

23RD NORDIC SOUND SYMPOSIUM

Training and Information Seminar For Audio People

BOLKESJØ TOURIST HOTEL,
27 – 30 SEPTEMBER 2007

Impulse Response Measurements

Angelo Farina¹

¹ University of Parma, Ind. Eng. Dept., Parco Area delle Scienze 181/A, 43100 PARMA, ITALY
farina@unipr.it

ABSTRACT

Impulse response measurements have been employed for assessing acoustical properties of rooms since more than 50 years. Recently, however, audio people learned to employ these impulse responses as numerical filters, initially for applying reverberation to dry recordings, and more recently also for creating virtual spatiality, particularly when creating surround soundtracks.

Initially a pulsive source was employed, but around 1985 the usage of electroacoustical sound sources started to provide better signal to noise ratio and flatter frequency response. In particular, the MLS method gained a lot of popularity, because it did require very little computational power. Also the TDS (linear sine sweep) method was employed, within the severe constraints of the limited computing capability available at the time.

Exponential Sine Sweeps were employed since long time for audio and acoustics measurements, but only in recent years (2000 and later) their usage became much larger, thanks to the computational capabilities of modern computers. Recent research results allow now for a further step in sine sweep measurements, particularly when dealing with the problem of measuring impulse responses, distortion and when working with systems which are neither time invariant, nor linear.

This paper presents a review of the history of impulse response measurements, and some of the more recent advancements. It describes experimental results aimed to quantify the improvement in signal-to-noise ratio, the suppression of pre-ringing, and the techniques employable for performing these measurements cheaply employing a standard PC and a good-quality sound interface, and currently available loudspeakers and microphones.

1. INTRODUCTION

The concept of impulse response is nowadays widely accepted as a physical-mathematical model of the behavior of a linear, time-invariant system, characterized with just one input port and one output port.

In acoustics, this concept is usually applied to the study of sound propagation from an emission point and a receiver point, located within the same environment.

This technique is usually implemented by means of an omnidirectional sound source, and by an omnidirectional receiver (pressure microphone). This way any spatial information is lost, both on the emission

pattern of real sources, and on the direction of arrival of the wavefronts arriving on the receiver.

In the past it was attempted to obtain partially some spatial information by means of directive transducers (both sources and receivers). But this happened without a rational basis, with just one significant exception, represented by the Ambisonics method derived by Gerzon in the seventies [1].

Recently, advanced impulse-response measurement techniques have been developed [2], capable of performances significantly better than previous methods; furthermore, it is now possible to build, at reasonable costs, multichannel sound systems making use of large arrays of loudspeakers and microphones.

At AES-Paris in 2000 a paper of the author [2] did disclose some "new" possibilities related to sine sweep measurements, triggering a wave of enthusiasm about this method. The usage of exponential sine sweep, compared with previously-employed linear sine sweeps or MLS, provided several advantages in term of signal-to-noise ratio and management of not-linear systems. Furthermore, the deconvolution technique based on convolution in time domain with the time-reversal-mirror of the test signal allowed for clean separation of the harmonic distortion products. And the release of the Aurora software package [3] made it possible to perform these measurements easily and cheaply for everyone.

In reality, nothing was really new, as other authors (Gerzon [4], Griesinger [5]) did already discover these possibilities. The fact that this approach was not successfully employed before is mainly due to the lack of computers with enough computational power and of easily-usable software tools.

In the following 6 years, many research groups and professional consultants started using exponential sine sweeps, and a lot of papers were published (particularly remarkable were the JAES papers of Muller/Massarani [6] and of Embrechts et al. [7]). The tradeoffs of this technique were understood much better, and it was recognized the need of further perfecting the measurement technique for dealing with some problems.

- pre-ringing at low frequency before the arrival of the direct sound pulse
- sensitivity to abrupt pulsive noises during the measurement

- skewing of the measured impulse response when the playback and recording digital clocks were mismatched
- cancellation of the high frequencies in the late part of the tail when performing synchronous averaging
- time-smearing of the impulse response when amplitude-based pre-equalization of the test signal was employed

All of the problems pointed out here have been investigated, and several solutions have been proposed.

This paper presents these "refinements" to the original exponential sine sweep technique, and divulgates the results of some experiments performed for assessing the effectiveness of these techniques.

The methods analyzed include:

- post-filtering of the time-reversal-mirror inverse filter for avoiding pre-ringing
- "exact" deconvolution by division in frequency domain with regularization
- development of equalizing filters to be convolved with the test signal for pre or post equalization.
- counter-skewing of the measured impulse response when the playback and recording digital clocks are mismatched
- employing running-time cross-correlation for performing proper synchronous averaging without cancellation effects

The experiments for assessing the behavior of these "enhanced" measurement techniques were performed employing a state-of-the-art hardware system, including a multichannel sound interface, a powerful PC, and modified versions of the Aurora plugins [3].

Various kinds of microphones were employed too, with the goal of assessing if the measurement of certain acoustical quantities, such as the "spatial parameters" described in ISO 3382, and namely LF, LFC and IACC, can be reliably measured with currently available top-brand microphones.

The results show that, whilst some of the proposed methods really improve substantially the exponential sine sweep measurement method, solving the problems shown above, on the other hand the weak part of the

measurement chain is still about transducers, and namely loudspeakers and microphones, which do not act always along our expectations, and which can cause severe artifacts in the measured quantities.

It is therefore concluded that any impulse response measurement chain can be used with confidence only after a set of careful preliminary tests and alignments. Without this, the results are prone to be at least suspicious, and significant errors have been found in the experimental tests. Of consequence, it appears necessary to further improve the current measurement standards, and mainly ISO 3382, for ensuring reliable and reproducible measurements employing this (and other) methods of measuring impulse responses.

2. OLD METHODS

This chapter describes three “old” methods employed for measurement of impulse response:

- Pulsive sources
- The MLS (maximum length sequence) method
- The TDS (linear sine sweep) method, a.k.a. “stretched pulse”

2.1. Pulsive sources

The original definition of a room impulse response was simply the recording of the sound in a room, being excited with a pulsvive source, such as the explosion of a balloon or a gun shot. Fig. 1 shows these sources.



Fig. 1 – Pulsive sound sources

The sound is usually recorded digitally, employing a small portable recorder, such as a DAT deck, or a PC equipped with a sound card.

Usually the recording is already usable without the need of any subsequent post-processing. Fig. 2 shows such an impulse response and the related octave-band spectral analysis.

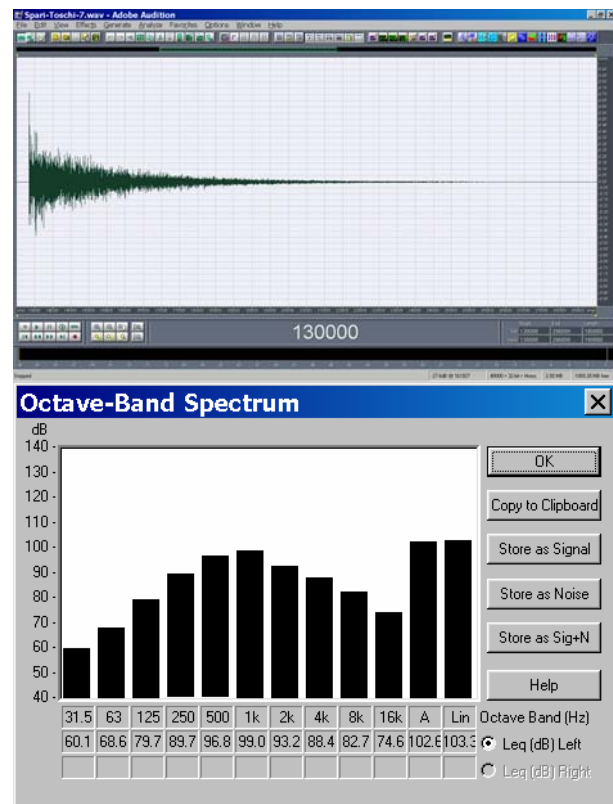


Fig. 2 – Pulsive impulse response

It can be noted that the spectrum is bell-shaped: this is usually not a big problem when the goal of the measurement is to derive acoustical parameters (reverberation time, clarity, etc.), but, of course, this heavily uneven spectrum makes these measurements unusable as numerical filters to be employed in convolution reverbs.

However, some digital equalization can be applied: this can flatten the spectrum, but still the signal-to-noise ratio will be poor at very low and very high frequencies.

2.2. The MLS method

The MLS signal is well known since at least two decades [8,9,10]: it is a binary sequence, in which each

value can be simply 0 or 1, obtained by a shift register as the one shown in fig. 3. The obtained signal is periodic, with period of length L given by:

$$L = 2^{N-1} \quad (1)$$

in which N is the number of slots in the shift register, also called the order of the MLS sequence. Thus an order $N=16$ means a sequence with a period of 65535 samples.

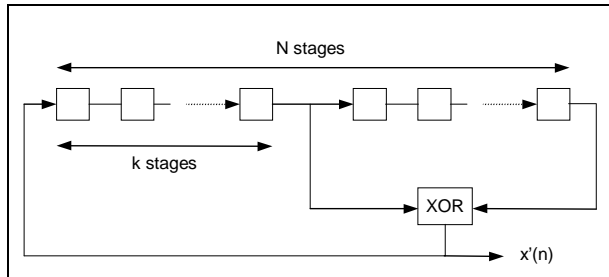


Fig. 3 – Shift register for the creation of the MLS signal

Another very important point is the position of the tap inside the shift register: it is possible to generate MLS sequences also with multiple taps, and the position of the taps influences the behavior of the sequence, particularly when it is used for the excitation of system which are not perfectly linear. The work of Vanderkooy [11] suggests that some sequences are better than others, reporting a list of known taps positions which works reliably for various orders of the MLS sequence.

The generated test signal resembles a square wave, but the length of the “flats” in the signal is not always equal to one sample, depending on the presence of several consecutive 1 or 0 in the MLS sequence.

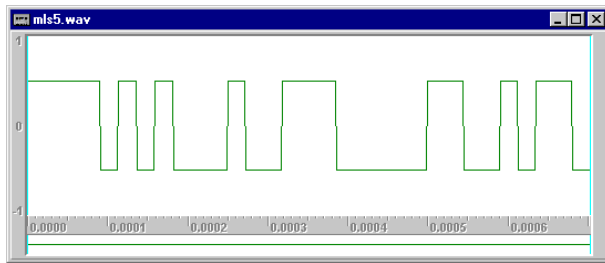


Fig. 4 – MLS sequence of order 5

After having sampled the response of the system to the test signal, this recording has to be processed, for extracting the system’s IR. Thanks to the favorable mathematical properties of the MLS test signal, the deconvolution of the IR can be made with the well known Fast Hadamard Transform (FHT), as originally

suggested by Alrutz [8], and clearly explained by Ando [9] and Chu [10]. The Ando formulation was employed here. The process is very fast, because the transformation happens “in place”, and requires only addition and subtractions. The computations are done in floating point math. Here a brief description of the method is presented.

When the periodic MLS signal $m(j)$ [$j=0..L-1$] is applied at the input of a linear system, characterized by an impulse response $h(i)$ [$i=0..P$], in which $P < L$, the measured output signal $y(k)$ can be interpreted as the convolution of the excitation signal with the system’s impulse response:

$$y(k) = \sum_{j=0}^{L-1} h(j) \cdot m(k-j) \quad (2)$$

Let we transform eq. 2 in matrix notation, defining a matrix M so that:

$$M(i, j) = m[(i+j-2) \bmod L] \quad (3)$$

Thus, eq. 2 becomes simply:

$$\{y\} = [M] \cdot \{h\} \quad (4)$$

To obtain the impulse response h , we find the inverse matrix of M , named \tilde{M} :

$$\tilde{M}(i, j) = M(i, j) - 1 \quad (5)$$

Thanks to the mathematical properties of the MLS sequence, it can be shown that the product of this inverse matrix with the original one produces a slightly modified unit matrix:

$$\tilde{M} \cdot M = (L+1) \cdot I \quad (6)$$

in which I is the identity matrix. Thus, for extracting h from y , it is required to compute:

$$h = \frac{1}{L+1} \cdot \tilde{M} \cdot y \quad (7)$$

The above process is effective, but computationally quite heavy. There is a trick for obtaining the same

result with a very little number of mathematical operations. Let us introduce a larger square matrix U , of order $L+1$:

$$U = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & M & \\ 1 & & \end{bmatrix} = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \tilde{M} & \\ 0 & & \end{bmatrix} + \begin{bmatrix} 1 & \dots & 1 \\ \vdots & 1 & \vdots \\ 1 & \dots & 1 \end{bmatrix} \quad (8)$$

This makes eq. 7 to rewrite as

$$\begin{Bmatrix} 0 \\ \mathbf{h} \end{Bmatrix} = \frac{1}{L+1} \cdot \left(U \cdot \begin{Bmatrix} 0 \\ \mathbf{y} \end{Bmatrix} - \begin{Bmatrix} \bar{y} \\ \vdots \\ \bar{y} \end{Bmatrix} \right) \quad (9)$$

in which \bar{y} is simply the sum of the L measured samples $y(k)$.

The matrix U can be expressed in terms of an Hadamard-type matrix H , thanks to a pair of permutation matrixes P and P^T :

$$U = p^T \cdot H \cdot p \quad (10)$$

The multiplication of the Hadamard matrix H with the measured vector y can be done very efficiently through the Fast Hadamard Transform (analogous to the Fast Fourier Transform) algorithm. Thus, the steps for deconvolving the impulse response from the measured signal y are:

1. generate the permutation vector p from one period of the MLS sequence with the well-known technique of Cohn and Lempel [12].
2. reorder the elements in y according to the permutation vector p
3. Add a zero element at the beginning of the permuted vector, for making it of length $L+1$; let us call y' this modified vector
4. apply the Fast Hadamard Transform to y' : the result is a "shuffled" version of h , called h' , in which the first element is always zero
5. throw away the zero at the beginning of h' , and reorder backward the elements following the inverse permutation described by p .

After subtracting the DC offset \bar{y} , and adjusting the scale factor dividing by $(L+1)$, the wanted impulse response is obtained.

It must be noted that the generation of the permutation vector p can be done once, and stored for every future use of the same MLS signal. Thus the only heavy computation is the Fast Hadamard Transform, which requires only $L \cdot \log_2(L)$ summations: on modern PCs, this task can be easily done in real time, even for very long MLS signals. This means that the continuous acquisition of the signal can be done simultaneously with its processing, and even with the visualization of the subsequently deconvolved impulse responses, making use simply of the computer's CPU and of a standard sound board (nowadays incorporated in any notebook computer).

The system employed for making impulse response measurements with the MLS method is conceptually described in fig. 5. A computer generates the test signal, which passes through an audio power amplifier and is emitted through a loudspeaker placed inside the theatre. The signal reverberates inside the room, and is captured by a microphone. After proper preamplification, this microphonic signal is digitalized by the same computer which was generating the test signal, and processed with the algorithm described above.

