A DIGITALLY CONTROLLED TWO DIMENSIONAL LOUDSPEAKER ARRAY

PAOLO MARTIGNON[1], FONS ADRIAENSEN[1], DANIELE TORELLI[1], ANGELO FARINA[1]

[1] *L.A.E., Laboratory of Acoustics and Electroacoustics, Parma, Italy*
martignon@laegroup.org

A "sonic chandelier" has been designed and manufactured to be installed in the new "Casa del Suono" in Parma. This museum will be an exhibition of sound reproduction devices from the 20th century, and a multi-channel audio laboratory as well. The chandelier is a 64-channel, dome shaped array of 228 speakers suspended in the center of the main hall of the museum, at a height of 4 meters. It creates, by means of Wave Field Synthesis, virtual sources moving above the listener's heads, and will reproduce an original spatial music composition. This listening experience plays the role of a very advanced music reproduction system, in contrast to the other, historic devices exhibited. An important requirement is that the sound produced must be confined to a restricted area beneath the system itself, in order not to interfere with other exhibits present in the hall. The physical structure and algorithm design are described, as well as some listening tests, performed for now on a reduced linear version of the array itself.

INTRODUCTION

While most of the existing WFS applications are in the form of surround horizontal line arrays, the application discussed in this paper is a 2D dome array hung from the ceiling as a "chandelier". It is used to create and move virtual sources above the listener's heads.

In this paper the logical and chronological outline of the design process of the sonic chandelier is reported. Particular attention is given to the relation between array shape, speaker type and disposition and the aliasing problem, which constituted the topics of the preliminary study. Then the algorithm design and implementation is described, with explanation of the solid and practical procedure employed with respect to WFS theory. In the end the first experimental results of this WFS application are described, obtained for now on a line array, reduced version of the 2D final one.

1 WFS AND SOUND FOCALIZATION

1.1 Basic concept

Sound focalization can be seen as a particular application of WFS, which allows creating a virtual source (the focus) between the speakers array and the listeners.

The formal description of this theory is given for line arrays by a modified version of the "Rayleigh 2 ½" integrals, which in the original version describes the synthesis of a source placed behind the array; this description can be found in [2][3][4].

The main principle of focalization is that *the signals feeding each speaker are each delayed by a constant time minus the fly-time from the speaker to the focus point.* In this way all speaker signals will arrive at the focus a the same time. This creates a concave wavefront which 'implodes' in the focus and 'explodes' again into a convex front.

A particular *gain* must also be set on each speaker channel for correct spatialization. Also a *common* (speaker independent) gain depending on focus position and a *common filter* derived from Rayleigh integrals; this will be treated more deeply in the following paragraph.

With this focalization technique, within a certain angle centred on the focus the proper field of a source placed in the focus is (approximately) created, in the plane containing the array and the focus itself. Figure 1 show the simulation of this effect for a pure tone reproduction.

The followings assumptions are important for the comprehension of this work.

A strict implementation of the WFS theory as expressed by Rayleigh integrals is basically not possible, since they consider theoretical sources like monopoles and dipoles, which would theoretically lead to different expressions and implementation, but real secondary sources (speaker with their boxes) do definitely not match those theoretical models.

In the experience of the author, for WFS the only thing which must be implemented directly from the theoretic model in a strict and precise way is the delay set, based on the concept explained so far, and this is valid for a line or surface array of any shape.

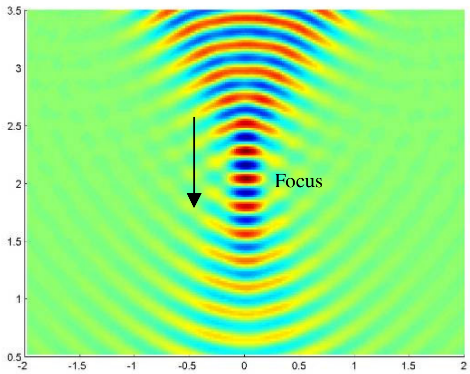


Figure 1: Sound focalization for a pure tone (instant pressure). Axes units are meters.

Moreover, when the geometry of a straight line array is abandoned for something more particular, no strictly dedicated formulas for implementation are available. In particular the author is not aware of algorithms for sound focalization with surface distribution of speaker. They may be studied and calculated starting from the general Kirkhoff-Helmoltz integral, [2][4], and using method like the "stationary phase": a fascinating task, but not particularly useful, in our opinion, for the present work. This is why some solutions adopted may well appear to be ad-hoc.

