ESPRO 2.0 – IMPLEMENTATION OF A SURROUNDING 350-LOUDSPEAKER ARRAY FOR SOUND FIELD REPRODUCTION

MARKUS NOISTERNIG, THIBAUT CARPENTIER, OLIVIER WARUSFEL

Acoustic and Cognitive Spaces Research Group,
UMR STMS IRCAM-CNRS-UPMC, Paris, France

The variable acoustics performance hall of IRCAM was designed and built for providing the largest variability possible with regard to form, volume, and surface material structure. Extending the remarkable flexibility of this room a surrounding 350-loudspeaker array has been recently installed. This system aims at the physically correct synthesis of acoustical wave fields applying wave field synthesis (WFS) and higher-order Ambisonics (HOA). This article reviews the theory of HOA with regard to the design of a feasible grid of loudspeakers for this room, followed by a discussion of methods to overcome the practical limitations of using non-uniform loudspeaker arrays.

INTRODUCTION

The variable acoustics performance hall of IRCAM (*Espace de projection*, ESPRO) provides a remarkable flexibility with regard to form, volume and acoustical properties. The walls and ceiling consist of remotely controlled and individually rotatable prisms with three different material surfaces to absorb, reflect or diffuse the incident sound. To vary the hall's volume and shape three ceiling panels can be raised or lowered independently, and two roller curtains allow for separating the different volumes; the volume can be changed in the range of 24 x 15.5 x (0.8 – 10.5) m^3. The variable acoustics is thus achieved 'passively'.

Electro-acoustic sound reproduction systems provide the means of addressing variable acoustics 'actively'. This approach is apart from what is traditionally considered amplification and provides acoustical environments tailored to each production and aims at varying and controlling the acoustics to a greater extent than to what is possible by passive variable acoustics.

Virtual acoustics extends this approach to real-time synthesis and modeling of three-dimensional sound fields. The artistic play of spatial sound adds a further dimension to expressivity and embodiment in music performance. It may be, for instance, used to re-define and distort spatial relationships, to violate the spatial realism of sound and to break up its natural relation to the radiating body, with the intention to confuse the listener's understanding of sound in space.

The rapid increase in available computing power and the fast evolution of audio interfacing technologies have led to a new generation of immersive audio systems with a high number of playback channels. Several approaches exist to reproduce spatial sound with surrounding loudspeaker arrays. These techniques can be categorized into hearing-related model approaches and physical sound field reproduction. The latter aims at holophonic reconstruction of 2D/3D sound fields. Wave field synthesis (WFS) and higher-order Ambisonics (HOA) fall, for instance, in this category.

This article reviews the theory with regard to the installation of a 350-loudspeaker HOA/WFS array in the ESPRO. The discussion is mainly focused on the definition of a feasible grid of loudspeakers for 3D HOA and the design of decoders to overcome the practical limitations of using non-uniform and/or hemispherical loudspeaker arrays.

THEORETICAL BACKGROUND

This section briefly discusses the fundamental solutions to the acoustic wave equation in spherical coordinates. For notational simplicity, the following equations are given in the frequency domain with respect to time; the dependency on the frequency variable ω is omitted in the notation. It will be clear from the context of the discussion if the quantity is in the frequency or in the time domain. The spherical direction vector is defined as $\Theta = \left[\cos(\phi)\sin(\theta), \sin(\phi)\sin(\theta), \cos(\theta)\right]^T$, where $[\cdot]^T$ denotes the transpose of a vector or a matrix.

In spherical coordinates the scalar Helmholtz wave equation is separable in equations with respect to zenith angle θ and azimuth angle ϕ, and with respect to radius r. The angular portions of this solution are conveniently combined into single functions called *spherical harmonics*, $Y_n^m(\Theta)$, which are defined as

$$Y_n^m(\Theta) = \sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta) e^{im\phi}, \tag{1}$$

where n and m denote the order and the degree of the spherical harmonics, respectively, P_n^m are the associated Legendre functions, and $i = \sqrt{-1}$.

