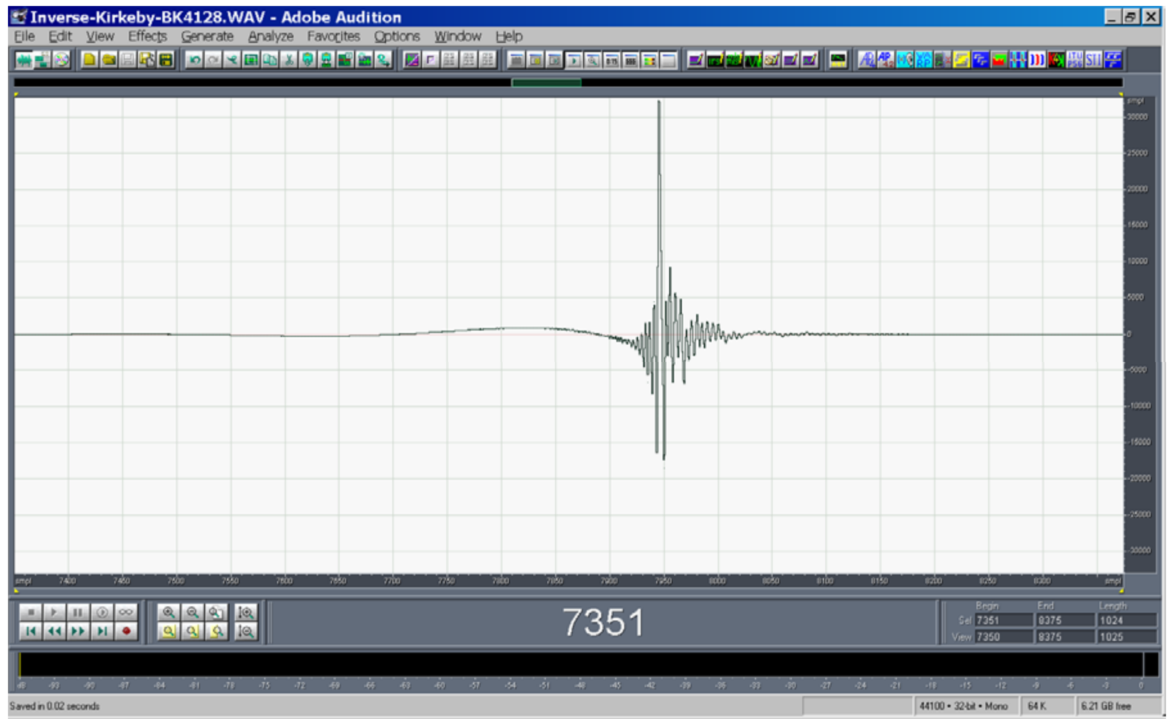
- **An anechoic measurement is first performed**



Equalization of the whole system

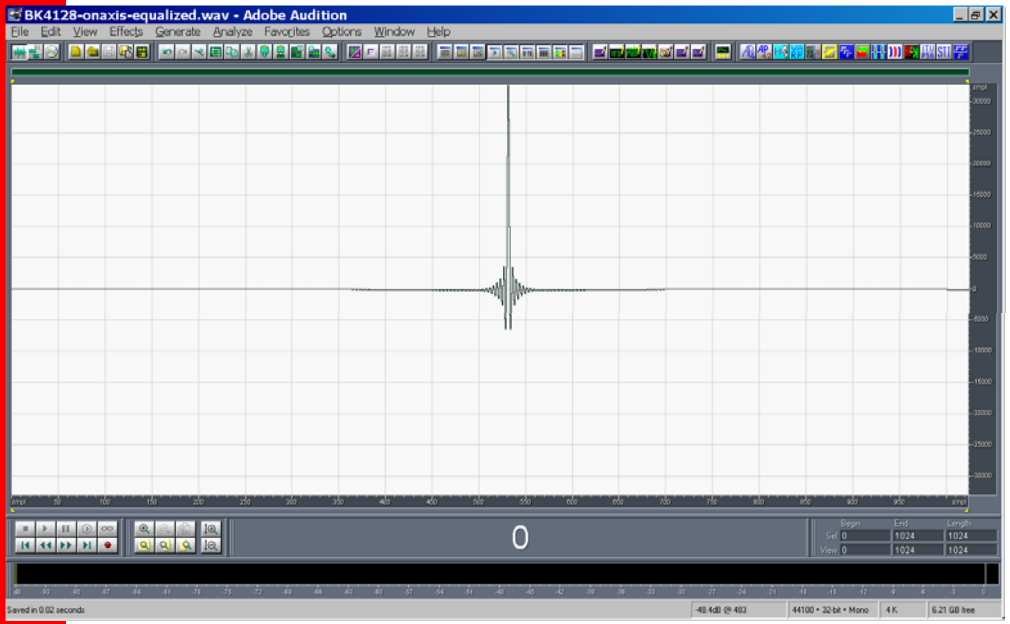
- **A suitable inverse filter is generated with the Kirkeby method by inverting the anechoic measurement**

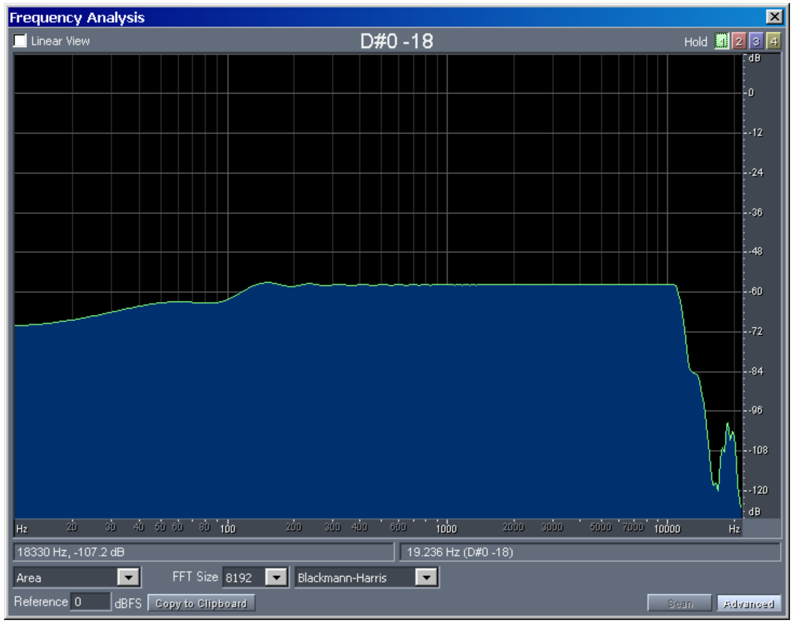




Equalization of the whole system

- The inverse filter can be either pre-convolved with the test signal or post-convolved with the result of the measurement
- Pre-convolution usually reduces the SPL being generated by the loudspeaker, resulting in worst S/N ratio
- On the other hand, post-convolution can make the background noise to become "coloured", and hence more perciptible
- The resulting anechoic IR becomes almost perfectly a Dirac's Delta function:



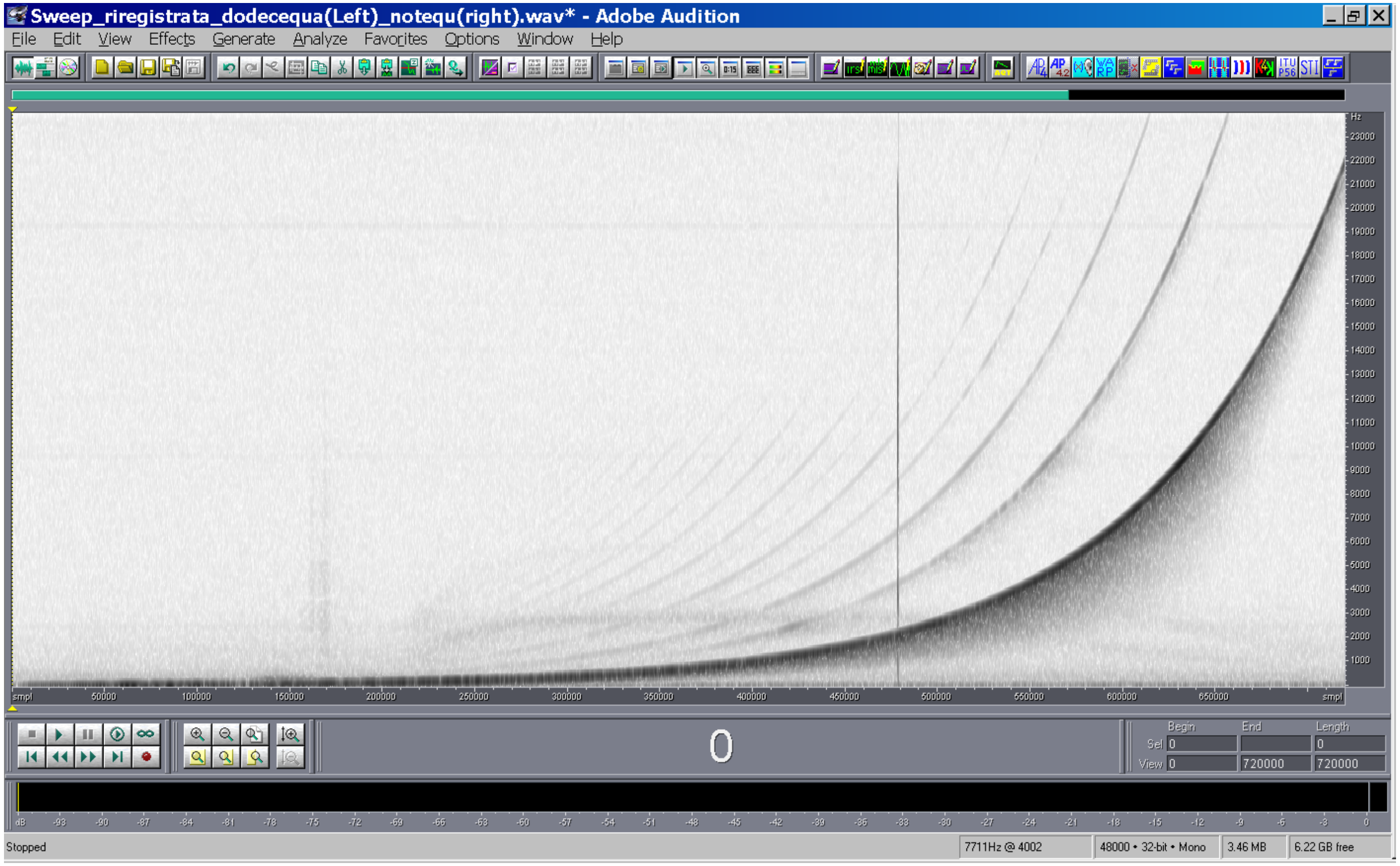


Problems with ESS measurements

- Pre-ringing at high and low frequency before the arrival of the direct sound pulse
- Pre/post equalization of the test signal performed in a way which avoids time-smearing of the impulse response
- Sensitivity to abrupt pulsive noises during the measurement
- Skewing of the measured impulse response when the playback and recording digital clocks are mismatched
- Cancellation of high frequencies in the late part of the tail when performing synchronous averaging

