Audio Engineering Society
Convention Paper 6375
Presented at the 118th Convention
2005 May 28–31 Barcelona, Spain

Reproducing multichannel sound on any speaker layout

Arnaud Laborie[1], Rémy Bruno[1], and Sébastien Montoya[1]

[1] *Trinnov Audio, Paris, France.*

Correspondence should be addressed to the authors (team@trinnov.com)

ABSTRACT

Consumers are more and more interested in multichannel sound. However, installing a surround system is still a headache for most users. The ITU recommendations are generally incompatible with homes arrangement and people install their system how they can, which generally results in large spatial sound distortions. This paper presents a system allowing to overcome this problem by adapting multichannel sound to the actual user loudspeaker layout. This system consists of a small calibration microphone measuring loudspeaker characteristics (3D position, frequency response) and a calculation process which remaps multichannel sound over the calibrated layout so as to compensate the measured loudspeakers misconfiguration, including full 3D position.

0. INTRODUCTION

Multichannel sound is achieving a great success among end consumers. A proof of this is the unprecedented growth of the DVD and multichannel systems market for the last years. Users are more and more interested in having multichannel sound at home and they are doubtlessly sensitive to the improvement brought by spatial sound.

However, the complexity of installing multichannel playback systems acts as a brake upon their expansion. One of their main constraints is loudspeakers positioning. Spatial sound quality is indeed limited by speakers arrangement in users' homes. Consumers generally do not know where they should place their loudspeakers, and ITU recommendations are generally incompatible with homes arrangement anyway. Placing the speakers on a circle around the listening position and at the specific angles is almost always impossible because of the furniture, of the doors, etc. Moreover, multichannel sound must compromise with the so-called "Wife Acceptance Factor" (WAF). As a result, the subtle work of the sound engineer is not correctly delivered to the end-user.

One of the next challenges concerning multichannel sound in the next years will be to bring to the users simplicity in multichannel system installation while preserving, or even improving, quality.

A couple of solutions already exist in the industry to compensate the distance differences between the loudspeakers. These solutions often use a small calibration microphone to determine the distances between each loudspeaker and the listening position. However, these solutions do not allow to compensate the full 3D positioning error, in particular angular errors in loudspeaker placement, which cause spatial distortions in the sound scene. The difficulty of faithfully reproducing spatial sound over an incorrect configuration is due to the nature of multichannel sound, which imposes in advance the positions of the loudspeakers.

Ambisonics [1,3,11,13,15–17,20–23] aims at disconnecting recording, channel transmission (for example music on the media), and playback. It allows to represent a sound environment without linking it to the recording or reproduction means and their characteristics. Sound capture and playback are achieved by, respectively, an encoder and a decoder that performs the conversion between the sound environment and the physical means to record or reproduce it. This most interesting approach is unfortunately generally limited to first order. In standard Ambisonic decoders, the number of loudspeakers must exceed the number of acoustic field components, locking the spatial precision to 1st order and leading to a suboptimal spatial rendering. Furthermore, designing a decoder for a non-regular loudspeaker layout is not straightforward given the numerous discussions about decoding on an ITU configuration. As a result no generic decoder can guaranty reliable performances on any loudspeaker layout.

This paper proposes a system that allows to reproduce multichannel sound on a loudspeaker layout that does not follow the standards, with limited impact of the incorrect configuration over the spatial properties of the reproduced sound environment. This system allows to bring flexibility and ease of use in multichannel sound while improving the repeatability and reliability of the surround sound experience. This system is based on an acoustic field approach which is first presented. The application to multichannel sound is then described.

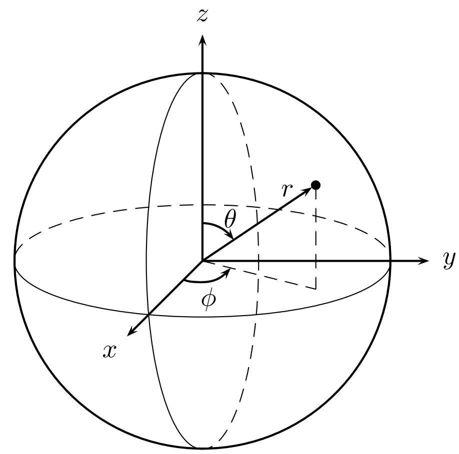


Fig. 1: Spherical Coordinate System.

1. PRINCIPLE

1.1. Representing a spatial sound environment

The way how a spatial sound environment is represented in a system is fundamental as it determines the systems' ability to adapt to various loudspeaker configurations.

The usual approach of spatial sound is to assign one channel to one loudspeaker. This multichannel sound approach leads to a sound space depending on:

- the knowledge of what is sent to which loudspeaker, which is the multichannel signal, and
- the knowledge of the position of each loudspeaker, given by the multichannel format standard.

The spatial sound environment is thus described by a means to produce it: the loudspeakers have to be placed at specific positions according to the standard, and they have to be fed by specific signals (a DVD stream, for example). This way to describe a sound environment is quite simple and easy to implement, provided the users loudspeaker layout follows the standard (cf. fig. 2).

However, representing a sound environment by a set of channels connected to loudspeakers does not correspond to the physical reality of spatial sound.

Three-dimensional sound can be described, from a physical point of view, by an acoustic field, which is defined for each point (x, y, z) in space and for each instant t using the pressure field $p(x, y, z, t)$ (cf. fig. 3). This acoustician approach will be kept throughout this article because, as will be shown later, it is compatible with the classical multichannel one while offering more flexibility.

Nevertheless, manipulating an acoustic field using its primary representation $p(x, y, z, t)$ is not easy because it would be necessary to know it for each value of (x, y, z, t). Therefore, an acoustic field is decomposed, in spherical coordinates (cf. fig. 1), into its Fourier-Bessel expansion (cf. appendix A) [4,14–16,24]. From the three dimensional continuous function $p(r, \theta, \phi, t)$, the Fourier-Bessel decomposition gives a set of signals called Fourier-Bessel coefficients of the acoustic field, denoted $p_{l,m}(t)$, where l and m are integers that satisfy $l \geq 0$ and $-l \leq m \leq l$. In the Fourier-Bessel formalism, l is called the order. In the frequency domain, $P(r, \theta, \phi, f)$ and $P_{l,m}(f)$ are the Fourier transforms of $p(r, \theta, \phi, t)$ and $p_{l,m}(t)$ respectively. This decomposition is as follows:

$$P(r, \theta, \phi, f) = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} P_{l,m}(f)\, j^l\, j_l(kr)\, y_l^m(\theta, \phi)$$

where $k = 2\pi f/c$ and c is the speed of sound, approximately 340 m/s.

