

RECORDING CONCERT HALL ACOUSTICS FOR POSTERITY

ANGELO FARINA¹, REGEV AYALON²

¹ *Industrial Engineering Dept., University of Parma, ITALY*
farina@unipr.it

² *K.S. Waves Inc., Tel Aviv, ISRAEL*
regev@waves.com

The title of this paper is the same as a famous contribution given by Michael Gerzon on the JAES Vol. 23, Number 7 pp. 569 (1975) [1]. After more than 25 years the problem is still open, particularly about the optimal technique for capturing the "spatial" characteristics of the sound inside an existing theatre. A novel technique is presented here, which is compatible with all the known surround formats.

INTRODUCTION

When the famous and renowned Gran Teatro La Fenice in Venice burned during the night of 29 January 1996, one of the best sounding opera houses in the world suddenly disappeared. Its sonic behaviour, however, was at least partially saved, because several acoustical measurements had been performed just two months before, employing the binaural impulse response technique [2].

The availability of these binaural impulse responses was very relevant during the design of the reconstruction of the theatre, and demonstrated the importance of recording and storing the acoustics of concert halls for posterity.

M.Gerzon [1] first proposed to start a systematic collection of 3D impulse responses measured in ancient theatres and concert halls, to assess their acoustical behaviour and preserve it for posterity. His proposal found sympathetic response only very recently, with the publication of the "Charta of Ferrara" [3] and the birth of an international group of researchers who agreed on the experimental methodology for collecting these measurements [4].

Only a small number of theatres have yielded a complete three-dimensional impulse response characterization up till now.

Nevertheless, the techniques proposed for recording "3D" impulse responses, containing both temporal and spatial information, are actually being criticized for then employing this measured data in surround reproduction, through the auralization technique (convolution).

In fact, the two currently employed methods (Binaural measurements with a dummy head facing the sound source, and B-format measurement employing a Soundfield microphone) are both unsuitable for effective high-quality reproduction over "standard" multichannel reproduction systems (ITU 5.1). Other "alternative" loudspeaker arrays have been developed

(based on cross-talk cancellation for the reproduction of binaural material, and on Ambisonics-like decoding for the reproduction of B-format material). In some cases, these two techniques can be coupled together, for a better 3D reproduction (Ambiophonics, [5]).

Recently, a completely alternative, 2.5-D technique was proposed, based on the Wave Field Synthesis theory (WFS) and the usage of a Soundfield microphone moved around on a rotating boom [6]. Also this technique, however, is unsuitable for direct employment of the measured impulse responses over a standard surround setup.

In this paper a new measurement method is proposed, which incorporates all the previously known measurement techniques in a single, coherent approach: three different microphones are mounted on a rotating boom (a binaural dummy head, a pair of cardioids in ORTF configuration, and a Soundfield microphone), and a set of impulse responses are measured at each angular position. Fig. 1 shows a schematic of this microphone setup.

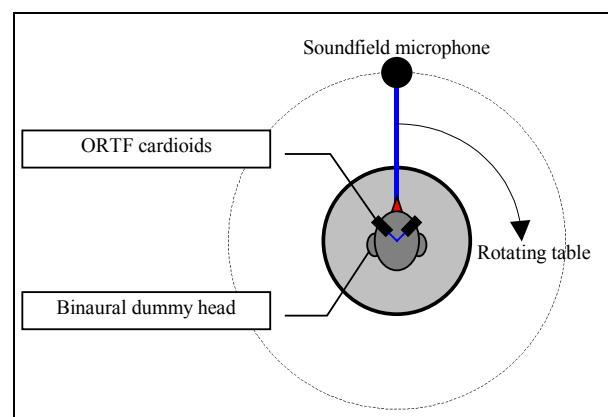


Figure 1: Scheme of microphones.

The results of this set of measurements are compatible with the already proposed methods for measurements in concert halls (binaural, B-format and WFS), but add the possibility to derive "standard surround" formats such as OCT and INA, and open the possibility to employ even the Binaural Room Scanning method [7] or the Poletti high-order circular microphones [8].

The paper describes the details of the implementation of the new measurement technique, and provides the first experimental results obtained by measurements performed in several halls.

1 MEASUREMENT METHOD

This chapter describes the details of the measurement method, the equipment (hardware and software), and the procedure.

Although most of these items are not inherently new, the combination of them in a coherent approach provides a general method from which all known multichannel formats can be derived.

1.1 Test signal and deconvolution

The excitation-deconvolution technique employed for the measurement of the impulse response is the log sine sweep method, as initially suggested by one of the authors [9]. Independent evaluations have shown that this method is superior to the previously employed ones [10,11].

A good compromise between measured frequency range, length of the sweep and signal-to-noise ratio has been reached, by choosing the following parameters:

Start frequency	22 Hz
End frequency	22 kHz
Length of the sweep	15 s
Silence between sweeps	10 s
Sweep type	LOG

The "unusual" length of the silence between sweeps is due to the traveling time of the rotating table. The rotation is triggered by a proper pulsive signal, automatically generated in the middle of the silence gap on the second channel of the sound card.

The choice of the above parameters allows for measurement of impulse responses which have wide frequency span, good dynamic range (approximately 90 dB) and are substantially immune from background noise eventually present during the measurements.

The deconvolution is obtained by linear (not circular) convolution with a proper inverse filter, which is automatically generated together with the test signal. As explained in [9], this inverse filter is simply the time reversal of the test signal, properly amplitude-equalized for compensating the 6 dB/oct falloff caused by the log sweep.

The linear deconvolution is effective in avoiding that not-linear behavior of the transducers can cause harmonic distortion artifacts affecting the measured impulse response.

As the playback-recording is performed at 96 kHz-24 bits, there is enough distance between the maximum generated frequency and the Nyquist frequency, that the ringing of the anti-aliasing filters is not excited, and the measured impulse response does not suffer from high-frequency phase distortion.

Also the amplitude of the emitted test signal has been properly amplitude-equalized, for compensating the uneven frequency response of the loudspeaker: this way, the emitted sound power has a reasonably flat spectrum over the whole frequency range.

Figs. 2 and 3 show respectively the equalized test signal (CoolEditPro was employed for playback & recording) and the user's interface of the software employed for the deconvolution. Thanks to the usage of the new, highly optimized Intel Integrated Performance Primitives v. 3.0 FFT routines, the deconvolution is now incredibly fast (approximately 20% of the duration of the recorded signal).

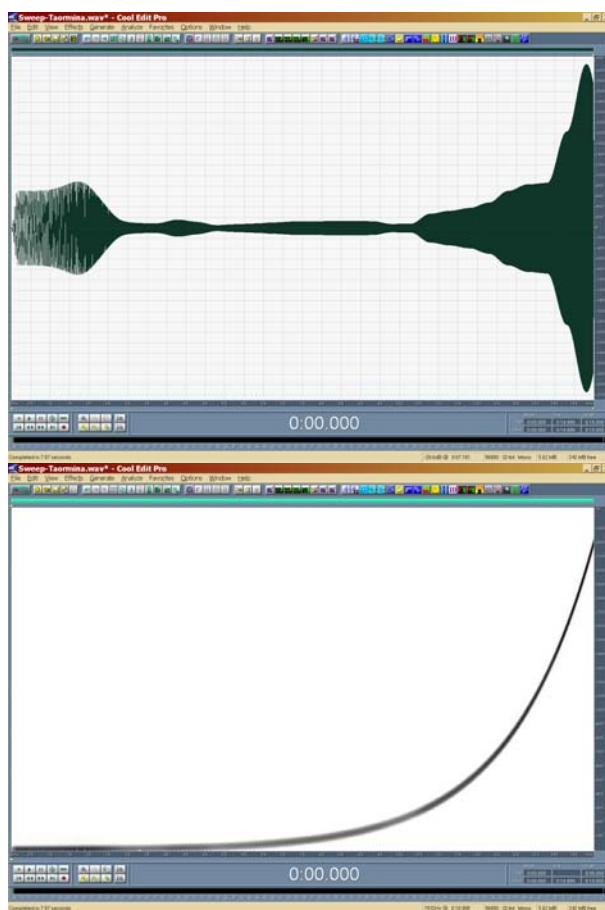


Figure 2: Equalized test signal.