1.2 Spatial aliasing

A problem which arise in every array application is spatial aliasing, which is generally bound to the impossibility of a correct "spatial sampling" for a wave having a wavelength shorter than twice the speaker spacing.
More specifically the important parameter is the wavelength as seen along the array direction. (see fig. 2)
Formally, this condition is described by the following equation:

$$\lambda_x = \frac{\lambda}{\sin\varphi} > 2\cdot\Delta x \qquad (1)$$
$$\Rightarrow f = \frac{c}{\lambda} < \frac{c}{2\Delta x \sin\varphi}$$

With c sound speed and other symbols explained by Figure 2. It expresses that in every point the array cannot handle waves which wavelength, divided by the sine of the incidence angle in that point, are more than twice the speaker spacing. This can be called "sampling condition".

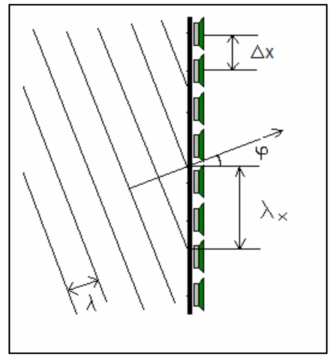


Figure 2: Sampling condition for a plane wave impacting a line array.

As for time sampling, along with a sampling condition, a reconstruction condition exists:

$$\sin\vartheta_{MAX}(f) = \frac{c}{2\Delta x f} \qquad (2)$$

with ϑ_{MAX} maximum angle at which the speakers radiate an appreciable power. This means that the speakers should have a frequency dependent directivity pattern.
The technical solution to fulfil the two conditions described so far are:

- Sampling filter: implementation of a low pass filter for each single speaker depending on the angle which the wave fronts of the sampled field form with the array in that point, based on equation (1).
- Reconstruction filter: use of speakers with a frequency dependent directivity given by equation (2).

As an example, a frontal plane wave (φ=0) reproduced by a line array doesn't need any sampling filter, but still need a reconstruction filter. In Figure 3 this is very schematically suggested: the rigid pistons on the right are a particular solution for the reconstruction filter (but just in the case of the frontal plane wave).
It's not a case that the directivity of a rigid piston gets narrower with the frequency increasing, in a way qualitatively similar to what imposed by equation (2).

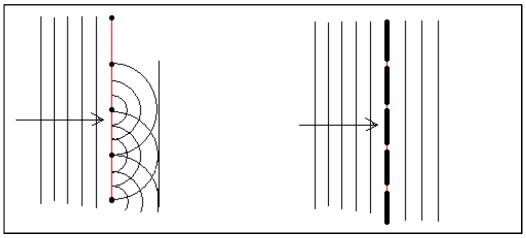


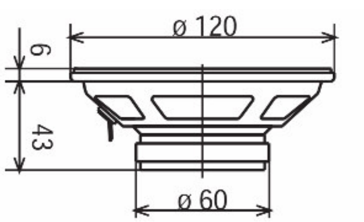
Figure 3: Frontal plane wave on a line array. Left: omnidirectional sources. Right: directive sources (rigid pistons).

Generally, for non-frontal waves, the use of a spatial (sampling and reconstruction) filter technique can help to avoid the phenomenon of spatial aliasing in the entire reproduction area. The drawback is that not all the frequencies will be reproduced everywhere: for each reproduced source there will be a cut off frequency depending on the angle under which the source is seen by the listener. This is the reason for which many WFS systems work according to theory only up to a certain frequency, say 2 kHz, and then use a “randomization approach” at higher frequencies. This permits to produce high frequencies everywhere with a realistic level and without noticeable aliasing effects, but unfortunately also without correct spatialization. This is surely a clever trade off for some applications. Our application is instead a very particular one as the area allowed for the focus and for the listeners is very small. This permits to make better use of spatial filtering, and consequently to implement the WFS technique on an extended frequency range.

2 PRELIMINAR STUDY (SIMULATIONS)

2.1 Speakers and aliasing

Considering the importance of the reconstruction filter, our aim from the start was to have a speaker disposition that would cover as much as possible of the available area, and a speaker type close to the model of a rigid piston.