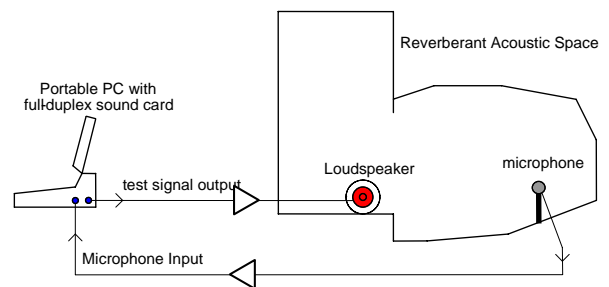


Fig. 5 – schematic diagram of the MLS measurement system

2.3. The TDS method

TDS stands for Time Delay Spectrometry, and this method was invented by Richard Heyser, in 1967 [13].

The method, initially, was implemented with analog equipment, and it was based on employing a tracking filter directly in frequency domain. Later, Poletti did show the limitations of this approach [14]; he also suggested that a direct deconvolution can remove these

limitations, providing a perfect impulse response independently of the sweep rate also in highly reverberant rooms.

The goal of the original Heyser's method was to measure just the initial part of the impulse response, ideally only the direct sound, as he was mainly interested in performing "anechoic" loudspeaker measurements inside a not-anechoic room.

Inside a reverberant room, any wave emitted by the loudspeaker will be reflected off surfaces in the room; we need a way to accept only the direct wave from the loudspeaker into our microphone, excluding the reflected waves. This means filtering out the reflected waves, which make a longer path, and consequently arrive on the microphone with some delay.

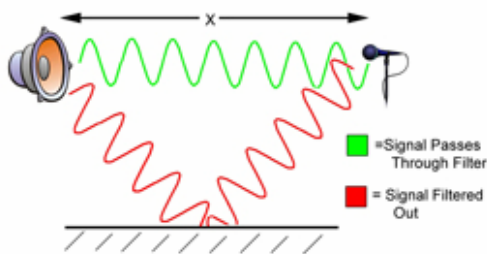


Fig. 6 – direct and reflected waves

We accomplish this by creating a bandpass filter tuned to a given frequency, leaving the system on for the time it takes for the wavefront to reach the microphone from the loudspeaker, and then turning the system off. We then shift the loudspeaker to a new frequency, re-tune the filter to the new frequency, and then repeat.

Alas, we would like to test the loudspeaker system response for a given range of frequencies (say 20 Hz to 20 kHz, the range of human hearing). Using this system, we would need to repeat the experiment for every frequency in the range. We need a practical method to measure every frequency in a given range.

One way around this obstacle is to use a *linearly swept sine wave* (also called an *FM chirp*) as our input signal. A linearly-swept sine wave can be expressed as:

$$s(t) = \sin(k \cdot t^2) \tag{11}$$

where **k** is called the *sweep rate*, and has units of *Hz/second*. The swept sine wave starts at zero frequency, and "sweeps" linearly upward toward higher

frequencies (theoretically up to the Nyquist frequency, in a digital system).

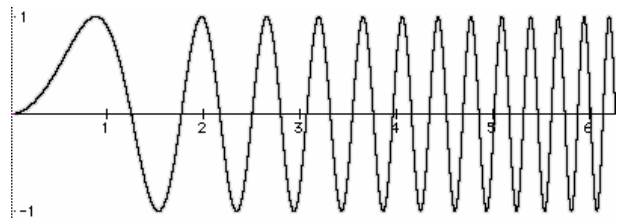


Fig. 7 – linear sine sweep

Since our input signal sweeps, our filter must also sweep upward at the same rate. However, our filter must always sweep a certain amount of time *behind* the input signal. In other words, we must account for the time it takes for the emitted wavefront to reach the microphone. This delay is usually constant, given that the microphone and speaker do not move relative to one another. This delay is a time offset, but it is usually referred to as the *frequency offset*, since the center frequency of the bandpass filter is always slightly behind the frequency of the signal emitted by the speaker.

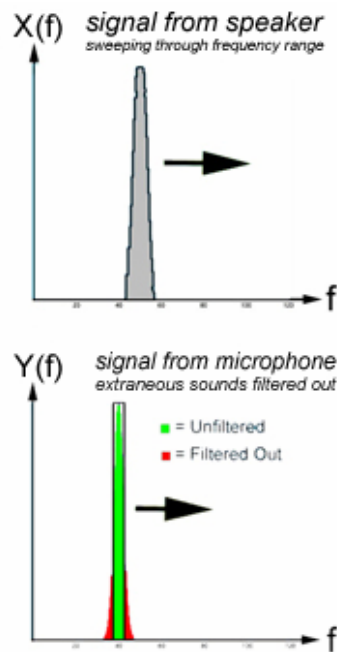


Fig. 8 – tracking filter is always behind

We can quantitatively determine certain variables within our system, including the time offset of our tracking filter and the bandwidth of the tracking filter. Assume that the speaker and the microphone are a distance **x**

apart, and let \mathbf{Ft} represent the instantaneous frequency of the signal being sent to the loudspeaker, and \mathbf{Fr} be the instantaneous frequency of the signal being received at the microphone. The difference between the frequencies of the transmitted and received signal is:

$$F_t - F_r = k \cdot \Delta t = k \cdot \frac{x}{c} \quad (12)$$

Here, c is the *speed of sound* and k is the *sweep rate* of the signal (usually in Hz/second).

If our tracking filter has a bandwidth \mathbf{B} Hz, then our sweep tone will travel some \mathbf{dx} in space while within the bandwidth of the filter. Since our filter is not ideal, we must further define \mathbf{dx} as the region in space where the signal power is at least half of its maximum value. We can call \mathbf{dx} the *space equivalent bandwidth*, and we can relate it to bandwidth \mathbf{B} , speed of sound c , and the sweep rate k :

$$\mathbf{dx} = \mathbf{B} \cdot \left[\frac{c}{k} \right] \quad (13)$$

If the closest reflecting wall is causing a reflected wave having a path longer than the direct path more than \mathbf{dx} , then the result of the measurement will be a frequency response unaffected by room reflections.

We can further expand the concept of time delay spectrometry to measure the *complete* system response of a sound system, *including the reverberant environment* – this is usually called a room system response. We must simply slow down the sweep signal so that some (or all) of the acoustical scattering from the room surfaces leaks through our tracking filter. Essentially, our goal is to let the room reach steady-state equilibrium at each frequency before moving to a new frequency. By taking measurements at progressive time intervals, we obtain a plot of sound pressure as a function of *both* frequency and time.

Since we are assuming that our system is LTI, we may think of each reflecting surface as a *loudspeaker image*. Our system is a series of loudspeaker images, each system having its own unique transfer function. Each transfer function is the product of a *spectral energy distribution* from each surface, $\mathbf{S}(\omega)$ and a linear phase coefficient, $e^{j\omega t}$. The room system response is the sum of all of these individual surface responses:

$$H(\omega) = \sum_k S_k(\omega) \cdot e^{-j\omega t} \quad (14)$$

The impulse response of the system is given by the Fourier transform of $\mathbf{R}(\omega)$.

$$h(t) = \frac{1}{2\pi} \cdot \int_{-\infty}^{+\infty} R(\omega) \cdot e^{j\omega t} \cdot d\omega \quad (15)$$

However, the above approach lends to an impulse response which is more-or-less “smoothed out”, due to the spectral shape of the tracking filter. And a very slow sweep rate is required when measuring the impulse response of an highly reverberant room.

Poletti did quantify these limitations, and he did show that it is possible to get a perfect impulse response simply by recording the unfiltered room response to the sine sweep, usually called $\mathbf{r}(t)$, and then apply a suitable deconvolution technique. This is feasible directly in time domain, by convolving with the time reversal of the sine sweep signal, $s(-t)$:

$$h(t) = r(t) \otimes s(-t) \quad (16)$$

More often, however, the deconvolution is performed in frequency domain (employing FFT and IFFT for going from time domain to frequency domain and vice versa):

$$H(\omega) = \frac{R(\omega)}{S(\omega)} \quad (17)$$

Thanks to the efficiency of the FFT operation, the latter approach was widely employed at the end of the nineties, particularly in Japan, where this method was known as “stretched pulse”.

However, a linear sweep has a “white” spectrum, and consequently provides a signal-to-noise ratio which is not good enough at very low frequencies. For this, often the measurement was spliced in two frequency ranges, with a slower sweep rate for the first measurement (at low frequency), followed by a second measurement with an higher sweep rate covering medium and high frequencies. This “double stretched pulse” was widely employed by Hidaka and Beranek in their huge survey of a collection of the most renowned concert halls all around the world [15].

3. QUICK REVIEW OF THE EXPONENTIAL SINE SWEEP (ESS) METHOD

This chapter is recalling the theory already presented in [2], so the reader has a consequential presentation of the “basic” method, before discussing problems and possible enhancements. The reader already knowing this method can skip directly to chapter 5.

When spatial information is neglected (i.e., both source and receivers are point and omnidirectional), the whole information about the room’s transfer function is contained in its impulse response, under the common hypothesis that the acoustics of a room is a linear, time-invariant system.

This includes both time-domain effects (echoes, discrete reflections, statistical reverberant tail) and frequency-domain effects (frequency response, frequency-dependent reverberation).

A first approximation to the above system is a “black box”, conceptually described as a Linear, Time Invariant System, with added some noise to the output, as shown in fig. 9.

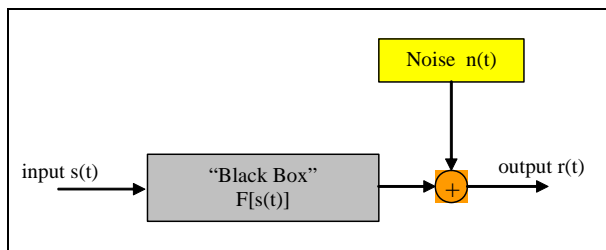


Fig. 9 – A basic input/output system

In reality, the loudspeaker is often subjected to non-linear phenomena, and the subsequent propagation inside the theatre is not perfectly time-invariant.

The quantity which we are initially interested to measure is the impulse response of the linear system $h(t)$, removing the artifacts caused by noise, non-linear behavior of the loudspeaker and time-variance.

The method chosen, based on an exponential sweep test signal with aperiodic deconvolution, provides a good answer to the above problems: the noise rejection is better than with an MLS signal of the same length, non-linear effects are perfectly separated from the linear response, and the usage of a single, long sweep (with no synchronous averaging) avoids any trouble in case the system has some time variance.

The mathematical definition of the test signal is as follows:

$$s(t) = \sin \left[\frac{\omega_1 \cdot T}{\ln \left(\frac{\omega_2}{\omega_1} \right)} \cdot \left(e^{\frac{t}{T} \cdot \ln \left(\frac{\omega_2}{\omega_1} \right)} - 1 \right) \right] \quad (18)$$

This is a sweep which starts at angular frequency ω_1 , ends at angular frequency ω_2 , taking T seconds.

When this signal, which has constant amplitude and is followed by some seconds of silence, is played through the loudspeaker, and the room response is recorded through the microphone, the resulting signal exhibit the effects of the reverberation of the room (which “spreads” horizontally the sweep signal), of the noise (appearing mainly at low frequencies) and of the non-linear distortion.

These “distorted” harmonic components appear as straight lines, above the “main line” which corresponds with the linear response of the system. Fig. 10 shows both the signal emitted and the signal re-recorded through the microphone.

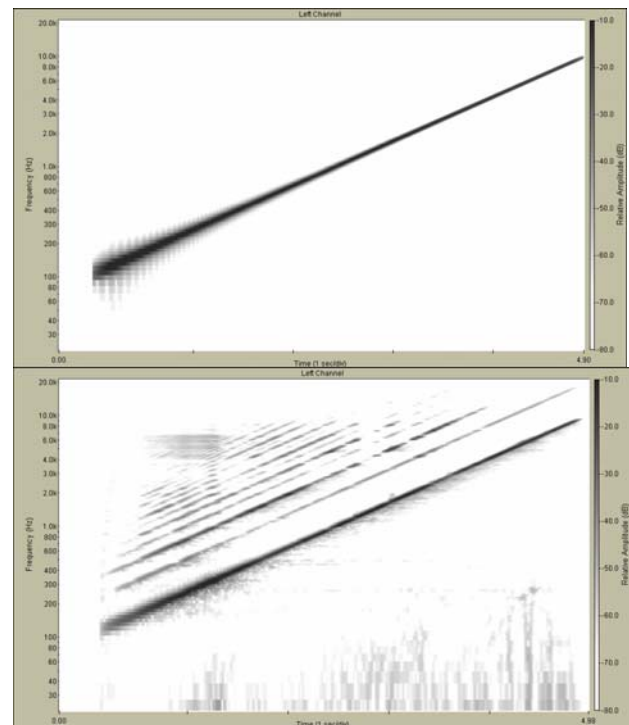


Fig. 10 – sonograph of the test signal $s(t)$ and of the response signal $r(t)$

Now the output signal $r(t)$ has been recorded, and it is time to post-process it, for extracting the linear system's impulse response $h(t)$.

What is done, is to convolve the output signal with a proper filtering impulse response $f(t)$, defined mathematically in such a way that:

$$h(t) = r(t) \otimes f(t) \quad (19)$$

The tricks here are two:

- to implement the convolution aperiodically, for avoiding that the resulting impulse response folds back from the end to the beginning of the time frame (which would cause the harmonic distortion products to contaminate the linear response)
- to employ the Time Reversal Mirror approach for creating the inverse filter $f(t)$

In practice, $f(t)$ is simply the time-reversal of the test signal $s(t)$. This makes the inverse filter very long, and consequently the above convolution operation is very "heavy" in terms of number of computations and memory accesses required (on modern processors, memory accesses are the slower operation, up to 100 times slower than multiplications).

However, the author developed a fast and efficient convolution technique, which allows for computing the above convolution in a time which is significantly shorter than the length of the signal. [16]

It must also be taken into account the fact that the test signal has not a white (flat) spectrum: due to the fact that the instantaneous frequency sweeps slowly at low frequencies, and much faster at high frequencies, the resulting spectrum is pink (falling down by -3 dB/octave in a Fourier spectrum). Of course, the inverse filter must compensate for this: a proper amplitude modulation is consequently applied to the reversed sweep signal, so that its amplitude is now increasing by +3 dB/octave, as shown in fig. 11.

When the output signal $y(t)$ is convolved with the inverse filter $f(t)$, the linear response packs up to an almost perfect impulse response, with a delay equal to the length of the test signal. But also the harmonic distortion responses do pack at precise time delay, occurring earlier than the linear response. The aperiodic deconvolution technique avoids that these anticipatory response folds back inside the time window, contaminating the late part of the impulse response.