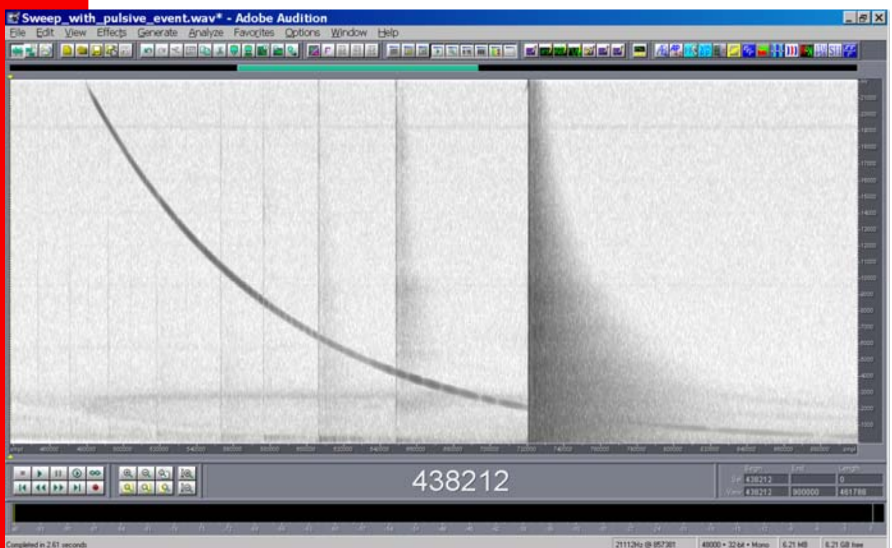
Sensitivity to abrupt pulsive noises

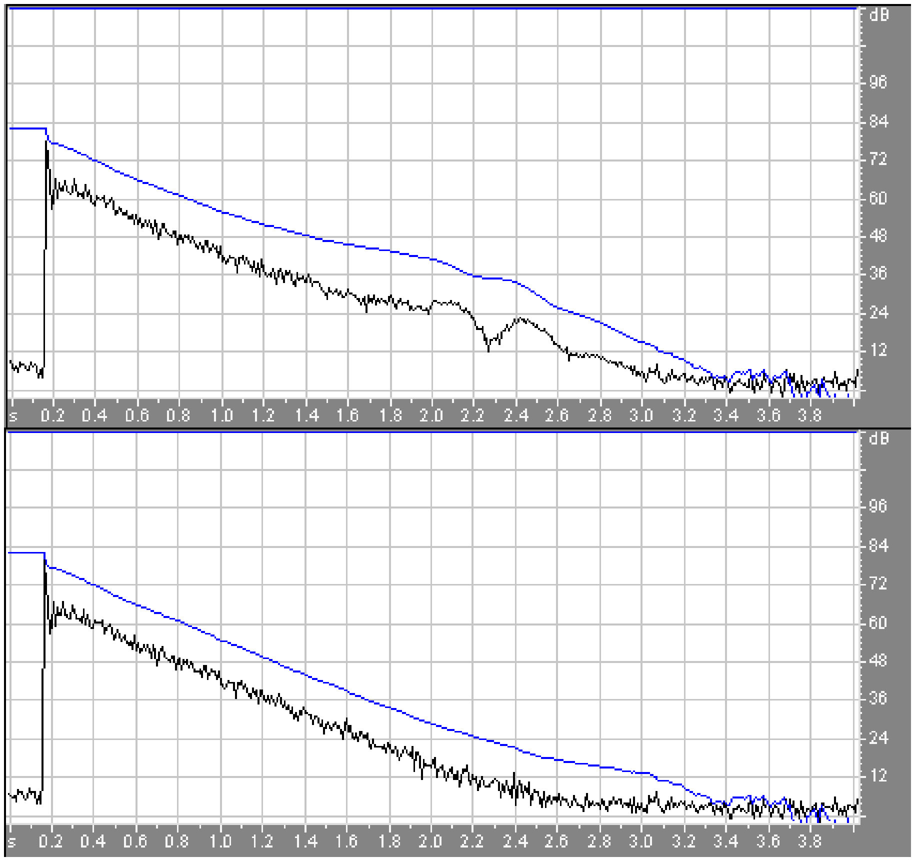
- **Often a pulsive noise occurs during a sine sweep measurement**



Sensitivity to abrupt pulsive noises

- **After deconvolution, the pulsive sound causes untolerable artifacts in the impulse response**

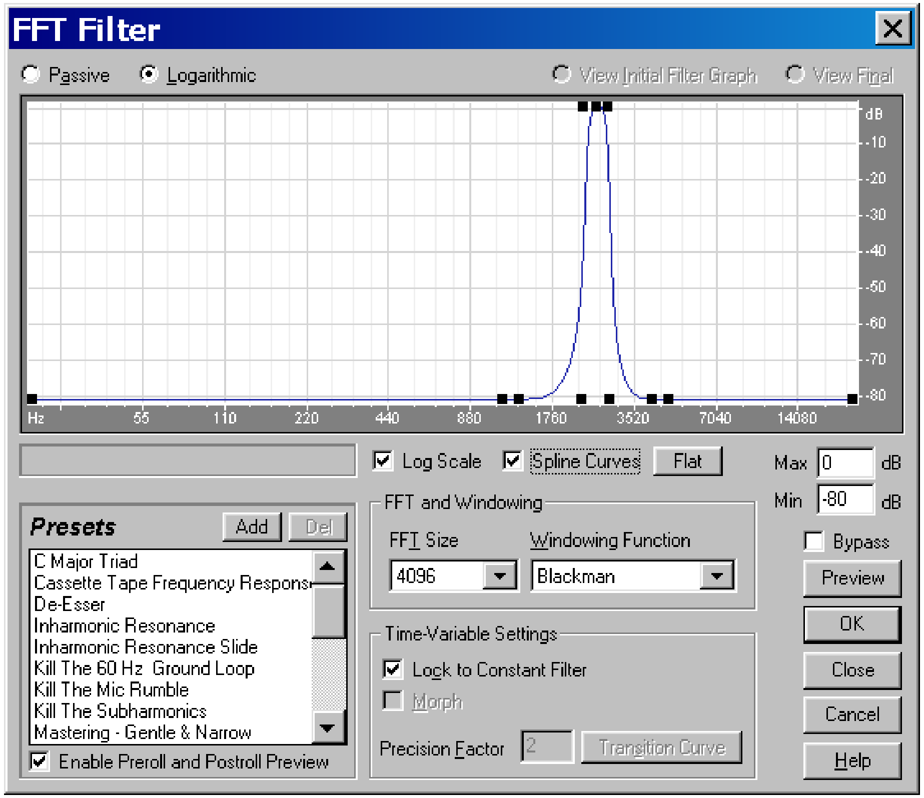


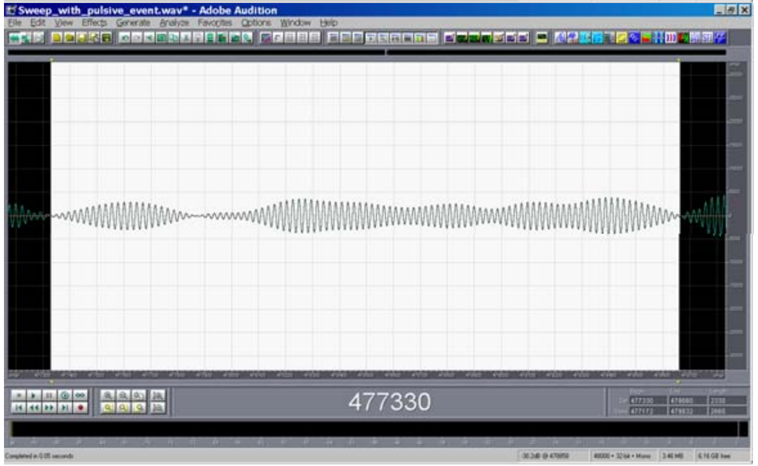


The artifact appears as a down-sloping sweep on the impulse response. At the 2 kHz octave band the decay is distorted, and the reverb. time is artificially increased from 2.13 to 2.48 s

Sensitivity to abrupt pulsive noises

- **Several denoising techniques can be employed:**
  - Brutely silencing the transient noise
  - Employing the specific “click-pop eliminator” plugin of Adobe Audition
  - Applying a narrow-passband filter around the frequency which was being generated in the moment in which the pulsive noise occurred
- **The third approach provides the better results:**



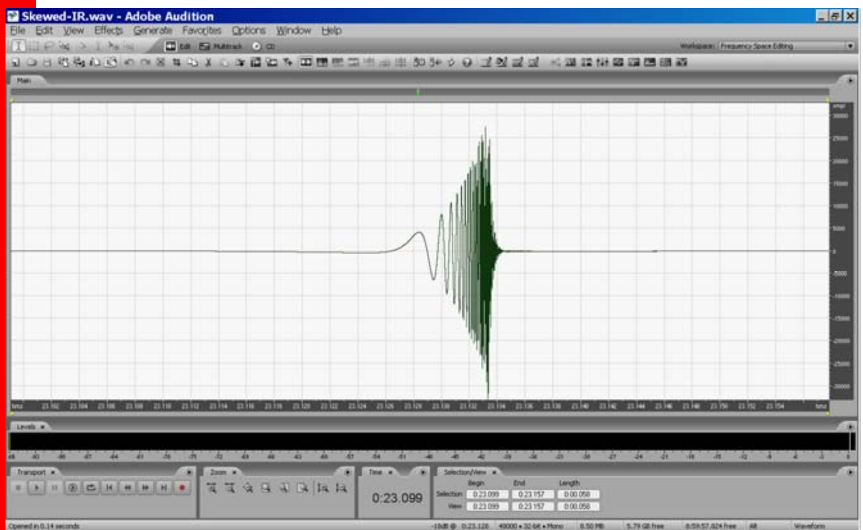


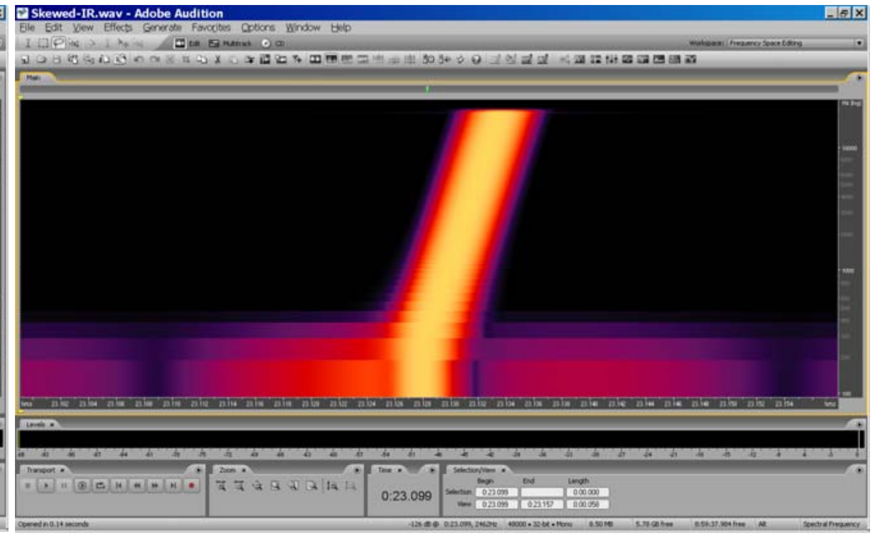
Problems with ESS measurements

- Pre-ringing at high and low frequency before the arrival of the direct sound pulse
- Pre/post equalization of the test signal performed in a way which avoids time-smearing of the impulse response
- Sensitivity to abrupt pulsive noises during the measurement
- Skewing of the measured impulse response when the playback and recording digital clocks are mismatched
- Cancellation of high frequencies in the late part of the tail when performing synchronous averaging

Clock mismatch

- **When the measurement is performed employing devices which exhibit signifcant clock mismatch between playback and recording, the resulting impulse response is “skewed” (stretched in time):**

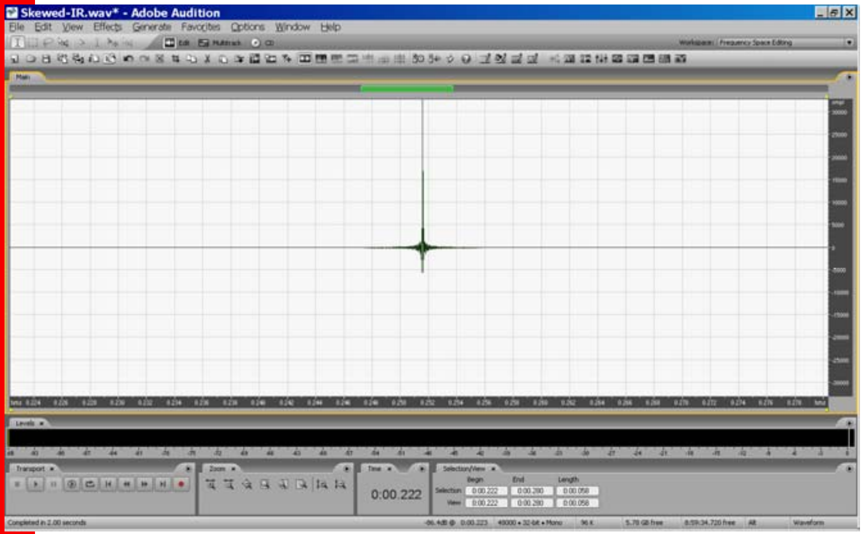


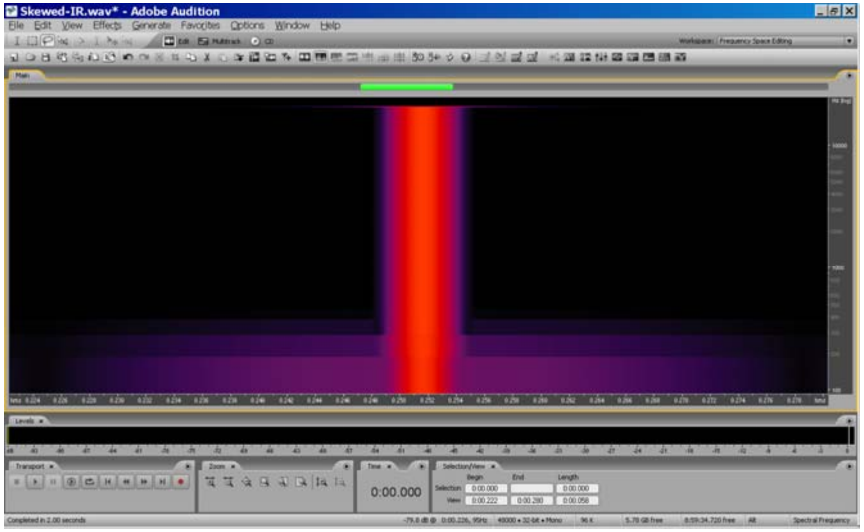


The pictures show the results of an electrical measuremnt performed connecting directly a CD-player with a DAT recorder