The functions appearing in this equation are defined in the appendix A.

The Fourier-Bessel expansion is generally truncated at some order L. This order determines the resolution of the acoustic field representation. The higher the order, the higher the acoustic field representation fidelity will be, but the more computation power and signals will be required.

This representation of an acoustic field leads to three different interpretations represented on figure 4:

- The *Directivity* visualization represents the directivity function associated with an acoustic field consisting of the coefficient (l, m), the other coefficients being null. This corresponds to the standard representation of spherical harmonics (obtained by inverse Spherical Fourier Transform of the Fourier-Bessel coefficients).
- The *Field* visualization represents the modulus of the amplitude of a sound field consisting of the corresponding Fourier-Bessel coefficient (l, m), the other coefficients being null. The field is represented in the horizontal plane $z = 0$.
- The *Coefficients* visualization represents the spatial spectrum of an acoustic field consisting of the coefficient (l, m), the other coefficients being null. The spectrum is represented as a triangle containing the amplitude of the Fourier-Bessel coefficients. Each line corresponds to a value of l (from top to bottom: 0, 1, 2) and each column corresponds to a value of m (from left to right: -2, -1, 0, 1, 2). The black coefficients correspond to invalid combinations of l and m outside the triangle, the gray coefficients correspond to a null value, and the white coefficients correspond to a value of 1.

The following formalism is subsequently used:

- The multichannel signals are transformed into an acoustic field (cf. § 2.1) which is independent of the multichannel format and of the loudspeaker layout. This is achieved using a *reference radiation* model.
- The acoustic field is then transformed into loudspeaker feeds adapted to the actual layout. This is performed by using a *decoder*.

1.2. Decoding principle

The representation of a sound environment using an acoustic field has the great advantage of allowing to represent spatial sound itself independently of recording or playback means. However, this acoustic field approach does not have the advantage of the standard multichannel one: it does not directly provide a way to reproduce the sound environment. Using acoustics laws, it is nevertheless possible, for a given loudspeaker layout, to determine loudspeaker feeds that will produce a sound environment close to the one described by a given acoustic field. This step of generating loudspeaker feeds from an abstract sound environment description is known as decoding. This principle is illustrated on figure 5.

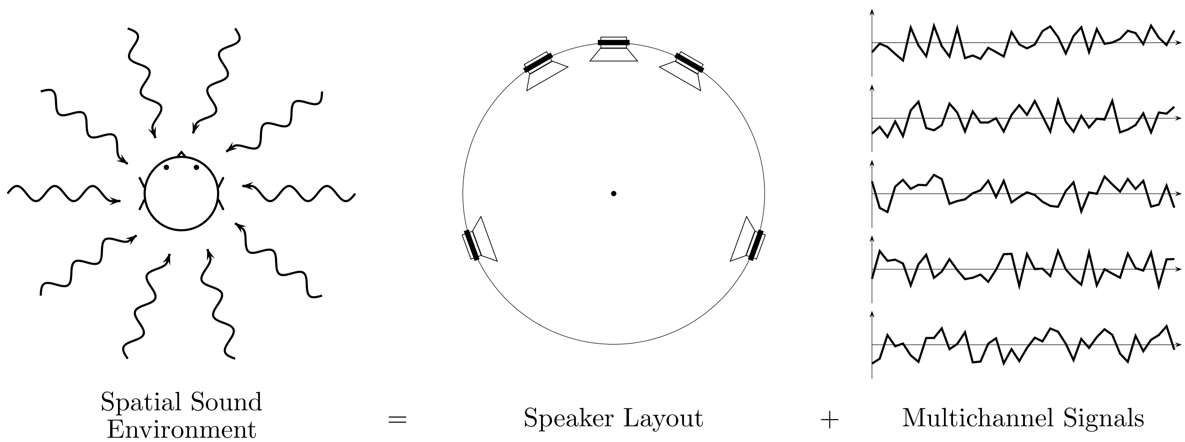


Fig. 2: Standard representation of a sound environment

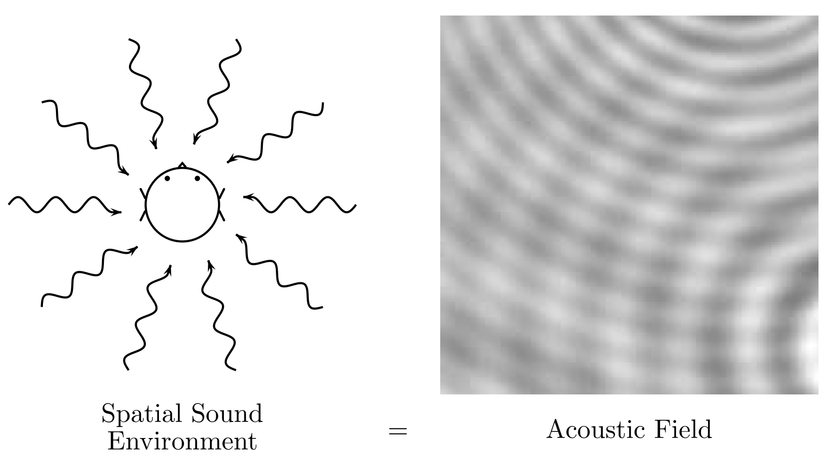


Fig. 3: Representation of a sound environment using an acoustic field

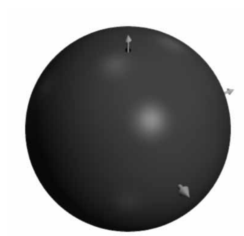
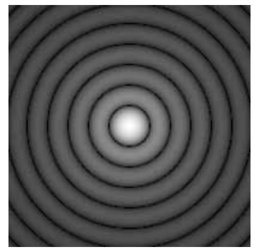
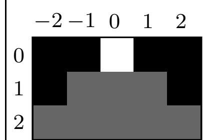
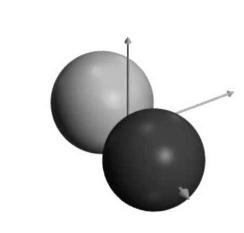
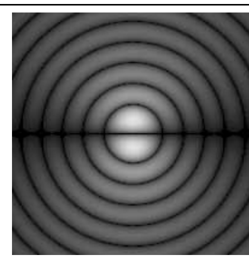
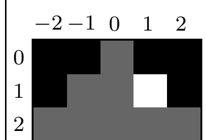
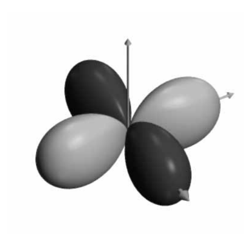
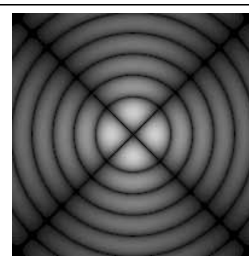
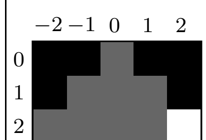
Type / Coef. #	Directivity	Field	Coefficients
$l = 0$, $m = 0$			
$l = 1$, $m = 1$			
$l = 2$, $m = 2$			

Fig. 4: Various representations of acoustic fields.