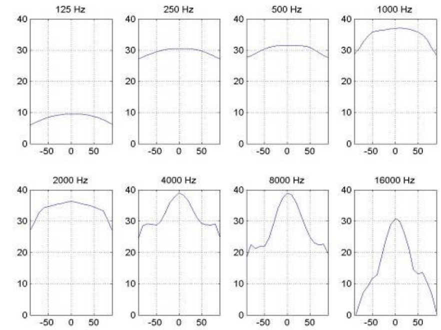


Figure 4: Speaker section and directivity measurement. Angle (Deg) and Decibels on axes.

This excluded for example the use of coaxial speakers, which surely would not have the desired beam narrowing at high frequencies.

Our final choice was the Ciare PM120. Figure 4 shows the speaker section and the results of our directivity measurements.

Figure 5 shows the results of two simulations on a 24 speaker line array, one with and one without taking speaker directivity into account. The delays are set as explained in the previous paragraph, while the gains are set according to the Rayleigh 2 ½ derived integrals [3].

The simulation algorithm is a phasor summation of all speaker contributions, finally the resulting SPL is plotted. The difference between the clearest and the darkest points is 15 dB.

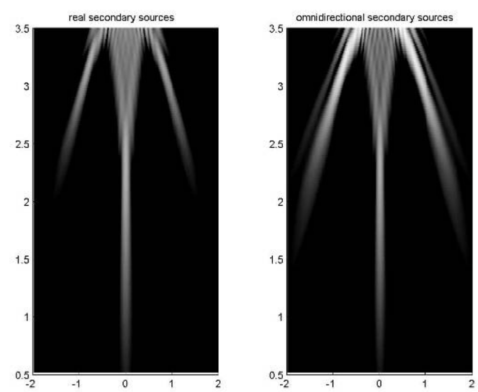


Figure 5: Line array simulations. Left: using measured speaker directivity. Right: using omni-directional sources. The axis units are meters.

First it should be noted that the aliasing effect does not affect the focus itself, but instead appears as a 'mustache' shaped leakage at the sides. This is typical of sound focalization, see also [3] and [4]. As expected, high frequency leakage due to spatial aliasing is considerably attenuated by the directivity of the speakers, which acts as reconstruction filter.

2.2 Choice of array shape

The system should create virtual sound source moving above the head of one or few listener placed beneath the center of the chandelier. The area allowed for good reproduction, for each source, as already suggested by Figure 1, is a cone (a triangle in section) with the vertex on the source itself and the sides aligned with the array border, as shown in figure 6.

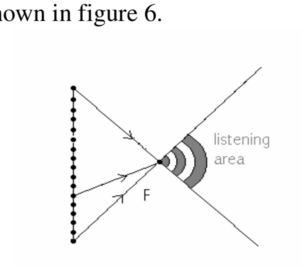


Figure 6: Area of good reproduction for a focused virtual source.

Consequently, to ensure that the listeners are always in the useful area, the sources can't be positioned outside the cone formed by the array border and the listener head. On the other hand, the most striking effect is achieved moving the sources close to the listener's head. Again, the more the source is placed away from the listener and closer to the array, the more sound will be radiated towards the sides of the hall, which is not desirable.

All this means that the movement of the sources should be restricted to a small volume, say a cubic meter, around a “main focus”, placed some decimeters above the head of a central listener.

Considering the aliasing problem, an optimal shape of the array was investigated. Actually, a dome shaped array, that is a part of a sphere centered on the main focus, allows us to neglect the sampling filter for virtual source creation in the focus itself. This is true since the incidence angle of the sampled (or reproduced) field, is roughly zero everywhere on the array. This is similar to the case of the plane wave and line (or planar) array of Figure 3.

This means that no antialiasing lowpass filter depending on the single speaker (the “sampling filter”) will be necessary; this is valid for focalization in the main focus, and, approximately, in the surrounding zone.

With a planar array (a disc), a sampling filter would be necessary complicating the signal processing. Moreover, in that case the reconstruction filter implemented by speaker directivity would result in a coloration depending on the source position.

3 ARRAY DESIGN AND SIMULATIONS

3.1 Structure and speaker grouping

Figure 7 shows some 3D models of the chandelier.