Clock mismatch

- **It is possible to re-pack the impulse response employing the already-described approach based on the usage of a Kirkeby inverse filter:**

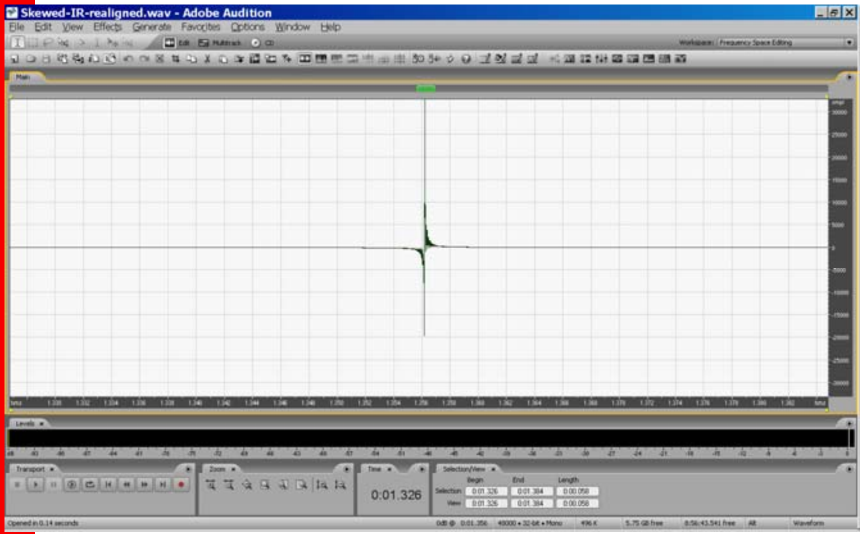


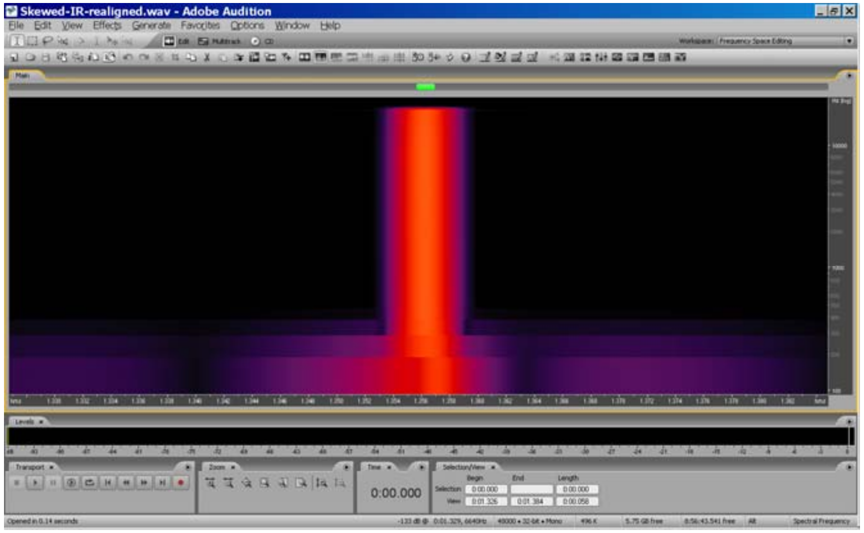


However, this is possible only if a “reference” electrical (or anechoic) measurement has been performed. But, in many cases, one only gets the re-recorded signals, and no reference measurement is available, so the Kirkeby inverse filter cannot be computed.

Clock mismatch

- **However, it is always possible to generate a pre-stretched inverse filter, which is longer or shorter than the "theoretical" one - by proper selection of the lenght of the inverse filter, it is possible to deconvolve impulse responses which are almost perfectly "unskewed":**





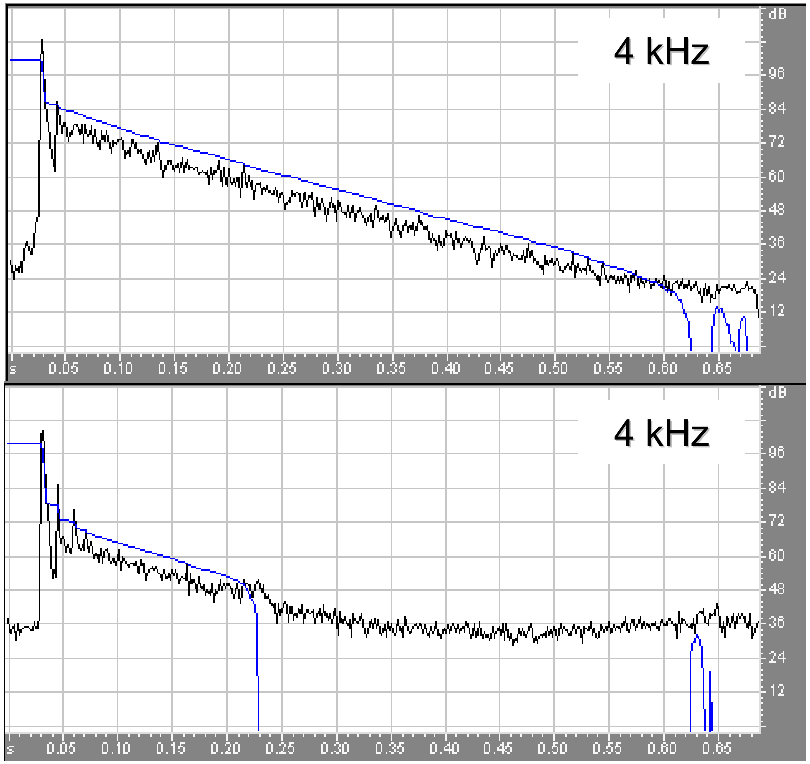
The pictures show the result of the deconvolvution of a clock-mismatched measurement, in which a pre-strecthed inverse filter is employed, 8.5 ms longer than the theoretical one.

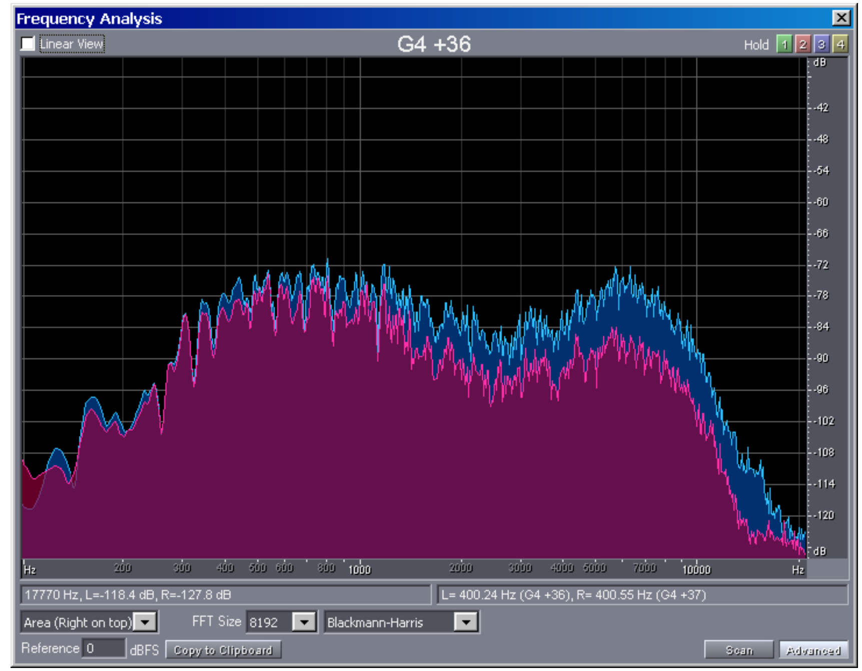
Problems with ESS measurements

- Pre-ringing at high and low frequency before the arrival of the direct sound pulse
- Pre/post equalization of the test signal performed in a way which avoids time-smearing of the impulse response
- Sensitivity to abrupt pulsive noises during the measurement
- Skewing of the measured impulse response when the playback and recording digital clocks are mismatched
- Cancellation of high frequencies in the late part of the tail when performing synchronous averaging

High-frequency cancellation due to averaging

- **When several impulse response measurements are synchronously-averaged for improving the S/N ratio, the late part of the tail cancels out, particularly at high frequency, due to slight time variance of the system**



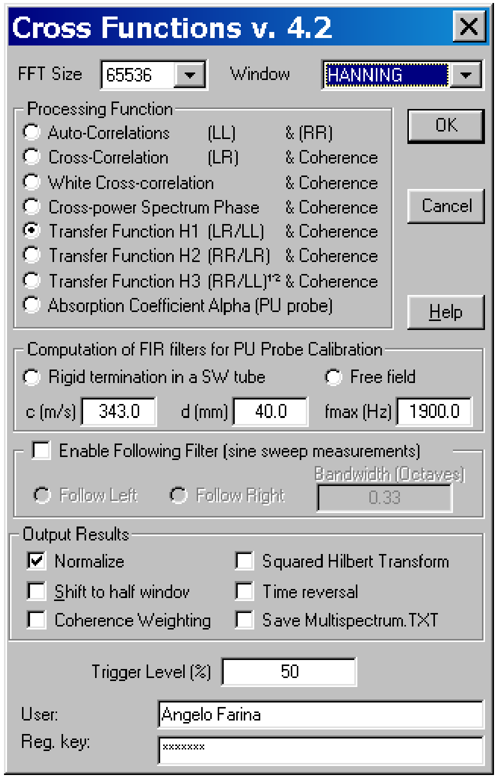


Spectrum of a single sweep of 50s (above) versus 50 sweeps of 1s (below)
short-FFT spectrum at 200 ms after direct sound

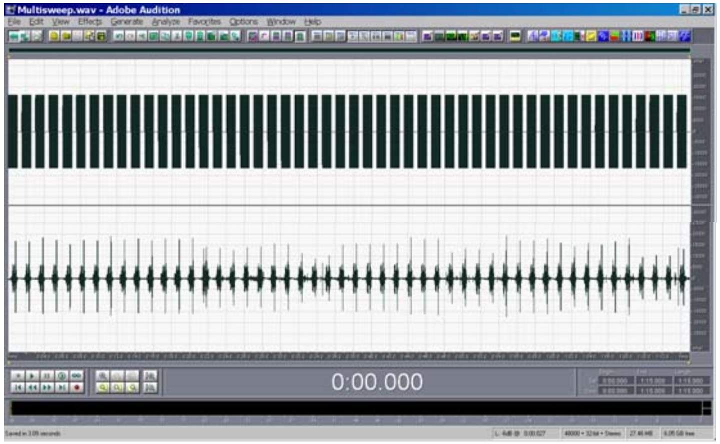
Comparison of a single sweep 50 s long with the synchronous average of 50 sweeps, 1 s long each.