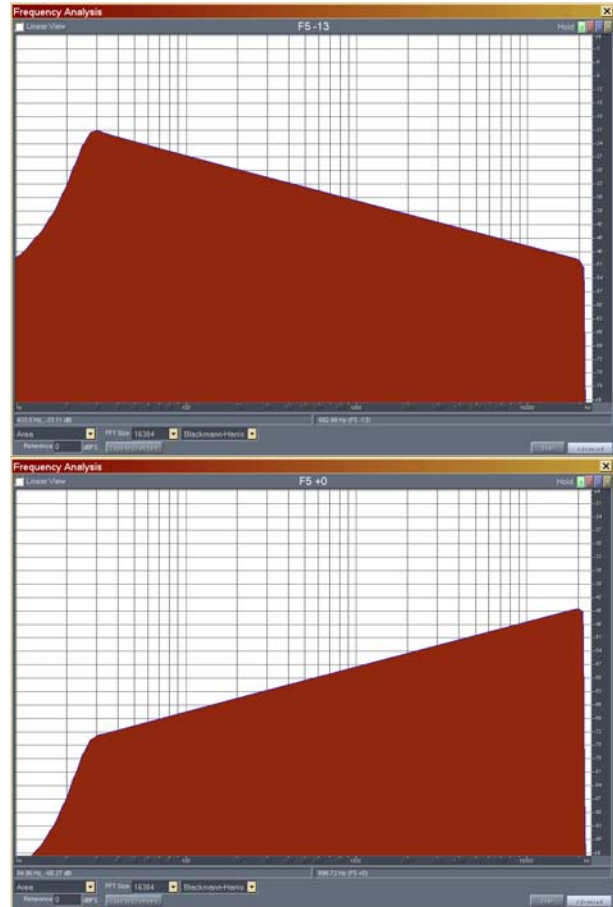


Fig. 11 – Fourier spectrum of the test signal (above) and of the inverse filter (below)

Fig. 12 shows a typical result after the convolution with the inverse filter has been applied.

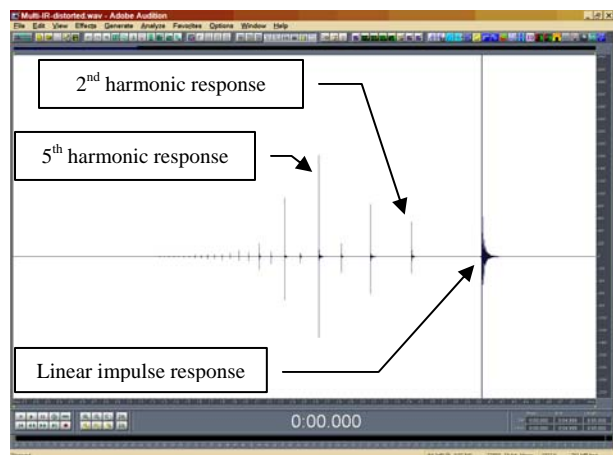


Fig. 12 – output signal $y(t)$ convolved with the inverse filter $f(t)$

At this point, applying a suitable time window it is possible to extract just the portion required, containing only the linear response and discarding the distortion products.

The advantage of the new technique above the traditional MLS method can be shown easily, repeating the measurement in the same conditions and with the very same equipment. Fig. 13 shows this comparison in the case of a measurement made in a highly reverberant space (a church).

It is easy to see how the exponential sine sweep method produces better S/N ratio, and the disappearance of those nasty peaks which contaminate the late part of the MLS responses, actually caused by the slow rate limitation of the power amplifier and loudspeaker employed for the measurements, which produce severe harmonic distortion.

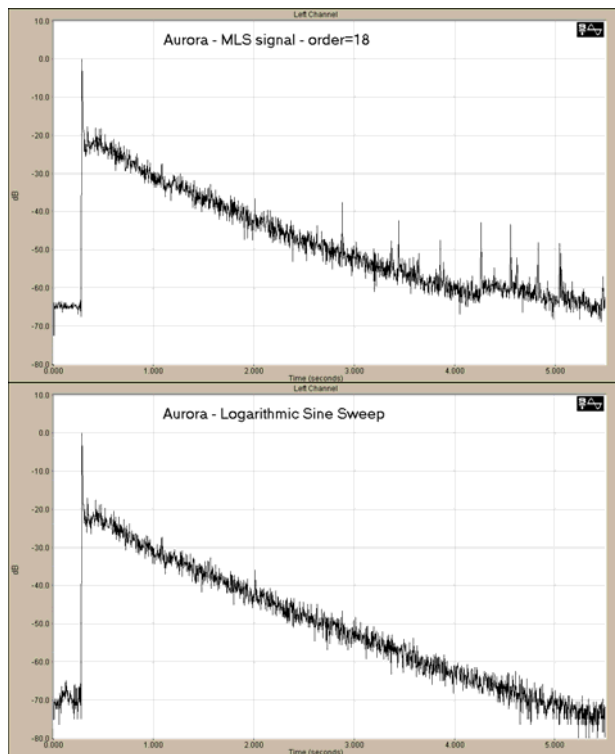


Fig. 13 – comparison between MLS and sine sweep measurements

This method has nowadays wide usage, and is often employed for measuring high-quality impulse responses which are later employed as numerical filters for applying realistic reverberation and spaciousness during the production of recorded music [17].

4. DIRECTIVE SOURCES AND RECEIVERS

When we abandon the restriction to omnidirectional sources and receivers, it becomes possible to get also spatial information. A first basic approach is to “sample” the room’s spatial response with a number of unidirectional transducers, pointing all around in a number of directions.

For example, fig. 14 shows a system employing a highly directive microphone mounted on a rotating table, employed for mapping the direction-of-arrival of reflections inside a theater.

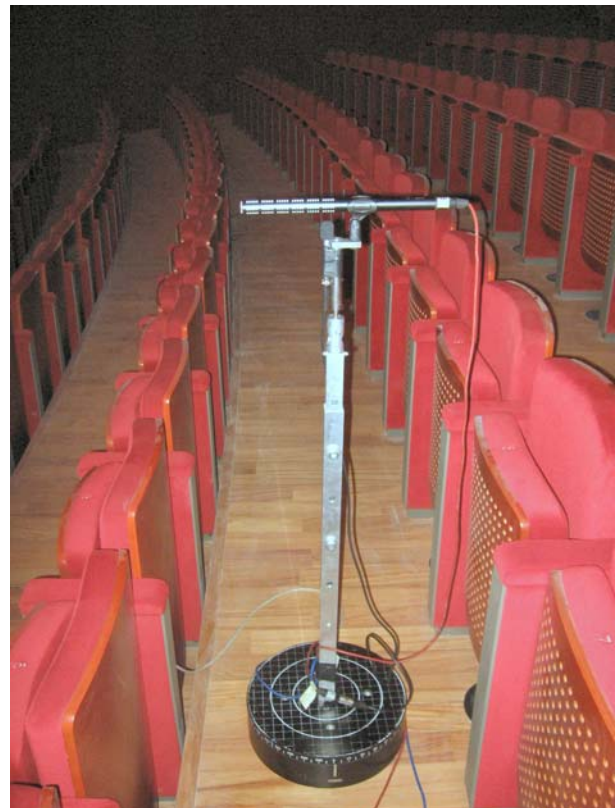


Fig. 14 – directive microphone mounted on a rotating table for sampling room reflections

However, such an approach often ends in repeating a large number of measurements while rotating the transducers in steps, resulting in long measurement times. The approach, furthermore, is not easily scalable: all the measurements need to be performed and analyzed for “covering” uniformly a notional sphere surrounding each transducer.

The approach proposed here is to employ a spherical harmonic expansion of the sound field around the source and receiver points. This corresponds to a two-

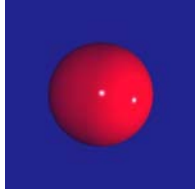
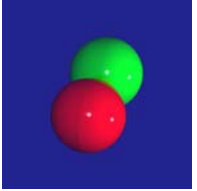
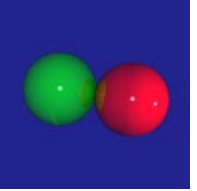

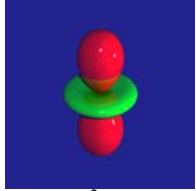
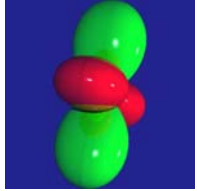
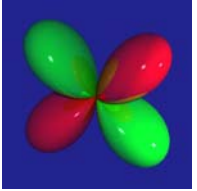
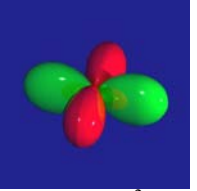
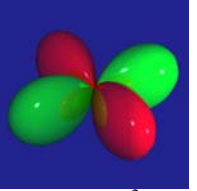
dimensional, spatial Fourier transform, conceptually similar to what is employed in image processing, but working in a spherical coordinate system instead of in a plane Cartesian one.

This approach is the basis of the Ambisonics method [18], initially employed with an expansion limited to 0th-order and 1st-order spherical harmonics around the microphone. Here this concept is extended to higher

orders, and adopted for describing both what happens at the source and at the receiver.

For the sake of concision, here we report the mathematical formulas in polar coordinates, as function of the Azimuth angle A and the Elevation angle E, and a pictorial representation for the spherical harmonics of order 0, 1, and 2 – the equations for higher orders are indeed quite common to find.

Table 1 – spherical harmonics up to 3rd order

Order 0	 0.707107	Order 1	 $\cos(A)\cos(E)$	 $\sin(A)\cos(E)$	 $\sin(E)$
Order 2	 $1.5\sin^2(E)-0.5$	 $\cos(A)\sin(2E)$	 $\sin(A)\sin(2E)$	 $\cos(2A)\cos^2(E)$	 $\sin(2A)\cos^2(E)$

Unfortunately, “native” loudspeakers or microphones having directivity patterns corresponding to the above spherical harmonic functions are available only for orders 0 and 1 (monopoles and dipoles).

However it is possible to “synthesize” the pattern of a spherical harmonic by combining the signals being fed to, or coming from, a number of individual transducers being part of a closely-spaced transducer array.

The recombination is possible with the following formula:

$$y = \sum_{i=1}^N f_i \otimes x_i \quad (20)$$

Where f_i are a set of suitable “matched” FIR filters, designed in such a way to synthesize the required spherical-harmonic pattern. The design of the filtering coefficients can be performed numerically (least-squares approach), starting from a huge number of impulse response measurements made in free field and with a source (or receiver) located in P different polar positions around the transducer array.

The system is solved with the least-squares approximation, imposing the minimization of the total squared error, obtained summing the squares of the deviations between the filtered signals and the theoretical signals v_k :

$$\varepsilon_{\text{tot}} = \sum_{k=1}^P \left[\sum_{i=1}^N (f_i \otimes x_{ki}) - v_k \right]^2 \quad (21)$$

The solution of an overconditioned system requires some sort of regularization. The Nelson-Kirkeby method [19] provides this solution (in frequency domain), which can be adjusted by means of the regularization parameter β :

$$F_i = \frac{X^T \cdot V}{X^T \cdot X + \beta \cdot I} \quad (22)$$

These inverse numerical filters have the advantage that they automatically compensate for the deviation between the responses of the individual transducers, and also for acoustical shielding or diffraction effects due to the mounting structure.

The most basic of such a closely-spaced transducer array is a spherical array. The following figure shows a source array and a microphone array.



Figure 15 – spherical arrays of 12 loudspeakers (above) and 32 microphones (below)

Once a set of spherical harmonics (in emission or in reception) has been measured, it is possible to recombine them for creating any three-dimensional polar pattern, with an error becoming smaller as the order increases. So it is possible to create the emission directivity pattern of a real musical instrument, or to synthesize the response of an ultra-directive virtual microphone, and to aim them in any direction wanted.

This recombination, again, is trivial: it is just matter of summing the signals coming from each of the spherical harmonics patterns with proper gains. This is already well known with reference to the “receiving” spherical

harmonics, which are employed for the reconstruction of a virtual sound field in the high-order Ambisonics method (HOA). The possibilities opened by the measurement of a set of impulse responses which are spatially-expanded in spherical harmonics both at the emission and reception ends is yet to be fully explored.

However, the measurements can be efficiently performed employing a PC equipped with a multichannel sound card. Nowadays a portable system capable of 32 simultaneous inputs and 32 simultaneous outputs can cost less than 3000 USD, all included. Such a system can be easily employed for performing measurements up to 3rd order (16 harmonics) both in emission and in reception: a sequence of 16 sine sweeps is played, each of them being simultaneously fed with different gains and polarities to the individual loudspeakers being part of the spherical emission array. The signals of the 32 microphones are recorded, and subsequently processed for the deconvolution of the impulse response, and for recomputing the 16 spherical harmonic signals. At the end of the measurement, which takes approximately 8 minutes if 15s-long sweeps are employed, a complete set of $16 \times 16 = 256$ impulse responses are obtained.

This set is a complete characterization of the room impulse response, containing both the time-frequency information, and the spatial information as “seen” both from the source and the receiver. It is therefore possible to derive subsequently, by post-processing the measured set of impulse responses, the virtual impulse response produced by a source having arbitrary directivity and aiming, as captured by a microphone also having arbitrary directivity and aiming.

The data measured also allow for spatial analysis, computation of spatial parameters, pictorial representation of the spatial information as colour maps, and high quality rendering of the recorded spatial information by projection over a suitable three-dimensional sound playback system.

As an example, it is shown here how it is possible to display the direction of arrival of early reflections, at different frequencies, performing a directive analysis based on the computation of the Sound Intensity vector.

The following figure, taken from a work of Merimaa, Peltonen and Lokki [], is an example of superposing the directional vectors over the first 100 ms of the sonograph taken from a B-format impulse response (1st order Ambisonics signals, corresponding to spherical harmonics of orders 0 and 1):

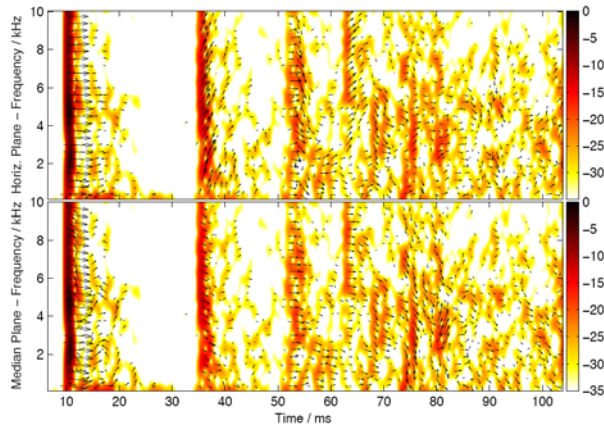


Fig. 16 – graphical display of a directional impulse response

5. PROBLEMS WITH THE ESS METHOD

Despite the significant advantages shown by the ESS method in comparison with all the other previously-employed methods, some problems can still be found, as already pointed out in chapter 1.

In the following subchapters, each of these problems is analyzed, and proper workarounds are presented.

5.1. Pre-ringing

The measured impulse response often shows some significant pre-ringing before the arrival of the direct sound.

