Spatial sound reinforcement using Wave Field Synthesis. A case study at the Institut du Monde Arabe.

Etienne Corteel*, Arnault Damien**, Cornelius Ihssen***

* sonic emotion labs, Paris, France, etienne.corteel@sonicemotion.com
** Euphonia, Paris, France, arnault.damien@euphonia.fr
*** sonic emotion, Obergllatt, Switzerland, cornelius.ihssen@sonicemotion.com

Abstract

Spatial audio in sound reinforcement remains an open topic, requiring good level coverage and at the same time good localization accuracy over a very large listening area, typically the entire audience. Wave Field Synthesis offers high localization accuracy over an extended listening area but the number of required loudspeakers, their placement on stage and the level coverage that results from it can be problematic.

The paper addresses these issues, presenting a case study of a sound reinforcement system based on Wave Field Synthesis. The system has been installed at the "Rafik Hariri" auditorium of the "Institut du Monde Arabe" in Paris, France, a 420 seats auditorium with a 25 m wide stage. The installed system comprises a total of 19 broadband loudspeakers and 6 subwoofers to address particular low frequency challenges of the room.

The paper describes the system design and the particular arrangement that has been chosen together with specific loudspeaker directivity characteristics. Simulations and measurements are performed so as to evaluate the level coverage throughout the audience. The paper also describes the procedure used onsite for the equalization of loudspeakers and tuning for optimum performance using the Sonic Wave I processor. Finally, the paper addresses the interfaces used for sound source positioning in the context of spatial sound reinforcement.

1. Introduction

Sound reinforcement in theaters and concert hall usually suffers from the lack of spatial sound rendering. Currently available systems usually consist of line arrays or powerful loudspeakers to either side of the stage that can be completed with a center loudspeaker and front fill loudspeakers in order to provide the best level coverage all over the audience. Such systems provide little or no reproduction of space. The only tool available for the sound engineer to reproduce a notion of space is often limited to the panpot of the mixing desk that relies on stereophonic principles for rendering spatial impressions. However, stereophony has three main drawbacks for spatial rendering over large audiences:

- the spatial rendering only works in a limited listening area, the so-called “sweet spot” that covers a very small portion of the audience;
• stereophony cannot reproduce the natural localization changes perceived in the audience, the parallax effect, that results from the true position of the actor/musician on stage compared to the position of the listener. This creates a localization mismatch between the visual and the auditory position;
• level coverage in intensity difference stereophony cannot be maintained because only the loudspeakers closer to the target direction are active.

The level coverage and stability problems are usually solved by using only very limited panning among speakers which results in practice in an essential mono feed to the loudspeaker system.

In contrast, Wave Field Synthesis (WFS) is a sound field rendering technique that allows for spatial sound reproduction over a large audience [1][2]. Unlike any other spatial audio reproduction technique, WFS also allows for the proper reproduction of the parallax effect thus providing a consistent visual and auditory impression for any seat in the audience [5]. WFS has long been restricted to horizontal reproduction although a practical formulation of 3D WFS has been recently presented by Corteel et al. in [6]. For sound reinforcement however, only horizontal reproduction is considered in the context of this study.

WFS has proved to provide improved spatial reproduction over stereophony, even at the sweet spot, for multiple source reproduction. Sanson et al. have shown in that listeners could discriminate the respective (left/right) position of two sources within a mixture of 3 sources (male speech) with spacing as small as 8 degrees whereas stereophony could only provide consistent spatial position discrimination for a perceived angle of 16 degrees [8].

The main drawback of traditional WFS is the very large amount of loudspeakers needed to fulfill the requirements defined in the literature. The number of required loudspeakers can reach several hundreds in typical sound reinforcement applications, limiting the applicability of the technique to experimental installations. Additionally, the required small spacing between loudspeakers imposes the use of very compact systems. Such small physical dimensions impose low sensitivity, limited dynamic characteristics and reduced frequency range.