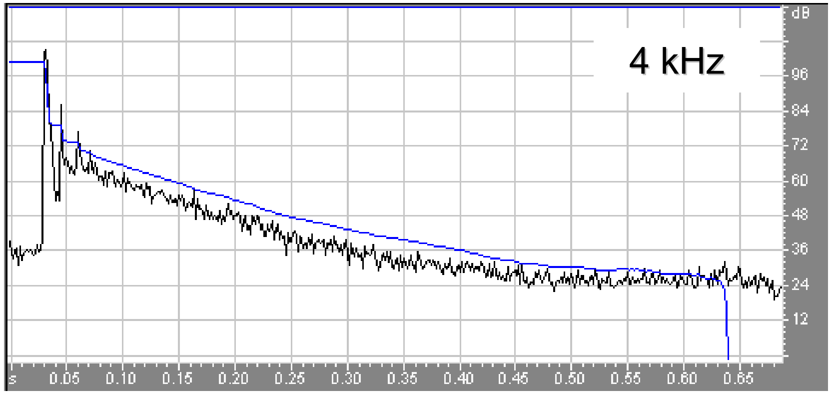
High-frequency cancellation due to averaging

- **However, if averagaing is performed properly in spectral domain, and a single conversion to time domain is performed after averaging, this artifact is significantly reduced**
- **The new “cross Functions” plugin can be used for computing H1:**



$$H_1(f) = \frac{\overline{G_{LR}}}{\overline{\overline{G_{LL}}}}$$





Result of transfer function H1, processing a sequence of 50 sine sweeps (above)

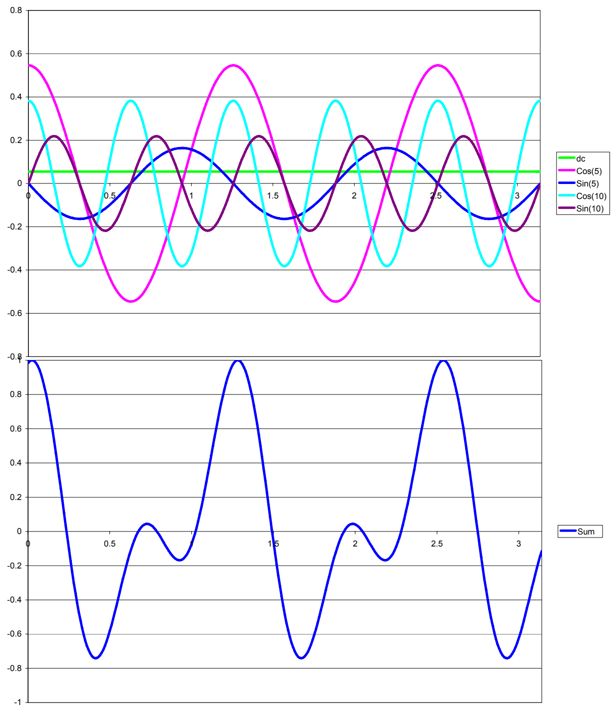
The Future

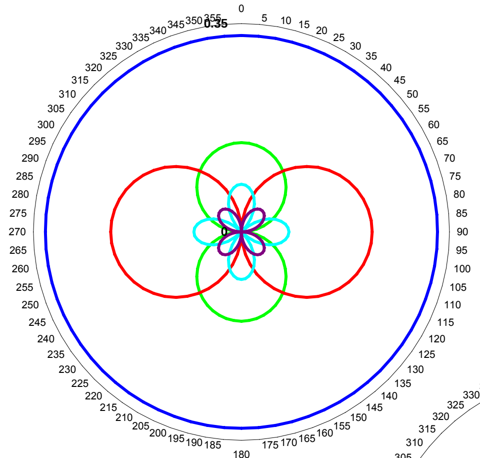
The Future 1 : better spatial information

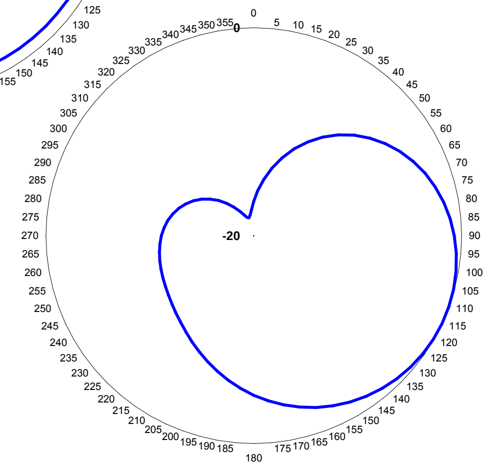
- Microphone arrays capable of synthesizing aribitrary directivity patterns
- Advanced spatial analysis of the sound field employing spherical harmonics (Ambisonics - 1° order or higher)
- Loudspeaker arrays capable of synthesizing arbitrary directivity patterns
- Generalized solution in which both the directivities of the source and of the receiver are represented as a spherical harmonics expansion

How to get better spatial resolution?

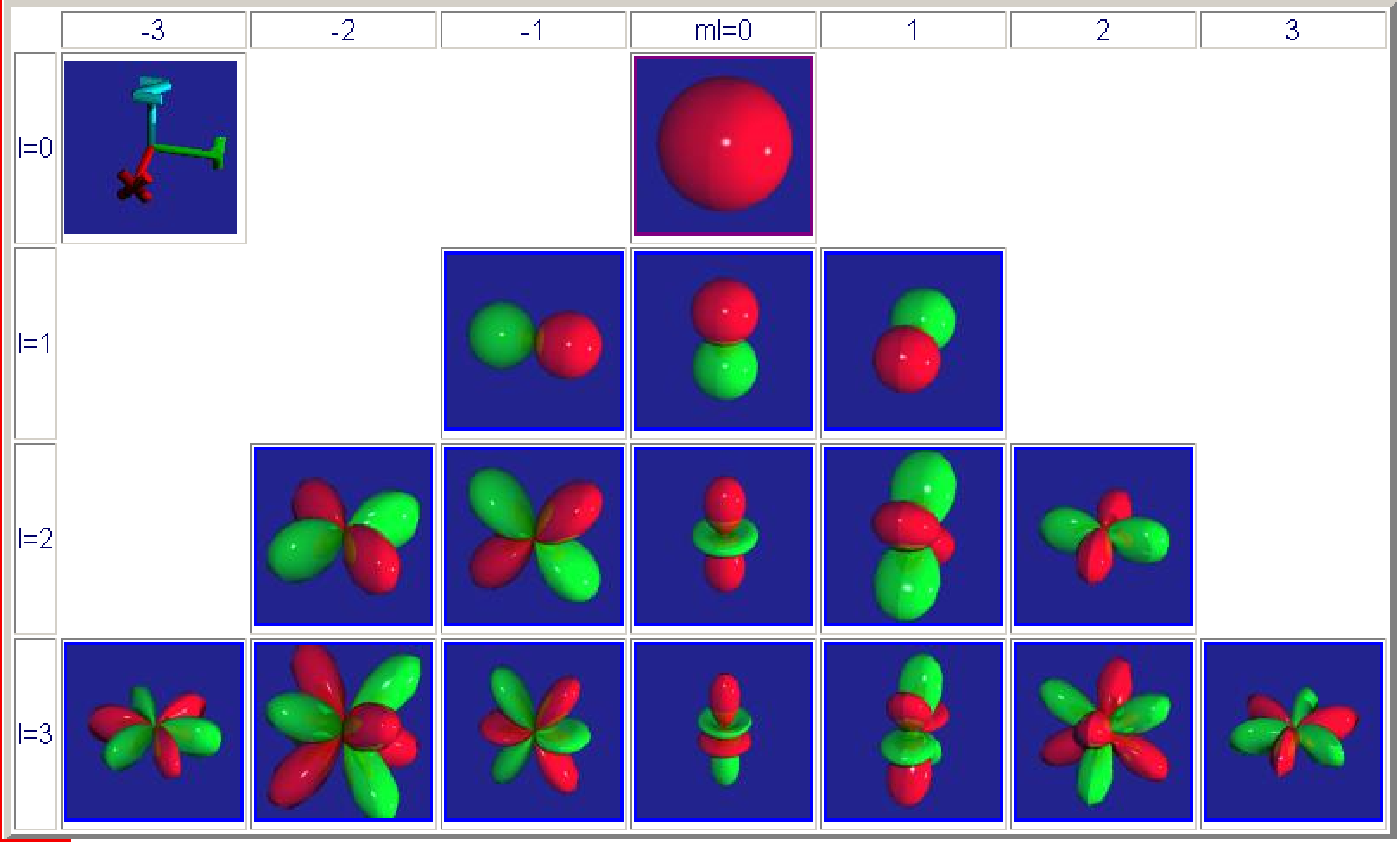
- The answer is simple: analyze the spatial distribution of both source and receiver by means of higher-order spherical harmonics expansion
- Spherical harmonics analysis is the equivalent, in space domain, of the Fourier analysis in time domain
- As a complex time-domain waveform can be though as the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





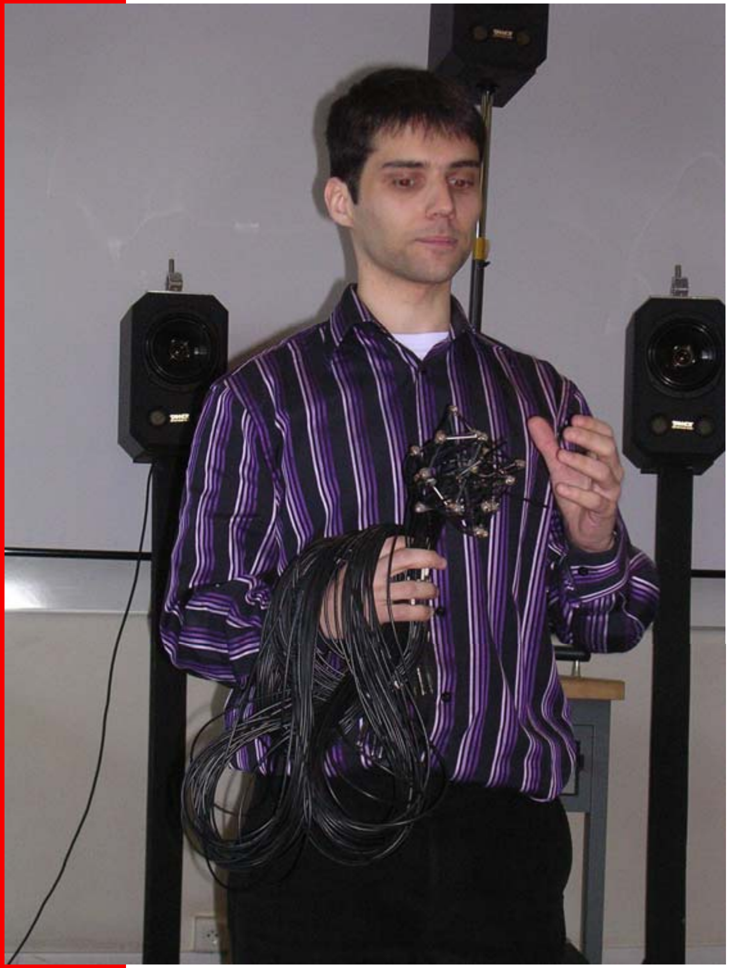


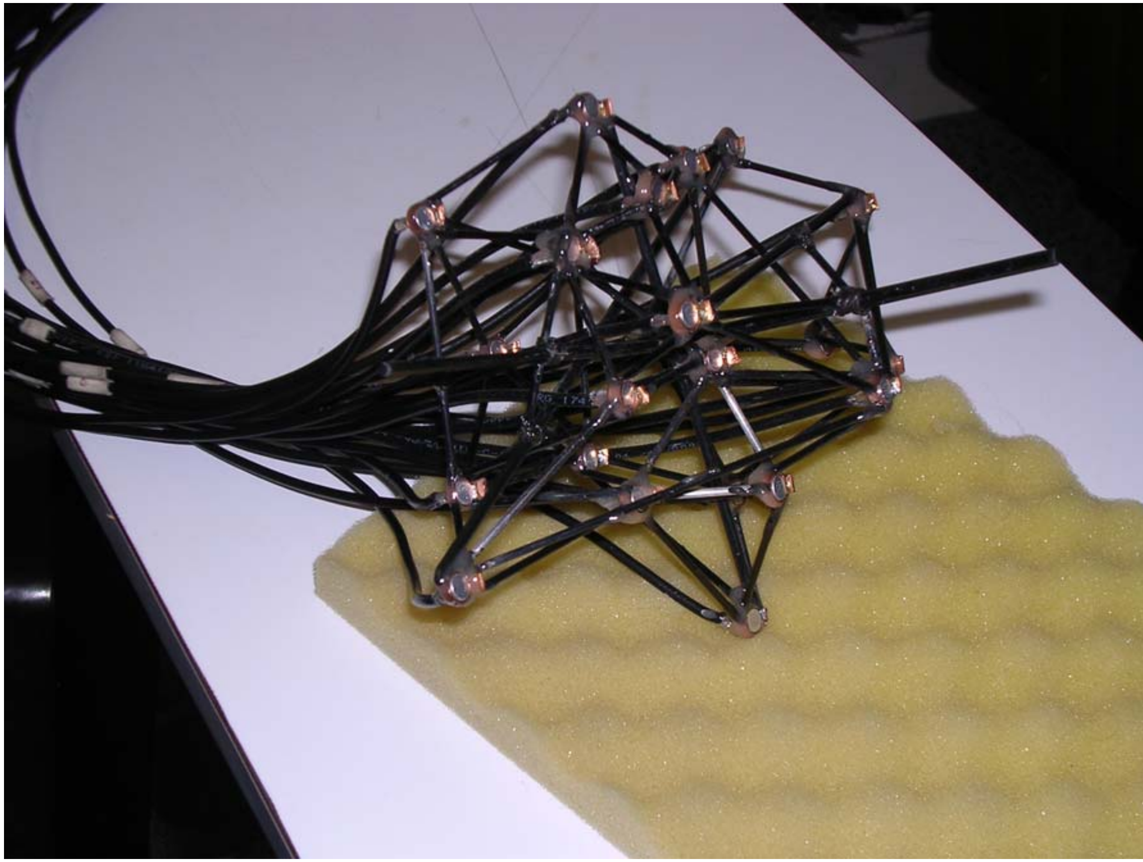
Higher-order spherical harmonics expansion