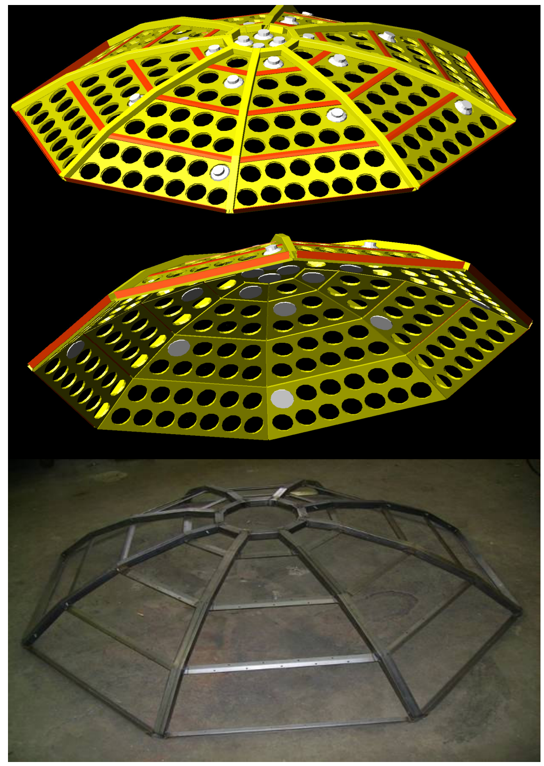


Figure 7: 3D model of the chandelier, two visions. Iron structure.

An iron structure supports trapezoidal wooden panels, on which the 228 speakers are mounted. The back side of the speakers will be covered by hard plastic boxes.

The base diameter is 2.6 meters, the radius of the sphere approximated by the polyhedron is two meters. The top will be hung at 4 meter of height so the main focus is at the height of 2 m.

The 228 speakers are divided in 64 groups of 3 or 4 near each (see the Switch64 gui of Figure 12).

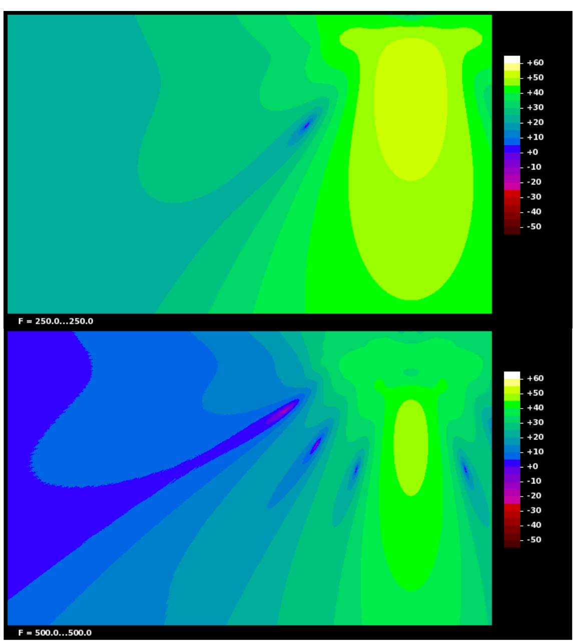
Speaker of the same group are connected in parallel. A special version of the speaker with an impedance of 32 Ohm has been manufactured for this project by Ciare.

This way we obtain a minimum load of 8 Ohm, which still provides acceptable damping, considering the expected resistance of the cables (roughly 20 meters long).

The average spacing among speaker groups seems to be very big for a WFS application. What should comfort us are the considerations previously made on aliasing vs. array shape and focus positions. With a spherically shaped dome full of ideal speakers, no delays nor gains would be necessary for focalization in the sphere centre. In this case only a high frequency leakage problem remains, which is limited as much as possible by the directivity of our speakers. To move the source, small delays and gains changes have to be introduced, considering the single secondary sources centered in the speaker groups centers. This make it advisable to evaluate how much the speaker groups can be considered as point sources from the point of view of the focus position; for this purpose we can investigate whether we are in a far field with respect to the speaker group, according to Fraunhofer definition, or not. At 20 kHz , for a source width of about 0.25 m, it can be showed that the Fraunhofer region limit is roughly 1.8 meters.

3.2 Chandelier simulations

To evaluate the possible performance of the array, some simulations were carried out. Only focalization in the center of the sphere is shown here, using the same signal on all speakers. The simulation, based of course on phasor summation, is limited to free field radiation, and takes the measured speaker directivity into account. The results for the frequencies 250Hz, 500Hz, 1kHz, 2kHz and 4 kHz are shown in Figure 9. The horizontal axis ranges from 5m to the left of the chandelier axis to one meter to the right.