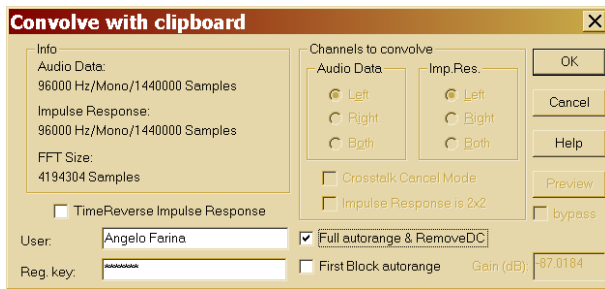


Figure 3: Fast convolver employed for deconvolution.

1.2 The sound source

An omnidirectional sound source is usually preferred for measurements of room impulse responses. Albeit this does not correspond to the effective directivity pattern of real-world sound sources (such as musical instruments or human talkers and singers), the usage of an omnidirectional sound source is predicated by current standards (ISO3382, for example), and avoids XX exploiting strange room effects, as can happen employing highly directive loudspeakers (abnormal energization of echoes and focalizations for selected orientations of the source).

A special, ultra-compact dodechaedron loudspeaker was built specifically for the purpose of this research, employing 12 full-range drivers installed on a small size enclosure (approx. diameter is 200 mm). This unit, of course, is not capable of producing significant acoustical power under 120 Hz; for extending the low frequency range a subwoofer was added, incorporating it inside the cylindrical transportation case, which also contains the power amplifier (300 W RMS) and serves as supporting base for the dodechaedron.



Figure 4: Dodechaedron loudspeaker and subwoofer.

Fig. 4 shows a photograph of this special omnidirectional sound source.

The acoustical performance of the loudspeaker was measured inside an anechoic room, averaging the radiated sound over a complete circumference. As the 1/3 octave spectrum measured when feeding the loudspeaker with perfectly flat pink noise was significantly uneven, a proper equalization of the test signal was necessary. Fig. 5 shows the comparison between the radiated sound power of the loudspeaker prior and after the equalization, which was performed applying directly to the test signal the graphical 1/3 octave filtering required for flattening the response.

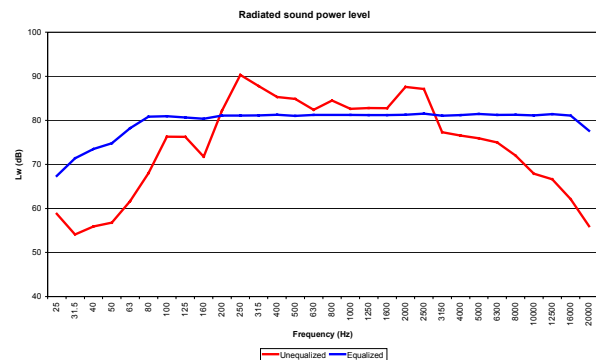


Figure 5: Spectra of the radiated sound power.

From the graph, it can be seen how the digital equalization was capable of flattening perfectly the loudspeaker's response between 80 and 16000 Hz, with a gentle roll-off outside this interval. After the equalization, the total radiated sound power level (with pink noise) was approximately 97 dB.

1.3 The microphones

Three different microphonic probes were employed:

- a pair of high quality cardioids in ORTF configuration (Neumann K-140, spaced 180mm and diverging by 110°);
- a binaural dummy head (Neumann KU-100);
- a B-format 4-channels pressure-velocity probe (Soundfield ST-250).

All these microphones were installed over a rotating table, in such a way that the rotation center passed through the center of the dummy head, and through the point at the intersection of the axes of the two cardioids (which were mounted just above the dummy head). alternatively the Soundfield microphone was displaced exactly 1m from the rotation axis, in front of the dummy head.

The rotating table (Outline ET-1) was programmed for stopping each 10°, and consequently along a complete rotation 36 discrete sets of impulse responses were measured at each position of the microphonic array.

Fig. 6 and 7 show photographs of the microphone setup.



Figure 6: The microphones over the rotating table.



Figure 7: Closeup of the microphones.

1.4 Computer and sound card

The measurement method required the usage of a top-grade sound card, equipped with 8 analog inputs at 24 bits / 96 kHz, incorporating digitally controlled mic preamplifiers (for ensuring accurate control of the input gain, and relative and absolute calibration of the recordings). At the moment, these requirements can only be fulfilled by external rack-mounted units, connected to the computer by means of a PCI card.

This impeded the usage of any portable computer, and forced the choice of the only currently-available fanless PC, which stands out for its completely silent design: the Signum Data Futureclient.

The model employed for this research mounts a α 1.8 GHz P-IV processor, and is equipped with 512 Mbytes Ram and an high speed (7200 RPM) hard disk. This allows for faultless operation when recording 8 channels and playing 2 channels at 96 kHz, 24 bits.

The sound card chosen for the task is an Aadvark Pro-Q10. Fig. 8 shows a picture of the equipment, which is installed inside a couple of fly-cases for easy transportation.



Figure 8: Liquid-cooled PC (FutureClient).

1.5 Measurement method

CoolEditPro was employed for the playback of the test signals and the simultaneous recording of the 8 microphonic channels. The test signal was looped 36 times, corresponding to the 36 steps of the rotating table along a complete rotation.

The following picture shows a multi-track session, resulting from a measurement with the above-described approach.



Figure 9: Multitrack session of a measurement

Each measurement takes approximately 15 minutes (25s x 36 repetitions); after the measurement is complete, Another 10 minutes are required for storing all the waveforms on the hard disk (in 32-bits format, for preserving all the available dynamic range); during this time, the source and/or the microphonic array are displaced into another position.

1.6 Measured data

At the time of writing, 9 famous theaters were measured with the previously described method, as reported in the following table.

N.	Theatre	N. sources/ receivers
1	Uhara Hall, Kobe, Japan	2/2
2	Noh Drama Theater, Kobe, Japan	2/2
3	Kirishima Concert Hall, Kirishima, Japan	3/3
4	Greek Theater in Siracusa, Italy	2/1
5	Greek-Roman Theater in Taormina, Italy	3/2
6	Auditorium of Parma, Italy	3/3
7	Auditorium of Rome (Sala 700), Italy	3/2
8	Auditorium of Rome (Sala 1200), Italy	3/3
9	Auditorium of Rome (Sala 2700), Italy	3/5

However the number of rooms being measured is increasing quickly, and it is planned to reach at least 30 different rooms in less than 6 months.

The goal of this paper is not to present a comprehensive comparative study of the measured data, which will follow when the collection of impulse response responses is complete, and all the results are fully analyzed.

However, the next figure shows a set of 36 impulse responses measured in the Auditorium of Parma, for giving an idea about the format in which the data are stored: for each microphone pair (Neumann ORTF in this case) the 36 impulse responses measured during the microphone rotation are stored one after the other, and the sequence is saved as a 32-bits float WAV file.