It can be shown [1-3] that the spherical harmonics are eigenfunctions for the two-dimensional surface of the sphere, which are mutually orthonormal when applying the normalization given in equation (1). The Condon-Shortley phase, a multiplicative phase factor of $(-1)^m$, is often included in the defining equation of $Y_n^m(\Theta)$. However, it is not necessary [4] and usually omitted in acoustics.

Consider now the non-homogeneous Helmholtz wave equation with continuous excitation $f(\Theta)$ on a sphere with a radius of r_L, notated as, cf. [5,6]

$$\left(\Delta + k^2\right) p = -f(\Theta)\frac{\delta(r - r_L)}{r_L^2}, \tag{2}$$

where Δ is the Laplacian, k is the wave number, p is the sound pressure, and $\delta(\cdot)$ is the Dirac delta distribution. The homogeneous solution to this equation derives to

$$p = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} b_{nm} j_n(kr) Y_n^m(\Theta), \tag{3}$$

with the spherical Bessel functions $j_n(kr)$ and the spherical harmonic spectral coefficients $b_{nm}(kr)$. The Fourier-Bessel expansion (FBE) of the wave field in equation (3) allows one to express any incident sound field. The particular solution to equation (2) is

$$p = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} -ik\psi_{nm} h_n(kr_L) j_n(kr) Y_n^m(\Theta), \tag{4}$$

with $h_n(kr_L)$ denoting the spherical Hankel functions. Assuming that the continuous excitation $f(\Theta)$ is known on the sphere, the spherical wave spectral coefficients ψ_{nm} can be determined utilizing forward harmonic transform analysis, see e.g. [7]

$$\psi_{nm} = \int_{S^2} f(\Theta) Y_n^m(\Theta)^* \, d\Theta, \tag{5}$$

where the superscript $(\cdot)^*$ denotes complex conjugation. A comparison of equations (3) and (4) shows that any incident sound field can be synthesized by

$$\psi_{nm} = \frac{i}{k}\frac{b_{nm}}{h_n(kr_L)}. \tag{6}$$

In vector-matrix notation equation (4) notates as, cf. [6]

$$p = -ik\,\mathbf{y}(\Theta)^T \operatorname{diag}\{\mathbf{j}(kr)\} \operatorname{diag}\{\mathbf{h}(kr_L)\}\Psi, \tag{7}$$

given the following definitions:

$$\Psi = \left[\psi_{00}, \dots, \psi_{nm}, \dots\right]^T,$$
$$\mathbf{y}(\Theta) = \left[Y_0^0(\Theta), \dots, Y_n^m(\Theta), \dots\right]^T,$$
$$\mathbf{j}(kr) = \left[j_0(kr), \dots, \underbrace{j_n(kr), \dots, j_n(kr)}_{2n+1}, \dots\right]^T,$$
$$\mathbf{h}(kr) = \left[h_0(kr), \dots, \underbrace{h_n(kr), \dots, h_n(kr)}_{2n+1}, \dots\right]^T,$$

The spherical wave spectrum Ψ decomposes the wave field (e.g. the sound pressure) on the sphere of radius r_L into its spherical wave components, which are also referred to as Ambisonic signals:

$$\Psi = \int_{S^2} f(\Theta)\mathbf{y}(\Theta) d\Theta. \tag{8}$$

Ambisonic Encoding

Given the above equations and the two-dimensional delta function on the sphere, which integrates to unity over the solid angle, we are now in the position to express a point source on a sphere with radius r_L as $f_S(\Theta) = s\,\delta(\Theta - \Theta_S)$. Inserting this expression into equation (8) results in the Ambisonic encoding equation:

$$\Psi_S = s\int_{S^2} \delta(\Theta - \Theta_S)\mathbf{y}(\Theta) d\Theta = s\,\mathbf{y}(\Theta_S). \tag{9}$$

Daniel has shown [8] that encoding sources at distances $r_S > r_L$ requires additional distance-dependent filters.

$$\Psi_S = s\,\mathrm{diag}\{\mathbf{h}(kr_L)\}^{-1}\,\mathrm{diag}\{\mathbf{h}(kr_S)\}\mathbf{y}(\Theta_S). \quad (10)$$

Even though these distance compensation filters are stable, in practical realization regularization is required. A detailed study on different regularization functions has been recently presented in [9].