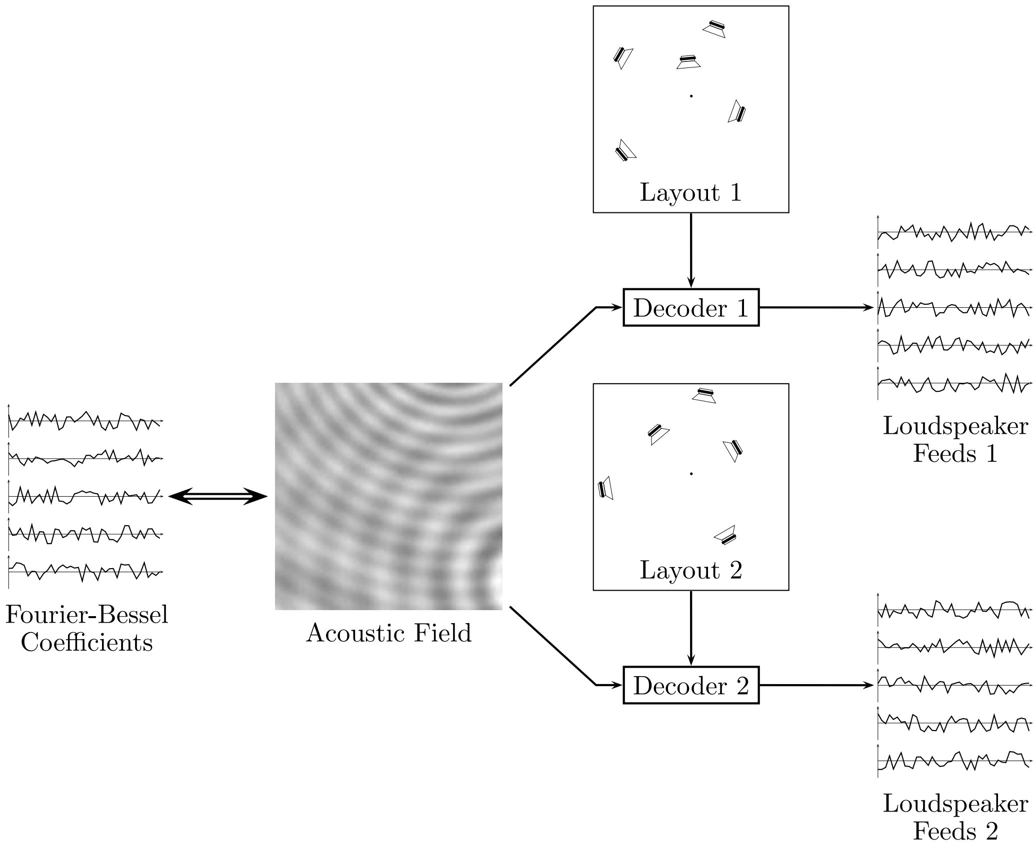


Fig. 5: Decoding principle: an acoustic field, consisting of its Fourier-Bessel coefficients, can be decoded for several loudspeaker layouts in order to be reproduced.

The decoding process is achieved using a decoding matrix D. This matrix is applied on the Fourier-Bessel coefficients of the acoustic field to be reproduced, arranged in a vector $\boldsymbol{p}$, in order to get the loudspeaker feeds, placed in a vector $\boldsymbol{v}$:

$$\boldsymbol{v} = D\boldsymbol{p} \tag{1}$$

The notations used are defined in appendix B.

As the elements of matrix D generally depend on frequency, equation (1), written in frequency domain, corresponds to a full matrix filtering, as illustrated on figure 6. This means that each Fourier-Bessel coefficient, corresponding to the sound environment to be reproduced, gets filtered and contributes to each loudspeaker feed.

1.3. Decoder determination

Acoustics laws allow to determine the acoustic field produced by a loudspeaker layout fed by known signals, assuming that each loudspeaker follows a known radiation model. This model gives, for each loudspeaker, the transfer function $p_n^M(r,\theta,\phi,t)$ between the loudspeaker feed $v_n(t)$ and the acoustic field $\hat{p}_n(r,\theta,\phi,t)$ produced by the loudspeaker when it is supplied with $v_n(t)$. The relation between $\hat{p}_n(r,\theta,\phi,t)$ and $v_n(t)$ is then a convolution:

$$\hat{p}_n(r,\theta,\phi,t) = \int_{-\infty}^{+\infty} v_n(\tau)\, p_n^M(r,\theta,\phi,t-\tau)\,\mathrm{d}\tau \tag{2}$$

The transfer function $p_n^M(r,\theta,\phi,t)$ is called spatio-temporal response of the loudspeaker n. It can be expanded into Fourier-Bessel series in order to obtain its Fourier-Bessel coefficients denoted $M_{l,m,n}(f)$. These coefficients are arranged in a matrix M called radiation matrix (cf. appendix B). The spatio-temporal response of each loudspeaker can be measured using methods described in [2, 3, 5, 9, 10, 14, 18, 23].

In the frequency domain, the loudspeaker feeds $V_n(f)$ and the Fourier-Bessel coefficients $\hat{P}_{l,m}(f)$ of the acoustic field produced by the whole layout can be arranged in two vectors, respectively $\boldsymbol{v}$ and $\hat{\boldsymbol{p}}$ in order to get the following equation, called radiation relation:

$$\hat{\boldsymbol{p}} = M\boldsymbol{v} \tag{3}$$

This relation is simply obtained from equation (2) by summing the contributions of each loudspeaker and expanding the resulting acoustic field in Fourier-Bessel coefficients.