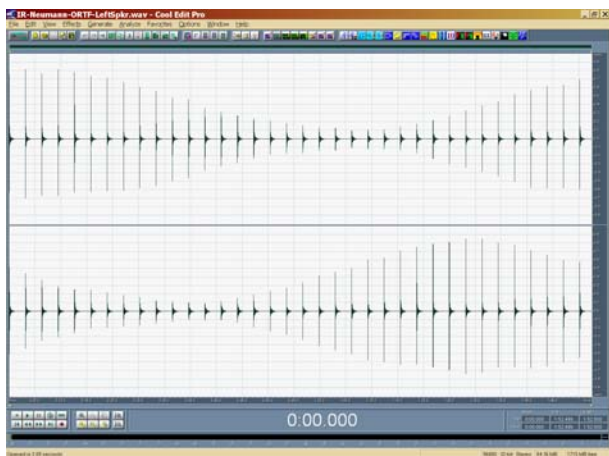


Figure 10: Measured impulse responses (36 microphone positions)

2 EXTRACTION OF OBJECTIVE ACOUSTICAL PARAMETERS

Basically, the computation of objective acoustical parameters is based on the ISO 3382-1997 standard. Most parameters are computed from an impulse response captured with an omnidirectional microphone, which is substantially the channel W of the Soundfield microphone, at the initial position (0 degrees).

However, the spatial parameters require processing stereo impulse responses: consequently, also the binaural and the WY pair had to be processed.

This research is devoted mainly to capturing and analyzing the spatial properties of the sound field, with the goal of creating realistic multichannel surround reconstructions: consequently the greater effort was reserved for the analysis of the spatial parameters.

The highly innovative result made available from the new measurement technique is the possibility to measure and display polar plots of the spatial acoustical parameters, showing their variation along with the rotation of the receiver.

2.1 Reverberation time

The W channel of the B-format impulse response is employed (omnidirectional). The impulse response is first backward-integrated, following the Schroeder method, and applying the noise-removal allowed by the ISO 3382 standard.

Then the reverberation time T30 is computed, by means of a linear regression over the decay curve in the range between -5 and -35 dB below the steady-state level before the decay. It must be noted that usually these impulse responses are so clean and noiseless that it would be possible to measure directly the T60 (in the range -5 to -65 dB), but the ISO3382-1997 standard does not allow for this (it was written when measurement of impulse responses with such high dynamic range was very difficult to obtain).

Fig. 11 shows a typical plot of the impulse response and of the backward-integrated decay curve obtained in one of the theaters objects of this research.

The picture shows that the total integrated sound pressure level is approximately 90 dB above the steady background noise present after the impulse response is finished.

2.2 Monophonic temporal criteria

Although the reverberation time is the most important criterion for evaluating the acoustical behaviour of a room, it is often advisable to get a better insight about the fine temporal distribution of the acoustical energy.

For this goal, the ISO 3382 standard suggests the usage of 4 temporal-monoaural criteria: C₅₀, C₈₀, D, T_s.

C⁵⁰ is the Clarity over 50ms, evaluated by applying the following formula over the measured omnidirectional pressure impulse response, and starting from the arrival

time of the direct sound:

$$C50 = 10 \cdot \lg \left[\frac{\int_0^{50\text{ms}} p^2(\tau) \cdot d\tau}{\int_0^{\infty} p^2(\tau) \cdot d\tau} \right] \quad (1)$$

C_{80} is similar, but the time boundary is moved from 50 ms to 80 ms. Usually C_{50} is considered more representative of the clarity of speech, whilst C_{80} is more relevant for assessing clarity of the instrumental music.

D is somewhat similar to C_{50} , but it is expressed in % instead of in dB, following this equation:

$$D = \frac{\int_0^{50\text{ms}} p^2(\tau) \cdot d\tau}{\int_0^{\infty} p^2(\tau) \cdot d\tau} \cdot 100 \quad (2)$$

Finally, the Center Time T_s is defined as:

$$T_s = \frac{\int_0^{\infty} \tau \cdot p^2(\tau) \cdot d\tau}{\int_0^{\infty} p^2(\tau) \cdot d\tau} \quad (3)$$

Which has the advantage of avoiding a steep separation between the “early” and “late” energy, inherent in the definition of C and D.

The computation of all the above parameters, and of the reverberation time, is made thanks to a proper plugin, developed with the goal of automatizing the computation of the ISO 3382 Acoustical Parameters. Fig. 11 shows the user’s interface of this plugin.

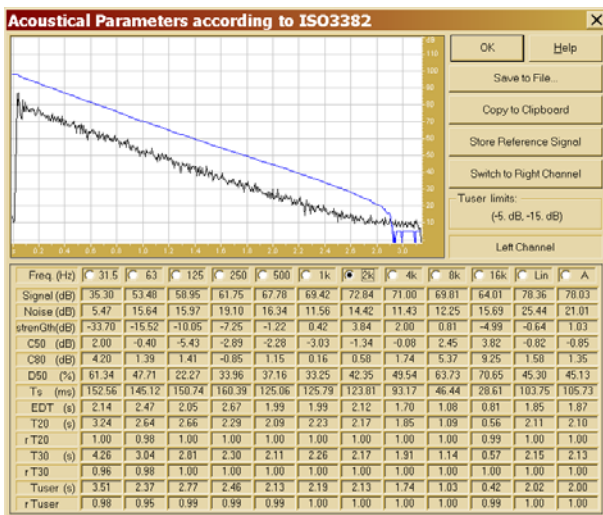


Figure 11: ISO 3382 acoustical parameters

2.3 Absolute and relative sound pressure level

As the acoustical power of the sound source was carefully calibrated thanks to the anechoic-room measurements, and having care of keeping track of the gain applied in the microphone preamplifiers, it is possible to know with reasonable accuracy (+/- 1 dB) the absolute sound pressure level captured during the measurement.

Furthermore, as the deconvolution of all the impulse responses of a given theater is done employing the same rescaling factor, the displayed amplitude of the impulse responses preserves a relative scaling.

The difference between the absolute SPL and the radiated sound power level L_w allows for the computation of a very relevant acoustical parameter, the Strength G:

$$G = \text{SPL} - L_w + 31 \quad \text{dB} \quad (4)$$

The corrective factor of +31 dB derives by the definition of G, which refers to the difference between the measured SPL inside the room and the theoretical SPL measured in free field, at a distance of 10m from the source.

2.4 Binaural spatial criteria (IACC)

Following Ando’s theory [12], the basic binaural parameter is the Inter Aural Cross Correlation (IACC), defined as the maximum value of the Normalized Cross Correlation function:

$$\rho(\tau) = \frac{\int_{-\infty}^{\infty} h_d(\tau) \cdot h_s(\tau + t) \cdot d\tau}{\sqrt{\int_{-\infty}^{\infty} h_d^2(\tau) \cdot d\tau \cdot \int_{-\infty}^{\infty} h_s^2(\tau + t) \cdot d\tau}} \quad (5)$$

Other related parameters are τ_{IACC} and w_{IACC} , defined respectively as the delay (in ms) of the maximum value of the normalized cross correlation function, and as the width of the peak (at 10% of the maximum) in ms.

A special plugin was created for measuring the IACC-based parameters. This plugin also computes the time delay gap between direct sound and first reflection, and the T_{sub} (subsequent reverberation time), conforming to the Ando’s theory. Fig. 12 shows the user’s interface of this plugin.

Traditionally, this measurement is performed when the binaural dummy head is pointed directly towards the sound source. In this case, however, the head is pointed in 36 different directions, with 10° steps. Consequently, 36 values of IACC are obtained, and it is possible to create a polar plot of IACC.