Ambisonic Decoding

For Ambisonic playback the surround loudspeakers are modeled as point sources at discrete positions $\{\Theta_l\}$, $l = 1 \ldots L$, with the corresponding driving signals g_l. Using the two-dimensional delta function the angular excitation $\hat{f}(\Theta)$ can be written as

$$\hat{f}(\Theta) = \sum_{l=1}^{L} \delta(\Theta - \Theta_l) g_l. \quad (11)$$

Applying the FBE similar to equation (9) the Ambisonic representation of the loudspeaker contributions becomes

$$\hat{\Psi} = \sum_{l=1}^{L} \mathbf{y}(\Theta_l) g_l = \mathbf{Y}\mathbf{g}, \quad (12)$$

with

$$\mathbf{g} = [g, \ldots g_L]^T,$$
$$\mathbf{Y} = [\mathbf{y}(\Theta_1), \ldots \mathbf{y}(\Theta_L)]^T.$$

With equation (12) the loudspeaker driving signals $\mathbf{g}$ can be computed from the Ambisonic signals Ψ by matching the sum of the FBEs of the contributing loudspeaker sources with the target field FBE, i.e. $\hat{\Psi} = \Psi_S$. This is commonly referred to as mode matching in the literature [8]:

$$\mathbf{g} = \mathbf{D}\,\Psi. \quad (13)$$

In practice, matching the FBE coefficients requires a spatial band limitation, i.e. a truncation to a maximum order $n \le N$, which results in a matrix $\mathbf{Y}_N \in \mathrm{R}^{KxL}$ with $K = (N+1)^2$. It is clear from equations (12) and (13) that determining the mode-matching decoding matrix $\mathbf{D}$ requires one to invert the matrix $\mathbf{Y}_N$. In many cases this matrix is badly conditioned and a direct inversion is likely to fail. A well-established solution method is the singular value decomposition (SVD), which computes the generalized inverse of a matrix. Non-uniform and incomplete loudspeaker arrangements can cause very small but non-zero singular values and the computation of $\mathbf{D}$ becomes numerically unstable. Reasonable mode matching requires regularization, which approximates solutions $\mathbf{Y}_N\mathbf{D} \approx \mathbf{I}$ that are less sensitive to noise and perturbations than the naïve solutions. Tikhonov regularization and truncated SVD are often applied to solve this numerically rank-deficient problem, cf. [10].

Depending on the dimensions of $\mathbf{Y}_N$ and the number of non-vanishing singular values, M, the decoding matrix has different properties, which can be summarized as discrete spherical harmonic transform (DSHT), discrete spherical harmonic interpolation (DSHI), and discrete spherical harmonic approximation (DSHA), cf. [11].

In practical implementations not only the existence of discretization schemes but also their numerical quality is of critical importance. The condition number provides a quality measure. As one would expect, well-distributed points on the sphere yield low condition numbers and are thus well suited for the design of HOA decoder.

In this work we investigated the following fundamental discretization schemes on the sphere, cf. [11]:

- Extremal points for hyperinterpolation (hi);
- Spiral points (sp);
- Equal-area partitions (eqa);
- HEALPix (healpix);
- Gauss-Legendre grid (gl);
- Equi-distant cylindrical grid (ecp);
- Equi-angle grid (equiangle).