Equation (3) gives the acoustic field produced by a loudspeaker layout from the loudspeaker feeds. This is just the opposite of the decoding process described in paragraph 1.2, which consists in determining the loudspeaker feeds from the acoustic field. Consequently, the determination of a decoder consists in inverting the radiation relation (3). However, inverting this relation raises a number of problems:

- The radiation matrix M is generally not square, so M^{-1} does generally not exist.
- When the acoustic field to be reproduced is a source in a given direction, carelessly generated loudspeaker feeds could lead to feed a loudspeaker which is far from the direction of the considered source, which could possibly create listening artifacts.
- The situation where loudspeakers emit signals in phase opposition should be avoided.
- The aim of decoding is to generate signals that make the loudspeaker layout emit an acoustic field close to the original one. But all Fourier-Bessel coefficients do not have the same importance: accurately reproducing the 5th-order coefficients is useless unless the lower-order components are correctly reproduced.

In order to overcome these problems, we may introduce the following parameters, that will allow to carry out several optimizations:

- the position of the loudspeakers (r_n, θ_n, ϕ_n)
- the frequency response of the speakers $H_n(f)$
- the spatio-temporal response $M_{l,m,n}(f)$
- a spatial weighting window given either in direct space $W(r,f)$, or with its Fourier-Bessel coefficients $W_l(f)$, or with its radius $R(f)$
- parameters $\left\{(l_k, m_k)\right\}_{1\leq k\leq K}(f)$ constituting, for each frequency, a list of Fourier-Bessel coefficients whose correct reconstruction is imposed

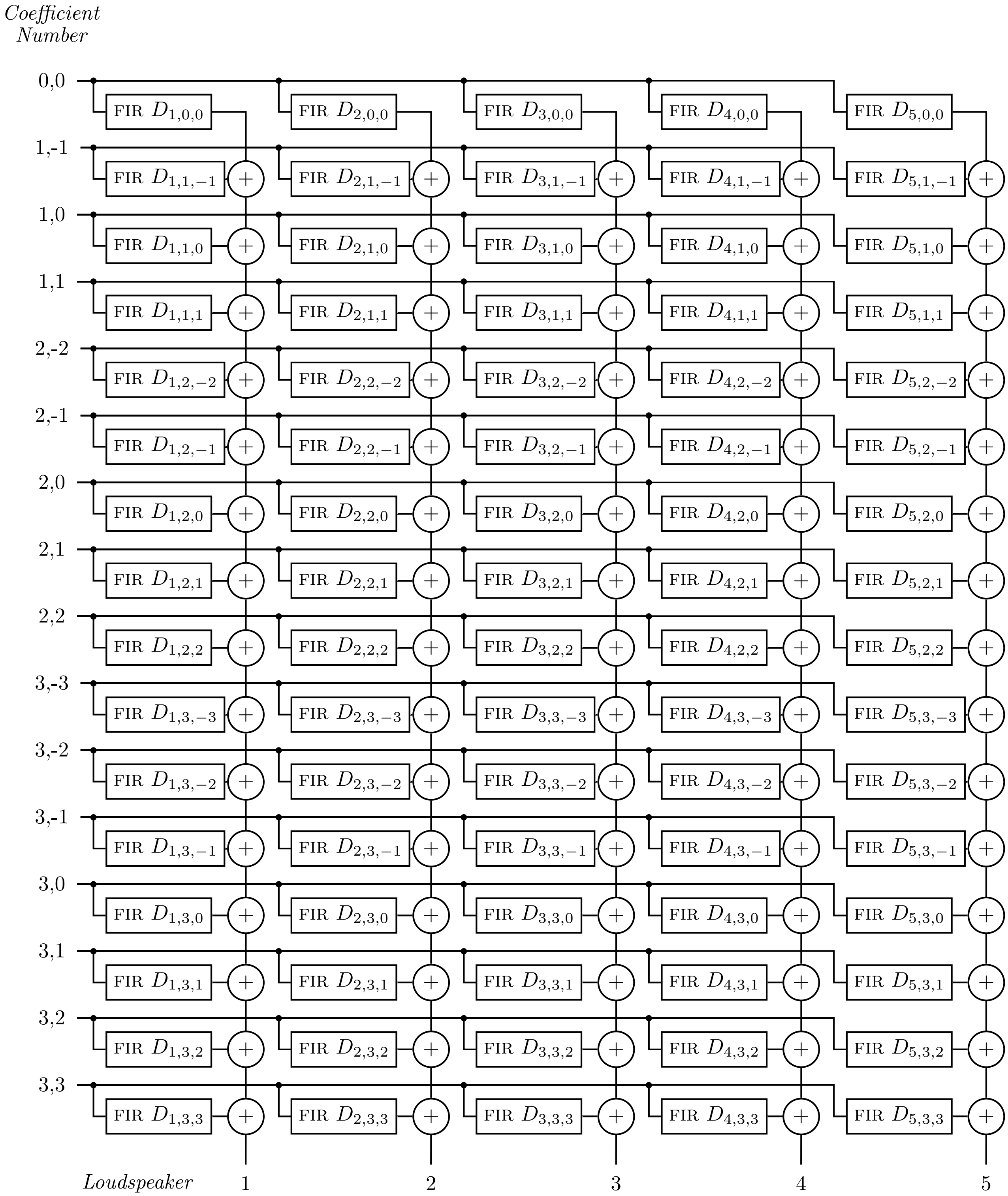


Fig. 6: Full Matrix Filtering.

- a parameter $\mu(f)$ giving the ability of the system to follow the irregularities of the loudspeakers configuration
- the maximal order $L(f)$

Model of the loudspeaker radiation (matrix M)

The parameters (r_n, θ_n, ϕ_n), $H_n(f)$ and $M_{l,m,n}(f)$ are not all required. The purpose of these three parameters is to determine the coefficients $M_{l,m,n}(f)$ forming the matrix M and corresponding to the radiation pattern of the loudspeaker n.

If the spatio-temporal response $M_{l,m,n}(f)$ of the loudspeaker n is known (either measured or indirectly determined), the coefficients of the matrix M are directly given by the spatio-temporal responses $M_{l,m,n}(f)$.