The availability of these polar plots is new, and it is yet to be evaluated what information can be extracted from them. What immediately appeared, however, is that

rooms with almost the same value of the “standard” IACC can have quite different polar plots, showing that the “surround” properties of the room are not completely described by the old-style, single-valued “standard” IACC.

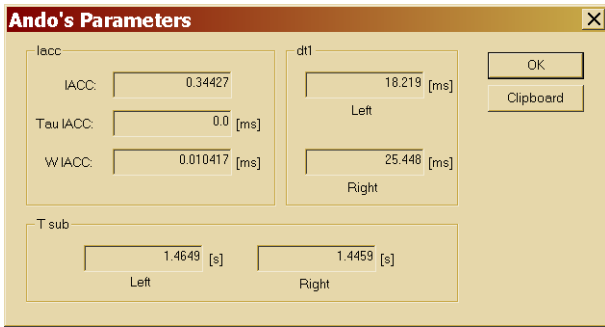


Figure 12: Ando’s parameters plugin

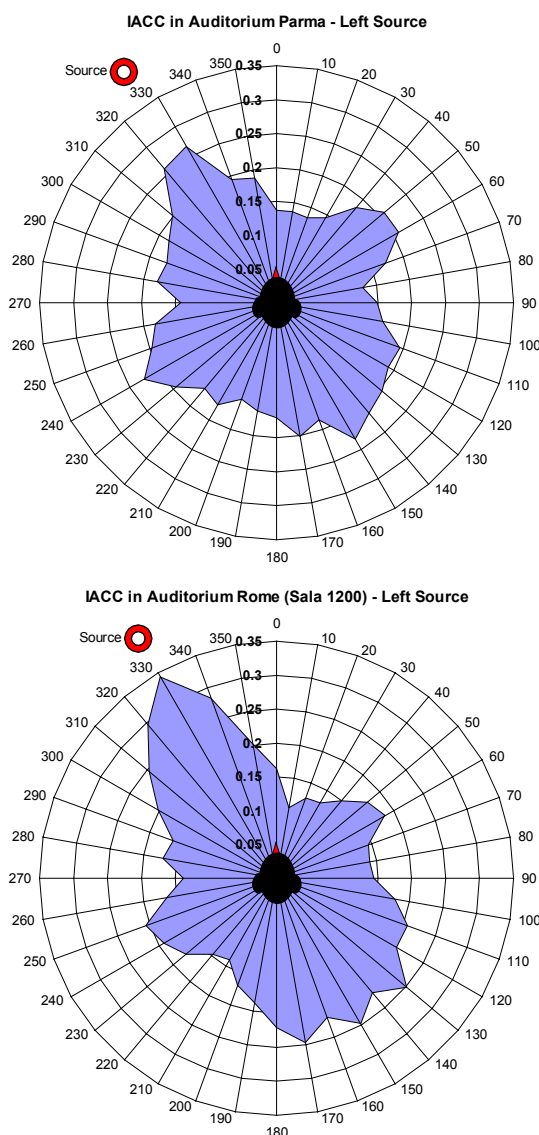


Figure 13: Polar Plots of IACC in Parma and Rome.

This is proven by the comparison of the polar plots reported in fig. 13, which refers to the Auditorium of Parma vs. the Auditorium of Rome. In the latter, the sound appears to be more strongly “polarized”, whilst in the Auditorium of Parma it is more diffuse.

2.5 B-format spatial criteria (Lateral Fractions)

The ISO 3382 standard defines two spatial descriptors derived by a B-format impulse response (more precisely, by the W and Y channels of a B-format impulse response), called respectively LF and LFC. LF is the ratio between the early lateral sound and the omnidirectional sound:

$$LF = \frac{\int_{0ms}^{80ms} h_Y^2(\tau) \cdot d\tau}{\int_{0ms}^{80ms} h_W^2(\tau) \cdot d\tau} \tag{6}$$

For the application of the above formula to the measurement with a Soundfield microphone, it must be noted that the X axis should be horizontal and pointing towards the sound source, the Y axis is horizontal and orthogonal to X pointing in the direction of the left ear, and the Z axis is pointing to the ceiling. Furthermore, it is necessary to compensate for the fact that the W channel (omni) has a gain 3 dB lower than X, Y and Z. The second parameter, LFC, is defined by:

$$LFC = \frac{\int_{0ms}^{80ms} h_Y(\tau) \cdot h_W(\tau) \cdot d\tau}{\int_{0ms}^{80ms} h_W^2(\tau) \cdot d\tau} \tag{7}$$

In this case the numerator equals the Sound Intensity, whilst the denominator equals the squared RMS sound pressure. In substance, LFC is a parameter quite close to the definition of the pressure-intensity index usually employed in applications of sound intensity measurement system (ISO9614).

Also for these B-format based parameters a special plugin was developed: its user interface is shown in the next figure.

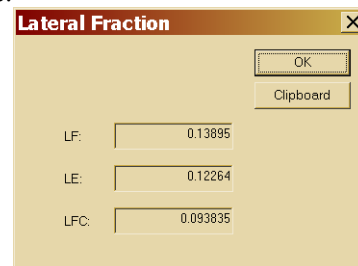


Figure 14: Lateral Fraction parameters plugin

It must be noted that the plugin also computes the Jordan’s LE (Lateral Efficiency) parameter [13], which definition resembles LF, but with a starting time limit for the integral at numerator equal to 25ms instead of 5ms.

As the Soundfield microphone can be “virtually rotated” around its axis, it is easy, from a single B-format impulse response, to compute a complete polar plot of LF. But the microphone was not simply rotated, it was displaced along a circumference with 1m radius. So, taking for each microphone position the radial orientation of the microphone, it is also possible to build a modified polar plot, which shows the variation of LF (or 1-LF) along the circumferential path described by the microphone.

The following picture shows these polar plots for the same two rooms already analyzed with the IACC.

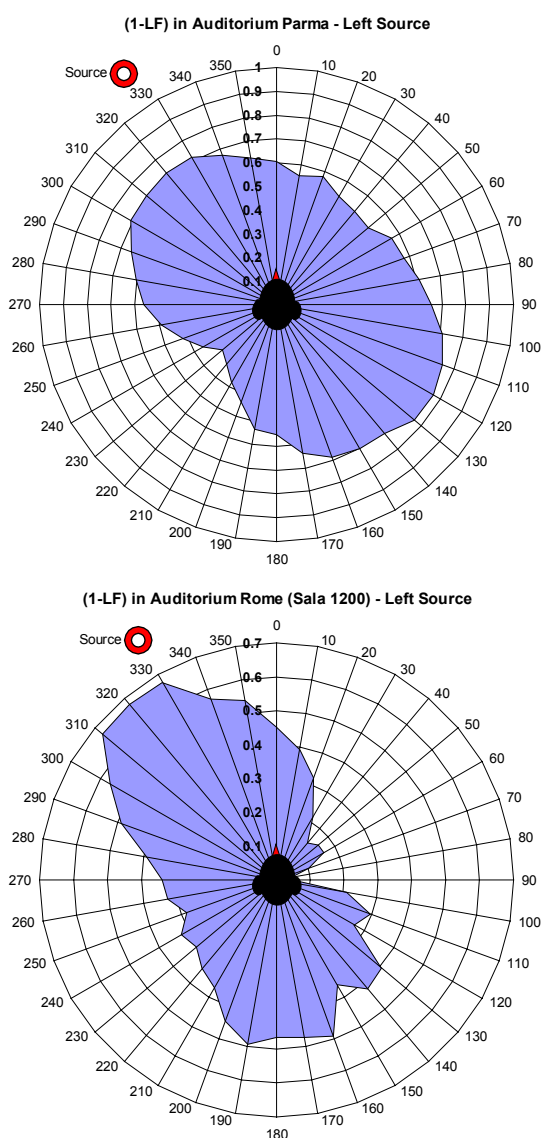


Figure 15: Polar Plots of (1-LF) in Parma and Rome.