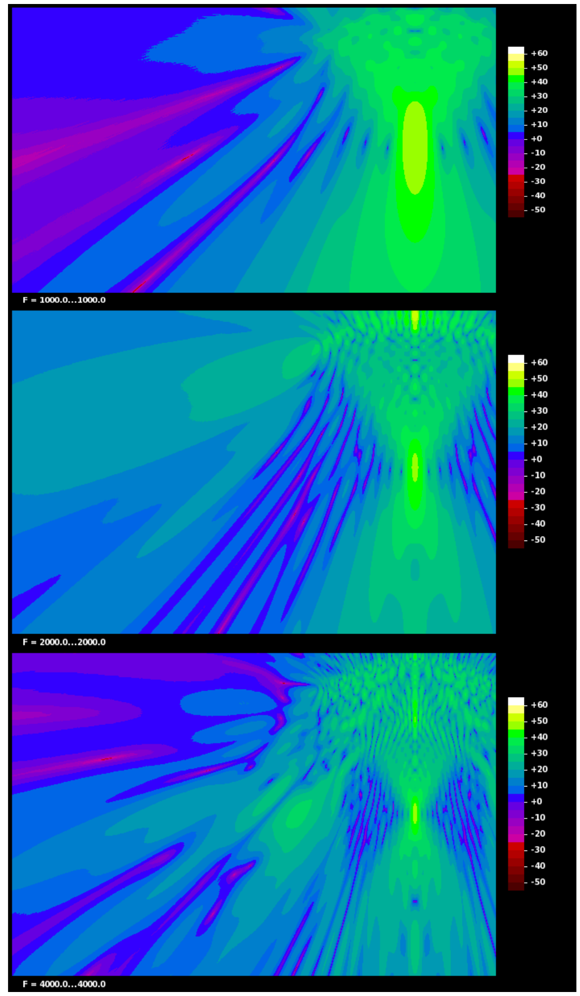


Figure 9: Chandelier simulations.

The SPL differences from the focus to 5 m away (where other sound messages are present) are strongly remarkable, already at low frequencies.

The aliasing effects is clearly visible, but only starting at 4 kHz. It may be distracting in the surroundings of the chandelier zone (where no particular attention has to be paid to any sound), but fades away before reaching the 5 meter limit, beyond which other sound messages are present. Results are very satisfactory, always considering that the influence of the (very reverberant) room acoustics is not taken into account.

A sound absorbing platform with a diameter equal to the chandelier will be constructed beneath the array. This should absorb most of the acoustic energy radiated by the device. It is hoped this will minimize the reverberant energy, and help to obtain a fast roll-off of the SPL when going from the center of the hall to the sides. Moreover the signal heard beneath the chandelier should have a remarkable direct to reverberant ratio, that is a high "clarity", which will sound very odd and striking inside a church.

4 MULTICHANNEL PROCESSING

4.1 Filter structures

The filter structure can be summarized in the scheme of Figure 10. A static input FIR for each focused source and an output FIR for each output channel are implemented. The dynamically updated part is limited to a gain-delay matrix, one gain-delay couple for each source-output channel couple. The simplification of the real time controlled part to a gain-delay structure is possible thanks to all the geometrical considerations made at the paragraph 2.2, and allows a much lighter computational load vs the case of a matrix of FIRs, necessary in case of "sampling filter" implementation.

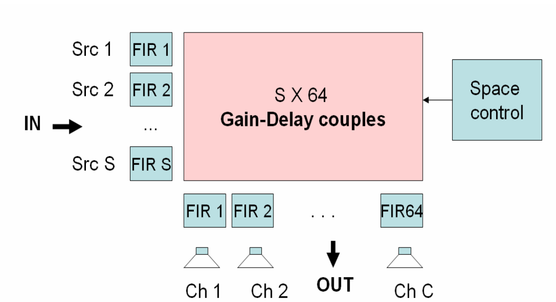


Figure 10: Filter structure.