Figure 1 summarizes the existence and kind of solutions for these point distributions on the sphere. It provides a classification into the above-mentioned three-part scheme for discrete spherical harmonics of order $N = 9$ and a varying number of sampling points L. In this study, condition numbers less than $\chi(\mathbf{Y}_9) < 20\text{dB}$ are considered as to provide a pseudo-inverse without the need for regularization (i.e. DSHT and DSHI); greater condition numbers $\chi(\mathbf{Y}_9) > 20\text{dB}$ require regularization and thus result in an approximate solution (i.e. DSHA).

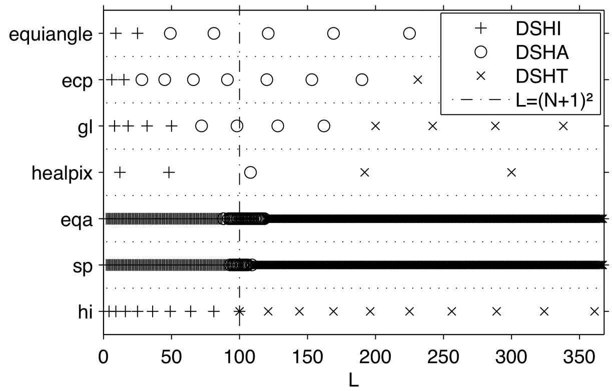

Figure 1 – Comparison of various discretization schemes on the sphere for order N=9 with respect to the number of loudspeakers L [11].

It can be seen from figure 1, that most discretization schemes only allow for DSHA near the critical number of sampling nodes, i.e. $L=(N+1)^2$; only the extremal system of points for numerical integration on the sphere (hyperinterpolation) provides an exact and well-conditioned inverse (DSHT and DSHI).

Figure 2 illustrates the condition numbers for different arrays with 75 loudspeakers on the upper hemisphere (including loudspeakers at the equator) applying the abovementioned discretization schemes. For computing the condition numbers the hemispherical array is completed to a full sphere by adding virtual speakers at positions mirrored across the equator of the sphere; see [12] for a more detailed description. This results in 121 points on the sphere and thus fulfills the requirements for applying spherical harmonic expansions for $N \leq 10$. The figure clearly shows that the performance with respect to the condition number of hyperinterpolation, spiral points, and equal-area partitions is qualitatively equivalent for orders lower than the critical sampling case. They result in well-conditioned matrices and clearly outperform the other discretization schemes.

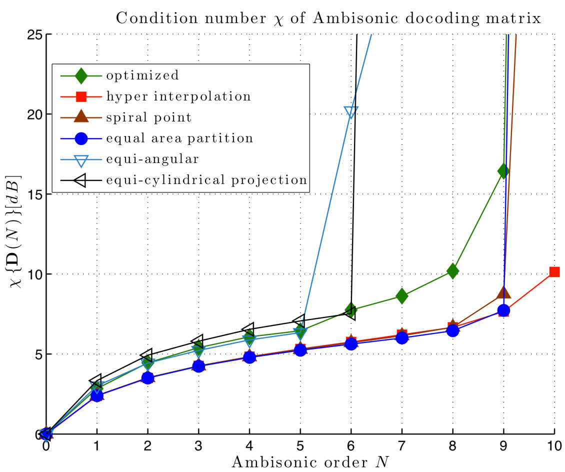

Figure 2 – Condition number χ of different fundamental spherical discretization schemes for a virtual set of 121 points on the sphere.

It should be noted that only hyperinterpolation is well conditioned in the critical sampling case, i.e. for $N=10$ in this example. This is also reflected in the results depicted in figure 1. Point sets for hyperinterpolation do not exist for all spherical harmonic orders, especially for $N>80$, and number of loudspeakers. However, in practice it is not feasible to implement 3D speaker arrays for very high orders. We thus consider the set of extremal points for hyperinterpolation as a lower bound for the condition number of loudspeaker arrangements for Ambisonic sound reproduction.

ARRAY IMPLEMENTATION

Figure 5 (see Appendix) illustrates the 350-loudspeaker array for sound field reproduction that has been installed in ESPRO. The four horizontal (linear) arrays consist of a total of 280 independently controlled speakers at a heigth of 2.20m; the distance between the speakers is 16cm for the front/back arrays and 29cm for the side arrays. For 3D sound reproduction the horizontal array is complemented by a rectangular loudspeaker array; a detailed discussion on the design of this array is given below.