It must be observed that employing (1-LF) the parameter has the same polarity as IACC, so the polar plots of fig.15 are directly comparable to those of fig.13. Also in this case it is quite evident how the sound field is much more diffuse in the Parma Auditorium, whilst in Rome Auditorium the sound is more polarized. In the second, furthermore, there is a small angular sector where LF is almost unitary (and consequently 1-LF is zero).

Analyzing the results, shows little significance for the parameter LFC (which is always very little, independent of the room and of the orientation of the probe) and the weak dependence on the orientation of the probe of LE. LF is confirmed to be the more sensitive parameter based on B-format impulse responses, although it is also clear how the ranking of the spatial impression based on LF does not necessarily correspond with the ranking based on IACC. The following table compares the values of IACC and (1-LF) for the two cases already reported on fig. 13 and 15:

Auditorium	IACC	1-LF
Parma	0.266	0.725
Rome	0.344	0.676

From the above table, looking at IACC Parma seems to have greater spatial impression than Rome, whilst looking at LF the opposite judgment is obtained.

This means that the information about sound diffusion derived from these two descriptors can be misleading, and that the true evaluation of the two rooms actually characterized by a more enveloping soundfield cannot be derived just by the parameters computed pointing the microphones towards the sound source, but instead requires one to analyze the variation of the spatial parameters when the microphones are rotated in all directions.

The subjective listening experience of the authors clearly indicates, in the above two cases, that the Parma Auditorium is significantly more diffuse than the “sala 1200” of the Rome Auditorium, and the same conclusion appears evident when comparing the polar plots, both in fig. 13 and in fig.15.

2.6 Criticism of ISO3382 parameters

Applying the ISO 3382 parameters to these high-end impulse responses has shown how this standard, albeit having been updated in 1997, already requires substantial revision. In practice, three main topics require refinement:

- The standard does not give proper indications for sweep-based measurements, nor discusses the issues which make the sweep method preferable to MLS (time invariance, non-linearity, clock mismatch tolerance, etc.)

- Almost all parameters are said to be related to the “acoustical energy”, but they are actually computed over the squared pressure. From a B-format measurement, instead, the true values of active intensity and sound energy density are available. And it is well known how, in a partially reactive sound field, the true energetic parameters can differ significantly from the estimates based on the squared pressure.
- The definition of the spatial parameters (either based on binaural or B-format impulse responses) assumes a specific orientation of the microphone, pointing to the sound source. This is meaningless in presence of multiple sources, or in rooms equipped with sound reinforcement systems. Also in case of a single point source, these parameters give contradictory results.

3 AURALIZATION OF THE MEASURED DATA

This chapter analyzes the possibility of employing the results of these measurements for creating audible presentations of the acoustical behavior of the original rooms, to listeners exposed to an artificial soundfield, by means of headphones or loudspeakers.

The basic method for auralization is convolution: the impulse responses are employed as very long FIR filters, applied to dry (anechoic) recordings of music or speech. Convolution is a very efficient filtering technique, particularly if implemented with proper (old) algorithms on fast (new) processors: as clearly demonstrated in [14], a PC equipped with a last-generation processor can perform the real-time, low-latency convolution of dozens of channels with multiple impulse responses of hundreds of thousands of coefficients each. And the performances obtained with the simpler algorithms initially developed in the sixties [15] are better than those obtained with more recent developments [16], which appear to be preferable from the point of view of the total number of multiplications required, but are much less optimized for the memory-management architecture of modern processors.

The goal of this research is to create sets of impulse responses suitable for being employed by these software convolvers, creating the results in any of the currently available formats suitable for multichannel reproduction, and attempting to recreate as faithfully as possible the spatial attributes of the original soundfield.

3.1 ORTF-stereo impulse responses

This is the most basic processing, aimed at the creation of a “standard” stereo presentation of the results of the auralization. The process is based on the availability of a number of dry mono recordings, one for each section of the orchestra or for each singer.

Each mono recording has to be convolved with a specific stereo impulse response, obtained by the pair of cardioid microphones in ORTF configuration. In principle, each of these impulse responses should be measured with the proper position of the sound source.

In reality, the measurements are typically performed with just three positions of the source on the stage (Left, Right, Center), and this limits the number of independent “virtual sources” which can be placed on the sonic scene.

In practice, however, it is possible to take advantage of the fact that, for each source position, the ORTF measurement was performed with 36 different orientations of the microphones (in 10 steps). This means that some minor adjustment of the virtual source position (by 10 or 20 degrees) can be obtained by selecting the ORTF impulse response coming from an orientation different than 0°. This of course is not perfectly rigorous, but is effective and subjectively undistinguishable from convolution with ORTF impulse responses measured with microphone orientation at 0° and true displacement of the source.

Of course, the results of the convolution of all the dry recordings are summed in a single stereo output file, which is suitable for reproduction in a normal stereo system (2-loudspeakers).

3.2 Binaural impulse responses (binaural room scanning)

The basic binaural approach is substantially the same as for the previous ORTF-based method, but employing the binaural IRs. This way, the result of the convolution is a 2-channels file, suitable for headphone reproduction.

However, two methods can be employed for substantially improving the surround effect obtained: for loudspeaker reproduction a proper cross-talk cancellation must be added, and for headphone reproduction an head-tracking sensor can drive a real-time convolver, switching the impulse responses being convolved as the listener rotates his head.

Regarding the creation of optimal cross-talk cancelling filters, and optimal layouts for the loudspeakers employed for the reproduction, several papers were published in recent years [17,18].

Regarding instead the head-tracking real-time processing, some solutions were proposed by LakeDsp [19] and Studer [7], but requiring dedicated and expensive DSP-based workstations. The authors are working at a new, low-cost system for real-time auralization, making use of a game-quality head tracking system and a new, high efficiency, low latency convolution software.

3.3 B-format impulse responses (Ambisonics)

In this case, each dry mono source is convolved with the proper B-format impulse response. So, after the mixing of all these convolutions, a 4-channels B-format output is obtained.

The reproduction of a B-format signal over a suitable array of loudspeakers requires an Ambisonics decoder, for computing the proper feed for each speaker.

The creation of a software-based decoder has been pioneered by one of the authors [20], and has been further perfected by colleagues at the University of York, who recently released for free a suite of VST plugins [21], allowing for manipulation and decoding of B-format signals over various loudspeaker rigs.

In conclusion, the Ambisonics auralization simply requires the availability of a multichannel convolver (with 1 input and 4 outputs), a B-format mixer, and a B-format Ambisonics decoder. The first tool is being developed by Waves, the second and third tools are already available from [21].

3.4 ITU 5.1 surround (from selected B-format impulse responses)

The basic approach for ITU 5.1 rendering is to first select a configuration of microphones to be employed, for driving the 5 main loudspeakers [22]. Many of these microphone arrangements have been proposed, and in a recent round-robin project, called the Verdi project, most of them were comparatively evaluated [23].

Here we consider just three of them, which got good results in the aforementioned comparative test: Williams MMA [24], OCT [22] and INA [25].