In this paper, we propose a new technique for Wave Field Synthesis reproduction for sound reinforcement that addresses these problems. We first present the existing installations using Wave Field Synthesis for sound reinforcement and analyze the potential and the possible deficiencies of these installations. We then propose the concept of multiple level WFS as a concrete solution for improving sound level coverage and vertical localization in sound reinforcement applications while reducing the number of required loudspeakers. Then, we present a permanent sound reinforcement system that has been installed in the “Rafik Hariri” auditorium at the “Institut du Monde Arabe” in Paris, France. The system comprises only 19 broadband loudspeakers distributed over two height levels (below and above the stage) and allows for WFS reproduction over the entire audience.

### 2. Wave Field Synthesis for sound reinforcement

In this part, we present the potential but also the limitations of currently available Wave Field Synthesis systems for sound reinforcement. We address potential issues related to latency, level coverage and array positioning for optimal rendering with the entire audience.
2.1. Research systems

Direct sound enhancement for Wave Field Synthesis has been evaluated already 15 years ago with the work of Evert Start at the Delft University of Technology [4]. Start used a system developed with Duran Audio comprising a total of 192 loudspeakers with 12.5 cm distance. Loudspeakers were mounted in 24 channels bars of 3 m long each. The system was driven with 12 floating point DSPs providing a limited number of 16 inputs and 96 outputs with 32 kHz sampling rate and restricted to 16 bits resolution. The 96 outputs were then passively distributed to the 192 loudspeakers to provide enhanced horizontal directivity characteristics to the system so as to improve the aliasing frequency.

Start defined two main loudspeaker geometries for sound reinforcement that are reproduced in Figure 1 for stage sound reinforcement:

- front stage configuration, where loudspeakers are located at the limit between the stage and the audience;
- embracing stage configuration, where loudspeakers are surrounding the stage, with a main array located on the back wall of the stage and two side loudspeaker arrays on either side of the stage.

In the first configuration, on stage sound sources are reproduced as virtual sources located onstage behind the loudspeaker array. In the embracing configuration, virtual sources should be synthesized as focused sources onstage to maintain consistency between auditory and visual positions.

![Figure 1: classical loudspeaker array configuration for WFS sound reinforcement. Front stage configuration (left), embracing stage configuration (right).](image)

Start conducted experiments in concert halls using his direct sound enhancement in an anechoic room and two concert halls using a front stage configuration. He showed that localization accuracy of virtual synthesized sources was similar to real sources in these environments. Moreover, although the apparent source width appeared larger for virtual sources synthesized by WFS than real sources in the anechoic room, the apparent source width in rooms is equivalent for real sources or virtual sources. This is due to the natural acoustics of the room that tends to extend the apparent source width in realistic environments.
Another system has been used at IRCAM since 2009 consisting of 128 closely spaced loudspeakers driven by sonic emotion™ rendering engines for sound reinforcement and electronic music playback. This system is typically installed in the embracing stage configuration.

2.2. Problems with the use of focused sources in sound reinforcement

Focused sources suffer from several artefacts. Focused sources result from the synthesis of wave fronts that converge at the position of the virtual source. The reproduced level is thus maximal at the virtual source position, which may be uncomfortable for the musician and result in increased risk of feedback.

Delays of side loudspeakers are typically smaller than loudspeakers closer to the target virtual source positions, which result in erroneous localization above the spatial aliasing frequency due to first wave front high frequency localization on the closest speaker. This commands the use of a large number of closely spaced loudspeakers of small size and limited acoustical performance. At best, an aliasing frequency of 1.5 kHz can be reached in practical applications, which might not be sufficient for proper localization in the entire audience.

Additionally, the synthesis of converging wave fronts results in an increased latency corresponding to the propagation time between the furthest loudspeaker active in the array and the virtual source position as illustrated in Figure 2. For large stages of 15 m wide, this value can easily exceed 50 ms. The latency of the complete processing chain may therefore easily exceed practical values for sound reinforcement.

![Figure 2: synthesis of focused sources with embracing stage configuration, illustration of latency issues](image)

In such embracing stage setup, the minimum latency can only be achieved by positioning virtual sources behind the rear stage portion of the loudspeaker array (i.e. beyond the back wall of the stage). In this case, the latency is reduced to the propagation time between the closest loudspeaker to the virtual source and the real acoustical source onstage.

Even in this scenario, the latency can remain unpractical when the stage is too deep (> 6/10 m) for actors/musicians positioned near the proscenium. Additionally, the virtual source position cannot match the one of the instrument/voice to be amplified, which results in audio-visual localization mismatch.