4.1.1 Output filters

The 64 output filters are 256 taps long in the present application (sample rate 48 kHz), but can be set to different lengths; they provide for "speaker alignment". They are calculated so to have a "perfect" focalization in the main focus. Actually they may be seen a set of small delays and gains at each frequency, which provides for compensation of the non perfect sphere shape and compensation for native speaker differences. Every speaker group is virtually set to respond the in same way and to be placed at the same distance from the main focus. This procedure and the preliminary measurement associated are treated deeper in paragraph 4.3.1.
It must be said that some speaker behaviors will deviate again from the desired common response, depending on temperature, humidity and usage. Still this approach is supposed to make the focalization more precise.

4.1.2 Input filters

The input filter, one for each virtual source, has the role of general equalization, in magnitude and phase. All filters are normally identical, unless different sources have a different role and require a different kind of spatialization. A typical filter length would be 4096 samples.
Kirkhoff Helmoltz and Rayleigh integrals would provide for an analytic form of the common (speaker independent) filter to be implemented in addition to gains and delays, filter which, generally speaking, compensates for the fact that different number of speakers interfere constructively at different frequencies. Actually neither integral strictly applies to our array: in both perfect point sources (monopole, dipole or both mixed) would be needed, in terms of transfer function and directivity; moreover, in the Rayleigh integrals a plane distribution is assumed. So this system is not a strict application of theory, but an implementation based on the following ideas:

- the particular geometry of the system allow us to make a good focalization by mean of just a delay set (strictly defined) and a gain set (more disputable).
- a common filter, which can be empirically designed basing on measurements, will be necessary to compensate for some WFS related spectral effects and speaker transfer function.

4.2 Implementation

4.2.1 Hardware

All the processing is performed by software running on a Linux PC. The 64 channels are output by means of a MADI PCI card; the MADI signal is split into 8 ADAT streams of 8 channels each, which then feed the D/A converters.

4.2.2 Processing program, algorithms

The basic unit of the processing software is the program "WFSfocus", written by two of the authors (Torelli and Adriaensen). One instance of WFSfocus is run for each input source. Referring to the filter structure of Figure 10, WFSfocus includes one input filter and one row of 64 elements of the gain-delay matrix.
The 48 kHz signal entering WFSfocus is convolved with the input filter which is calculated off-line. The convolution step is used as well to upsample the signal, by actually performing the convolution operation twice, the second time with a filter that has been delayed by

half a sample. Interleaving the two outputs then produces the filtered signal at twice the input sample frequency. This in turn will allow the use of linear interpolation in the following step.
The upsampled signal then enters the row of gains and delays, which are controlled in real time by external commands (using either MIDI or the OSC protocol) to make the dynamic spatialization.
Delays set for each virtual source position are calculated as variation with respect to the set theoretically necessary for the main focus, since the main focus is automatically created by the output FIRs. For this purpose, the spatial coordinates of every speaker group center have been accurately measured.
For a smooth movement (as well as for a natural implementation of the Doppler effect) a fractional delay is necessary. This is done by linear intepolation. Since this step is performed on a signal at twice the original sample frequency, the resulting worst case HF loss is limited to about 1dB at the original Nyquist frequency, which is perfectly acceptable for this application. After impolation the signal again has its original sample rate.
The output signals from all WFSfocus instances are summed together and then processed by BruteFir (a convolution engine created by Anders Torger), which implements the 64 output FIRs.

4.2.3 Control

WFSfocus can be controlled in real time via MIDI or OSC (Open Sound Control). It accepts three 32 bit floating point numbers which define the position of the virtual source in the 3D space. Considering the particular geometry of our system, in the current version the three numbers represent spherical coordinates centered on the main focus:

- distance from the main focus
- azimuth angle
- angle with the vertical array axis.

A range checks are applied tn this coordinates based on the geometrical limits described in paragraph 2.2.
The spatialization can be performed by the user either in in real time via a MIDI controller or OSC commands, or off-line, using data from a MIDI sequencer or DAW automation data.
One possible controller (using OSC) is the WFSgui, program written specially for this application by one of the authors (Torelli); it allows to move several virtual sources by moving cursors (height) and mouse dragging (horizontal position). It can also receive OSC messages, to be used in passive mode as a monitor only. A screen dump is shown in Figure 11.

4.3 Settings

As already mentioned, both the output and the input FIRs are calculated from measurements.

