Audio Engineering Society
Convention Paper 5717
Presented at the 114th Convention
2003 March 22–25 Amsterdam, The Netherlands

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

A New Comprehensive Approach of Surround Sound Recording

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

[1] *ImmersiveSound project, Paris, France.*

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

ABSTRACT

Many techniques have been developed concerning surround sound recording and the issue has turned out to be a challenge without a comprehensive theory. This paper presents an approach based on a full 3D acoustic field theory and the use of a 3D microphone array. Our research work has lead to a new spatial digital processing technique which allows the use of freely positioned capsules of any type such as omnidirectionals, bidirectionals, or cardioids. This technique can be seen as an extension of full sphere, generalized Ambisonics which provides a spatial resolution never achieved before, but requires neither high order directivity capsules nor assumes that all capsules are coincident. The theory has been validated with a full sphere 3rd-order prototype using 24 omnidirectional capsules. The theory also allows for a 5th order spherical harmonic, multichannel 5.0 microphone.

0 INTRODUCTION

Spatial sound recording has been an issue for a long time. Beginning with stereo, important steps have been made. However, high spatial resolution sound recording is still a challenge [9, 14, 18] mainly because of the lack of good high order elementary microphones and of a comprehensive theory [20]. Fundamental acoustics has been understood for a long time but has long been unused. It does not currently offer ready-to-use tools except for Wave Field Synthesis. Surround sound reveals the need for further application of this theory.

In the early 70's, Michael Gerzon [8] developed Ambisonics, a recording technique based on the encoding of the incoming direction of the sound within the sound itself [1, 2, 10, 13]. This approach considers the three dimensional sound field as a source distribution around a central point and was an innovative extension of stereo in the same spirit as its inventor, Blumlein. Ambisonics allows separation of the recording stage from the reconstruction stage via the so-called encoding and decoding processes. The Ambisonic microphone contains four cardioid capsules which deliver the A-format. An encoder transforms these signals to the B-format [7]. The B-format signals have the advantage of being independent of the characteristics of either the recording or the reconstruction systems, contrary to stereophonic, quadraphonic or today's other multichannel schemes. During the reconstruction stage, a decoder generates the loudspeaker's feed signals from the B-format. The decoder depends on the loudspeaker's configuration and several decoders have been designed for various configurations, such as quadraphony or 5.1 multichannel format. The spatial resolution is limited, however, to first order spherical harmonics because the capsules used to record the sound are first order. More precisely, the tetrahedral microphone array known as the SoundField microphone, patented by Gerzon, has cardioid capsules [4].

Many extensions and developments concerning Ambisonics have been studied to reach higher orders [5, 6, 11, 12, 15, 16] but these extensions have been limited by the fact that there do not exist any higher order microphones with adequate characteristics. Moreover, the theory assumes that all microphones are physically at the same point, which leads either to imperfections or to requiring compensation of the signals.

In a recent article [17], M. A. Poletti developed a new theory for a circular microphone array using only omnidirectional or cardioid capsules which leads to a technique that does not require that the capsules be all at the same point and also provides high order directivity in the horizontal plane. He achieves this by no longer considering a source distribution around the origin, but rather by considering the physical sound field in a horizontal plane near the origin. This has the advantage of not requiring the knowledge of high order derivatives of the sound field at the origin to obtain high order directivity in the horizontal plane. This theory still has some drawbacks. The capsules are necessarily on a circle, and omnidirectional capsules can not be used because they lead to infinite gains. Moreover, the capsule noise is not taken into account and could be amplified during the encoding process. Finally, this theory uses a two dimensional model, which constitutes a significant departure from reality.

In this paper, a new spatial recording technique based on a three dimensional theoretical approach is presented. The goal of this technique is to extract and deliver as much spatial information as possible from the set of signals delivered by a microphone array. In order to do this, we use fundamental acoustics [3, 19] and develop new tools to apply it to spatial sound recording. This leads to a sound representation in a form compatible with full sphere, high order, generalized Ambisonics. However, this work is not derived from Ambisonics, as we do not consider the source repartition around the origin. This work can be seen as an extension to Ambisonics that overcomes its limitations. Our technique makes it possible to obtain high order Ambisonic signals from an array of capsules which can be of any type (omnidirectional, bidirectional, cardioid, etc.) and can have any position and orientation in space. Nothing is presumed a priori about the position of the capsules, so they are not necessarily on a plane or on a sphere. The resulting microphone consists of an array of capsules and a processing system based on a DSP or a PC.