The following pictures show the microphone configurations for these three setups:

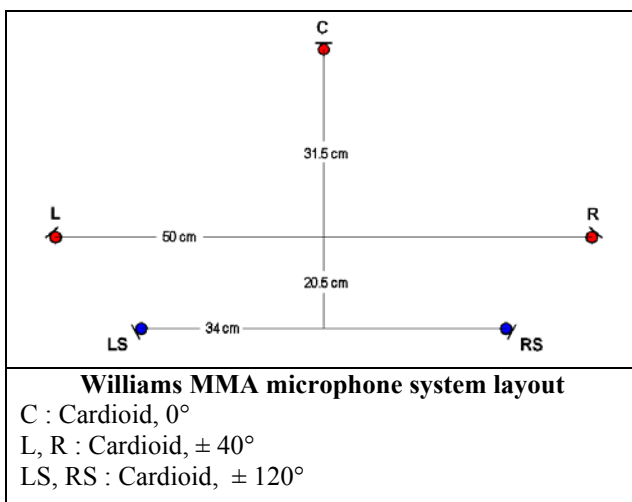


Figure 16: Layout of microphones (Williams MMA)

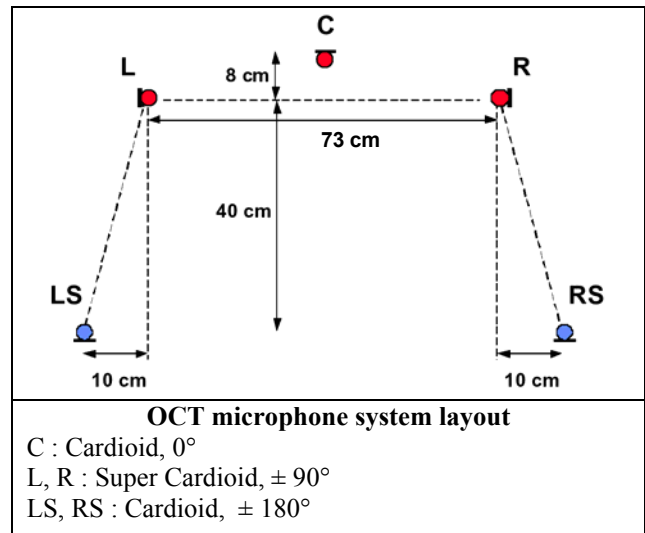


Figure 17: Layout of microphones (OCT)

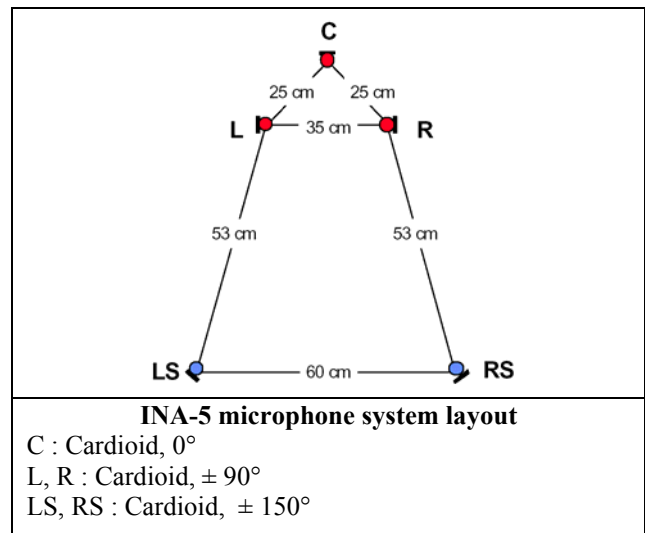


Figure 18: Layout of microphones (INA)

For each of the above setups, it is possible to select a subset of 5 of the 36 positions where the Soundfield microphone was displaced, corresponding as close as possible to the intended positions of the chosen setup. Then, from the B-format impulse response measured in each of these 5 selected positions, a single (mono) impulse response is extracted, thanks to the program Visual Virtual Microphone, developed by David McGriffy and freely available on the Internet [26]. Fig. 19 shows the user's interface of this program, when employed for extracting the hypercardioid response for the R channel of an OCT setup from the B-format impulse response coming from the 20° position, and with the sound source on the left of the stage.

It must be noted that the measurements performed with the rotating Soundfield microphone inherently assume a clockwise angle (due to the fact that the rotating table

only turns in this way), whilst usually in surround-sound applications a counter-clockwise angle is employed.

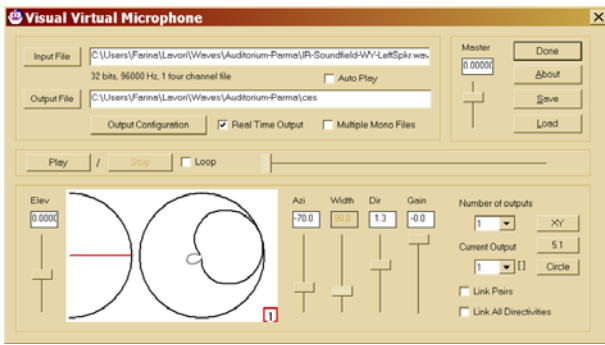


Figure 19: Visual Virtual Microphone

As the microphone in this position was already tilted 20° on the right, and OCT mandates for an orientation of the Right supercardioid of 90°, a further rotation of 70° has to be implemented in the Visual Virtual Microphone program.

In the case where the chosen microphonic setup requires a microphone position which is not actually lying over the 1m-radius circumference, it is possible to use the WFS method (par. 3.6) for extrapolating the impulse response in the required position.

Finally, each mono dry source is convolved with the 5-channels impulse response derived from the corresponding sound source position over the stage, and the results of all these convolutions are mixed in a single final 5-channels track, which is suitable for reproduction over a standard ITU loudspeaker rig.

3.5 Mark Poletti's high-directivity virtual microphones

During the rotation of the microphonic assembly, the two cardioids employed for ORTF recordings also describe a small circumference, with a radius of approximately 110 mm, as shown in fig. 20.

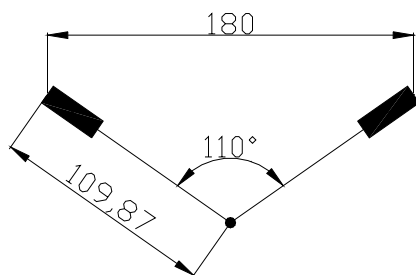


Figure 20: geometry of ORTF microphones

Looking for simplicity to just one of the two microphones, it samples 36 impulse responses during its complete rotation. From this set of data, it is possible to derive the responses of a set of various-orders coincident microphones, ideally placed in the center of

rotation, making use of a modified version of the Poletti's theory [8].

The basis of this method is to define a class of multileaf-shaped horizontal directivity patterns of various orders. The order 0 is an omnidirectional, order 1 are two crossed figure-of-eight microphones (as in horizontal-only Ambisonics); then order 2 and 3 are added, with directivity patterns corresponding respectively to the cosine of twice and three times the angle:

$$\begin{aligned}
 D_0 &= 1 \\
 D_{1,n} &= \cos\left(\vartheta + n \cdot \frac{\pi}{2}\right) \quad n = 0,1 \\
 D_{2,n} &= \cos\left(2 \cdot \vartheta + n \cdot \frac{\pi}{4}\right) \quad n = 0,1 \\
 D_{3,n} &= \cos\left(3 \cdot \vartheta + n \cdot \frac{\pi}{6}\right) \quad n = 0,1
 \end{aligned}
 \tag{8}$$

The responses of these virtual microphones can be thought of as a cylindrical harmonics decomposition of the sound field at the center position, or as a spatial Fourier analysis of the soundfield done along the angular coordinate ϑ .