This is easily shown performing directly the deconvolution of the IR from the original test signal, without having it passing through the system-under-test.

This way, one should get a theoretically-perfect Dirac's delta function. The old MLS method is perfect in this case, providing exactly a theoretical pulse. The following figure shows instead what happens with the standard ESS method.



Fig. 17 – pre-ringing artifact with fade-out

As shown in fig. 17, the peak is in reality some sort of Sync function, and it shows a number of damped oscillations both before and after the main peak. This is due to the limited bandwidth of the signal (22 Hz to 22 kHz, in this case) and to the presence of some fade-in and fade-out on the envelope of the test signal (0.1s in this example, employing a 15s-long ESS). These two factors define substantially a trapezoidal window in the frequency-domain, which becomes the Sync-like function in time domain.

However, the situation ameliorates significantly if we remove the fade-out. The following figure shows the results obtained with exactly the same settings as in the previous case, but with a length of the fade-in set to 0.0s (fade-in is still 0.1s).

Albeit the appearance of the waveform looks the same (due to the “analogue waveform” display of Adobe Audition), looking carefully at the digital values (the small squares along the waveform) one now sees that the results are very close to a theoretical Dirac's Delta function, and that no pre-ringing or post-ringing are anymore significantly present.

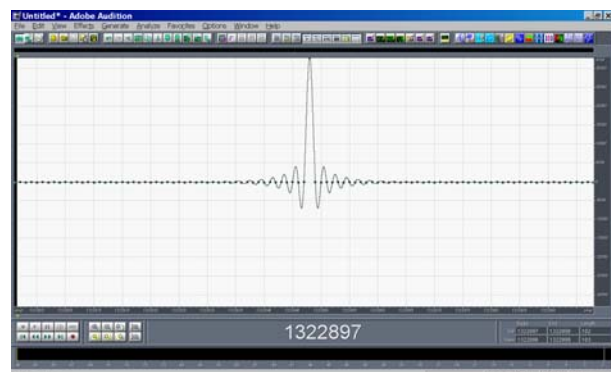


Fig. 18 – reduced pre-ringing artifact without fade-out

However, it is not a good idea to remove completely the fade-out: at the end of the sweep, the final value computed could be not-zero, and consequently the sound system will be excited with a step function, which spreads a lot of energy all along the spectrum.

A solution alternative to removing the fade-out is to continue the sweep up to the Nyquist frequency (22050 Hz, in our example, as the sampling rate was 44.1 kHz), and cutting it manually at the latest zero-crossing before its abrupt termination. This way, no pulsive sound is generated at the end, and the full-bandwidth of the sweep removes almost completely the high-frequency pre-ringing.

However, in some cases, also low frequencies can cause a significant pre-ringing. This is shown easily employing a “loopback” connection, that is, connecting a wire directly from the output to the input of the sound card.

The following figure shows the result of a “loopback” measurement, employing the same parameters as for the previous example ($f_s=44100$ Hz, sweep from 22 Hz to 22050 Hz, 15s long, 0.1s fade-in, no fade-out).



Fig. 19 – low-frequency pre-ringing artifact

Removing the fade-in does not provide any benefit, in this case. So, the way of controlling this type of pre-ringing (due to the analog equipment) is to create a proper time-packing filter, and to apply it to the measured IR.

A packing filter is a filter capable of compacting the time-signature of the impulse response. Various methods for creating a numerical approximation to an ideal packing filter have been proposed in the past. The method employed here is the one developed by Ole Kirkeby, when working at the ISVR with prof. Nelson [20].

Although Kirkeby did propose this method for multichannel inversion (cross-talk cancellation), it can be successfully employed also just for the purpose of packing in time the transfer function of a single-input, single-output system.

The Kirkeby algorithm is as follows:

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

$$H(f) = \text{FFT}[h(f)] \quad (23)$$

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

$$C(f) = \frac{\text{Conj}[H(f)]}{\text{Conj}[H(f)] \cdot H(f) + \varepsilon(f)} \quad (24)$$

Where $\varepsilon(f)$ is a small regularization parameter, which can be frequency-dependent, so that the inversion does not operate outside the frequency range covered by the sine sweep

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

$$c(t) = \text{IFFT}[C(f)] \quad (25)$$

Usually the regularization parameter $\varepsilon(f)$ is chosen with a very small value inside the frequency range covered by the sine sweep, and a much larger value outside that frequency range, as shown in fig. 20:

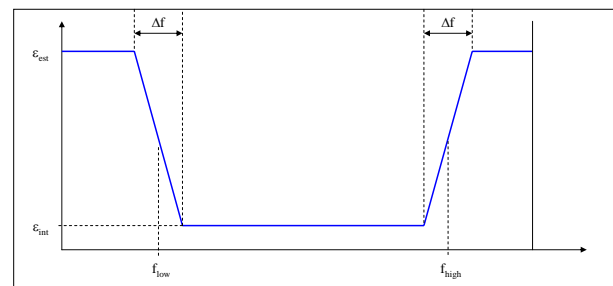


Fig. 20 – frequency-dependent regularization parameter

The following figure shows the inverse filter computed for compacting the “loopback” IR shown in fig. 19:

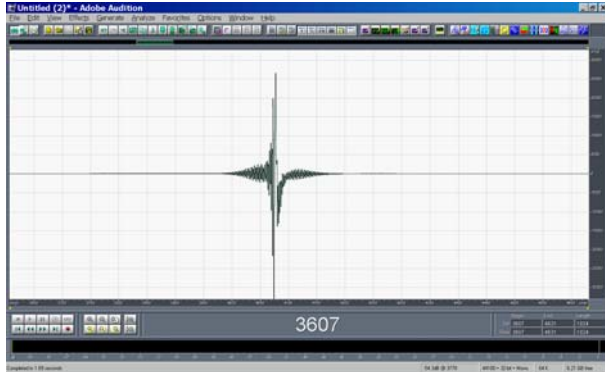


Fig. 21 – “compacting” inverse Kirkeby filter

When this filter is convolved with the measured “loopback” IR shown in fig. 19, the result is the one shown in the next figure:

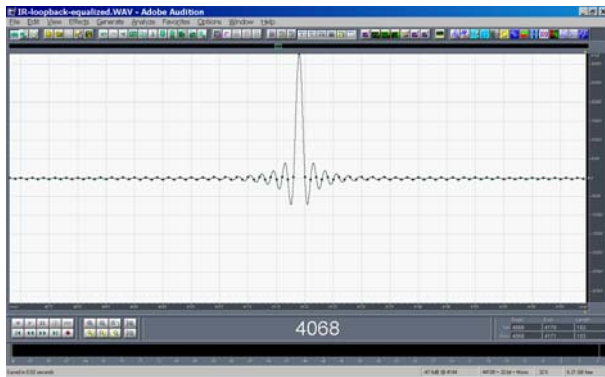


Fig. 22 – “loopback” IR convolved with the “compacting” inverse Kirkeby filter

It can be seen that the usage of the inverse filter managed to re-pack the measured IR back to an almost perfect Dirac’s Delta function.

In conclusion, pre-ringing artifacts can be substantially avoided by combining the usage of a wide-band sweep running up to the Nyquist frequency, without any fade-out, and the usage of a suitable “compacting” inverse filter, computed with the Kirkeby method from a “reference” impulse response.

In the example shown here, the “reference” measurement for computing the inverse filter has been performed electrically, so it does not contain the effect of power amplifier, loudspeaker and microphones. This makes sense if the goal of the measurement is to get information about the behaviour of these electroacoustics components (in most cases, for measuring the performances of the loudspeaker).

5.2. Equalization of the equipment

In other cases, in which the goal of the measurement is just to analyze the acoustical transfer function between an “ideal” sound source and an “ideal” receiver, also the effect of the electroacoustical devices should be removed. In this case, the “reference” measurement is a complete anechoic measurement including power amplifier, loudspeaker and microphone, and the Kirkeby inverse filter will remove any time-domain and frequency-domain artifact caused by the whole measurement system.

For example, the following figure shows the anechoic measurement of the transfer function of a loudspeaker + microphone setup:

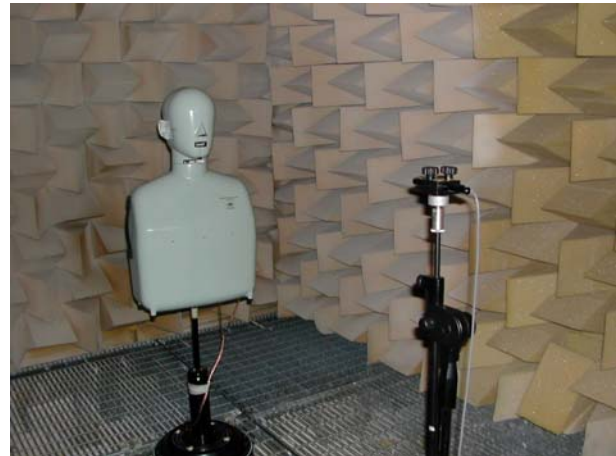


Fig. 23 – anechoic measurement of the “reference” IR of an artificial mouth and an omnidirectional microphone

This example refers to a small, limited-range loudspeaker, employed in a head-and-torso simulator. The measured IR and its frequency response are shown in the following pictures:

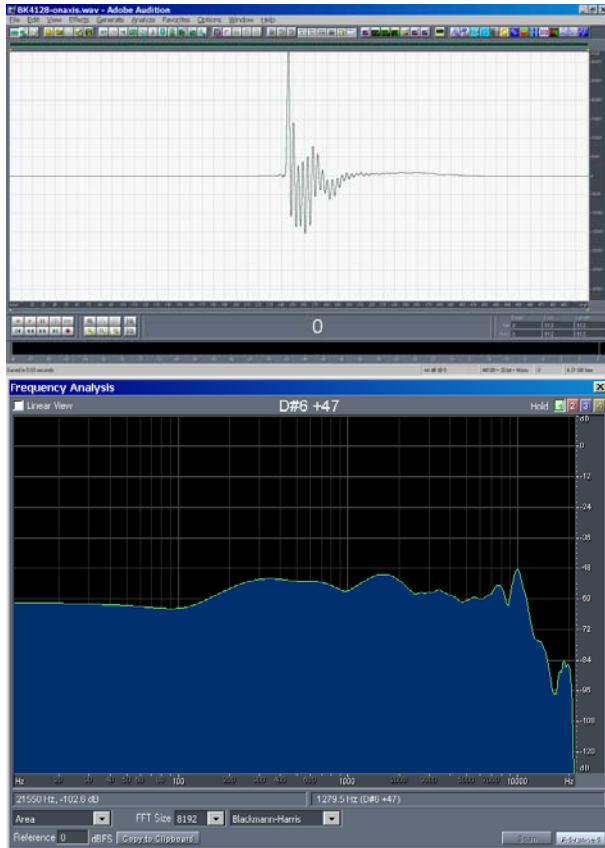


Fig. 24 – measured IR (above) and frequency response (below) of the artificial mouth system

Again, a Kirkeby inverse filter is computed, for correcting the transfer function of the whole measurement system (this time the usable frequency range has been narrowed to 100-11000 Hz):



Fig. 25 – “equalizing” inverse Kirkeby filter

When this inverse filter is applied (by convolution) to the measured IR of this artificial mouth system, we get an IR and a frequency response as shown here below:

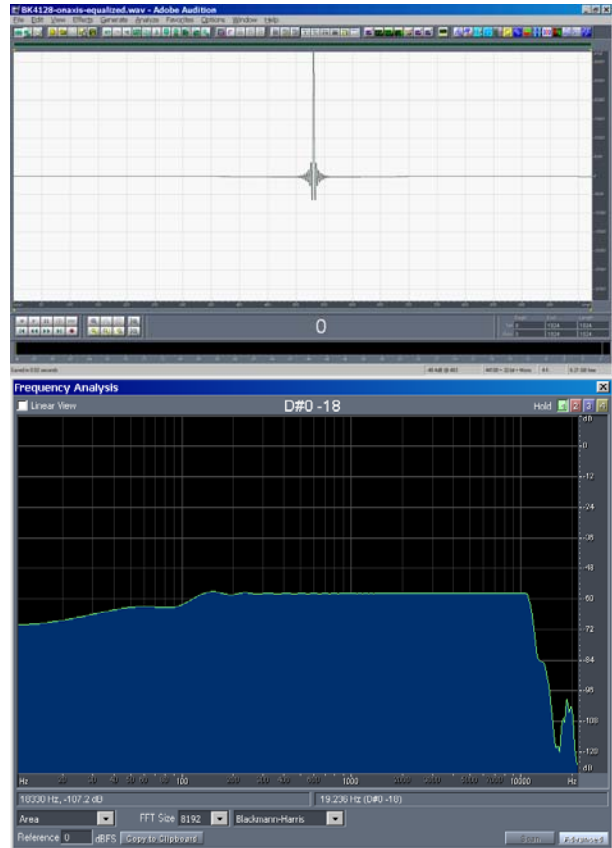


Fig. 26 – measured IR (above) and frequency response (below) of the artificial mouth system after equalization with the inverse filter

Although in this case the inverse filter did not manage to provide a “perfect” result, it still caused the transfer function of the system to closely approach the “ideal” one. This way, the sound system can be employed for measurements without any significant biasing effect.

The latter point to be discussed is if it is better to apply this equalizing filter to the test signal before playing it through the system, or to the recorded signal (indifferently before or after the deconvolution).

Both approaches have some advantages and disadvantages. Applying the equalizing filter to the test signal usually results in a weaker test signals being radiated by the loudspeaker, and in clipping at extreme frequencies (where the boost provided by the equalizing filter is greater).

On the other hand, the usage of the filter after the measurement is done results in “colouring” the spectrum of the background noise, which can, in some case, become audible and disturbing.

In practice, as it often happens, the better strategy revealed to be hybrid: the test signal is first roughly equalized, employing one of the standard tools provided by Adobe Audition (for example Graphic Equalizer). This allows to limit the boost at extreme frequencies and the gain loss at medium frequencies, but however the radiated sound becomes already almost flat.

Then, as usual, a reference anechoic measurement is performed (employing the pre-equalized test signal); a Kirkeby inverse filter is thereafter computed, with the goal of removing the residual colouring of the measurement system. This inverse filter is applied as a post-filter, to the measured data, ensuring that the total transfer function of the measurement system is made perfectly flat. This is the approach successfully employed in the Waves project, as described in more detail in [17].

5.3. Pulsive noises during the measurement

When long sweeps are employed for improving the signal-to-noise ratio, the risk that some pulsvive noise occurs during the measurement increases, as it is difficult to keep people perfectly still for more than a few seconds. Typical sources of pulsvive noise are objects falling on the floor, seats being moved, or “cracks” caused by steps over wooden floors.

The following sonogram shows a recorded sweep contaminated by an evident spurious pulsvive event (the vertical line), caused by an object falling on the floor.