To maintain good localization accuracy, a minimum overall latency of the system and a proper match between visual and auditory position, it is preferred to install a loudspeaker array in front of the stage. In this configuration, the propagation time between loudspeakers
and source to be reinforced is not an issue and the virtual source can be positioned at the same position than the actor/musician on stage.

2.3. Level coverage/vertical localization with Wave Field Synthesis

The last choice that needs to be done concerning the positioning of loudspeaker arrays is the vertical location of the array. In theory, virtual sources, loudspeaker array and listeners should be located in the same horizontal plane for WFS reproduction. However, a linear horizontal loudspeaker array, as used in WFS for sound reinforcement creates a sound field that exhibits a cylindrical symmetry around the axis of the loudspeaker array. Therefore, the loudspeaker array can be positioned at a different height without affecting the wave front reconstruction in the horizontal plane. Therefore, the reproduced virtual sources are perceived at the correct azimuth but with an incorrect elevation.

Figure 4 shows simulations of level coverage and vertical localization with horizontal loudspeaker arrays located either in the horizontal plane or at 6 m height (i.e. above the stage) as shown in Figure 3. The tilt of the microphone distribution represents typical floor tilt in theaters similar to the physical configuration of the audience at IMA. All levels are averaged along lateral positions considering an ensemble of potential virtual positions. The perceived elevation is simply calculated as the apparent elevation of the loudspeaker array at various seat positions.

It can be seen that there is a tradeoff to find between localization accuracy in height and level coverage. The loudspeaker array located onstage typically exhibits a rather fast decay in its vicinity. This is typical for WFS where level attenuation can be seen to be a combination of the natural attenuation of the target virtual source and the attenuation of the linear loudspeaker array. Furthermore, this simulation is based on a simple model that does not account for typical additional attenuation from seats and the audience for on-stage loudspeaker arrays. The lower array provides the best vertical localization accuracy but at the cost of poorer level coverage.

Figure 3: loudspeaker/microphone configuration for testing, 150 omnidirectional loudspeakers at each height (~15 cm spacing)
3. Multiple level Wave Field Synthesis

The concept of multiple level WFS tends to combine advantages of vertical localization accuracy allowed by lower loudspeaker array with improved coverage of higher loudspeaker array positioning. This concept relies on the use of two loudspeaker arrays positioned at different heights. In contrast to 3D WFS, multiple level WFS only targets the reproduction of horizontal source positioning with improved level coverage over the entire audience. Thanks to specific optimizations of WFS provided by sonic emotion™, the number of loudspeakers can be reduced to a limited amount with large spacing for each array leading to a reasonable cost with optimum performance.

3.1. Loudspeaker distribution and spacing

The algorithm developed in the Sonic Wave I processor enables to provide a homogeneous sound field reproduction at a minimum listening distance that exceeds once to twice the loudspeaker spacing for omnidirectional loudspeakers or typical wide dispersion loudspeakers. The only restriction to the standard use of WFS is that focused sources cannot be synthesized with this approach. However, the front stage positioning of the loudspeaker array and the latency issues outlined in part 2.2 heavily reduce the use of focused sources in sound reinforcement applications. The typical spacing for onstage speakers would then exceed 1 m enabling the use of professional quality loudspeakers and not limiting WFS reproduction to compact loudspeakers with limited dynamic and frequency range performances. The upper array can be formed with even increased loudspeaker spacing since loudspeakers are located further away from the audience. Spacing of 3 to 4 m can be easily employed, heavily reducing the total number of required loudspeakers.

3.2. Loudspeaker directivity choice

The choice of the loudspeakers in multiple level WFS can be crucial in order to optimize level coverage throughout the audience. Figure 5 shows level attenuation with distance of a two-level WFS system with a similar configuration as displayed in Figure 3. It shows the level provided by each loudspeaker array individually and the level reproduced by the combination of both loudspeaker arrays for omnidirectional loudspeakers and more directive
loudspeakers (80° conical dispersion at -6 dB) for the upper loudspeaker array. It can be seen that the use of directive loudspeakers at the upper level enables to significantly reduce the level reproduced by this array for distances close to the stage while maintain a high reproduced level at distant positions. The combination of both arrays thus enables to reach high SPL all over the audience.