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)**



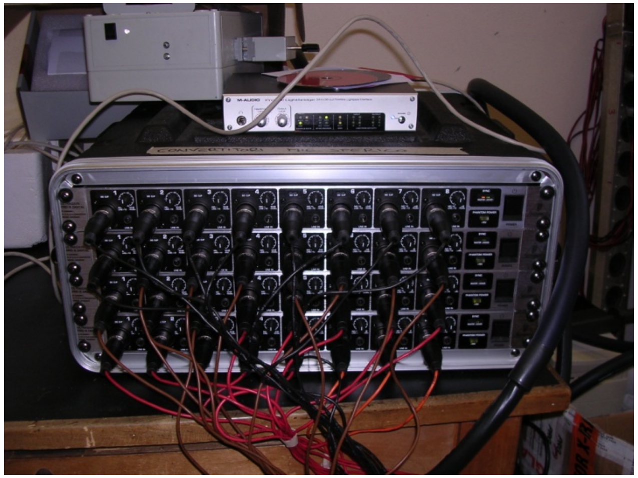


4°-order microphone (France Telecom)

- **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)**

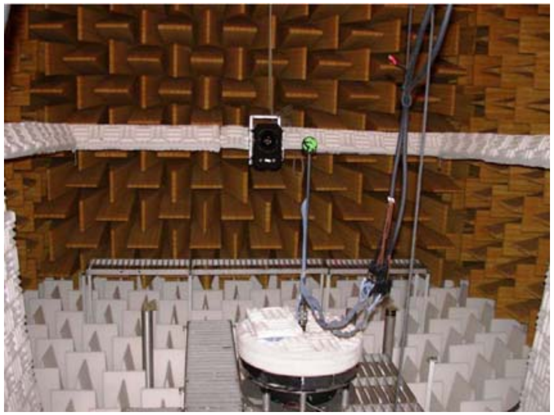
4°-order microphone (University of Parma)

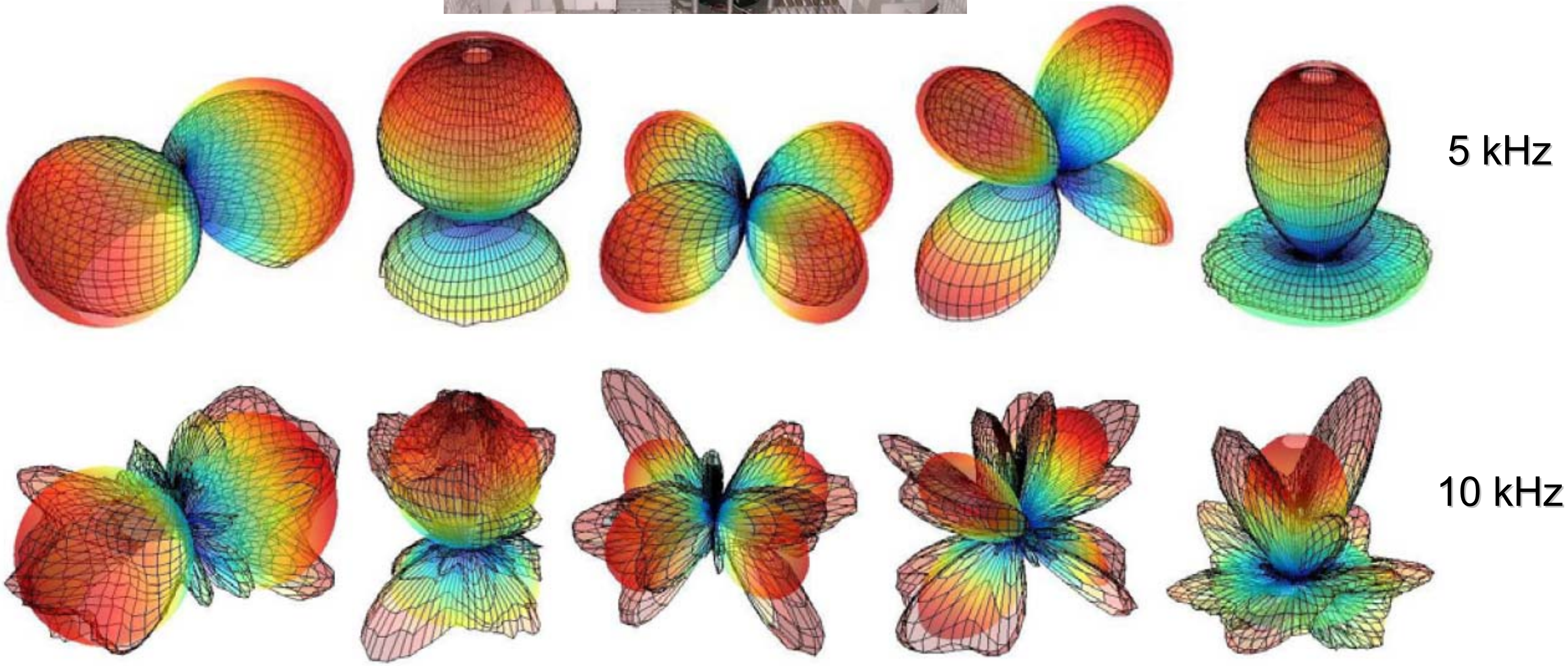
- **A spherical array of 32-capsules connected with a portable A/D conversion system**



Verification of high-order patterns

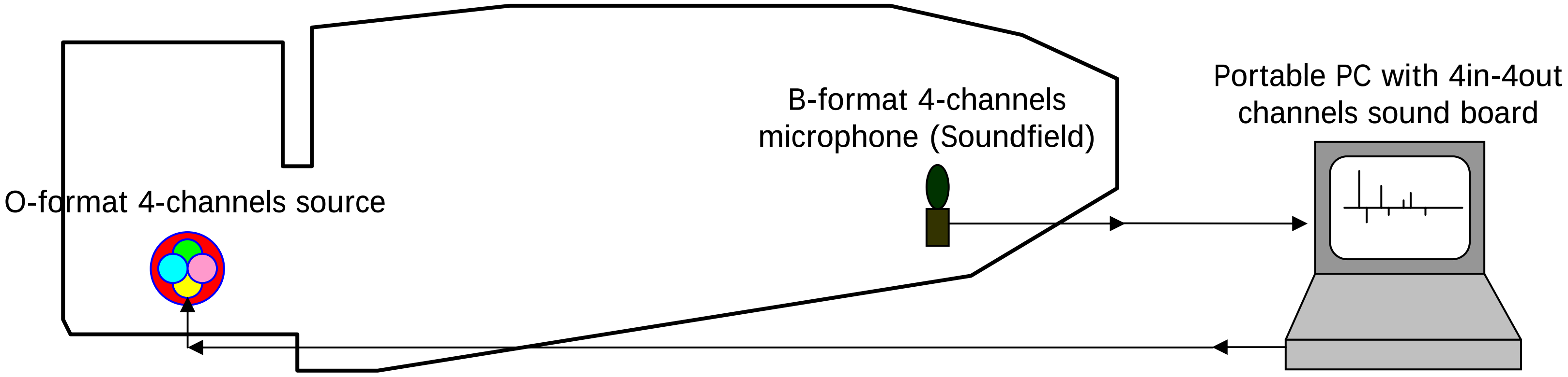
- **Sebastien Moreau and Olivier Warusfel verified the directivity patterns of the 4°-order microphone array in the anechoic room of IRCAM (Paris)**





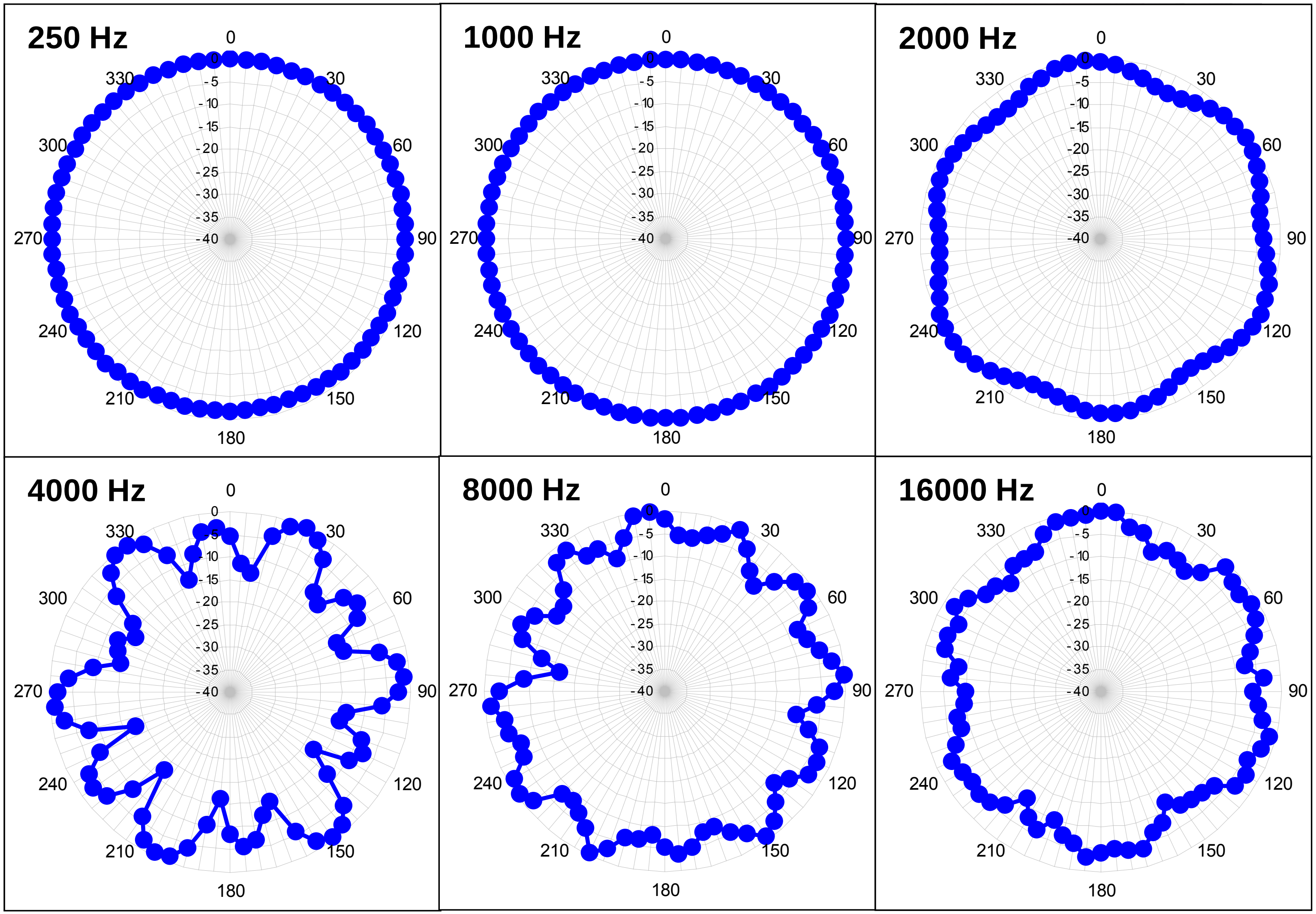
What about source directivity ?