In the more general case, coefficients $M_{l,m,n}(f)$ are computed from the loudspeakers characteristics and position. For an omnidirectional loudspeaker at position (r_n, θ_n, ϕ_n) with spherical radiation, these coefficients are given by:

$$M_{l,m,n}(f) = H_n(f)\, \frac{e^{-jkr_n}}{r_n}\, y_l^m(\theta_n, \phi_n)^*\, \xi_l(kr_n)$$

with

$$\xi_l(x) = \sum_{k=0}^{l} \beta_{l,k}\, (jx)^{-k}$$

$$\beta_{l,k} = \frac{(l+k)!}{2^k\, k!\, (l-k)!}$$

Model of the spatial window (matrix W)

The spatial weighting window allows to select the region in space where the acoustic field is optimized. This window is the spatial equivalent of time windows for time signals such as Hamming, Hanning, etc. As time windows, spatial windows can be adjusted to control unwanted ringing effect.

The weighting window is represented by a weighting matrix W, which is a diagonal matrix containing elements $W_l(f)$ (cf. appendix B).

These parameters can be directly given in order to weight the reproduction of a given order. For example, a value of 0 means that we do not care about the reproduction of the corresponding order.

As an alternative, weighting can be determined from a window in direct space $W(r, f)$. Weighting the acoustic field in direct space (r, θ, ϕ) is indeed equivalent to weighting its Fourier-Bessel coefficients. An example of direct space weighting window is a weighting ball of radius $R(f)$. In this case, $W(r, f)$ is given by

$$W(r, f) = \begin{cases} 1 & \text{if } r \leq R(f) \\ 0 & \text{if } r > R(f) \end{cases}$$

The computation of the elements $W_l(f)$ from the weighting window in direct space $W(r, f)$ is given, using Fourier-Bessel expansion, by:

$$W_l(f) = 16\pi^2 \int_0^\infty W(r, f)\, j_l(kr)^2\, r^2\, \mathrm{d}r$$

In the case of a weighting ball of radius R, the above equation gives:

$$W_l(f) = 8\pi^2\, R^3 \left(j_l(kR)^2 + j_{l+1}(kR)^2 - \frac{2l+1}{kR}\, j_l(kR)\, j_{l+1}(kR) \right)$$

Selecting the optimal spatial resolution (order L)

The maximal order L depends on the loudspeaker layout. For example, second order is sufficient for standard stereo, but 5th order is required in order to take advantage of the 5.0 layout. This maximal order L for a layout with N loudspeakers can be determined from the two speakers forming the smallest angle:

$$L = \lfloor \pi / \gamma_{\min} \rfloor$$

with

$$\gamma_{\min} = \min_{1 \leq n_1 < n_2 \leq N} \gamma_{n_1,n_2}$$

$$\gamma_{n_1,n_2} = \arccos\big(\sin\theta_{n_1}\, \sin\theta_{n_2}\, \cos(\phi_{n_1} - \phi_{n_2}) + \cos\theta_{n_1}\, \cos\theta_{n_2} \big)$$

Contrary to Ambisonics, the order is no longer determined by the number of loudspeakers but by the smallest angle between loudspeakers corresponding to the highest spatial resolution the loudspeaker layout can provide. This determination of the order leads to a more efficient use of the loudspeaker layout. Selecting a spatial resolution higher than this order doesn't significantly improve the quality.

Imposed Fourier-Bessel coefficients

Generally, the acoustic field reproduced by the loudspeakers will not be the same as the original one, because it is impossible to faithfully reproduce a real given acoustic field using only a few speakers. However, some of the Fourier-Bessel coefficients of the acoustic field can be precisely reproduced, as long as the number of these coefficients is not too high in regard to the number of loudspeakers.

The list $\{(l_k, m_k)\}_{1 \le k \le K}(f)$ contains the suffixes (l, m) for which the Fourier-Bessel coefficients are required to be faithfully reproduced. A matrix F of size $K \times (L+1)^2$ is constituted from this set (cf. appendix B). The number K of imposed coefficients has to be smaller than the number N of loudspeakers.

Adaptation parameter μ

This parameter allows to have a cursor between the reproduction of only imposed coefficients and the other coefficients. It acts as a normalization parameter in the mathematical inversion problem. A value of 0 means that only the imposed coefficients of the acoustic field are correctly reproduced, whereas a value of 1 means that the other coefficients of order less than L are also, but less faithfully, reproduced.

This parameter is representative of the "mathematical effort" that the system does to reproduce the acoustic field. High effort means high mathematical matching but at the expense of side-effects such as negative lobes.

This approach is different from Ambisonics, which imposes the exact reproduction of Fourier-Bessel coefficients, and thus guaranties the same spatial resolution for all directions. Here, spatial resolution can vary with direction, which allows to more optimally use an irregular layout. For example, in the case of an ITU configuration, the resolution is of order 5 on the front, 3 on the side and 1 on the back. The Ambisonics approach gives order 1 for all directions.

Determination of the loudspeaker feeds (matrix D)

The decoding matrix is determined from the above matrices and parameters. It is obtained by applying less squares methods:

$$D = \mu A M^\top W \quad + \\ A M^\top F^\top \left(F M A M^\top F^\top\right)^{-1} F \left(I_{(L+1)^2} - \mu M A M^\top W\right)$$

with

$$A = \left((1-\mu) I_N + \mu M^\top W M\right)^{-1}$$

The decoding process itself is achieved by applying the decoding matrix on the Fourier-Bessel coefficients of the acoustic field to be decoded, as shown on figure 6.

2. APPLICATION TO MULTICHANNEL SOUND

As stated before, multichannel sound requires the user to place his loudspeakers at positions corresponding to the standard of the considered multichannel sound. Unfortunately, this is rarely the case, and people listen to sound that is spatially distorted because of the wrong positioning of the loudspeakers.

In part 1, a method has been described that allows to correctly reproduce an acoustic field over any loudspeaker layout. However, in order to apply this method to multichannel sound, two more issues have to be dealt with:

- the input signals of the decoder must be the Fourier-Bessel coefficients of the acoustic field to reproduce,
- the characteristics of the loudspeakers used (at least their positions) have to be known.

2.1. Sound environment conversion

In multichannel audio, the sound environment is provided under the form of Q signals denoted $c_q(t)$ having to feed Q loudspeakers placed on a circle in predefined directions denoted (θ_q^c, ϕ_q^c). As a consequence, this sound environment has to be converted into an acoustic field that can then be used by a decoder as determined in part 1 in order to obtain the loudspeakers feeds adapted to the actual loudspeaker layout.