Figure 5: level coverage with two loudspeaker arrays, influence of loudspeaker directivity: omnidirectional (left) and directive (right, 80° -6 dB conical dispersion)

3.3. Vertical localization optimization

Vertical localization in multiple level WFS might be depending on listener distance to stage. It is expected that, at positions close to the stage, localization be achieved on the lower loudspeaker array. However, at further distances, where the level reproduced by the upper loudspeaker array increases relative to the level of the lower loudspeaker array, localization might shift towards the upper loudspeaker array and localization error might increase.

In order to evaluate localization in multiple level WFS, one should also account for the natural propagation delay of the wave fronts generated by each array. This is displayed in Figure 6, in which propagation time from array to listening position is displayed. It can be seen that the upper loudspeaker array is naturally delayed at closer positions to the stage. Thanks to the precedence effect [7], also known as “Haas” effect, vertical may be more easily shifted towards the lower loudspeaker array. Applying a delay to the upper loudspeaker array can further enhance this “natural” precedence effect so as to favor localization on the lower loudspeaker array further away from the stage. A delay of 10 to 20 ms is typically applied.

For final localization judgment, one may also consider the fusion between audio and visual modalities. A shift of audio-visual localization can be expected towards the visual object onstage thanks to the so-called “ventriloquism” effect. This effect results from integration of audio and visual localization cues into a unique perceptual object that may result from optimal combination of auditory and visual information [9]. Localization accuracy in elevation is known to be less precise than in the horizontal plane. Localization of auditory visual objects is thus more easily localized towards the position of the visual stimuli.
In this part, we present in more details the sound reinforcement system that has been installed at the “Rafik Hariri” auditorium of the “Institut du Monde Arabe” in Paris. The room hosts a very diverse program ranging from conference to Arabic music with high sound pressure required throughout the entire audience for certain events. The stage is as wide as the room (~25 m) with a deeper and mainly used central part.

The main challenge consisted in providing a homogeneous sound field reproduction of the stage within an as wide room. There was also the requirement to provide properly distributed low frequencies in this almost square room.
4.1. Description of the installation

The installation consists in a total of 19 broadband loudspeakers distributed over 2 distinct height levels. The lower loudspeaker array is mounted at the edge of the proscenium at fixed positions. It consists of 12 loudspeakers with an irregular spacing having a higher density in the middle of the array. The upper array is mounted above the stage, about 1 m in front of the stage. This array has an almost regular spacing of ~3.5 m with a total of 7 loudspeakers. The whole system is completed by a total of 6 subwoofers that are installed on the audience floor next to the stage. Such a large number of subwoofers enable to properly spread the bass frequency rendering over the entire width of the room. The subwoofer array is driven with a specific version of the WFS algorithm that is dedicated to low frequency rendering and completes the lower loudspeaker array.

All loudspeakers have been designed by Taylor Made Systems (France) and manufactured by APG (France) and equipped with PHL audio (France) drivers. The lower array consists of 12 units of model MX2N. These are broadband loudspeakers with a 6,5 inch coaxial loudspeaker. More precise loudspeaker characteristics are detailed in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>MX2N</th>
<th>CX1514F low section</th>
<th>CX1514F high section</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power handling</td>
<td>150 W AES</td>
<td>800 W AES</td>
<td>110 W AES from 2 kHz, 70 W AES from 1 kHz</td>
</tr>
<tr>
<td>Sensitivity / 1 W @ 1m</td>
<td>95 dBSPL</td>
<td>100 dBSPL</td>
<td>110 dBSPL</td>
</tr>
<tr>
<td>Dispersion @ 1 kHz</td>
<td>110°</td>
<td>80°</td>
<td></td>
</tr>
<tr>
<td>Dimensions (H x W x D)</td>
<td>360 x 240 x 220 mm</td>
<td>540 x 870 x 350 mm</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: loudspeaker characteristics

The MX2N and the high section of the CX1514F are driven by Labgruppen C16:4 amplifiers that deliver 4 x 400 W @ 8 Ohms. The lower section of the CX1514F are driven by Labgruppen C88:4 amplifiers that deliver 4 x 1250 W @8 Ohms. The subwoofers (model S154) have 15 inch loudspeaker and integrated amplifiers (Powersoft DigiMod) that deliver 1500 W @8 Ohms (mono bridged).
All loudspeakers driven by a Sonic Wave I processor from sonic emotion™. All loudspeaker management (Crossover, EQ, level/delay/phase adjustment of crossover sections) is achieved in the Sonic Wave I processor using a total of 32 outputs. The Sonic Wave I processor uses a MADI soundcard. A DirectOut Technologies Andiamo converter is used for D/A and A/D conversions.