- 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:



Directivity of transducers

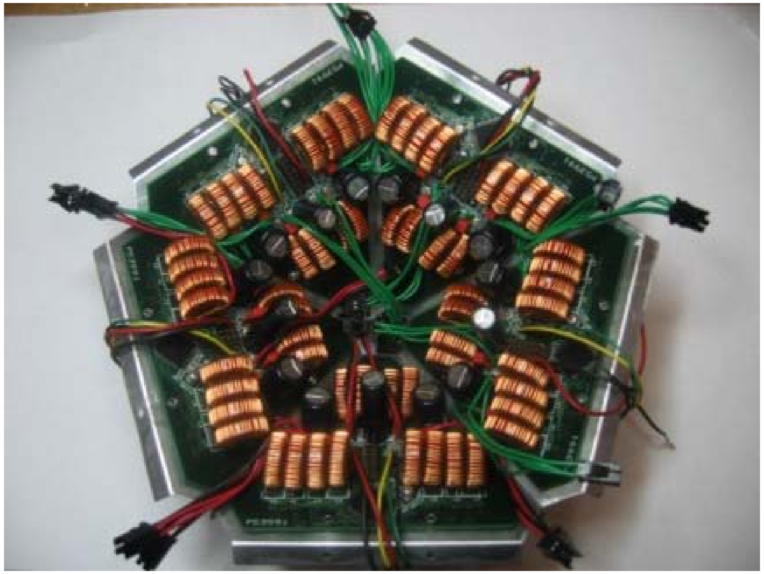
LookLine D200 dodechaedron



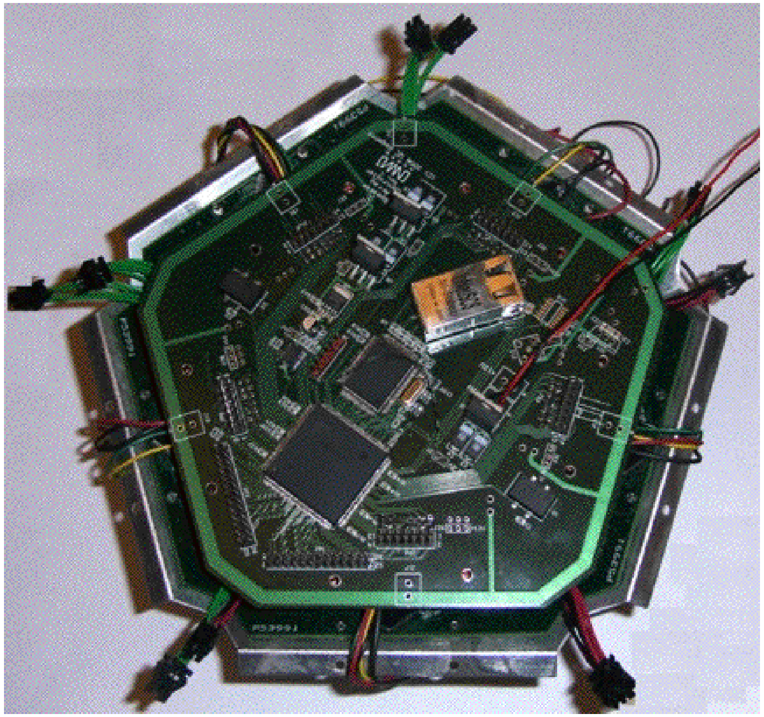
High-order sound source

- **Adrian Freed, Peter Kassakian, and David Wessel (CNMAT) developed a new 120-loudspeakers, digitally controlled sound source, capable of synthesizing sound emission according to spherical harmonics patterns up to 5° order.**

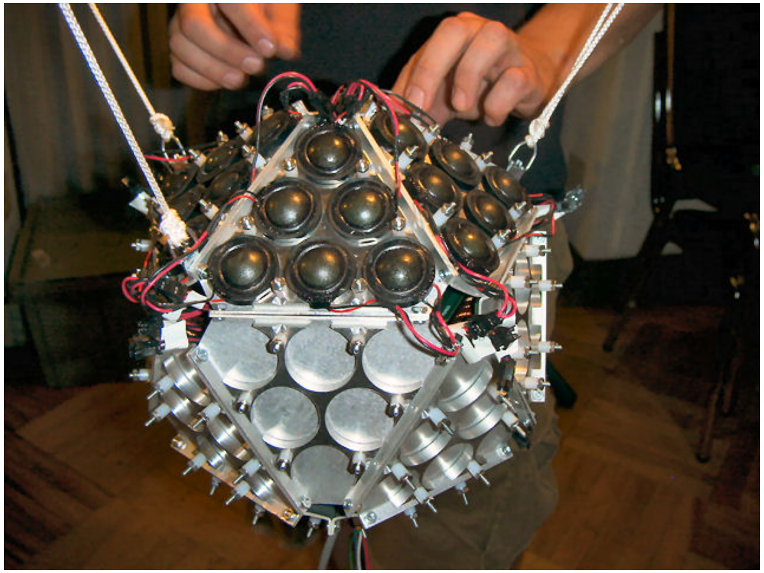
Technical details of high-order source



- Class-D embedded amplifiers



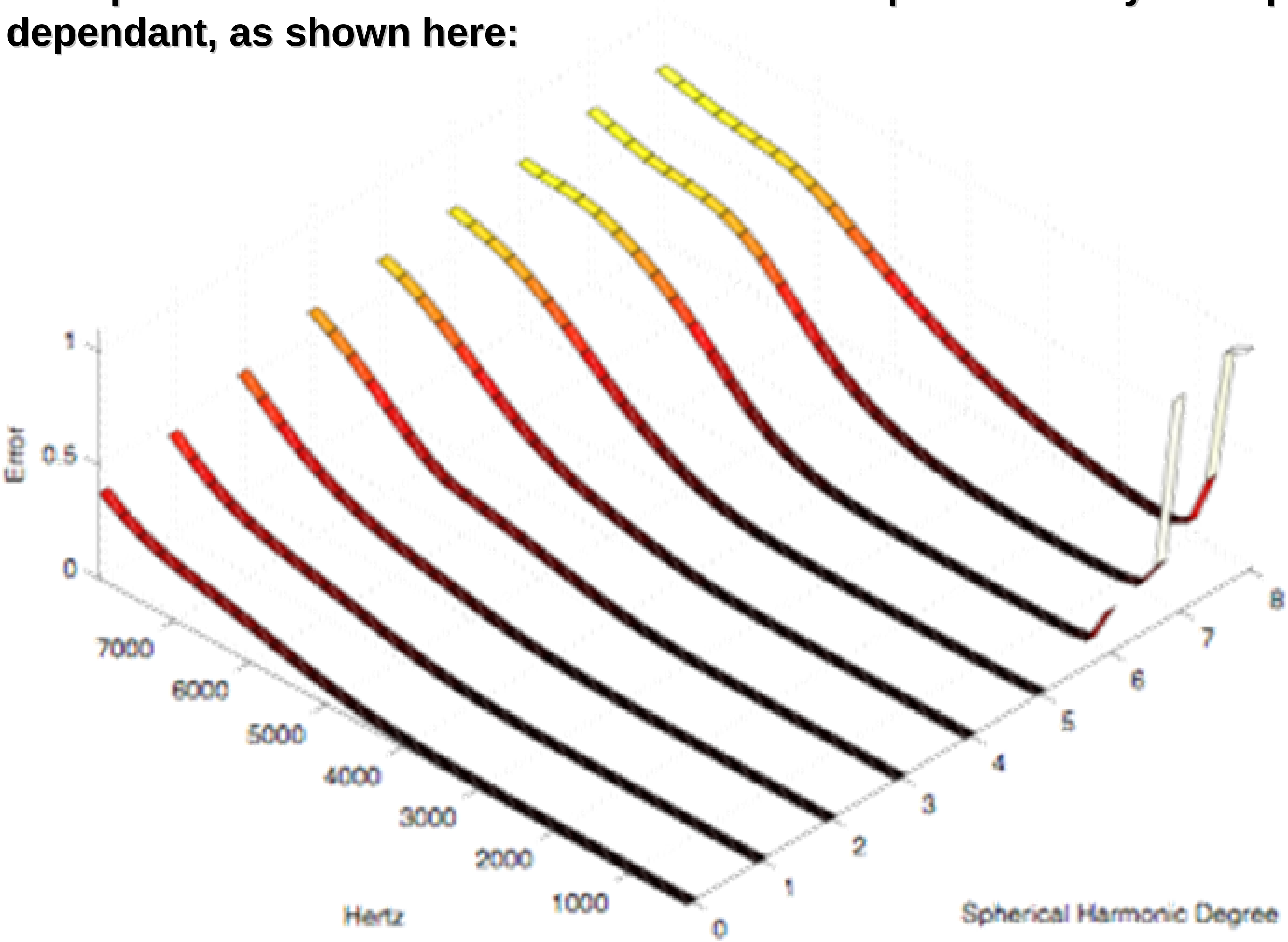
- Embedded ethernet interface and DSP processing



- Long-excursion special Meyer Sound drivers

Accuracy of spatial synthesis

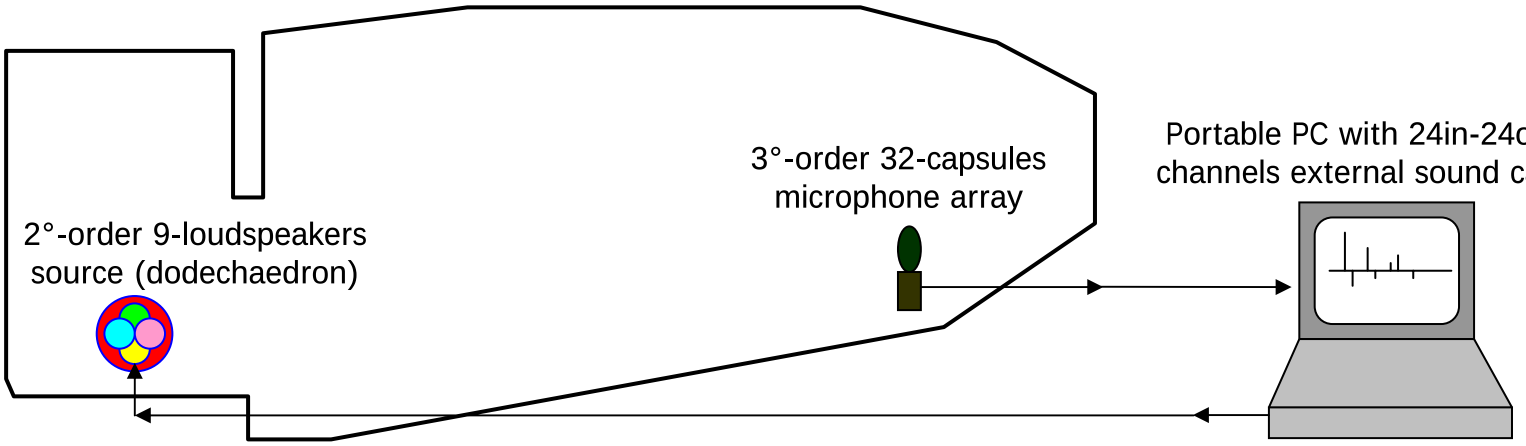
- **The spatial reconstruction error of a 120-loudspeakers array is frequency dependant, as shown here:**



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

Complete high-order MIMO method

- **Employing massive arrays of transducers, it will be 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.**



The Future 2 : not linear systems

- Often impulse responses are measured for being employed in auralization systems (i.e. Waves)
- Linear convolution is employed for this
- This method indeed does not sound realistic, as it removes any not-linear effect
- We can now exploy the results of an ESS measurement for performing a not-linear convolution
- For this, indeed, the measured “harmonic orders IRs” have to be transformed into corresponding Volterra kernels

Theory of nonlinear convolution

- The basic approach is to convolve separately, and then add the result, the linear IR, the second order IR, the third order IR, and so on.
- Each order IR is convolved with the input signal raised at the corresponding power:

$$y(n) = \sum_{i=0}^{M-1} h_1(i) \cdot x(n-i) + \sum_{i=0}^{M-1} h_2(i) \cdot x^2(n-i) + \sum_{i=0}^{M-1} h_3(i) \cdot x^3(n-i) + ...$$