Amadeus PMX-5 ES/A loudspeakers are used for the implementation of the array. The PMX-5 is a compact coaxial speaker with a 5.25-inch neodymium woofer 1.5-inch diaphragm compression driver. For full-band sound reproduction, audio signals below 100 Hz are sent via crossover to 8 subwoofers (db-audio Q-series).

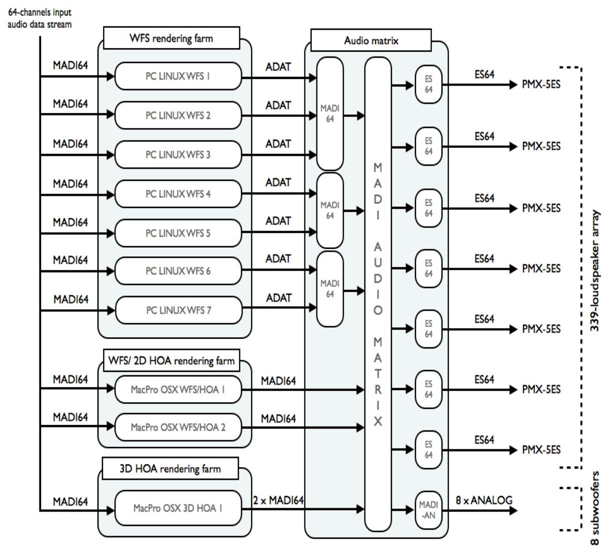

Figure 3 – Audio processing and signal routing structure.

The loudspeaker array is interfaced to a computer cluster in a sound isolated room using Ethersound audio networking; for audio connections in the server room MADI and ADAT is used. A 512 x 512 channel (8-port) MADI routing matrix distributes the audio output channels to the loudspeakers. It allows to combine individual channels to any output destination and thus to use different audio systems simultaneously, e.g. WFS for two-dimensional and HOA for three-dimensional spatial sound reproduction. Figure 3 illustrates the audio processing and signal routing structure.

Ambisonic loudspeaker array

The previous sections showed that the numerical quality of discretization schemes is of critical importance for the practical implementation of a loudspeaker array for

HOA. In addition, the movable ceiling panels, roller curtains, lighting bridges, and rotatable absorber prisms put strong constraints on the speaker positions. Figure 4 illustrates the design process for an optimal speaker array. First, an architectural model of the room with all possible speaker positions and orientations is created. Second, an optimal source distribution grid, which is defined on a circumscribed sphere, is projected to the walls of the room. Third, a nearest neighbor search algorithm is used to find the closest possible speaker positions and the condition number is computed for this array. Simulations for different discretization schemes on the sphere have been performed in order to find a speaker distribution that is well suited for HOA playback. For simulations the number of loudspeakers was limited to 75 in the upper hemisphere including speaker positions at the equator, i.e. the horizontal speaker arrays. Figure 2 shows that the implemented array structure is well conditioned for 3D HOA reproduction up to $N = 9$ and that its condition number is qualitatively equivalent to those of the optimized source distributions on the sphere.

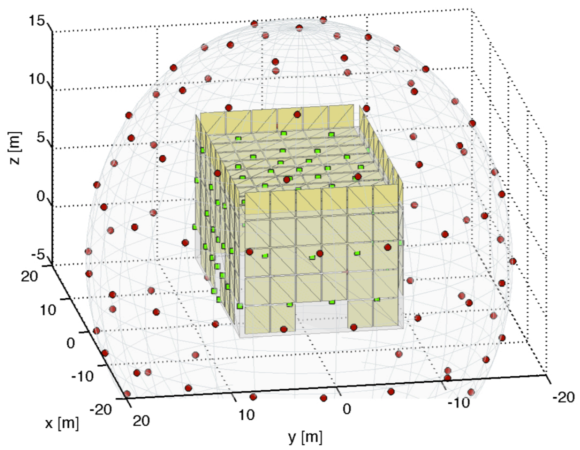

Figure 4 – Projection of an optimal source distribution on a circumscribed sphere to the walls of ESPRO.