The second explanation suggests a simple way of computing the required responses: the signals coming from the 36 microphones are simply multiplied for a set of 36 weighting factors, obtained by the eqn. 8 above, and summed.

This of course does not provide the wanted frequency-independent, linear-phase result: as clearly demonstrated by Poletti, these “raw” virtual microphones will exhibit strongly uneven magnitude and phase response, which can however be compensated afterwards.

Poletti also derived the theoretical expressions of the transfer functions, which can be used for creating the proper equalizing filters. However, a more clever and practical solution is simply to measure these “raw” transfer functions in an anechoic chamber, and then derive, for each virtual microphone, the proper inverse filter by means of the Kirkeby inversion method [18]. This has the added advantage of compensating also for the specific response of the microphone employed, and for its frequency-dependent directivity pattern (which will only roughly correspond to the theoretical cardioid pattern).

Once the response of the high-order microphones are obtained, they can be employed as convolution filters applied to the mono dry signals corresponding to the discrete source positions. After mixing of the results, an high-order Ambisonics decoder is required for deriving the feeds for a multichannel regular array of loudspeakers (typically arranged regularly around a circle surrounding the “sweet spot”), which provides

much better localization and channel separation than “standard” (1st order) Ambisonics.

A second possible way of employing these high-order signals is to drive a standard 5.1 ITU array, by synthesizing 5 proper asymmetrical directivity patterns, as suggested in [27].

3.6 Circular WFS approach

The 36 B-format measurements made along the 1m-radius circumference are exactly the set of data required for employing the WFS method described in [6].

The basis of this method is the Huygens principle: knowing the sound pressure and particle velocity on a closed surface makes it possible to recreate inside it the same sound field which was present in the original space, employing a suitable array of loudspeakers exactly corresponding to the positions of the microphone. The theory, however, also allows to “expand” or “shrink” the geometry of the transducer array, provided that the soundfield is decomposed in traveling wavefronts.

The WFS is a 2D reduction of this general theory, where the microphones are placed along a closed curve around the listening area, and consequently the expansion/shrinking can only be done in the horizontal plane. This also limits the amount of “movement” which can be applied. However, starting with a 1m-radius array, it is quite easy to derive the feeds for a loudspeaker array suitable for a medium-sized listening room, and to “stretch” the array so that the loudspeakers are arranged in 4 linear arrays instead of in a circular array. The next figure (partially taken from [27]) shows a schematic of this process.

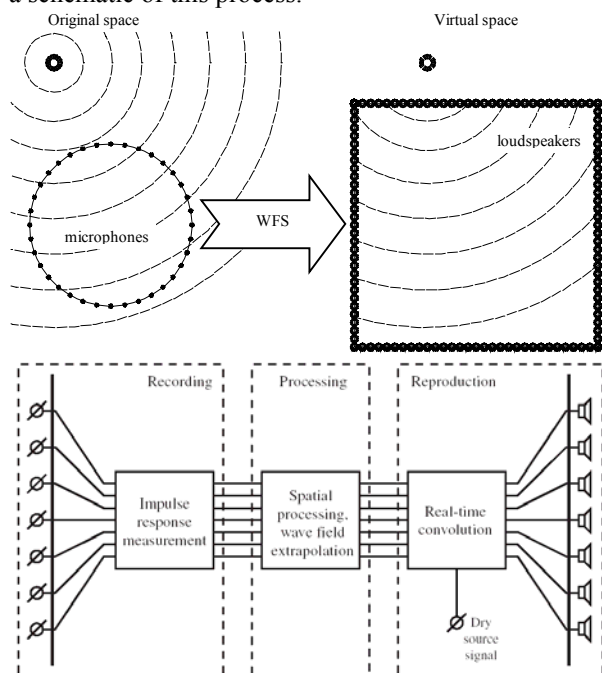


Figure 21: WFS processing scheme

The “spatial processing” required for deriving the reproduction impulse responses from the measured impulse responses is not trivial, and can be understood only after a deep study of the material published (and unpublished) at the Technical University of Delft. Till now the authors were not yet able to create a simple plugin for performing easily this spatial transformation, although this development is planned for the future.

Of course, this theory requires a little spatial step between consecutive microphone positions, for reducing the spatial aliasing which occurs when sampling the wavefronts. As in this case the number of microphone positions is quite limited (36), this translates in a severe limitation of the frequency range which does not cause spatial aliasing. Above this threshold (which is around 1 kHz for the geometry employed here), it is not possible anymore to reconstruct faithfully the wavefronts. For avoiding artifacts and coloration, it is then advisable to randomize the phases, so that the summation of the output of the various loudspeakers constituting the array does no longer cause interference, and reduces to simple energy summation (as in Ambisonics).

The phase randomization can be obtained by convolution of the signal driving each loudspeaker with a different burst of white noise, or by employing phase-incoherent loudspeakers (distributed-mode loudspeakers).

3.7 Hybrid methods (Ambiophonics)

The Ambiophonics method is an hybrid solution, aimed to mask the defects of two basic systems: cross-talk cancelled reproduction of binaural material over closely-spaced loudspeakers (Stereo Dipole) and 3D surround driven by convolution of corresponding oriented virtual microphones.

The following figure shows a typical Ambiophonics array, (frontal stereo dipole, plus 8-loudspeakers surround rig).

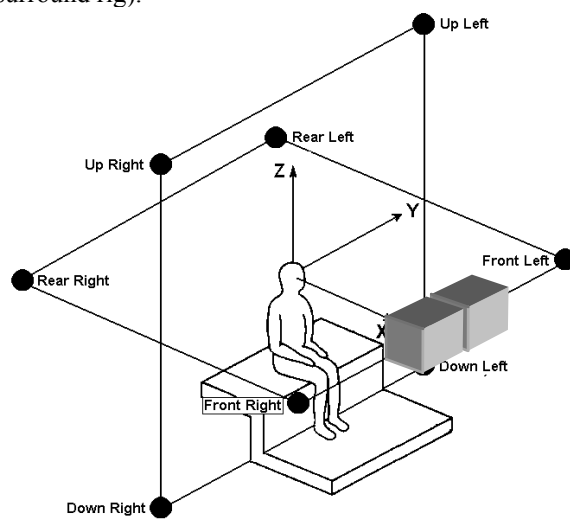


Figure 22: Ambiophonics array

The theory for deriving the signals for these loudspeakers has been already presented in the previous chapters, and the assembly of the whole system has been thoroughly described in [5]. The only point which deserves discussion here is the fact that, in an Ambiphonics system, the Stereo-Dipole loudspeakers should provide only the direct sound and early reflection from the stage enclosure, whilst the other “surround” loudspeakers should provide the late reflections and the reverb.

This means that the measured impulse responses need to be properly edited: the ORTF ones, which are employed for the Stereo Dipole, need to be cut smoothly just after the direct sound. On the other hand, the B-format impulse responses, from which the surround channels are derived, need to be deprived of the direct sound.

The management of this editing is quite delicate, because, if it is done improperly, it can cause an improper merging between the two basic systems, or can introduce artificial delays which alter the temporal distance between the direct sound and the subsequent reverberation.

The final remark regards the selection of the impulse responses for driving the “surround” array. In [5], these IRs were all derived from a single B.-format impulse response, simply employing Visual Virtual Microphone and pointing the virtual microphone in a direction corresponding to that of the corresponding loudspeaker. Now, the availability of many B-format impulse responses along a circle, makes it possible to select, for any “surround” loudspeaker, not only the direction of the virtual microphone, but also a corresponding position of it along the circumference.