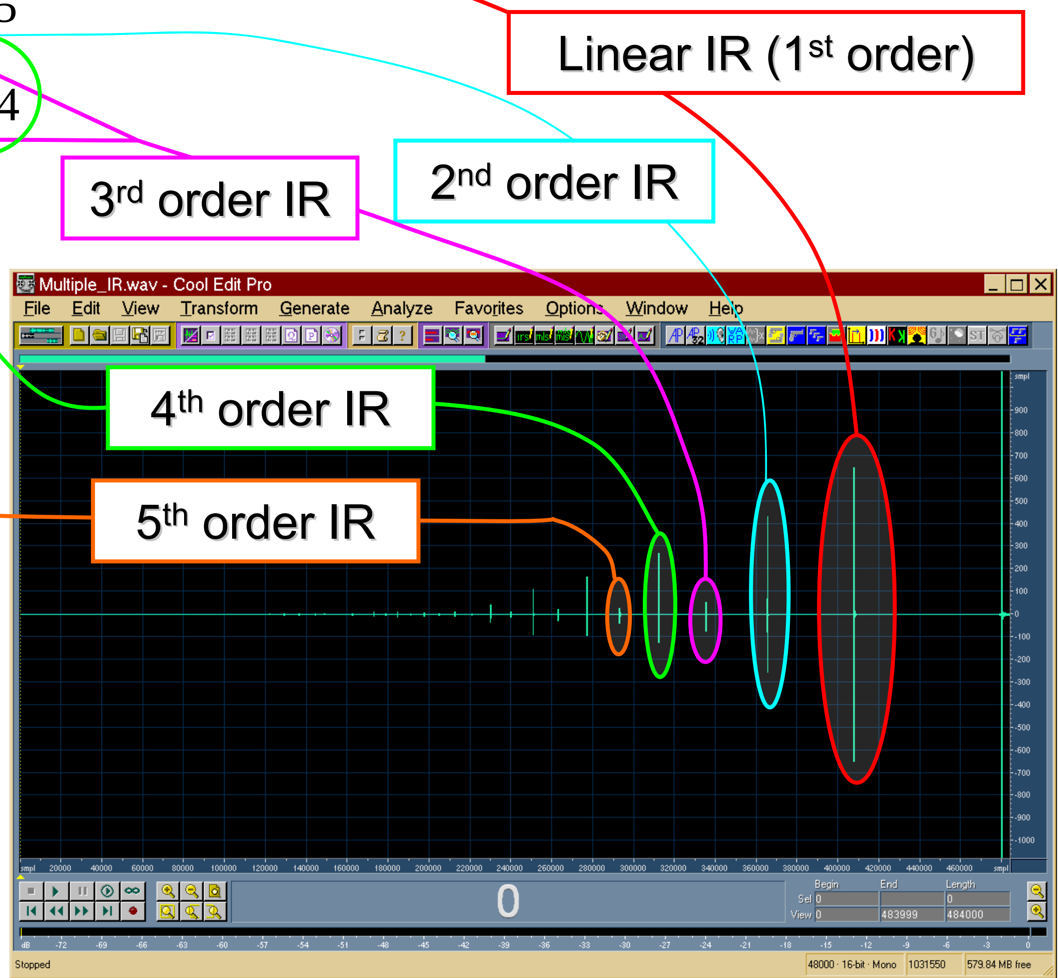
The problem is that the required multiple IRs are not the results of the measurements: they are instead the diagonal terms of Volterra kernels

From measured IRS to Volterra Kernels

- **A simple linear system allows for computation of Volterra Kernels starting from the measured “harmonic orders” IRs**

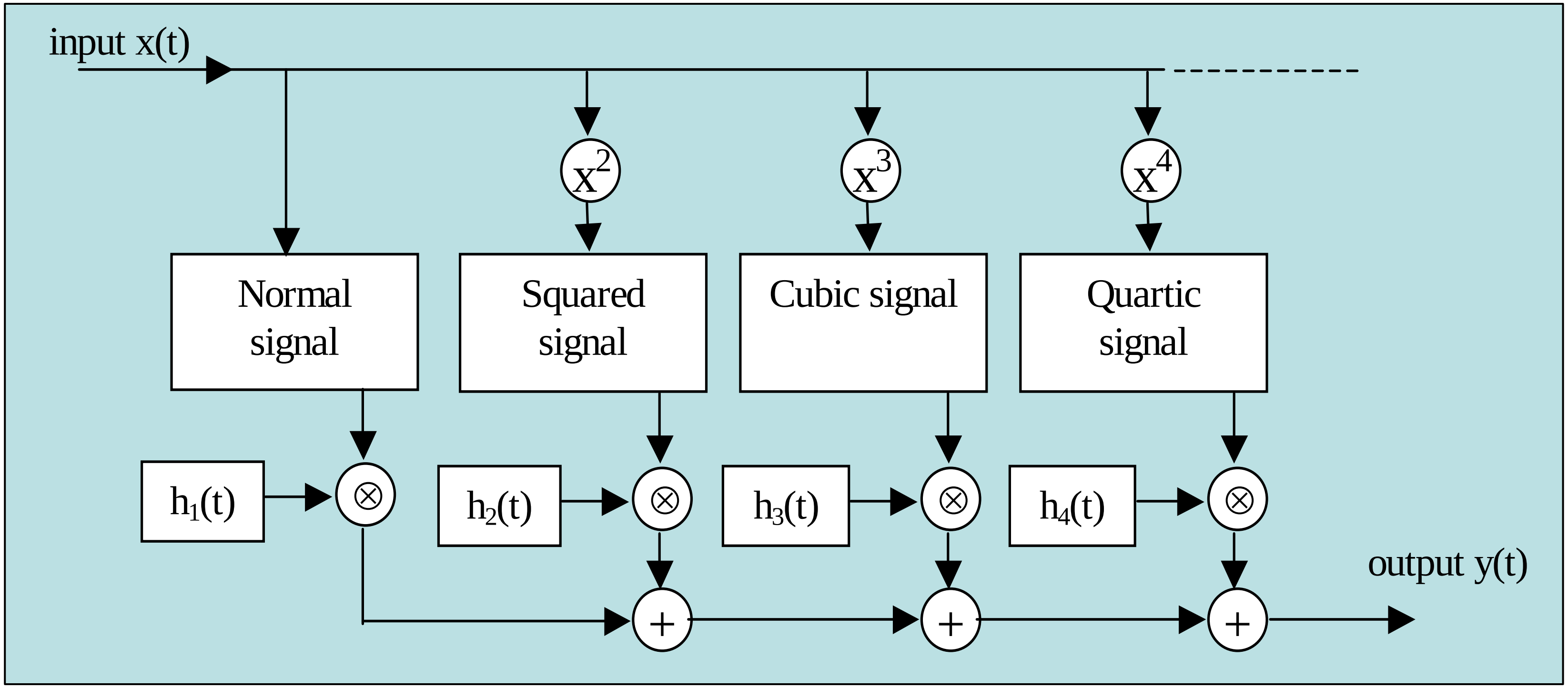
$$
\begin{cases}
\overline{H_1} = \overline{H'_1} + 3\cdot\overline{H'_3} + 5\cdot\overline{H'_5} \\
\overline{H_2} = 2\cdot j\cdot\overline{H'_2} + 8\cdot j\cdot\overline{H'_4} \\
\overline{H_3} = -4\cdot\overline{H'_3} - 20\cdot\overline{H'_5} \\
\overline{H_4} = -8\cdot j\cdot\overline{H'_4} \\
\overline{H_5} = 16\cdot\overline{H'_5}
\end{cases}
$$

Volterra Kernels



Efficient non-linear convolution

As we have got the Volterra kernels already in frequency domain, we can efficiently use them in a multiple convolution algorithm implemented by overlap-and-save of the partitioned input signal:



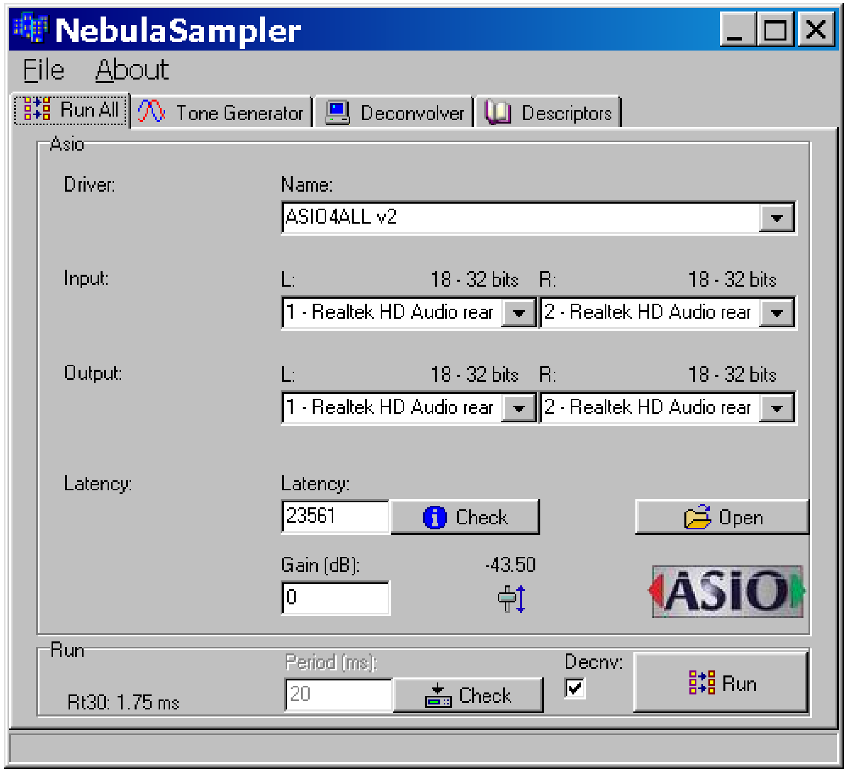
Software implementation

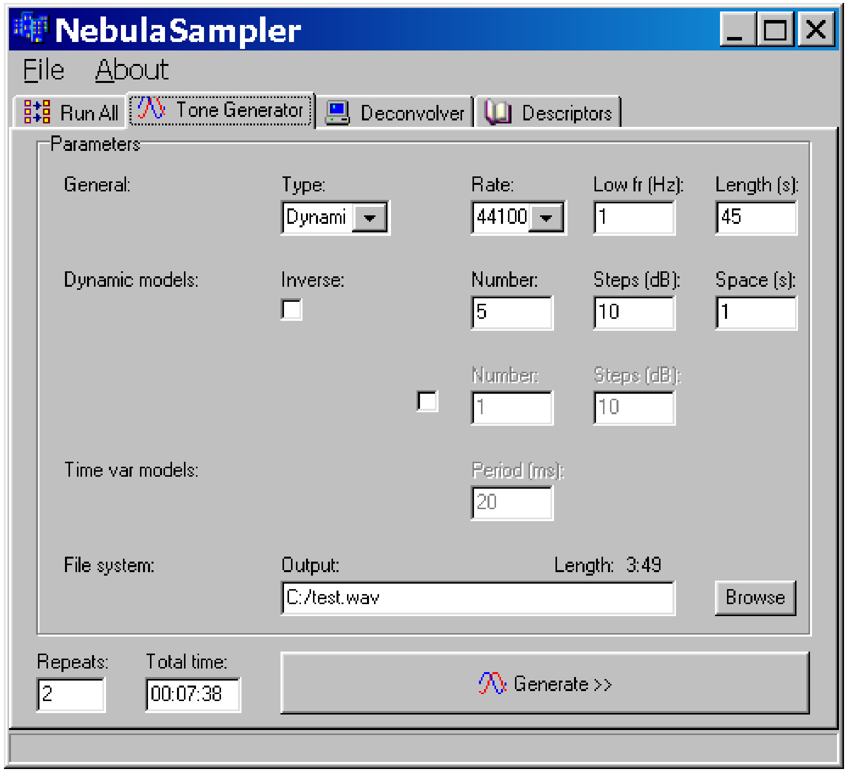
A small Italian startup company, Acustica Audio, developed a VST plugin based on the Diagonal Volterra Kernel method, named Nebula

This is capable of real-time operation even with a very large number of filter coefficients