Ambisonic decoder

The two common approaches for designing HOA decoders are mode matching and direct sampling. The first calculates the loudspeaker driving signals by matching the sum of the FBE coefficients, i.e. the modal source strengths, of the contributing loudspeakers to those of the target sound field, cf. equation (13). The second samples the continuous spherical harmonic excitation at the loudspeaker positions. When decoding Ambisonics to incomplete or partial spherical loudspeaker arrays, both mode matching and sampling decoders result in loudness variations for regions with sparse loudspeaker coverage. In the given loudspeaker setup the southern hemisphere is not considered for playback; thus the energy loss for virtual sources from directions below is no problem.

HOA only allows for error-free sound reproduction in a very restricted area, called the sweet spot, cf. [13] and [14]. Three-dimensional holophony is not feasible for the entire auditorium. For this reason, energy-preserving decoding methods have been recently presented [6], which interpret Ambisonics decoding as a panning law and preserve the energy over the entire angular domain. They have been shown to be numerically stable and feasible on the hemisphere. Alternative methods for hemispherical Ambisonics decoding, e.g. using a modified set of basis functions, have been presented in [12]. The horizontal 280-loudspeaker array provides a very dense grid of speakers and is used for two-dimensional sound field reproduction with both WFS and HOA.

The different decoding methods are implemented in IRCAM's real-time signal processing library for sound spatialization. When evaluating the spherical harmonic expansions for higher-degrees the computation of the associated Legendre functions (ALF) might fail. An existing recursion method for computing ALFs, c.f. [15], has been implemented and first numerical tests suggest that this approach is precise and numerically stable. In addition, near-field compensation (NFC) filters for the reproduction of monopole sources, cf. [8], with different regularization functions, cf. [9], have been implemented.

CONCLUSION

In this paper we discussed various aspects of the design and implementation of a 350-loudspeaker array in IRCAM's variable acoustics concert hall. This array aims at 2D/3D holophonic sound reproduction using WFS and/or HOA. Different discretization schemes and their effect on the numerical quality of the HOA decoder have also been discussed. The results have then been applied to the design of a feasible and well-conditioned grid of loudspeakers for this room. It should be noted that we are currently evaluating the performance of the real-time processing library using the full array. Measurements and sound field reproduction error analysis is ongoing.

ACKNOWLEDGMENTS

The authors are grateful to Etienne Corteel and Joseph Sanson for fruitful discussions and various contributions to this work. This project was realized with the financial support of the Région Ile-de-France (SESAME, Convention N°I-06-207/R), the Centre national de la recherche scientifique (CNRS), and the Université Pierre et Marie Curie (UPMC); it was supported in part by French ANR RIAM 004 02 "EarToy".

REFERENCES

[1] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, McGraw-Hill (1953)

[2] E. G. Williams, *Fourier Acoustics*, Academic Press (1999)

[3] N. A. Gumerov and R. Duraiswami, *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions*, Elsevier Science (2004)

[4] G. Arfken, *Mathematical Methods for Physicists*, 3rd ed., Academic Press (1985)

[5] F. Zotter, H. Pomberger, and M. Frank, *An alternative Ambisonics formulation: modal source-strength matching and the effect of spatial aliasing*, in Proc. 126th AES Conv., Munich (2009), preprint 7740

[6] F. Zotter, H. Pomberger, and M. Noisternig, *Energy-preserving Ambisonic decoding,* Acta Acust United Ac, 98:1 (2012) 37-47

[7] J. Driscoll and D. Healy, *Computing Fourier Transforms and convolutions on the 2-Sphere*, Adv Appl Math, 15:2 (1994) 202–250