The complete installation enables to reach 108 dB Lin with pink noise input signal and only +/- 3 dB variation throughout the entire audience.

4.2. Installation

This section describes the installation and tuning phase of the system. A first configuration phase provides a physical description of loudspeaker positioning so as to compute a filter set according to the installation. Loudspeakers are then tuned to account for their own frequency response and local environment in the room. Finally, level and delay of lower and higher loudspeaker arrays are adjusted so as to optimize vertical localization accuracy and power distribution throughout the audience.

All these operations are realized in the WaveDesigner interfaces that can be used only by installers for configuring the system.

4.2.1. loudspeaker assignment and configuration

Loudspeakers are placed in the WaveDesigner interface at their actual position in the room. Figure 1 shows a top view of the installation where all loudspeakers are displayed. Each loudspeaker is described by its position and orientation. Loudspeakers are then described in the Output Assignment tool. Loudspeakers are first assigned to a given system that corresponds to a dedicated listening area (room). At the IMA, there is only one system, the main PA system. Then, systems are assigned to subsystems that describe individual loudspeaker arrays for rendering. All loudspeakers of the lower array are thus assigned to the first subsystem, the upper array to a second subsystem and subwoofers to a third one. Up to 4 subsystems can be declared in the current version of the Sonic Wave I processor.
The rendering mode of loudspeakers should be declared. There are three rendering modes:

- main WFS: main WFS array that defines source positioning possibilities (used for lower array);
- support WFS: WFS rendering for loudspeaker arrays used as secondary arrays for power distribution, used here for the upper loudspeaker array, could be used also for under balcony area;
- subwoofer: specific WFS rendering dedicated to low frequency channels.

Each loudspeaker can be described as passive, 2-way, or 3-way speaker. Sections of loudspeakers (main/low, mid and high) should be assigned to a physical output of the processor. Further, the amplifier power and sensitivity of the speaker can be described to account for dissimilar loudspeaker characteristics within subsystems in order to speed up adjustments.

![Figure 9: loudspeaker positioning and output assignment](image)

4.2.2. loudspeaker tuning

Once loudspeakers are all in place and the configuration has been transferred to the Sonic Wave I processor, the loudspeakers can be tuned to optimize sound quality. The Sonic Wave I processor comprises a complete output loudspeaker management section that can handle cross-over for multi-way systems and advanced equalization options for each loudspeaker of the setup.

The first step consists in setting the crossover of the multi-way systems. The current version of the system implements Linkwitz-Riley 4th order crossovers for optimum crossover performance. The user can tune the crossover section by choosing the corner frequency, modifying the gain of each section, inverting phase and adding a delay for each section. Finally, a limiter can be enabled to avoid overdriving amps and loudspeakers.

In a second step, it is possible to equalize all loudspeakers manually. The goal of the tuning phase is to create a uniform loudspeaker setup with similar timbre characteristics. Test signals from conventional measurement equipment can be routed to each speaker individually. This measurement can be used to equalize the loudspeakers so as to compensate for their individual deficiencies and the acoustics of their surrounding environment. It is recommended to measure each loudspeaker at a distance of ~1 m on axis in order to compensate for variations of LSPs efficiency, then calibrate output SPLs using the make-up gain of the Equalizer. For this efficiency calibration step, pink noise may be used as an input signal.
The Equalizer comprises 8 parametric filters, one low-shelf filter, one high-shelf filter, and low/high-pass sections with controllable slope (from 12 to 48 dB per octave in steps of 12 dB). Loudspeakers can be equalized individually or globally, by subsystem, by selecting the target subsystem and enabling the “control all” check box. This avoids replicating all settings to similar loudspeakers in a setup having similar acoustical surrounding. After this equalization step, the system should provide a uniform sound color all over the listening area and independently of the source position. A global sound color adjustment is then offered to the sound engineer at the input stage of the Sonic Wave I processor through the WavePerformer interface. This input stage equalization is completely independent from the output equalization used for tuning the system in the WaveDesigner. Like so, the visiting sound engineer can only adapt the sound color of the system to her/his taste without affecting the loudspeaker/room equalization performed by the installer at setup phase. The input equalization of the WavePerformer can be tuned during operation either for each input separately or globally for all inputs.