Software implementation

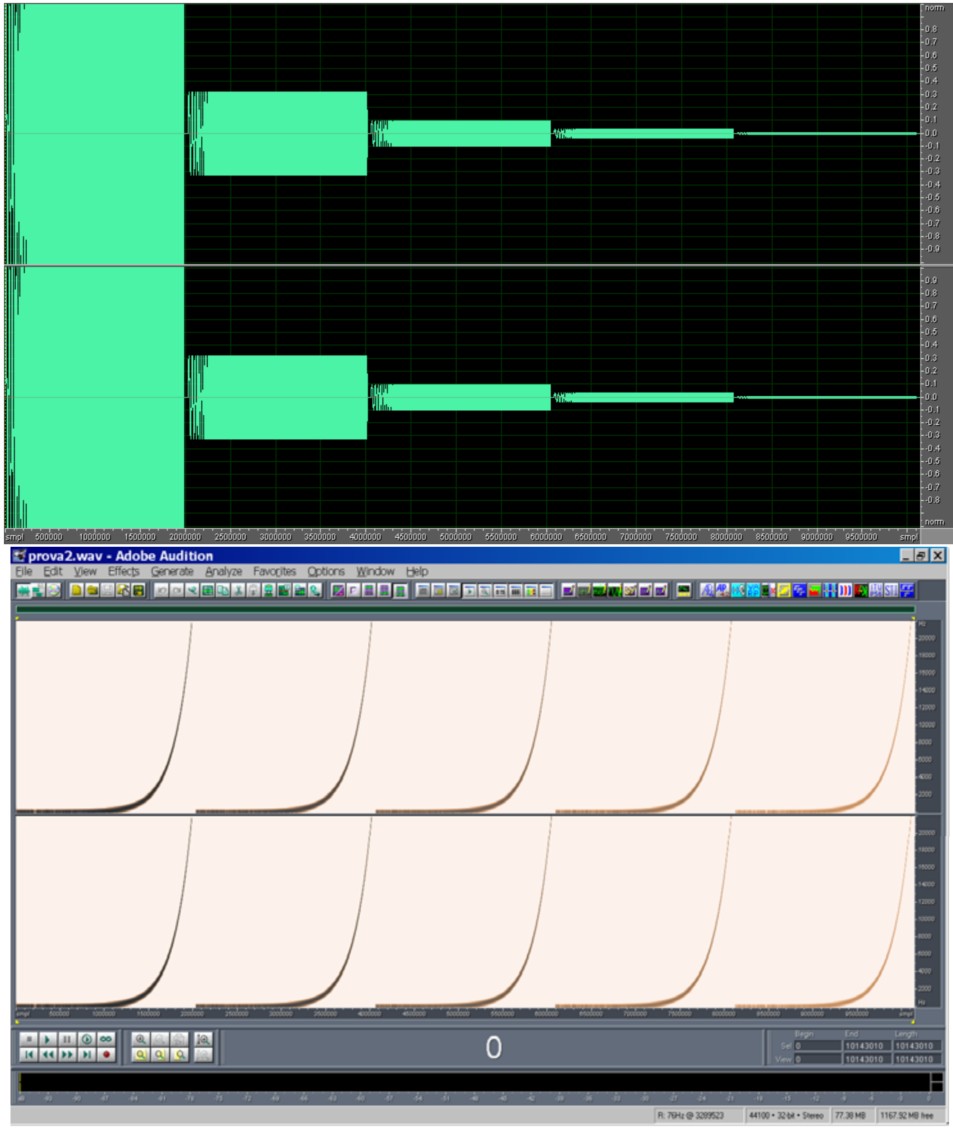
Nebula is also equipped with a companion application, Nebula Sampler, designed for automatizing the measurement of a not linear system with the Exponential Sine Sweep method:





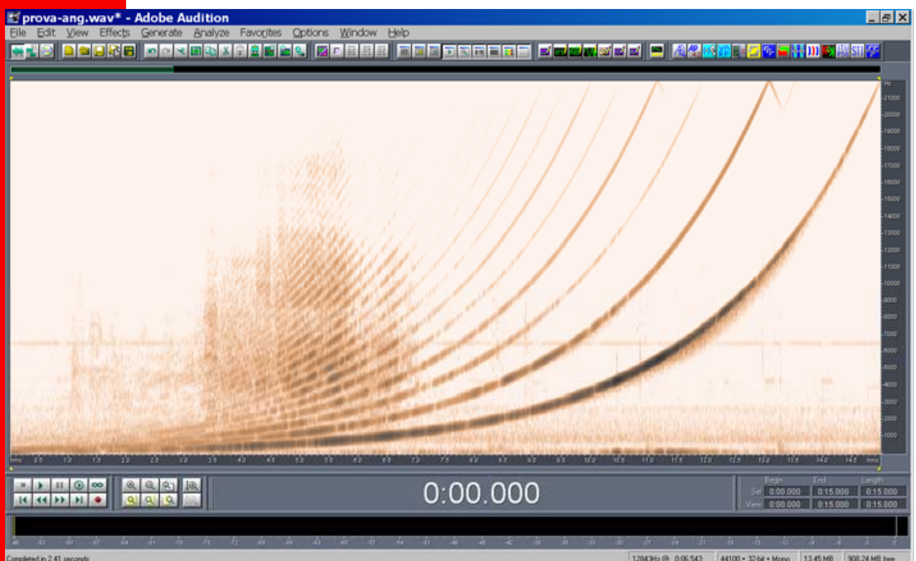
Time-variant systems

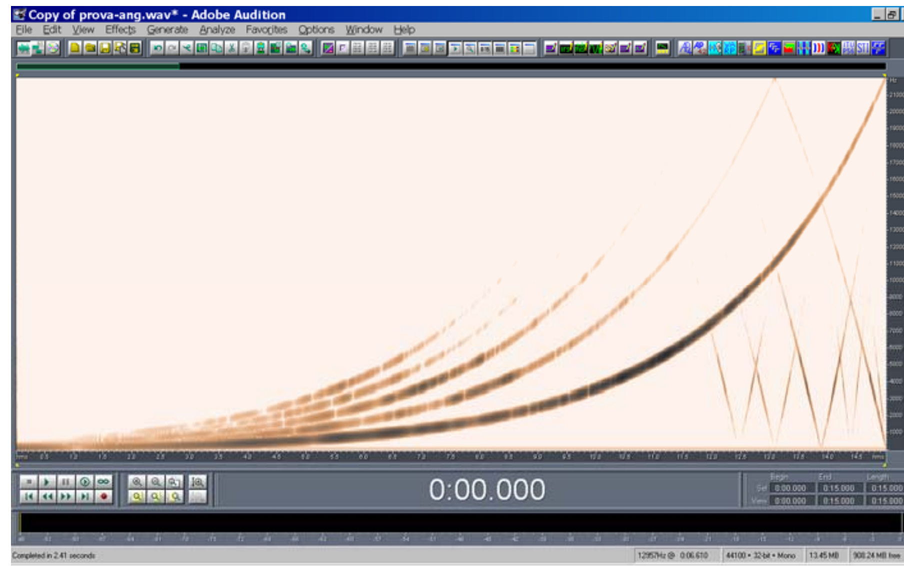
Nebula can sample also time-variant systems, such as flangers or compressors, by repeating the sine sweep measurement several times, along a repetition cycle or changing the signal amplitude



Reconstruction accuracy

Nebula is actually limited to Volterra kernels up to 5th order, and consequently does not emulates high-frequency harmonics:



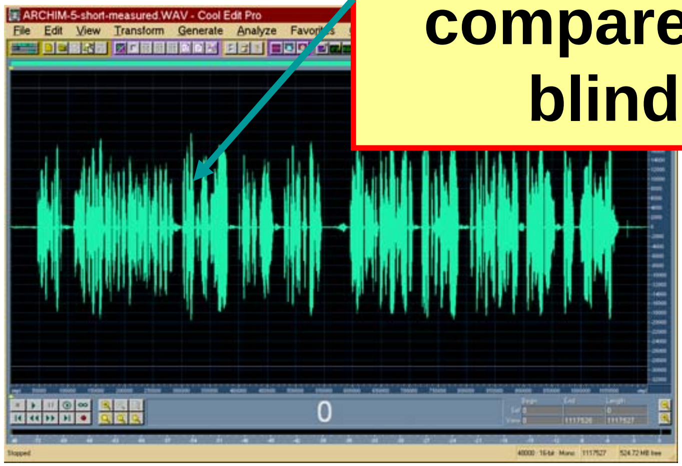


Audible evaluation of the performance

Original signal

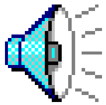
Linear convolution

These last two were compared in a formalized blind listening test



Live recording

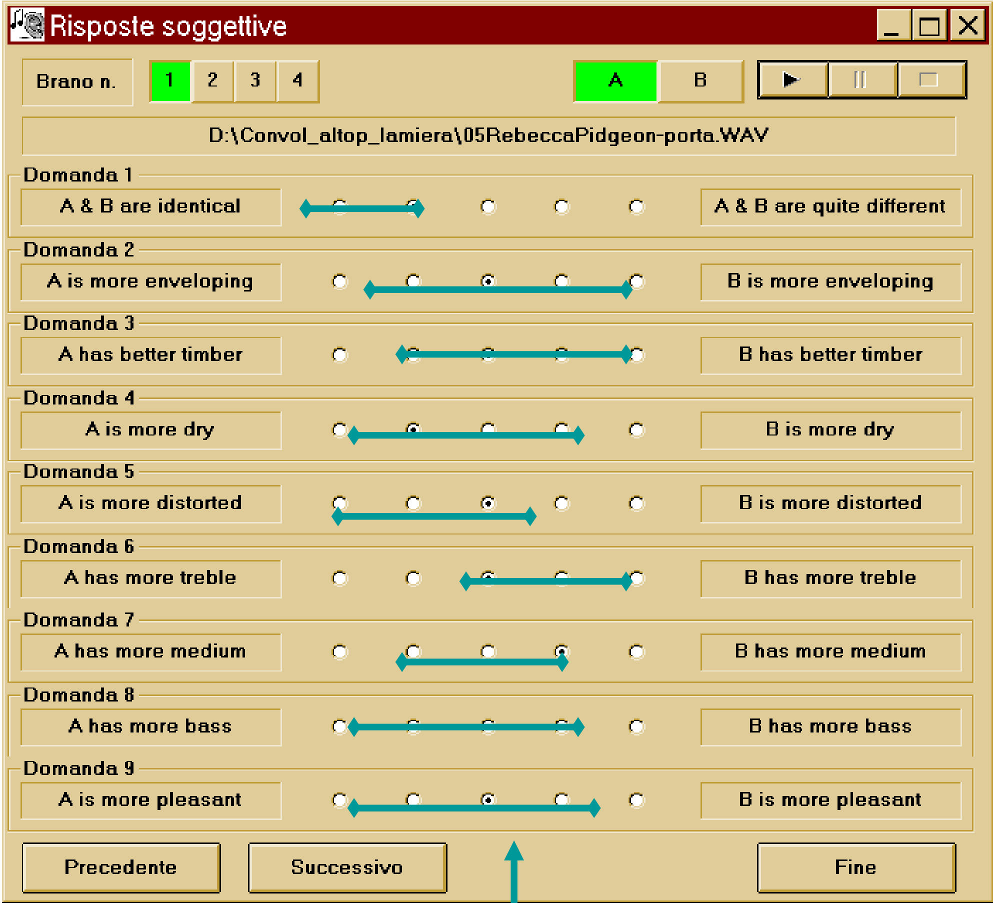
Non-linear
Diagonal Volterra Kernel



Stop

Subjective listening test

- A/B comparison
- Live recording & non-linear auralization
- 12 selected subjects
- 4 music samples
- 9 questions
- 5-dots horizontal scale
- Simple statistical analysis of the results
- A was the live recording, B was the auralization, but the listener did not know this



95% confidence intervals of the answers

Results

Statistical parameters – more advanced statistical methods would be advisable for getting more significant results

Question Number	Average score	2.67 * Std. Dev.
1 (identical-different)	1.25	0.76
3 (better timber)	3.45	1.96
5 (more distorted)	2.05	1.34
9 (more pleasant)	3.30	2.16

Comments

- **Most listeners judged the two samples identical**
- **However, sample B, on average, has slightly "better timber" (less distortion at high frequency), whilst sample A is "more distorted".**
- **Despite of the slight reduction in perceived distortion, the not-linear emulation was slightly preferred to the real-world recording.**

Another example

Original signal

Linear convolution

Live recording

Non-linear
Diagonal Volterra Kernel

Stop

Conclusions

- The sine sweep method revealed to be systematically superior to the MLS & TDS methods for measuring electroacoustical impulse responses
- The ESS method also allows for measurement of not-linear devices and to assess harmonic distortion
- Current limitation for spatial analysis in room acoustis is due to transducers (loudspeakers and microphones)
- A new generation of loudspeakers and microphones, made of massive arrays, is under development.
- The “harmonic orders” impulse responses obtained by the exponential sine sweep method can be used for not-linear convolution, which yields more realistic auralization