This ameliorates significantly the results, because this way the impulse responses are sampled in different positions, and are mutually incoherent. This avoids interference and artifacts due to the interaction of signals coming from many loudspeakers, all fed with strictly correlated signals.

4 CONCLUSIONS

This paper has described a new, advanced measurement technique, which allows for capturing the widest possible acoustical information inside an existing theatre. The method is based on the measurement of a huge number of impulse responses, by means of a rotating microphonic set-up.

From the set of data measured, it is possible to derive subsets of impulse responses suitable for the reproduction of the virtual acoustic space, following the currently available reproduction technologies. Referring in particular to the reproduction of the spatial properties of the sound field, it is noticeable that the measured data allow for the auralization of the results employing:

- Standard stereo reproduction over a pair of loudspeakers;

- Binaural reproduction over headphones, with head tracking;
- Reproduction over closely-spaced loudspeakers (by means of cross-talk cancelling filters);
- Ambisonics reproduction over a 2D or 3D regular array of loudspeakers
- ITU 5.1 “surround” reproduction conforming to “standard” microphonic setups (OCT, INA, etc.)
- High directivity, multichannel reproduction by means of Mark Poletti’s circular-array method.
- Wide-area auralization by means of the Wave Field Synthesis approach (WFS)
- Any combination of the above methods, resulting in hybrid, higher level surround methods (Ambiphonics, Panoramiphonics and derivations).

Consequently, this method provides the best available approach for storing the acoustical properties of famous and valuable rooms, such as concert halls and theatres, and preserving them for the posterity. The resulting data can be used for audible reconstructions (auralization) by means of today’s surround systems, without limiting the future usage by sticking to the limited reproduction technology currently available.

On the other hand, the measured sets of data can immediately be employed for high-quality processing of dry recordings, outperforming current “artificial” reverberation and spatialization units, if employed together with a state-of-the-art convolution software.

ACKNOWLEDGMENTS

This research was funded and logistically supported by Waves (www.waves.com), as part of the development of a new reverberation tool based on sampled acoustical impulse responses and capable of surround multichannel processing.

The calibration of the loudspeaker and the measurements performed in the theatres in Japan were possible only thanks to the support of prof. Yoichi Ando and colleagues of the University of Kobe, Japan (Kosuke Kato, Takuya Hotehama, Yosuke Okamoto), who allowed the authors to employ their laboratories and who helped during the measurements. Furthermore, useful discussion and exchange of technical information with these colleagues allowed the authors to improve the measurement technique.

The study of various rendering methods and of advanced hybrid multichannel solutions has been actively supported by the Ambiphonics Institute, where the listening experiments with various formats were performed.

The authors want to express their gratitude to the owners of the 9 theatres where the measurements were performed, who kindly also gave permission to publish the measured data, and to L. Tronchin and A. Avanzini, for their help during the measurements.

REFERENCES

- [1] Michael Gerzon - "Recording Concert Hall Acoustics for Posterity", *JAES* Vol. 23, Number 7 p. 569 (1975)
- [2] L. Tronchin, A. Farina - "The acoustics of the former Teatro "La Fenice", Venice", *JAES* Vol. 45, Number 12 p. 1051 (1997)
- [3] "Carta di Ferrara", *CIARM*, <http://acustica.ing.unife.it/ciarm/Carta.htm>
- [4] "Guidelines for acoustical measurements inside historical opera houses: procedures and validation", *CIARM*, <http://acustica.ing.unife.it/ciarm/download.htm>
- [5] A. Farina, R. Glasgal, E. Armelloni, A. Torger - "Ambiophonic Principles for the Recording and Reproduction of Surround Sound for Music" - *19th AES Conference on Surround Sound, Techniques, Technology and Perception* - Schloss Elmau, Germany, 21-24 June 2001.
- [6] E.Hulsebos, D.de Vries, and E. Bourdillat - "Improved Microphone Array Configurations for Auralization of Sound Fields by Wave-Field Synthesis", *JAES* Vol. 50, Number 10 p. 779 (2002)
- [7] A. Karamustafaoglu, U. Horbach, R. Pellegrini P. Mackensen, G. Theile - "Design and Applications of a Data-based Auralisation System for Surround Sound", *106th AES Convention*, pre-print n. 4976 (1999).
- [8] M. A. Poletti - "A Unified Theory of Horizontal Holographic Sound Systems", *JAES* Vol. 48, Number 12 p. 1049 (2000).
- [9] A. Farina - "Simultaneous measurement of impulse response and distortion with a swept-sine technique", *110th AES Convention*, Paris 18-22 February 2000.
- [10] S. Müller, P. Massarani - "Transfer-Function Measurement with Sweeps", *JAES* Vol. 49, Number 6 pp. 443 (2001).
- [11] G. Stan, J.J. Embrechts, D. Archambeau - "Comparison of Different Impulse Response Measurement Techniques", *JAES* Vol. 50, No. 4, p. 249, 2002 April.
- [12] Y. Ando, "Concert hall acoustics". Springer Series in electrophysics, Berlin, 1985.
- [13] V.L. Jordan, "A group of objective acoustical criteria for concert halls", *Applied Acoustics*, vol. 14 (1981)
- [14] 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.
- [15] T. G. Stockham Jr., "High-speed convolution and correlation", *AFIPS Proc. 1966 Spring Joint Computer Conf.*, Vol 28, Spartan Books, 1966, pp. 229 - 233.
- [16] W.G. Gardner, "Efficient convolution without input-output delay", *JAES* vol. 43, n. 3, 1995 March, pp. 127-136.
- [17] 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.
- [18] O.Kirkeby, P.A. Nelson, P. Rubak, A. Farina - "Design of Cross-talk Cancellation Networks by using Fast Deconvolution" - *106th AES Convention*, Munich, 8-11 may 1999.
- [19] Lake DSP Huron Workstation, [HTTP://www.lakedsp.com](http://www.lakedsp.com)
- [20] A. Farina, E. Ugolotti, "Software Implementation Of B-Format Encoding And Decoding", *Pre-prints of the 104rd AES Convention*, Amsterdam, 15 - 20 May, 1998.
- [21] A.Field, "B-dec High resolution First Order Ambisonic B-format decoder", University of York, http://www.york.ac.uk/inst/mustech/3d_audio
- [22] G. Theile - "Multichannel Natural Music Recording Based on Psychoacoustic Principles" - *AES 19th International Conference*, May 2001.
- [23] Roland Jacques, MultiMedia Projekt VERDI, TU Ilmenau Laboratory, Germany 2002 - <http://www.stud.tu-ilmenau.de/~proverdi/daten/umlen.html>
- [24] Williams, M.; Le Du, G. - "Multichannel Microphone Array Design", *108th AES Convention*, 2000, Preprint 5157.
- [25] Herrmann, U., Henkels, V., Braun, D. - "Comparison of 5 surround microphone methods", *Proceedings 20th Tonmeistertagung*, 1998, (ISBN 3-598-20361-6), pp. 508-517.
- [26] D. McGriffy, "Visual Virtual Microphone", [HTTP://mcgriffy.com/audio/ambisonic/vvmic](http://mcgriffy.com/audio/ambisonic/vvmic)
- [27] E. Hulsebos, T. Schuurmans, D. de Vries and R. Boone - "Circular microphone array for discrete multichannel audio recording", *114th AES Convention*, Amsterdam 22-25 March 2003, pre-print n. 5716.