This conversion can be done by associating to each channel a radiation pattern. This consists in determining Q transfer functions $p_q^R(r,\theta,\phi,t)$, each one allowing to compute an elementary acoustic field $\tilde{p}_q(r,\theta,\phi,t)$ from the multichannel signal $c_q(t)$. These transfer functions should be determined in such a way that the sound environment corresponding to the contribution of all these elementary acoustic fields is the same as the sound environment corresponding to the original multichannel signals. The relation between $\tilde{p}_q(r,\theta,\phi,t)$ and $c_q(t)$ is then a convolution:

$$\tilde{p}_q(r,\theta,\phi,t) = \int_{-\infty}^{+\infty} c_q(\tau)\, p_q^R(r,\theta,\phi,t-\tau)\,\mathrm{d}\tau$$

The transfer function $p_q^R(r,\theta,\phi,t)$ can be expanded into Fourier-Bessel series in order to get its Fourier-Bessel coefficients $R_{l,m,q}(f)$. These coefficients are arranged in a matrix R called reference radiation matrix (cf. appendix B). This matrix is much like the matrix M defined in section 1.3, except that the transfer functions $p_q^R(r,\theta,\phi,t)$ do not necessarily correspond to acoustic fields produced by existing loudspeakers. These transfer functions are indeed models corresponding to an ideal multichannel layout. Furthermore, even if $p_q^R(r,\theta,\phi,t)$ generally mainly depends on (θ_q^c,ϕ_q^c), an advanced transfer function $p_q^R(r,\theta,\phi,t)$ might also depend on the directions next to (θ_q^c,ϕ_q^c) in order to widen the corresponding virtual source so as to fill holes in the sound space.

For a plane wave reference radiation pattern, the coefficients $R_{l,m,q}(f)$ are obtained by Fourier-Bessel expansion:

$$R_{l,m,q}(f) = y_l^m(\theta_q^c,\phi_q^c)$$

In the frequency domain, the multichannel signals $C_q(f)$ and the Fourier-Bessel coefficients $\tilde{P}_{l,m}(f)$ of the acoustic field corresponding to the same sound environment as the multichannel signals can be arranged in two vectors, respectively $\boldsymbol{c}$ and $\tilde{\boldsymbol{p}}$ in order to get the following equation, called reference radiation relation:

$$\tilde{\boldsymbol{p}} = R\boldsymbol{c} \tag{4}$$

The acoustic field is thus determined using full matrix filtering from the multichannel signals. The resulting acoustic field can then be sent to the decoder. The combination of equations (4) and (1) gives:

$$\boldsymbol{v} = A\boldsymbol{c}$$

with

$$A = DR$$

The loudspeaker feeds $\boldsymbol{v}$ are thus the result of applying full matrix filtering of A on the multichannel signals $\boldsymbol{c}$. It is consequently possible to combine R and D to get a single matrix A. This is computationally more efficient because only one full matrix filtering is required instead of two. Matrix A is called adaptation matrix.

2.2. Automatic calibration

Decoder determination requires loudspeaker positions to be known. However, loudspeaker placement measurement is something difficult, and a sound system can hardly require the user to measure and input the loudspeaker positions into the sound system.

A sound system including a decoder that allows to adapt multichannel sound to the user's actual configuration should consequently include a way to automatically measure the positions, and possibly the frequency response, of the loudspeakers. This can be achieved by using a calibration microphone.

Loudspeaker distance can be determined from the delay between the loudspeaker signal and the calibration microphone signal by using sound speed (approximately 340 m/s).

If the calibration microphone is an Ambisonic-like one, the loudspeakers direction can be determined from the X, Y and Z signals. Descriptions of such microphones are quite numerous in the literature [2, 3, 5–10, 12–14, 17–23] and will not be given here.

2.3. Implementation

The above described multichannel optimization system has been implemented and consists of:

- a small 4 electret capsule microphone arranged in a tetrahedron, in order to calibrate the loudspeaker layout,
- a computation program that sends the test signals in order to calibrate the system, computes the adaptation matrix and applies it to the multichannel signals in order to deliver the loudspeaker feeds,

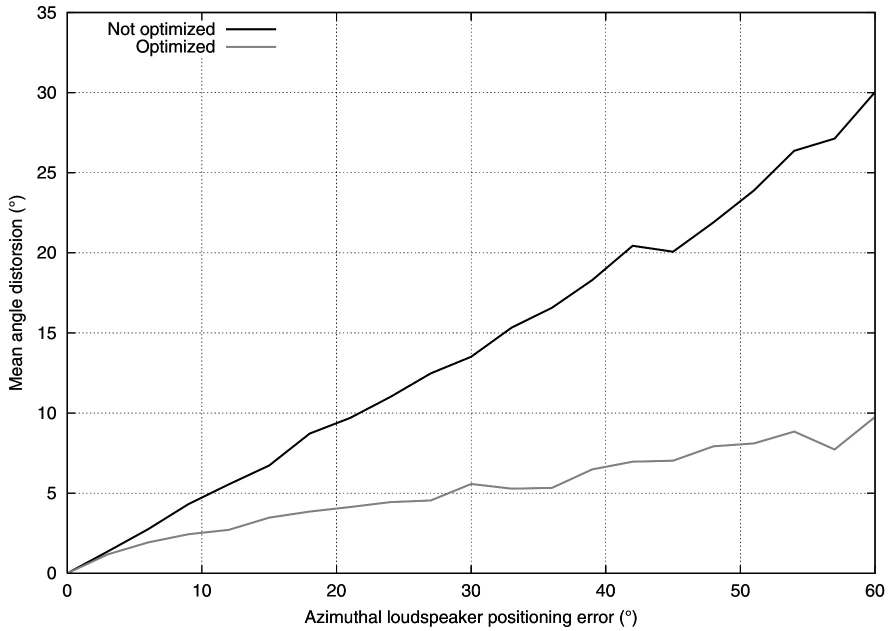


Fig. 7: Angle distortion with and without optimization.

- an interface that allows to visualize the calibrated layout and to change the optimization mode in order to compare between:
  - a multichannel sound played on a correct configuration (called *ideal sound*),
  - the same multichannel sound played on an incorrect configuration (called *incorrect sound*),
  - the same multichannel sound optimized and played on the same incorrect configuration (called *optimized sound*).

This prototype has been presented to several people, including sound professionals, and they all agreed that the optimized sound was much closer to the ideal sound (sometimes if distinguishable) than to the incorrect sound. Listening tests with objective and subjective evaluation criteria still have to be set up in order to quantify and qualify the improvement.