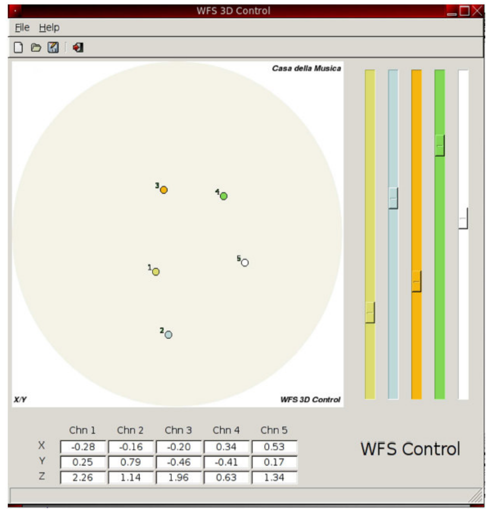


Figure 11: WFSgui screen dump.

4.3.1 Output filters

For the output FIRs calculation, the impulse responses of each of the 64 speaker groups are acquired in the main focus, and windowed so to represent as much as possible a free field IR. The IR stimulus used is an exponential sine sweep (Farina, 2000).
64 inverse filters are calculated with a method by Ole Kirkeby [5], using as target function the IR of the central group. In this way every speaker is forced to behave in a standardized but not ideal way. The equalization function, instead, is left to the common input filter (next subparagraph).

The measurement procedure is fully automatic, using the *Aliki* IR-measurement application created by one of the authors (Adriaensen) and a dedicated program called *Switch64* which is used to route test signals to any of the 64 inputs of the array. The Switch64 GUI is shown in figure 12; it allows manual switching during system verification, but can also be remotely controlled by Aliki, enabling a fully automated measurement sequence for all channels.

4.3.2 Input filter

For the input FIR, an inverse filtering approach is also employed. An impulse response of the whole array, set with the output filters working, is acquired, in a point close to the main focus.

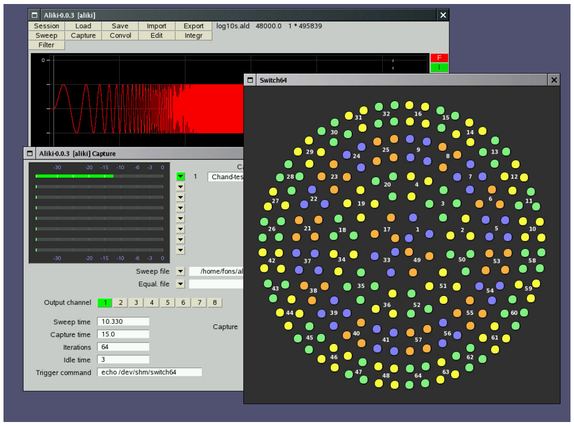


Figure 12: Aliky and Switch64.

An inverse filter, 4096 or more taps long (considering a 48 kHz sample rate), is calculated with the Kirkeby method using as a target function a 40 Hz -15 kHz Sinc, for a definite magnitude and phase equalization.

All the calculation of the 64 output inverse filters, as well as the input one, is performed using a Matlab program written by one of the authors (Martignon).

5 LINE ARRAY IMPLEMENTATION

The complete implementation has been possible just on the line array prototype up to now (Figure 13).
Note that, with the line array implementation:

- Source position is allowed only in the plane including the array and the listener, producing in that plane a 2D synthesis of the field (no spherical fronts, just circular ones on the plane).

Figure 13: The 24 speaker line array.

- In the 3D space a cylindrical symmetry around the array theoretically arises, if speaker directivity is neglected. So the focus spot is actually a sort of ring.
- No “front-back” movement are reproducible for a listener placed beneath the array, but just “left-right” and “far-close” movement.

5.1 Settings

The whole measurement–setting procedure has been carried out as outlined in the previous paragraphs.
The geometry of the system (listener position and source allowed area) are similar to the 3D shape described in paragraph 2.2, but projected on a plane including the line array and the listener.
Figure 14 and 15 shows the impulse responses acquired in the main focus and the respective calculated inverse filters. In this case the target function is the central (12^{th}) speaker. It is evident that the filters calculated for the extreme speakers of the array have greater amplitude; a frequency analysis would show how differences are stronger for high frequencies. This is due to the grater distance from the main focus and to the wider angle (less high frequency), which inverse filtering tends to compensate.