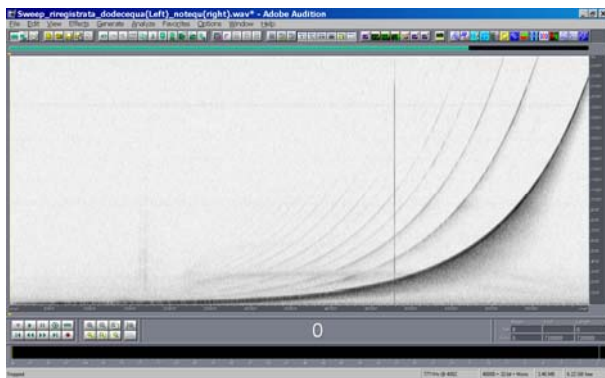


Fig. 27 – pulsvive event contaminating an ESS measurement

After convolution with the inverse filter, this pulsvive event causes a quite evident artifact on the IR:

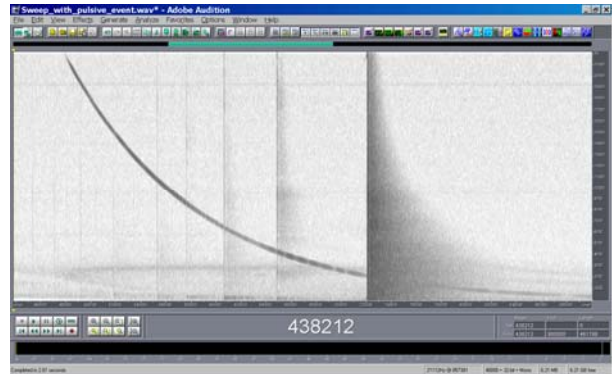


Fig. 28 – Artifact caused by a pulsvive event

In practice, the artifact is a sort of frequency-decreasing sweep, starting well before the beginning of the linear impulse response, and continuing after it. The first part is practically irrelevant on the linear IR, as it will be cut away together with the harmonic distortion responses.

However, the part of this spurious sweep occurring in the late part of the measurement can cause severe problems. The following figure shows a comparison between the octave-band-filtered IR with and without contamination by the spurious pulsvive noise.

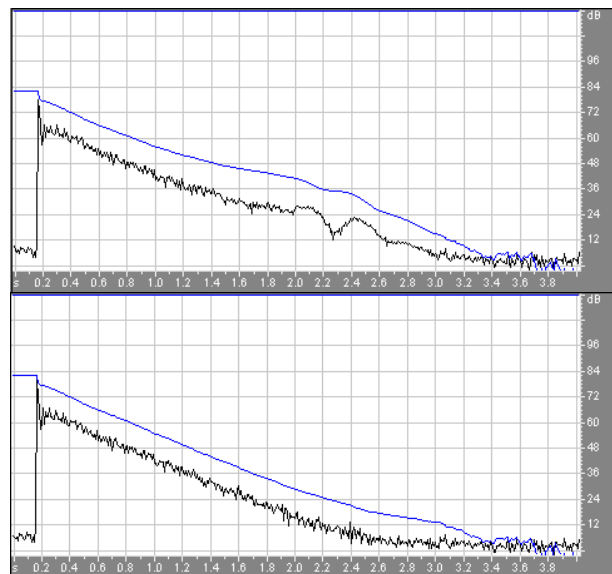


Fig. 29 – octave-band filtered IR (at 1 kHz) contaminated from pulsvive noise (above) and without contamination (below)

The presence of the spurious effect generated by the pulsvive noise is causing an overestimate of T30 (2.48 s instead of 2.13 s). Also Clarity C80 and Center Time are affected, but more slightly.

One way of removing this artifact consists in silencing the recording signal in correspondence of the pulsive event, as shown in the following figure:

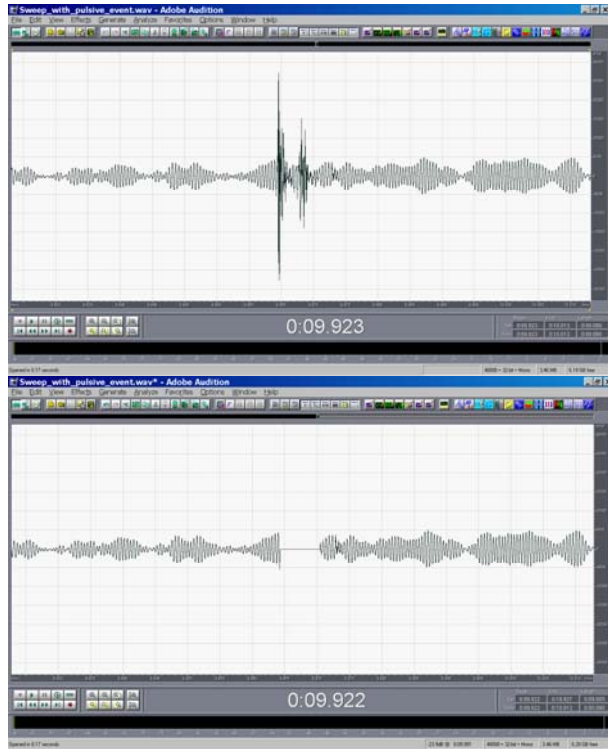


Fig. 30 – silencing the spurious event

After deconvolving the edited signal, the following IR is obtained:

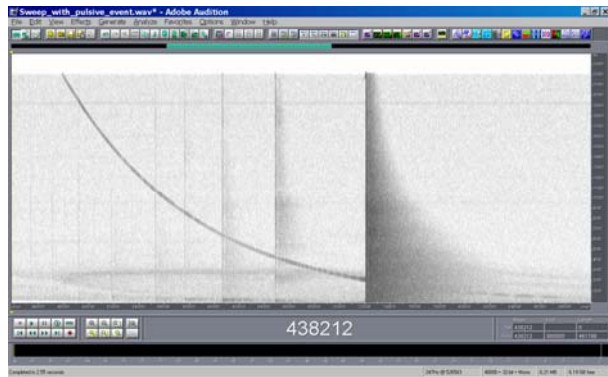


Fig. 31 – effect of the silenced pulsive event on the deconvolved IR

Despite silencing the event, the artifact is still there, albeit with reduced amplitude. The analysis of the reverberant tail still shows some effect of the pulsive artifact, as shown here:



Fig. 32 – octave-band filtered IR with silenced pulsive event

A much better removal of the pulsive event is obtained by employing the Click/Pop Eliminator provided by Adobe Audition. The following picture shows how it works:

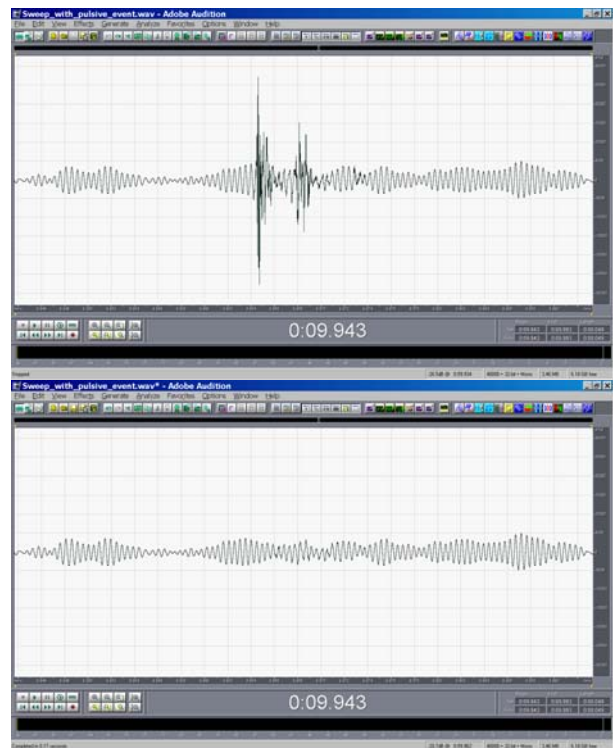


Fig. 33 – effect of the Auto Click/Pop Eliminator

In this case, the result of the deconvolution is the following:

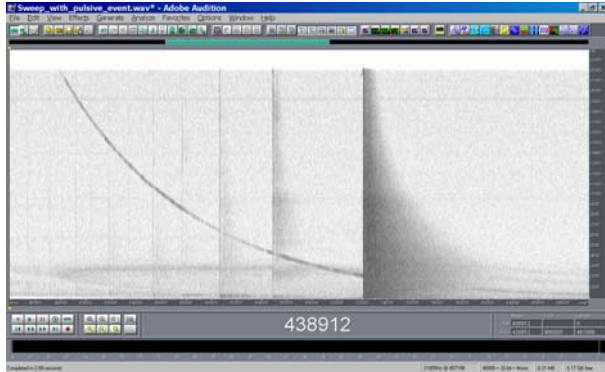


Fig. 34 – effect of the pulsvive event on the deconvolved IR after click/pop Eliminator

The artifact has been further reduced, but it is still there.

Finally, an even better way of removing the artifact is based on the knowledge of the frequency of the sine sweep at the moment in which the pulsvive event did happen. In the case presented here, the instantaneous frequency was 2159 Hz. So, applying a narrow-passband filter at this exact frequency, all the wide-band noise is removed, and a “clean” sinusoidal waveform is restored, as shown in the following figures:

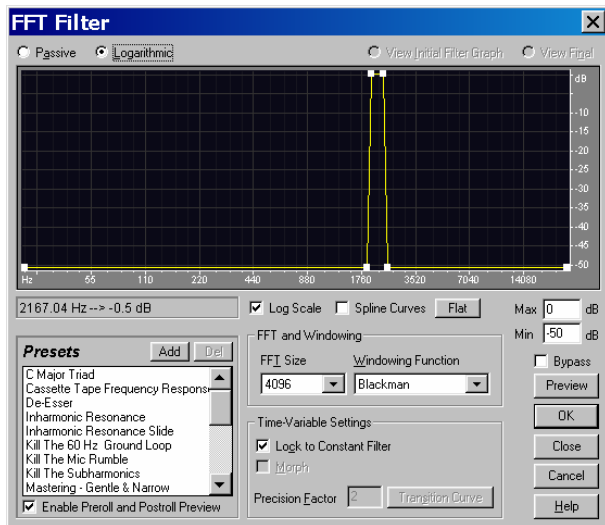


Fig. 35 – usage of FFT Filter for removing the pulsvive artifact

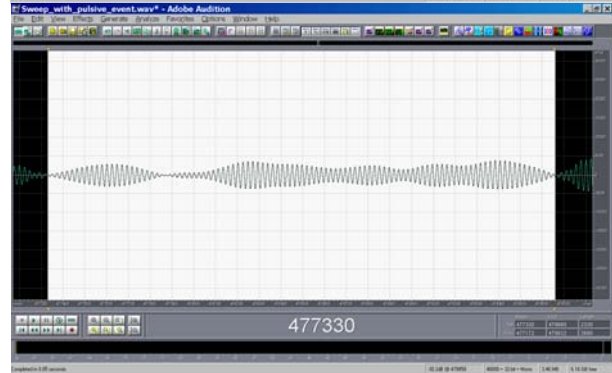
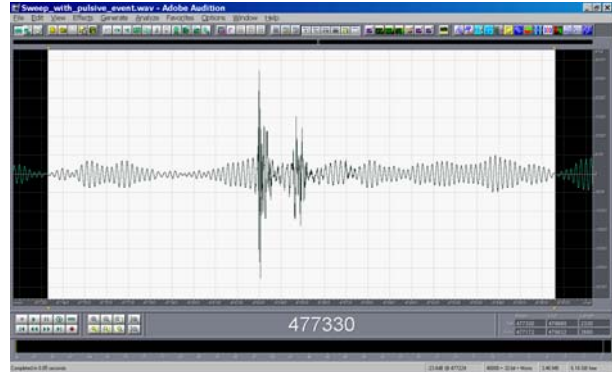


Fig. 36 – effect of FFT filter for removing the pulsvive artifact

After deconvolution, the measured impulse response is as follows:

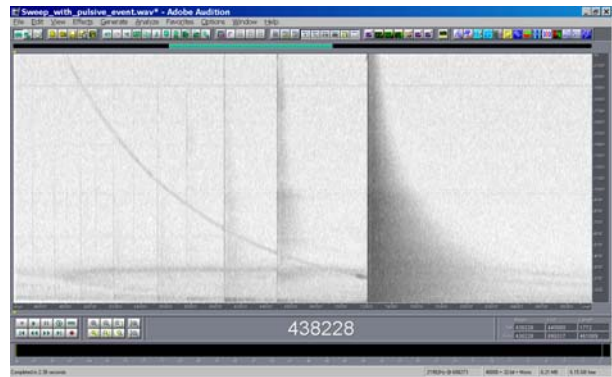


Fig. 37 – result of the FFT filter

Now the artifact amplitude has been reduced so much that there is no more distortion of the reverberant tail, as shown here:



Fig. 38 – octave-band filtered IR with pulsive event removed with FFT filter

So it can be concluded that the best way of removing a pulsive artifact from a sweep measurement is to apply a narrow-band filter just around the instantaneous frequency at which the event occurred.

5.4. Clock mismatch

One of the great advantages of the ESS method over other methods for measuring the impulse response is that a tight synchronization between the playback clock and the recording clock is not required.

In fact, even if two completely independent hardware devices are employed, and no clock synchronization is employed, usually the impulse response obtained is perfectly clean and without observable artifacts. However, when the mismatch between the two clocks becomes significant, the deconvolved impulse response starts to be “skewed” in the frequency-time plane.

For example, the following figure shows the result of a purely-electrical measurement, obtained playing the test signal with a portable CD player, directly wired to a computer sound card, employed for recording.

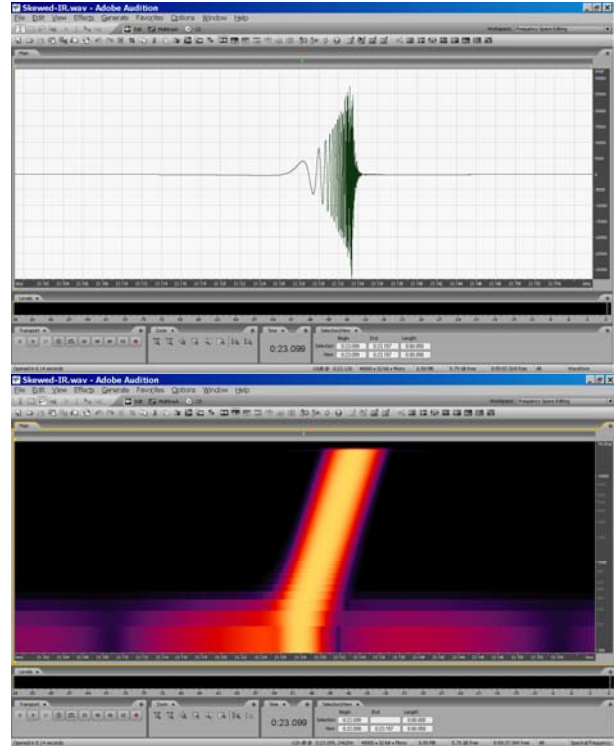


Fig. 39 – a skewed IR

The waveform clearly shows that low frequencies are starting earlier than high frequencies, and the sonograph demonstrates that, with a logarithmic frequency scale, the IR does not have a vertical (synchronous) appearance, but a sloped (skewed) appearance.