Angle distortion correction has been simulated using a large amount of incorrect virtual 5.0 configurations. The simulations consist in taking an ideal virtual configuration of loudspeakers and moving each loudspeaker randomly around its nominal position. The resulting distortion in sound direction is then calculated for various standard deviations in loudspeaker positioning. Figure 7 shows the results of these simulations. The x axis of this figure corresponds to the standard deviation in the simulated loudspeaker layout. The y axis gives the resulting error in sound direction. This is computed by moving a virtual source around the configuration. This figure show that when the loudspeaker feeds are not optimized, the distortion is about the half of the standard deviation, which is statistically expected. These simulations bring to the fore the fact that the optimization greatly reduces the distortion, by about 2/3.

3. CONCLUSION

Spatial sound distortion on users bad configurations is not a fatality, as it is possible to adapt a multichannel signal to an actual loudspeaker layout. Such an adaptation can be achieved by following an acoustic field approach that is completely transparent to the user.

Multichannel sound optimization first needs to calibrate the users' actual loudspeaker layout. It then converts the multichannel sound into an acoustic field, and decodes it for the actual system configuration.

Such a system can easily be inserted in a standard audio chain and requires very few user intervention. One only needs to push a button in order to calibrate and use the configuration with optimized loudspeaker feeds. The obtained results are quite satisfying, as the resulting optimized multichannel sound is close to the one obtained using a configuration following the recommendations.

As a result, the process described in this paper makes multichannel systems easier to use and more reliable for delivering in homes the original sound experience created in studios.

APPENDICES

A. THE FOURIER-BESSEL EXPANSION

The Fourier-Bessel expansion gives the acoustic field $p(r, \theta, \phi, t)$ as a function of its Fourier-Bessel coefficients $p_{l,m}(t)$:

$$P(r, \theta, \phi, f) = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} P_{l,m}(f)\, j^l\, j_l(kr)\, y_l^m(\theta, \phi)$$

where $P(r, \theta, \phi, f)$ and $P_{l,m}(f)$ are the time Fourier transforms of respectively $p(r, \theta, \phi, t)$ and $p_{l,m}(t)$. The functions $j_l(x)$ and $y_l^m(\theta, \phi)$ are respectively the spherical Bessel functions of the first kind, and the real spherical harmonics. These are given by

$$j_l(x) = \sqrt{\frac{\pi}{2x}}\, J_{l+\frac{1}{2}}(x)$$

$$y_l^m(\theta, \phi) = \frac{1}{\sqrt{2\pi}}\, P_l^{|m|}(\cos\theta)\, \mathrm{trg}_m\, \phi$$

where $J_\nu(x)$ is the (cylindrical) Bessel function of the first kind and order ν, and with

$$\mathrm{trg}_m\, \phi = \begin{cases} \sqrt{2}\, \cos m\phi & \text{for } m > 0 \\ 1 & \text{for } m = 0 \\ \sqrt{2}\, \sin m\phi & \text{for } m < 0 \end{cases}$$

The functions $P_l^{|m|}(x)$ are the associated Legendre functions and are given by

$$P_l^m(x) = \sqrt{\frac{2l+1}{2}} \sqrt{\frac{(l-m)!}{(l+m)!}}\, (1 - x^2)^{m/2}\, \frac{\mathrm{d}^m}{\mathrm{d}x^m} P_l(x)$$

where $P_l(x)$ are the Legendre polynomials :

$$P_l(x) = \frac{1}{2^l\, l!}\, \frac{\mathrm{d}^l}{\mathrm{d}x^l} (x^2 - 1)^l$$

Note that the normalization quantity in $P_l^m(x)$ varies from one publication to another. For instance, an additional $(-1)^m$ factor can sometimes be used.

B. NOTATIONS

In this paper, the following notations are used:

- $\boldsymbol{p}$ denotes a vector containing the time Fourier transforms of Fourier-Bessel coefficients of the considered acoustic field, and is arranged in this way:

$$\boldsymbol{p} = \begin{pmatrix} P_{0,0}(f) & P_{1,-1}(f) & P_{1,0}(f) & P_{1,1}(f) & P_{2,-2}(f) & \dots & P_{L,-L}(f) & \dots & P_{L,L}(f) \end{pmatrix}^{\mathrm{t}}$$

- $\boldsymbol{v}$ denotes a vector containing the time Fourier transforms of the loudspeaker feeds:

$$\boldsymbol{v} = \begin{pmatrix} V_1(f) & V_2(f) & \dots & V_N(f) \end{pmatrix}^{\mathrm{t}}$$

- D is the decoding matrix containing the filter parameters $D_{n,l,m}(f)$ of the full matrix filtering:

$$D = \begin{pmatrix} D_{1,0,0}(f) & D_{1,1,-1}(f) & D_{1,1,0}(f) & D_{1,1,1}(f) & \dots & D_{1,L,-L}(f) & \dots & D_{1,L,L}(f) \\ D_{2,0,0}(f) & D_{2,1,-1}(f) & D_{2,1,0}(f) & D_{2,1,1}(f) & \dots & D_{2,L,-L}(f) & \dots & D_{2,L,L}(f) \\ \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ D_{N,0,0}(f) & D_{N,1,-1}(f) & D_{N,1,0}(f) & D_{N,1,1}(f) & \dots & D_{N,L,-L}(f) & \dots & D_{N,L,L}(f) \end{pmatrix}$$

- M is the radiation matrix containing the transfer functions $M_{l,m,n}(f)$ between the loudspeaker feeds and the acoustic field they produce:

$$M = \begin{pmatrix} M_{0,0,1}(f) & M_{0,0,2}(f) & \dots & M_{0,0,N}(f) \\ M_{1,-1,1}(f) & M_{1,-1,2}(f) & \dots & M_{1,-1,N}(f) \\ M_{1,0,1}(f) & M_{1,0,2}(f) & \dots & M_{1,0,N}(f) \\ M_{1,1,1}(f) & M_{1,1,2}(f) & \dots & M_{1,1,N}(f) \\ \vdots & \vdots & & \vdots \\ M_{L,-L,1}(f) & M_{L,-L,2}(f) & \dots & M_{L,-L,N}(f) \\ \vdots & \vdots & & \vdots \\ M_{L,L,1}(f) & M_{L,L,2}(f) & \dots & M_{L,L,N}(f) \end{pmatrix}$$

- W is the weighting matrix:

$$W = \begin{pmatrix} W_0(f) & 0 & \cdots & & & & & 0 \\ 0 & W_1(f) & \ddots & & & & & \vdots \\ \vdots & \ddots & W_1(f) & \ddots & & & & \\ & & \ddots & W_1(f) & \ddots & & & \\ & & & \ddots & W_2(f) & \ddots & & \\ & & & & \ddots & W_L(f) & \ddots & \\ \vdots & & & & & \ddots & \ddots & 0 \\ 0 & \cdots & & & & & 0 & W_L(f) \end{pmatrix}$$