[8] J. Daniel, *Spatial sound encoding including near field effect: introducing distance coding filters and a viable, new Ambisonic format*, in Proc. 23rd AES Conf., Copenhagen (2003), preprint 16

[9] S. Favrot and J. M. Buchholz, *Reproduction of nearby sound sources using higher-order Ambisonics with practical loudspeaker arrays*," Acta Acust United Ac, vol. 98:1 (2012) 48–60

[10] P. C. Hansen, *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*. Society for Industrial and Applied Mathematics, Philadelphia (1998)

[11] M. Noisternig, F. Zotter, and B. F.G. Katz, *Reconstructing sound source directivity in virtual acoustic environments*, in Principles and Applications of Spatial Hearing, World Scientific Publishing (2011) 357–373

[12] F. Zotter, H. Pomberger, M. Noisternig, *Ambisonics decoding with and without mode matching: a case study using the hemisphere*. Proc. 2nd Int Symp Ambisonics Spherical Acoust, Paris, France (2010)

[13] D. B. Ward, T. D. Abhayapala, *Reproduction of a plane- wave sound field using an array of loudspeakers*, IEEE Trans. on Speech and Audio Proc. 9 (2001) 697-707

[14] M. A. Poletti, *Three-dimensional surround sound systems based on spherical harmonics*, J. Audio Eng. Soc. 53 (2005) 1004–1025

[15] S. A. Holmes and W. E. Featherstone, *A unified approach to the Clenshaw summation and the recursive computation of very high degree and order normalised associated Legendre functions*, Journal of Geodesy, 76:5 (2002) 279–299

APPENDIX

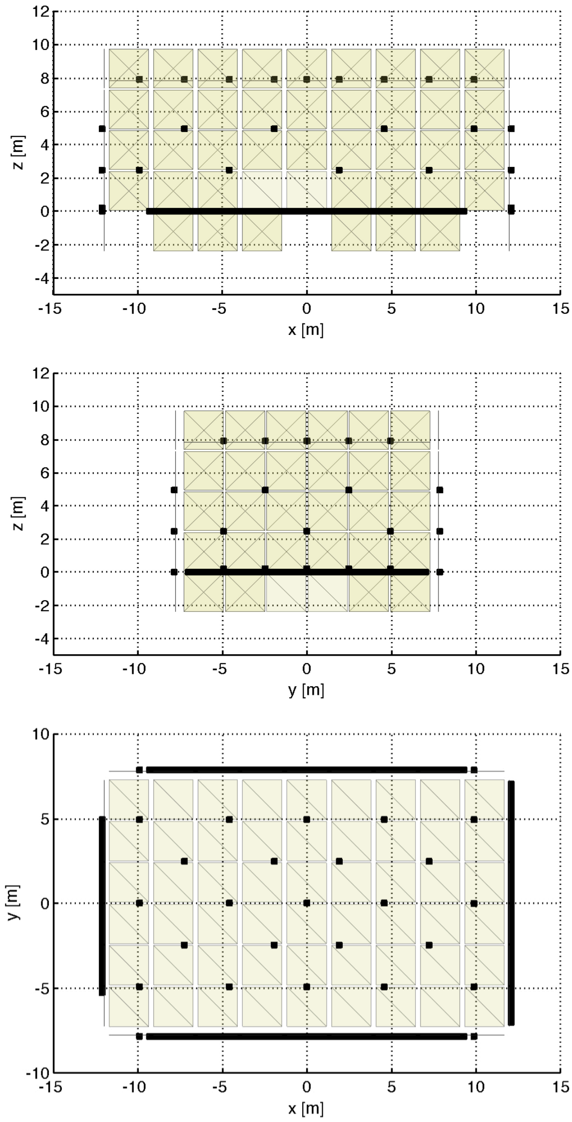

Figure 5 – 350-loudspeaker WFS/HOA array in ESPRO: vertical sections of the north wall (upper), the east wall (middle), and top view (lower)