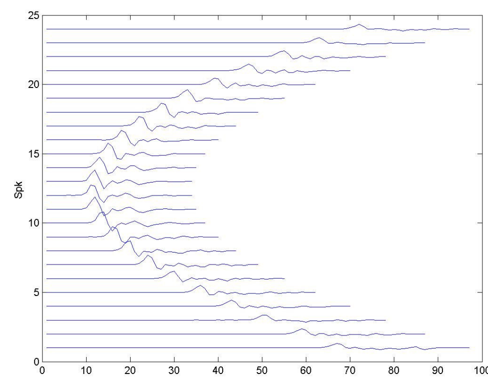
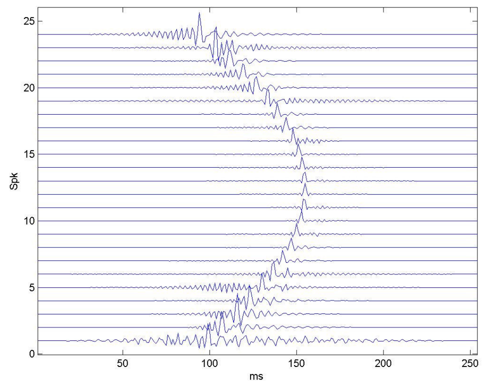


Figure 14: IR measurement in the main focus

The gain sets for WFSfocus are correctly arranged to obtain a final implementation which respects the “Rayleigh 2½ derived” integral for line array focalization.
The stress on high frequencies for peripheral speakers looks like a not very desirable effect, and in general is not. Actually, for the line array no particular negative effects (ringing, distortions, space dependent coloration (see paragraph 2.2)) arise.



Figur15: Inverse filters for speakers "alignment" in main focus.

But, if for example speakers next to each other were coupled (12 channels) - so to simulate in one dimension the group spacing of the real array - such problems would much probably arise. This prevision validates the design idea of the real chandelier, which will be free from these problems thanks to its spherical shape. The input filter has been calculated as well.
The WFSfocus program is adapted for the particular gain and delay set and the number of channels necessary for the line array implementation.

5.2 Listening

Both real time and previously recorded virtual source moving have been produced. As a source signal some synthetic sound where used, with several frequency contents, mainly concentrated in the medium-high frequency range.

Some impressions by the listening panel are given below.

- The acoustic scene is "objective", like in a correct WFS implementation it should be; that is, for example, a source placed above the listener moves to the right if the listener moves to the left and vice versa. This give precious perceptive information on the virtual source distance, since the source angle varies with listener movement with a sensitivity depending on its distance.
- The distance perception is partially achieved also not relying on transversal movements, that is: the source distance is somehow perceived also moving the source on a fixed direction passing through the (fixed) listener. In other word, moving the source farther and closer is not taken in from the listener just like a volume increase-decrease, but as a real movement (actually this can be said only by a quite focused and expert listener). Logically thinking, this effect depends on the front curvature reproduction; so the "real distance" effect is supposed to be improved by the 3D implementation with the dome array, with which real spherical fronts are synthesized.
- The source movement is "smooth" and "click free", and includes the Doppler effect.

6 CONCLUSIONS

A digitally controlled 2D dome shaped loudspeaker array, thought to be hung like a "chandelier", has been designed and manufactured, with the target of making vertical sound focalization with a Wave Field Synthesis like approach. Accurate preliminary studies have been done to evaluate the array shape and kind of speakers employed, with respect to the problem of spatial aliasing. The test on a reduced version of the chandelier has given very good result. The final 2D array should perform even better, thanks to the 3D spatialization, and despite the fact that secondary source are more spaced.

REFERENCES

[1] Berkhout A.J., "*A holographic approach to Acoustic Control*", J.Audio Eng.Soc. (1988)

[2] E.V.Start: "*Direct sound enhancement by Wave Field Synthesis* "(Thesis, Delft University of Technology)

[3] Boone M.M., Verheijen E.N.J, Jansen G. *"Virtual reality by sound reproduction based on Wave Field Synthesis"* AES (1996)

[4] Martignon P.*"Caratterizzazione e messa in opera di un array di altoparlanti per applicazioni di Wave Field Synthesis"* Thesis, (2001)

[5] Ole Kirkeby, Per Rubak, and Angelo Farina. *"Analysis Of Ill-Conditioning Of Multi-Channel Deconvolution Problems"* 1999, IEEE.