Various methods can be applied for re-aligning the clocks. For example, if a “reference” measurement can be performed, we could try to use a Kirkeby inverse filter for fixing the mismatch, as already shown in chapters 3.1 and 3.2.

The following figure show the result of such an inverse filter applied to the electrical measurement performed.

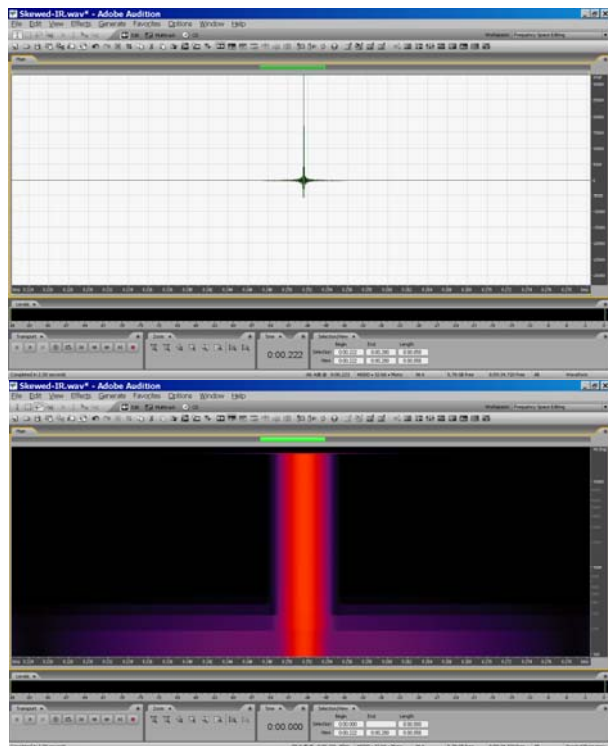


Fig. 40 – correction of a skewed IR employing a Kirkeby inverse filter

The result obtained employing the inverse filter is quite good; and it is also correcting for the magnitude of the frequency response of the system, not only for the frequency-dependent delay.

Nevertheless, this approach requires the availability of a clean reference measurement, performed either electrically or under anechoic conditions.

Whenever a reference measurement is not available, the inverse filter approach cannot be employed. Another possible solution is the usage of a pre-stretched inverse filter for performing the IR deconvolution.

For example, in this example it can be seen how the original inverse filter is too short. If we now create an inverse filter slightly longer than the original one, we can correct for the skewness of the sonograph.

Looking again at fig. 39, we see that the skewness is approximately 8.5 ms long. So we generate a new sine sweep, and its inverse sweep, 8.5 ms longer.

When we convolve this longer inverse sweep with the recorded signal, the deconvolution produces the following result:

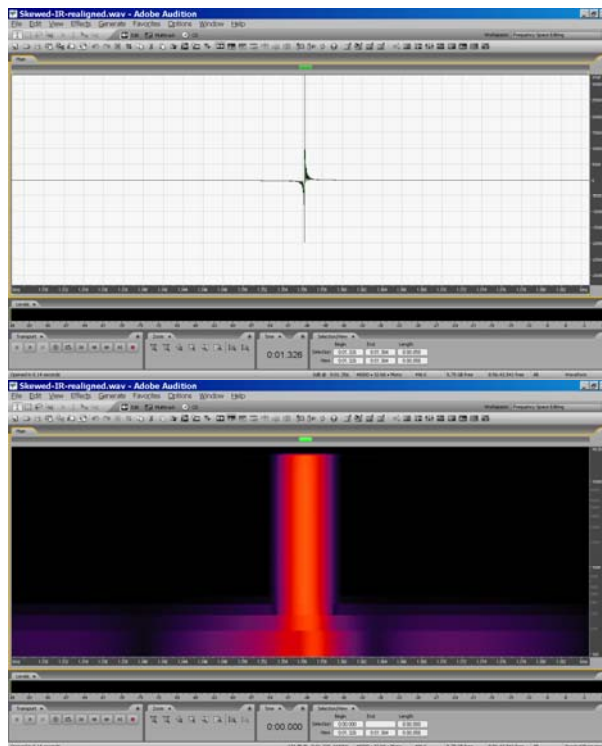


Fig. 41 – correction of a skewed measurement employing deconvolution with a longer inverse sweep

This result is not so clean as the one obtained with the Kirkeby inversion, but now we have got a quite good clock realignment without the need of a reference measurement.

It must be said, however, that a skewed impulse response, although bad to see and to listen, is still quite usable for computing acoustical parameters. It is nevertheless always useful to correct for the clock mismatch, as this significantly improves the peak-to-noise ratio. For example, with the data presented here, the usage of the longer inverse sweep for the deconvolution provides an amelioration of the peak-to-noise ratio by 12.45 dB, which is quite significant.

5.5. Time averaging

The usage of averaging several impulse responses for improving the signal-to-noise ratio is a deprecated technology when working with the ESS method.

Synchronous time averaging works only if the whole system is perfectly time-invariant. This is never the case when the system involves propagation of the sound in

air, due to air movement and change of the air temperature. So, the preferred way for improving the signal to noise ratio is not to average a number of distinct measurements, but instead to perform a single, very long sweep measurement, as clearly recommended in the ISO 18233/2006 standard.

However, in some cases the usage of long sweeps is not allowed (for example, when the method is implemented on small, portable devices equipped with little memory), and so time-synchronous averaging is the only way for getting results in a noisy environment.

Unfortunately, even a very slight time-variance of the system produces substantial artifacts in the late part of the reverberant tail, and at higher frequencies.

This happens because the sound arriving after a longer path is more subject to the variability of the time-of-flight due to unstable atmospheric conditions. Furthermore, a given differential time delay translates in a phase error which increases with frequency.

The following picture compares the sonographs of two IRS, the first comes from a single, long sweep of 50s, the second from the average of a series of 50 short sweeps of 1s each.



Fig. 42 – single sweep of 50s (above) versus 50 sweeps of 1s (below)

Although from the above picture it is not very easy to see the difference, it can be noted that the energy of the reverberant tail is significantly underestimated, at high frequency, in the second measurement. This can be seen easily displaying the spectrum of the signal in the range 100 ms to 300 ms after the direct sound, as shown here:

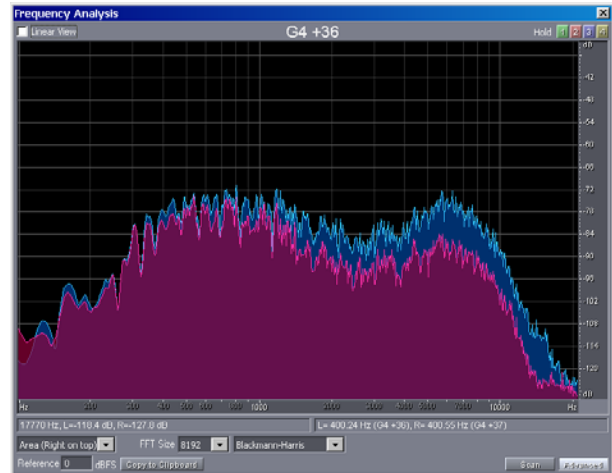


Fig. 43 – spectrum of single sweep of 50s (above) versus 50 sweeps of 1s (below)

It can be seen how, above 350 Hz, the synchronously-averaged IR is systematically underestimated. Around 5-6 kHz the underestimation is more than 10 dB.

This of course affects also the slope of the decay curve, and the estimate of reverberation times. The following figure shows the comparison between the octave-band filtered impulse response and decay curves at 4 kHz:

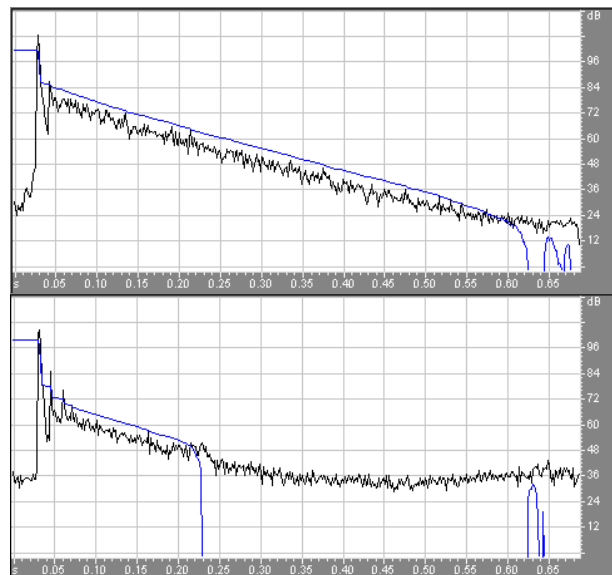


Fig. 44 – octave-band-filtered impulse response of a single sweep of 50s (above) versus 50 sweeps of 1s (below)

It can be seen how the single-sweep measurement is providing a perfectly linear decay with quite good dynamic range (63 dB), whilst the synchronously-

averaged IR exhibit strong underestimate of the energy of the reverberant tail, and simultaneously a much worst signal-to-noise ratio (43 dB).

It can be concluded that synchronously-averaging a number of subsequent IRs obtained with the ESS method is causing unacceptable artifacts.

However, an alternative technique can be used, in these cases, for processing the data.

It is necessary to create a stereo file, containing the test signal in the left channel, and the recorded signal in the right channel, as shown here:

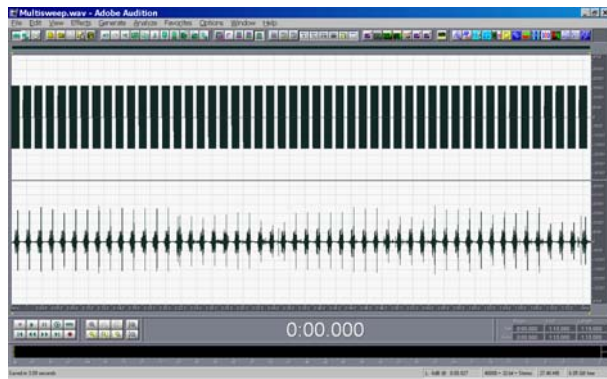


Fig. 45 – multisweep signal (test and response)

Now this stereo waveform is processed with the new Aurora plugin named Cross Functions, which is employed for computing the transfer function H_1 , by performing complex averaging in spectral domain:

$$H_1(f) = \frac{G_{LR}}{G_{LL}} \quad (26)$$

Where GLR and GLL are the averaged cross-spectrum and autospectrum, respectively

This is the user’s interface of this plugin:

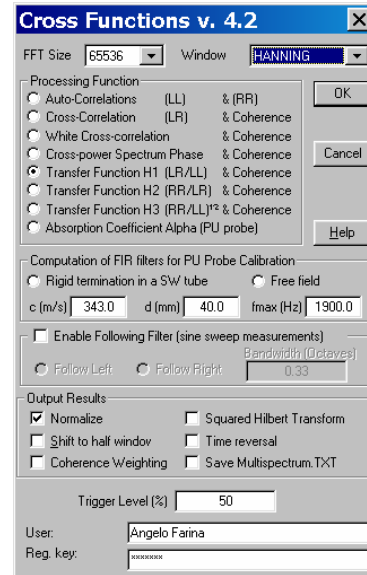


Fig. 46 – Computation of H_1

Only the first half of the resulting transfer function is kept, for removing most of the effects of the Hanning window. The following figure shows the recovered impulse response, compared with the single-sweep one:

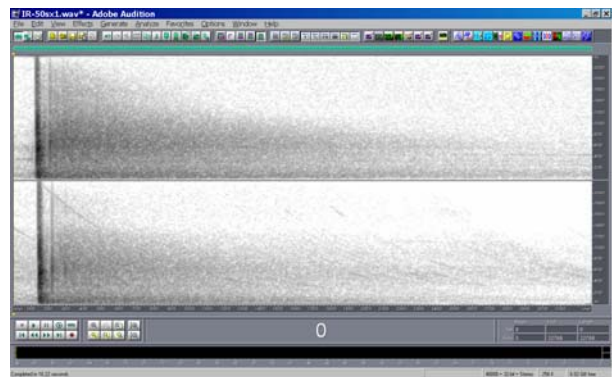


Fig. 47 – single sweep of 50s (above) versus 50 sweeps of 1s (below) processed with the Cross Functions module

Analyzing the octave-band-filtered impulse response (at 4 kHz), the following is obtained:

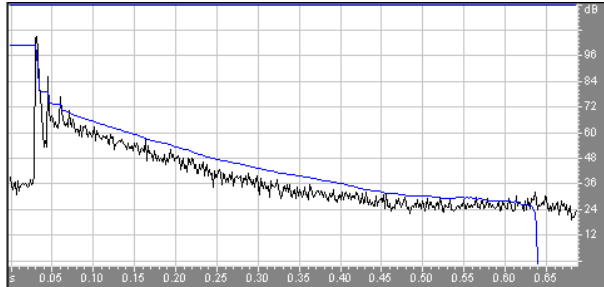


Fig. 48 – octave-band-filtered impulse response of a 50 sweeps of 1s (Cross Functions)

It can be seen that the situation is now significantly better than with “standard” time-synchronous averaging: the frequency-domain processing provided an impulse response with better signal-to-noise ratio and with a reverberant tail only slightly underestimated. The single sweep method is still better, but now the difference is not so large, and the measurement result is still usable.

So, in practice, the employment of a number of independent sweeps can provide almost acceptable results, provided that the deconvolution and averaging of the impulse response are performed in reversed order (first averaging, then deconvolution), and in the frequency domain.

6. PERFORMANCE OF ELECTROACOUSTIC TRANSDUCERS

For room acoustics measurements, it is common to employ:

- An omnidirectional loudspeaker (dodecahedron)
- An Omni + Figure of Eight microphone
- A binaural microphone (dummy head)

In the previous chapter it has been already discussed how to measure the impulse response and frequency response of a measurement chain containing also

loudspeakers and microphones, and how to reasonably equalize it. However, the problem still arises of the spatial properties (directivity) of these transducers.

It will be shown here that the measured directivities of loudspeakers and microphones differ significantly from the nominal ones, causing errors which are orders of magnitude greater than those described in the previous chapter.

6.1. Dodecahedron loudspeakers

These loudspeakers are usually employing single-way, wide-band transducers, and require heavy equalization for providing flat sound power response. However, the equalization cannot correct the polar patterns of these loudspeakers, which deviate significantly from omnidirectional starting at frequencies above 1 kHz.

Here we present the results of polar patterns measured in anechoic conditions for three dodecahedrons. The first one is a standard-size (40cm diameter) employing for building acoustics measurements (LookLine D-300); the second one is a smaller version (25 cm diameter) specifically developed for measurement of impulse responses in theaters and concert halls (Look Line D-100). Finally, the third one employs waveguides for reconstructing a more uniform spherical wavefront (Omnisonics 1000).