- F is the imposition matrix built from the $\left\{(l_k, m_k)\right\}_{1\leq k\leq K}(f)$ list. It is of size $K \times (L+1)^2$ and is given by:

$$F = \begin{pmatrix} F_{1,0,0}(f) & F_{1,1,-1}(f) & F_{1,1,0}(f) & F_{1,1,1}(f) & \dots & F_{1,L,-L}(f) & \dots & F_{1,L,L}(f) \\ F_{2,0,0}(f) & F_{2,1,-1}(f) & F_{2,1,0}(f) & F_{2,1,1}(f) & \dots & F_{2,L,-L}(f) & \dots & F_{2,L,L}(f) \\ \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ F_{K,0,0}(f) & F_{K,1,-1}(f) & F_{K,1,0}(f) & F_{K,1,1}(f) & \dots & F_{K,L,-L}(f) & \dots & F_{K,L,L}(f) \end{pmatrix}$$

with

$$F_{k,l,m}(f) = \begin{cases} 1 & \text{if } (l,m) = (l_k, m_k)\text{, with } (l_k, m_k) \text{ being the element } k \text{ of the list } \left\{(l_k, m_k)\right\}_{1\leq k\leq K}(f) \\ 0 & \text{if } (l,m) \neq (l_k, m_k)\text{, with } (l_k, m_k) \text{ being the element } k \text{ of the list } \left\{(l_k, m_k)\right\}_{1\leq k\leq K}(f) \end{cases}$$

- R is the reference radiation matrix containing the transfer functions $R_{l,m,q}(f)$ between the multichannel signals and the acoustic field corresponding to the same sound environment:

$$R = \begin{pmatrix} R_{0,0,1}(f) & R_{0,0,2}(f) & \dots & R_{0,0,Q}(f) \\ R_{1,-1,1}(f) & R_{1,-1,2}(f) & \dots & R_{1,-1,Q}(f) \\ R_{1,0,1}(f) & R_{1,0,2}(f) & \dots & R_{1,0,Q}(f) \\ R_{1,1,1}(f) & R_{1,1,2}(f) & \dots & R_{1,1,Q}(f) \\ \vdots & \vdots & & \vdots \\ R_{L,-L,1}(f) & R_{L,-L,2}(f) & \dots & R_{L,-L,Q}(f) \\ \vdots & \vdots & & \vdots \\ R_{L,L,1}(f) & R_{L,L,2}(f) & \dots & R_{L,L,Q}(f) \end{pmatrix}$$

REFERENCES

[1] Ambisonic website. URL: `http://www.ambisonic.net`.

[2] T. D. Abhayapala and D. B. Ward. Theory and design of high order sound field microphones using spherical microphone array. In *proceedings of the IEEE ICASSP-02, Orlando, Florida*, May 2002.

[3] Jeffery Stephen Bamford. *An Analysis of Ambisonic Sound Systems of First and Second Order*. PhD thesis, University of Waterloo, Ontario, Canada, 1995.

[4] Michel Bruneau. *Manuel d'acoustique fondamentale*. Hermes, 1998.

[5] P. Cotterell. *On the Theory of the Second-Order Soundfield Microphone*. PhD thesis, University of Reading, UK, 2002.

[6] Peter Graham Craven and Michael Anthony Gerzon. Coincident microphone simulation covering three dimensional space and yielding various directional outputs, August 1977. US Patent number 4,042,779.

[7] Jérôme Daniel. *Acoustic field representation, application to the transmission and the reproduction of complex sound environments in a multimedia context*. PhD thesis, University of Paris 6, July 2000. In French, English abstract available.

[8] Gary W. Elko. A simple first-order differential microphone. Technical report, Bell Labs, Lucent Technologies.

[9] Gary W. Elko, Robert A. Kubli, and Jens Meyer. Patent application, publication US 2003/0147539, Audio system based on at least second-order eigenbeams, August 2003.

[10] Gary W. Elko and Jens Meyer. A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield. In *proceedings of the IEEE ICASSP-02, Orlando, Florida*, May 2002.

[11] Angelo Farina and Emanuele Ugolotti. Software implementation of B-format encoding and decoding. In *proceedings of the 104th AES Convention*, May 1998.

[12] Michael A. Gerzon. Ultra-directional microphones: Part 1, part 2 et part 3.

[13] Michael A. Gerzon. Periphony: With-height sound reproduction. *J. Audio Eng. Soc.*, 21:2–10, January 1973.

[14] Arnaud Laborie, Rémy Bruno, and Sébastien Montoya. A new comprehensive approach of surround sound recording. In *proceedings of the 114th AES Convention, preprint 5717*, March 2003.

[15] Arnaud Laborie, Rémy Bruno, and Sébastien Montoya. Designing high spatial resolution microphones. In *proceedings of the 117th AES Convention, preprint 6231*, October 2004.

[16] Arnaud Laborie, Rémy Bruno, and Sébastien Montoya. High spatial resolution multichannel recording. In *proceedings of the 116th AES Convention, preprint 6116*, May 2004.

[17] Dave G. Malham. Higher order ambisonics systems for the spatialisation of sound. In *proceedings of the ICMC99*, October 1999.

[18] J. Meyer and T. Agnello. Spherical microphone array for spatial sound recording. In *proceedings of the 115th AES Convention, preprint 5975*, October 2003.

[19] James A. Moorer. Ultra-directional microphones: Part 4. Sonic Solutions, 2000–2001.

[20] James A. Moorer. Extension of the method of spatial harmonics to three dimensions. 2000-2001.

[21] Rozenn Nicol and Marc Emerit. Reproducing 3D-sound for videoconferencing: a comparison between holophony and ambisonic. In *proceedings of the DAFX98*, November 1998.

[22] Rozenn Nicol and Marc Emerit. 3D-sound reproduction over an extensive listening area: A hybrid method derived from holophony and ambisonic. In *proceedings of the 16th AES International Conference*, April 1999.

[23] M. A. Poletti. A unified theory of horizontal holographic sound systems. *J. Audio Eng. Soc.*, 48:1155–1182, December 2000.

[24] E. Skudrzyk. *The Foundations of Acoustics: Basic Mathematics and Basic Acoustics*. Springer-Verlag, 1971.