4.2.3. subsystem tuning

The final tuning phase of the system consists in adjusting the subsystems (contributions of the different loudspeaker arrays). This phase is simply achieved by assigning gains and delays globally to each subsystem. These gains and delays are then applied to all loudspeakers of the corresponding subsystem for optimum balance adjustment between subsystems.

The upper loudspeaker array should typically be delayed to benefit from the Haas effect and favor vertical localization on the lower array (i.e. on stage).
4.3. User interfaces

A second interface, the WavePerformer, can be used during normal operation of the system once the installation phase is finalized. This interface offers a different set of controls that enables the resident or visiting sound engineer to interact with the system during a live event.

Figure 13 is a screenshot of the “source positioning” window of the WavePerformer interface where the sound engineer can freely adjust or visualize the position of the 24 virtual sources. The loudspeaker positioning is automatically imported from the processor but speaker positions remain fixed and cannot be modified. Only a subset of sources can be selected for display and the view can be zoomed for clarifying the visualization.

The WavePerformer also provides a complete equalizer section for each input (virtual source) of the processor. The controls are similar to the output-based equalizer for loudspeaker/room equalization but are completely independent from it. The equalization settings can be grouped for all inputs so as to modify the global sound color to the taste of the sound engineer.

![Figure 12: WavePerformer interface](image)

The WavePerformer also provides an advanced routing panel allowing any input to be dynamically routed to subsystems with gain and delay control for each input and each subsystem individually. This enables high-level matricing of the input sound to the different subsystems of the installation. Furthermore, in each subsystem, a direct out mode can be selected for either, muting completely the subsystem for the given input, or assigning the input to only one loudspeaker.

All parameters available in the WavePerformer, including routing options, can also be stored as presets. Presets can then be recalled at any time during operation, without any click or unwanted noise, for new stage positioning of virtual sources or different routing options.

The basic preset consists of a stereo pair for inputs 1 and 2 that are reproduced by two virtual sources located at a large distance on either sides of the stage (plane waves). It creates a virtual stereophonic setup virtually located at a far distance, a classical configuration for stereo reproduction in WFS. This basic virtual stereophonic setup can be used when
only limited time is available for setting up the show. The visiting sound engineer can then use the panpot of the mixing desk the traditional way. It can also be used for the playback of background instruments, reverb and recorded material for which the exact onstage positioning and the match with a musician/actor onstage is not critical. The virtual stereophonic setup of the basic preset can extended to the popular 7.1 setup in rooms equipped with full surrounding capabilities. Other virtual sources can be used from the direct outputs or, preferably, output buses of the mixing desk, feeding additional virtual sources that can be freely positioned onstage by the sound engineer.

More interfaces are available in the processor such as an AU plugin (validated for Logic 9, VST plugin is upcoming), a MaxMSP interface (OSX), an OSC interface, an even a low level C++ library for direct communication. Additionally, tracking systems may be connected to the Sonic Wave I for live onstage tracking and assignment of actors/musicians movements to virtual source positions. More interface developments are ongoing and will be available in 2013.

![Routing panel for each input, high-level matricing for each input to subsystems](image)

**Figure 13:** Routing panel for each input, high-level matricing for each input to subsystems

5. **Conclusions**

This article reviewed the current possibilities and deficiencies of classical WFS installations for sound reinforcement. These deficiencies are addressed by using the concept of multiple levels Wave Field Synthesis. This concept was implemented in a sound reinforcement installation for the “Rafik Hariri” auditorium at the “Institut du Monde Arabe” in Paris. This installation is an example of multiple levels WFS that opens new for possibilities in sound reinforcement in auditoria and theaters.

6. **References**