The following figure shows the three dodecahedrons analyzed:



Fig. 49 – 3 dodecahedron loudspeakers

The above loudspeakers have been measured inside an anechoic chamber over a turntable, so the horizontal polar patterns have been obtained, in octave-bands.

The following three figures compare these polar patterns at 1000, 2000 and 4000 Hz.

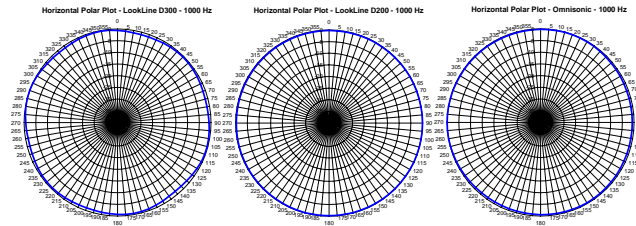


Fig. 50 – directivity patterns at 1 kHz

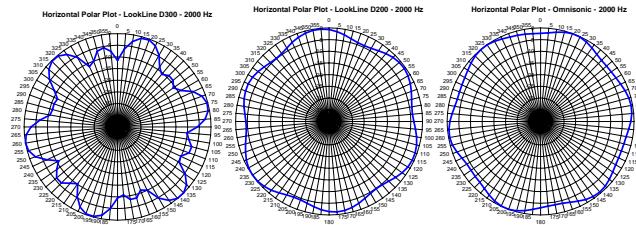


Fig. 51 – directivity patterns at 2 kHz

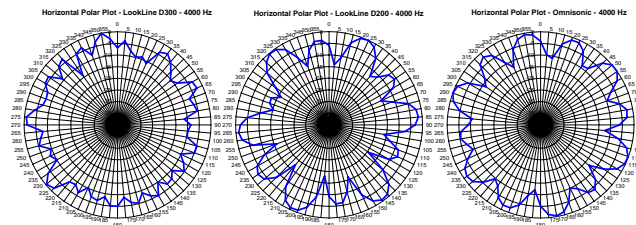


Fig. 52 – directivity patterns at 4 kHz

It can be seen how all three these dodecahedrons exhibit quite irregular polar patterns at medium-high frequency.

6.2. Omni + Figure of 8 mics

Although the usage of small-size measurement microphones does not pose any significant problem (as a B&K 1/2" capsule is almost perfectly omnidirectional and with flat frequency response up to 20 kHz), when spatial parameters such as LE, LF or LFC need to be measured it is necessary to employ a variable-directivity-pattern mike, providing both omnidirectional and figure-of-8 patterns.

For this purpose, it is common to employ not-measurement-grade probes, often manufactured by top-quality makers such as Neumann or Schoeps. However, the values of spatial parameters measured with different microphonic probes are often quite irreproducible.

So it was decided to perform a comparative experiment among 4 of these dual-pattern probes, including these mikes:

- Soundfield ST-250
- Bruel & Kjaer sound intensity kit type 3595
- Schoeps CMC5
- Neumann TLM 170R

The following image shows some of the probes being compared, during the measurements performed inside the Auditorium of Parma:

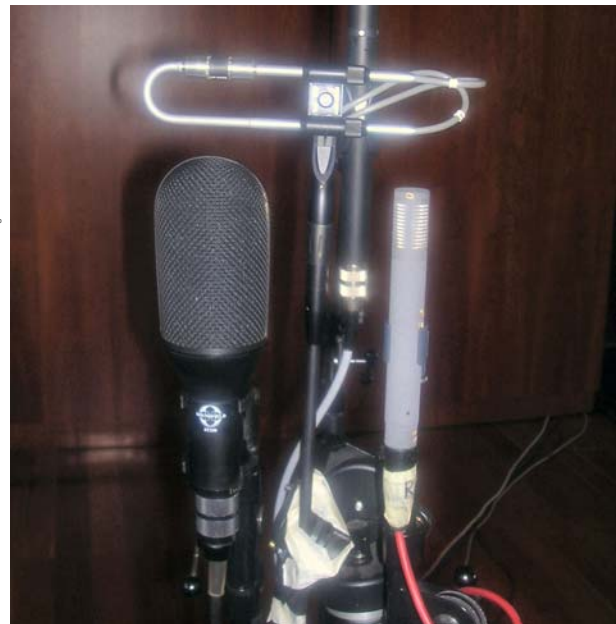


Fig. 53 – 3 microphonic probes

A stereo impulse response has been measured with each probe, containing the Omni response on the left channel, and the figure-of-8 response in the right channel. Each of these 2-channels IRs have been processed with the Aurora plugin named Acoustical Paramaters, specifying the type of probe being employed, as shown here:

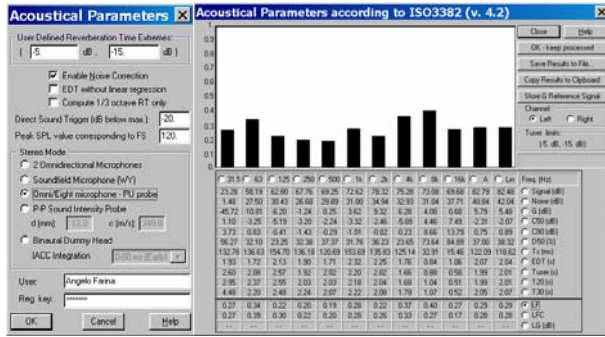


Fig. 54 – the Acoustical Parameters plugin

This way, the LF parameter has been measuring for all 4 probes, in octave bands, and at two distances from the sound source (7.5m and 25m). The following figure shows the results at 25m:

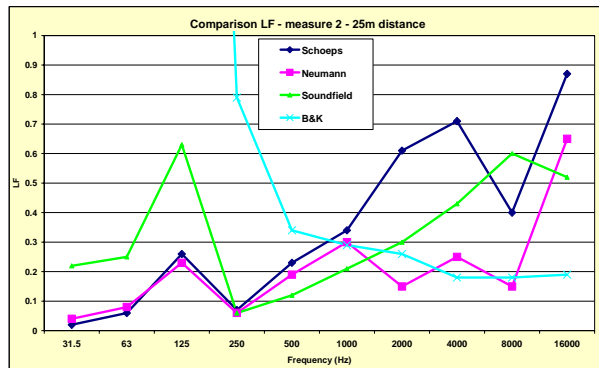


Fig. 55 – LF measured at 25m

It can be seen how the results are completely diverging; it is impossible to establish what of the 4 probes was measuring correctly, albeit the Schoeps looks more “reasonable” than the other three.

These deviations are caused by the polar patterns of the probes. As an example, here we report a couple of polar patterns of the Soundfield ST-250, measured on a turntable inside an anechoic room:

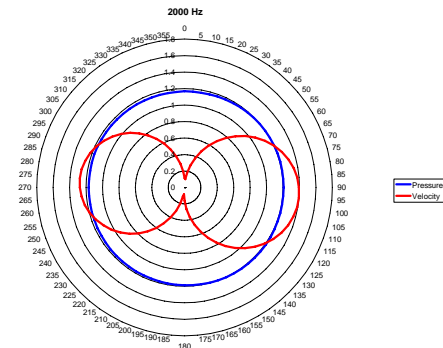
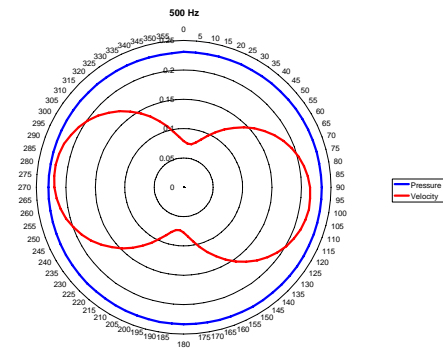


Fig. 56 – ST-250 – polar patterns at 500 Hz and 2 kHz

It can be seen that, even at medium frequencies, the figure-of-8 pattern is distorted, and is not properly gain-matched with the omnidirectional one. These deviations are even greater at very low and very high frequencies:

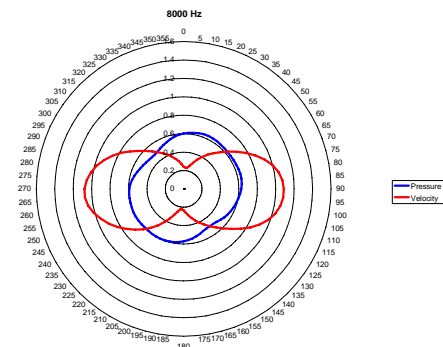
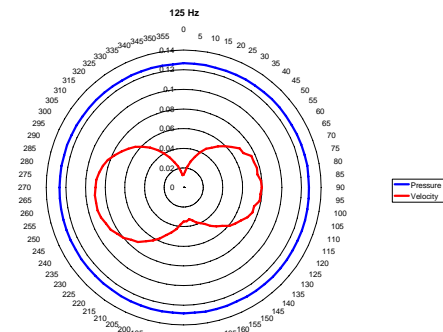


Fig. 57 – ST-250 – polar patterns at 125 Hz and 8 kHz

It can be concluded that actually no available microphonic system can be used for assessing reliably the values of spatial acoustical parameters such as LE, LF or LFC.

6.3. Binaural microphones

Another way of assessing the spatial properties of a room is by means of the IACC parameter (inter aural cross correlation), also defined in ISO-3382, and measurable employing a binaural microphone and the Aurora Acoustical Parameter plugin.

However, various makers of dummy heads produce quite different microphone assemblies. For checking comparatively their performances, a set of impulse response measurements have been performed in a large anechoic chamber, employing a turntable controlled by the sound card, as shown in the following figure:

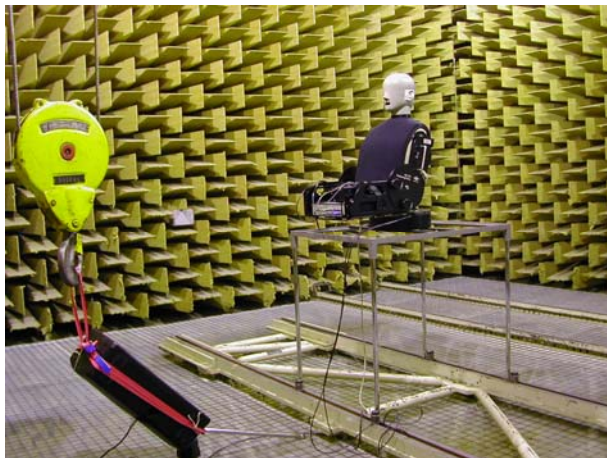


Fig. 58 – anechoic measurements on dummy heads

Also in this case 4 different binaural microphones have been tested:

- Bruel & Kjaer type 4100
- Cortex
- Head Acoustics HMS-III
- Neumann KU-100

A synthetic diffuse sound field has been generated, employing a number of loudspeakers surrounding the dummy head and feeding them with uncorrelated pink noise.

In principle, given the fact that the sound field was exactly the same, all the dummy heads should have given the same value of IACC. Instead, as shown in the following figure, the results have been quite diverging:

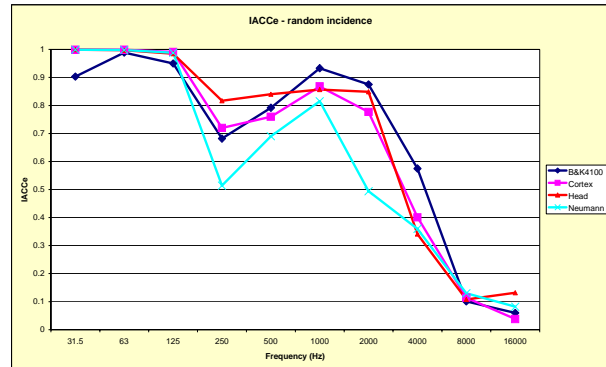


Fig. 59 – IACC measured with the 4 dummy heads

The deviations, however, are not so bad as those obtained in the previous chapter for the measurement of LF. It can be concluded that, with currently available systems, the measurement of IACC is slightly more reproducible than that of LF.

7. MEASURING AND EMULATING NON-LINEAR AND TIME-VARIANT SYSTEMS

When the system being measured (and later reconstructed by convolution with the measured impulse response) is not linear and time-variant, all the methods based on the assumption of a LTI system do fail.

We need a different mathematical framework for dealing with not-linear, time-variant system. The most known and widely applicable one is the Volterra series expansion. Unfortunately, this method is computationally too expensive for being performed in real time even on modern computers, so a subset of the complete Volterra scheme has been developed by the author.

7.1. The Diagonal Volterra Kernel method

This chapter provides a quick review of the digital-filtering method known as Diagonal Volterra Kernel convolution (also known as Vector Volterra Kernel). A more extended explanation is given in [21] and [22].

This method can be seen as a “reduced version” of the general Volterra representation of a not-linear system: while the general approach requires the usage of matrix kernels, having a dimensionality equal to the order of the kernel (so the second-order Volterra kernel is a square matrix, the 3rd-order one is a cubic matrix, and so on), the Diagonal Volterra Kernel approach employs mono-dimensional vectors at any orders.

The general Volterra formulation represents the output y of a not-linear system, being fed with the input signal x , as:

$$y(n) = \sum_{i_1=0}^{M-1} h_1(i_1) \cdot x(n-i_1) + \sum_{i_1=0}^{M-1} \sum_{i_2=0}^{M-1} h_2(i_1, i_2) \cdot x(n-i_1) \cdot x(n-i_2) + \sum_{i_1=0}^{M-1} \sum_{i_2=0}^{M-1} \sum_{i_3=0}^{M-1} h_3(i_1, i_2, i_3) \cdot x(n-i_1) \cdot x(n-i_2) \cdot x(n-i_3) + \dots \quad (27)$$

In this formulation, it is assumed that all the kernels are of the same size (M samples). The input signal x is linearly-convolved with the first-order kernel h_1 , the squared values of the input samples are convolved with the second-order matrix h_2 , and so on.

This approach provides, in theory, a complete representation of the not-linear behavior of the system, under the hypothesis that it is still time-invariant, and that the Volterra expansion is reaching an order high enough and that the kernel size M is large enough for capturing the whole “memory” of the system.

However, the computational load required tends to explode beyond any reasonable limit as the order of the expansion is more than 3, whilst it is known that at least 5 orders are required for perceptually-acceptable simulation of harmonic-distorting devices such as loudspeakers or musical instruments.

The simplified formulation does only take into account the Volterra coefficients located along the diagonal of the higher-orders matrixes, so that the computational load now scales linearly with the order up to which the not-linear simulation is performed:

$$y(n) = \sum_{i=0}^{M-1} h_1(i) \cdot x(n-i) + \sum_{i=0}^{M-1} h_2(i) \cdot x^2(n-i) + \sum_{i=0}^{M-1} h_3(i) \cdot x^3(n-i) + \dots \quad (28)$$

In principle, this could provide a strong degradation in the accuracy of the simulation. In practice, instead, most of the Volterra coefficients outside the diagonal are usually very close to zero, and hence this “reduced” Volterra convolution is generally perceptually very good. In most cases, the advantage provided by the possibility to extend the simulation to higher orders outperforms the error due to discarding the coefficients locate outside the diagonal.

Furthermore, as explained in [21], the experimental measurement of the diagonal Volterra kernels can be performed easily, employing as a test signal an Exponential Sine Sweep, and performing the deconvolution through aperiodic linear convolution with a suitable inverse filter. Very simple linear math is required for transforming the measured impulse responses in the Volterra vectors.

The capability of analyzing the not-linear properties of a system employing the Exponential Sine Sweep signal were first described by Gerzon [4], and only much later “rediscovered” by Griesinger [5] and Farina [2].

The basics of the method have already been given in chapter 3, so here we can assume that the reader already knows how to extract the system’s linear response employing the ESS method.

When the Exponential Sine Sweep signal is introduced in the non-linear system, its output also contains harmonic distortion products, as shown here:

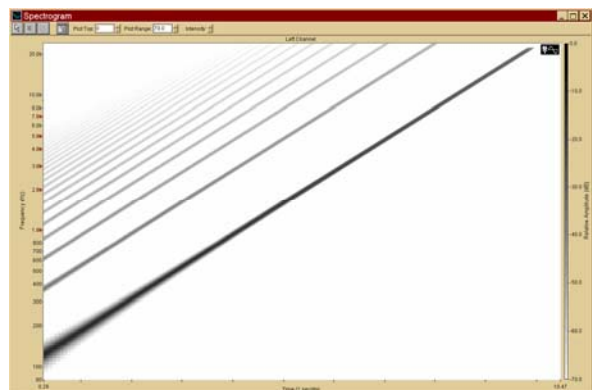


Figure 60. Spectrogram of the system’s response

It is possible to deconvolve the impulse response by applying to this response, by convolution, a proper inverse filter, which is simply the time-reversal of the excitation signal (3), equalized with a slope of 6dB/oct (time-reversal mirror plus whitening filter). This is the result:

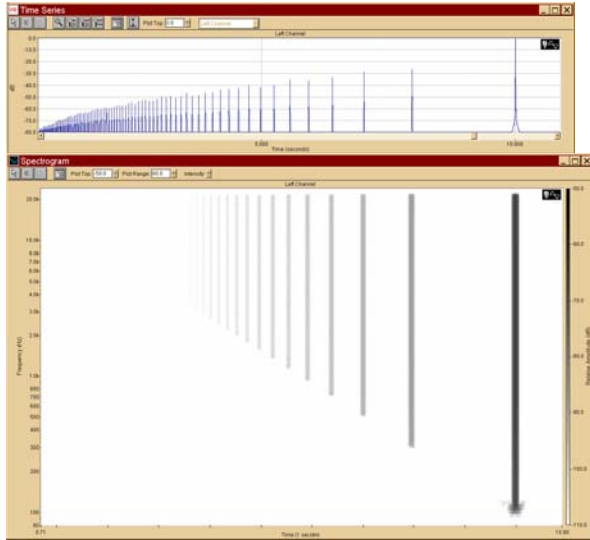


Figure 61. Spectrogram of deconvolved impulse responses

The rightmost impulse response is the linear one, which is preceded by the second-order harmonic response, and so on. The measured impulse responses are not directly the Volterra kernels, but these are easily computed by solving a linear equation system.

The solution of this system allows for the computation of the unknown diagonal Volterra kernels h_i starting from the measured ordered impulse responses h'_i .

This is the solution for the first 5 orders, formulated in frequency-domain:

$$\begin{cases} \overline{H}_1 = \overline{H}'_1 + 3 \cdot \overline{H}'_3 + 5 \cdot \overline{H}'_5 \\ \overline{H}_2 = 2 \cdot j \cdot \overline{H}'_2 + 8 \cdot j \cdot \overline{H}'_4 \\ \overline{H}_3 = -4 \cdot \overline{H}'_3 - 20 \cdot \overline{H}'_5 \\ \overline{H}_4 = -8 \cdot j \cdot \overline{H}'_4 \\ \overline{H}_5 = 16 \cdot \overline{H}'_5 \end{cases} \quad (30)$$

After computing the values of the kernels from the measured multiple impulse responses, the non-linear convolution can be efficiently implemented following eq. 28.

7.2. Efficient partitioned convolution

For real-time applications, indeed, the computational load required for performing all the multiplications and sums required by eq. 28 is still too large for an entry-level PC. However, a very efficient convolution scheme is known, making use of frequency-domain processing, which is very convenient when employing highly-optimized FFT subroutines.

By proper partitioning the kernels to be convolved, it is possible to simultaneously reduce the computational load and to meet the expectations of a reasonably low processing latency.

Some years ago, it was thought that the “optimal” partitioned strategy had to be searched with the goal of minimizing the total number of multiplications, as at that time this was the most costly operation for a digital processor. Under these assumptions, three different algorithms were proposed (and patented) by David McGrath [23], Bill Gardner [24] and Guillermo Garcia [25]. However, one of the authors subsequently proved [16] that these “advanced” algorithms, based on the non-uniform partitioning of the kernel, are systematically outperformed, on modern CPUs, by the previously-known, “suboptimal” algorithm initially proposed in 1966 by Stockham [26], which makes use of uniform partitions, thanks to the “trick” suggested by Barry Kulp in 1988 [27], consisting in performing the accumulation of the results of the convolution of the N blocks in frequency domain, prior of a single IFFT operation.

7.3. Time-Variant systems

Real-world systems often exhibit some degree of time-variance. Acoustic propagation in rooms, for example, is affected by air temperature and air movement; loudspeakers change their radiation properties due to temperature variations, and most electronic devices contains “adaptive” circuitry performing various form of automatic gain control, compression, expansion, limiting, soft and hard clipping. In some cases the variation of the system’s transfer function is driven by the amplitude of the signal, in others it is driven by an external control, which can be manual or based on an oscillator (for example, in a flanger an oscillator periodically changes the tuning of a parametric filter).

The most obvious approach for emulating time-variant devices is the usage of a convolver employing variable filtering coefficients. This can be seen to have some point in common with the approach known as “dynamic convolution”, developed by Mike Kemp [28] with the

goal of representing the behavior of not-linear device. This approach revealed to be suboptimal for the original goal, but it is indeed a great idea for emulating linear, time-variant system: if each sample being processed is convolved with the “proper” set of filtering coefficients, this can produce a faithful emulation of a time-variant system.

However, “dynamic convolution” requires to employ the old, inefficient time-domain convolution algorithm, whilst we want to employ the computationally-efficient partitioned convolution scheme. However, it is trivial to see the bridge between the two techniques: a time-variant device can be emulated by changing the set of filtering coefficients for each data block being processed, instead of for each sample. If the blocks are small enough, this still allows to emulate faithfully a system which is changing over time.

If one has to measure, for example the set of impulse response of a device which changes its transfer function depending on the amplitude of the input signal, such as, for example, a compressor, one has simply to generate a test signal which is made of a number of exponential sine sweeps, each with a different amplitude, as shown in the following figure.

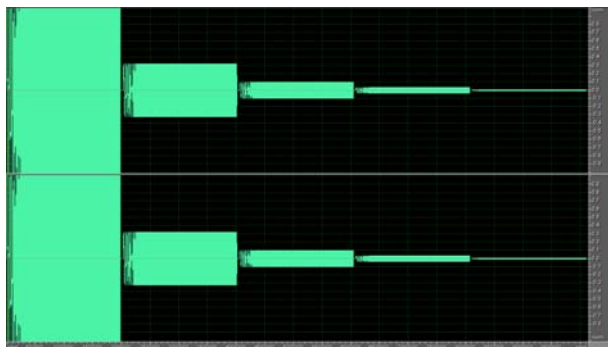


Figure 62. Waveform view of the test signal

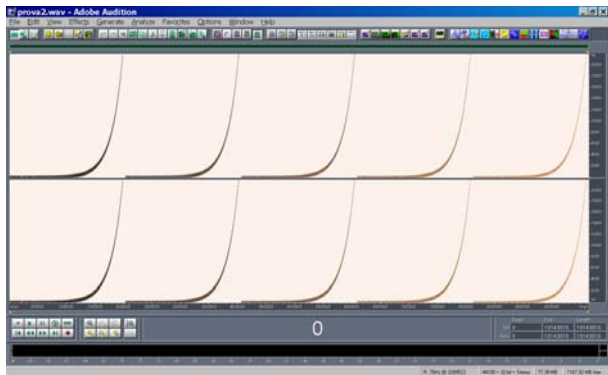


Figure 63. Spectral view of the test signal

After deconvolving the system’s response, one gets a number of consecutive sets of not-linear impulse response, one set for each amplitude of the input signal. During convolution, the set being employed will be changed according to the average amplitude of the signal being processed, evaluated block by block...

This real-time processing is nowadays feasible thanks to a recently-released software package Nebula, by Acustica Audio [29]. More details about the features of this software will be presented at the forthcoming 123 AES Convention [30].

8. CONCLUSIONS

In this paper a comprehensive review of the techniques employed for measuring room impulse response is given.

Each technique is described in detail, and its advantages and problems are discussed.

The author is strongly advocating the usage of the Exponential Sine Sweep method, both for measuring room impulse responses, for electroacoustical measurement of loudspeaker and loudspeaker arrays, and for electrical characterization of the transfer function of audio processing equipment, including strongly not-linear and not-time-invariant devices such as compressors, flangers, etc...

This paper shows, after 7 years of experience employing the ESS method, that it is superior to all the other known methods, and provides capabilities completely impossible with the pother methods.

9. ACKNOWLEDGEMENTS

This work was supported by LAE (www.laegroup.org).

10. REFERENCES

- [1] Michael Gerzon - "Recording Concert Hall Acoustics for Posterity", *JAES* Vol. 23, Number 7 p. 569 (1975)
- [2] A. Farina – "Simultaneous measurement of impulse response and distortion with a swept-sine technique", 110th AES Convention, February 2000.
- [3] www.aurora-plugins.com
- [4] P.Craven, M.Gerzon - "Practical Adaptive Room And Loudspeaker Equaliser for Hi-Fi Use" - 92nd AES Convention, March 1992
- [5] D.Griesinger - "Beyond MLS - Occupied Hall Measurement With FFT Techniques" - 101st AES Convention, Nov 1996
- [6] S. Müller, P. Massarani – "Transfer-Function Measurement with Sweeps", *JAES* Vol. 49, Number 6 pp. 443 (2001).
- [7] G. Stan, J.J. Embrechts, D. Archambeau – "Comparison of Different Impulse Response Measurement Techniques", *JAES* Vol. 50, No. 4, p. 249, 2002 April.
- [8] H. Alrutz, M.R. Schroeder, "A fast Hadamard transform method for the evaluation of measurements using pseudorandom test signals" *Proc. of 11th International Congress on Acoustics*, Paris, **6** (1983), pp. 235-238.
- [9] Y. Ando, *Concert Hall Acoustics*, Springer-Verlag, Berlin 1985, appendix
- [10] W.T. Chu, "Impulse response and Reverberation Decay measurements made by using a periodic pseudorandom sequence", *Applied Acoustics*, **vol. 29**, pp. 193-205 (1990).
- [11] J. Vanderkooy, "Aspects of MLS measuring systems", *JAES* **vol. 42, n. 4**, 1994 April, pp. 219-231.
- [12] M. Cohn, A. Lempel, "On fast M-sequence transforms", *IEEE Trans. Inf. Theory* **IT-23** (1977), pp. 135-137.
- [13] R. Heyser – "Acoustical measurements by time delay spectrometry" - *Journal of the Audio Engineering Society*, 15(4):370--382, October 1967.
- [14] M. Poletti – "Linearly swept frequency measurements, time-delay spectrometry, and the Wigner distribution" – *JAES* vol. 36, n. 6, 1988 June, pp. 457-468.
- [15] T. Hidaka, L.L. Beranek – "Objective and subjective evaluations of twenty-three opera houses in Europe, Japan, and the Americas" - *The Journal of the Acoustical Society of America* - January 2000 - Volume 107, Issue 1, pp. 368-383
- [16] A. Torger, A. Farina – "Real-time partitioned convolution for Ambiophonics surround sound", *2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* - Mohonk Mountain House New Paltz, New York October 21-24, 2001.
- [17] A. Farina, R. Ayalon – "Recording concert hall acoustics for posterity" - 24th AES Conference on Multichannel Audio, Banff, Canada, 26-28 June 2003
- [18] M.A. Gerzon "Ambisonics in Multichannel Broadcasting and Video" - *J. Audio Eng. Soc.*, vol. 33 no. 11, pp. 859-871 (1985 Nov.)
- [19] J. Merimaa, T. Peltonen, and T. Lokki - "Concert hall impulse responses - Pori, Finland: Analysis Results", 2005. Available at: <http://www.acoustics.hut.fi/projects/poririrs/>
- [20] O. Kirkeby, P. A. Nelson, H. Hamada, "The "Stereo Dipole" - A Virtual Source Imaging System Using Two Closely Spaced Loudspeakers" – *JAES* vol. 46, n. 5, 1998 May, pp. 387-395.
- [21] A. Farina, A. Bellini, E. Armelloni – "Non-Linear Convolution: A New Approach For The Auralization Of Distorting Systems" - *Proceedings of the 110th AES Convention*, 2001 May 12-15 Amsterdam, The Netherlands
- [22] A. Farina, E. Armelloni – "Emulation of Not-Linear, Time-Variant devices by the convolution technique" - *Congresso AES Italia 2005*, Como, 3-5 November 2005
- [23] P. Single, D. McGrath – "Implementation of a 32768-tap filter using real time fast convolution" – 87th AES Convention, New York, 18-21 october 1989.
- [24] Gardner, W.G., - "Efficient convolution without input/output delay." - *Journal of Audio Engineering Society*, 43(3):127-136, March 1995.
- [25] G. Garcia – "Optimal Filter Partition for Efficient Convolution with Short Input/Output Delay" – 113 AES Convention, Los Angeles, 5-8 October 2002.
- [26] T. G. Stockham Jr. – "High speed convolution and correlation" – *AFIPS Proc. 1966 Spring Joint Computer Conf.*, Vol. 28, Spartan Books, 1966, pp. 229-233.
- [27] Barry D. Kulp – "Digital equalization using Fourier Transform techniques" – *Pre-Prints of the 85th AES Convention*, Los Angeles, 3-6 November 1988.
- [28] Michael J Kemp - "Analysis and Simulation of Non-Linear Audio Processes using Finite Impulse Responses Derived at Multiple Impulse Amplitudes", presented at the 106th AES convention in Munich, Germany, 8 – 11th May 1999. preprint number 4919.
- [29] www.acusticaudio.com
- [30] Angelo Farina, Adriano Farina – "Realtime auralization employing a not-linear, not-time-invariant convolver" – *Pre-prints of the 123th AES Convention*, New York, 5-8 October 2007
-