This paper will first present the theoretical principle including the spatial sampling and encoding aspects. The practical results we have arrived at will then

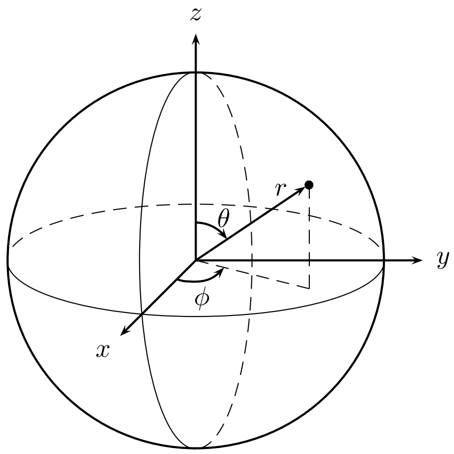


Fig. 2: Spherical Coordinate System.

be discussed, and last we will present a microphone prototype we have realized.

1 PRINCIPLE

Our approach is based on a model of the three dimensional acoustic field which allows us to describe any sound environment through a discrete number of signals. With this model of the sound environment and the knowledge of the microphone array properties (type, position and orientation of the capsules) it is possible to theoretically calculate the signals delivered by a microphone array. A relation we call the sampling relation theoretically yields the signals provided by a microphone array as a function of its properties and of the sound environment in which it is located. The aim of spatial sound recording is the opposite of this sampling. The capsule's signals are known and the goal is to estimate the acoustic field model that corresponds to the three dimensional sound environment. The principle of our method is to invert the sampling relation to obtain what we call the encoding relation which finally leads to the encoding process. This approach can be represented with the diagram given on figure 1. The aim is to determine a field representation as near as possible to the unknown field model with only the knowledge of the capsule's signals and their characteristics.

1.1 Three-dimensional modeling

An acoustic field can be modeled using a discrete number of signals in a source free region. This allows processing a spatial sound environment without having to consider a continuum of signals in space. We use a three dimensional model in spherical coordinates (cf. Fig. 2) called the Fourier-Bessel expansion. We consider sound pressure as the physical unit representing the acoustic field. This scalar field is defined in the region under consideration by the function $p(r, \theta, \phi, t)$. From this three dimensional continuous function the Fourier-Bessel decomposition [3, 19] gives a set of signals called Fourier-Bessel coefficients, 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:

$$P(r, \theta, \phi, f) = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} P_{l,m}(f)\, j^l\, j_l(kr)\, y_l^m(\theta, \phi) \tag{1}$$

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.

This decomposition is to be compared with that of the Fourier series of a periodic signal. Spherical coordinates have the advantage of revealing the periodicity of the sound pressure with θ and ϕ, which allows the acoustic field to be described using a discrete number of coefficients.

Ambisonics only considers the angular part of this decomposition, $y_l^m(\theta, \phi)$, which are the spherical harmonics. In order to fully describe the three dimensional field, and to overcome this limitation of Ambisonics, we use the spherical Bessel functions, $j_l(kr)$, in equation (1) to describe the radial behavior of the field.

Figure 3 represents 3 basis functions using 3 different visualization tools:

- The Directivity visualization represents the directivity function associated with an acoustic field consisting of the coefficient (l, m), the other

Fig. 1: Principle diagram.

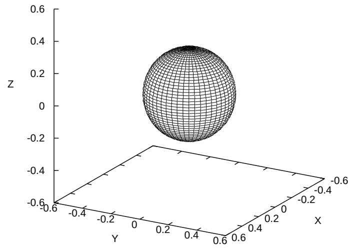
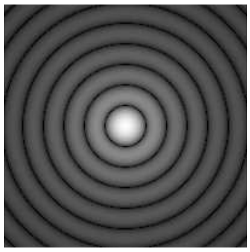
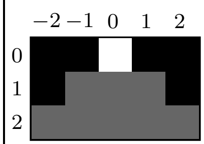
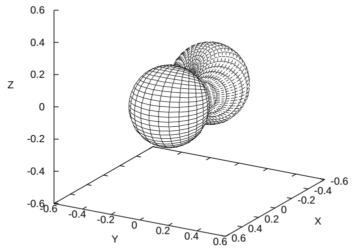
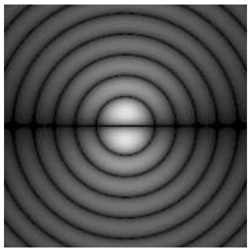
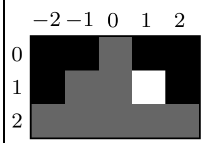
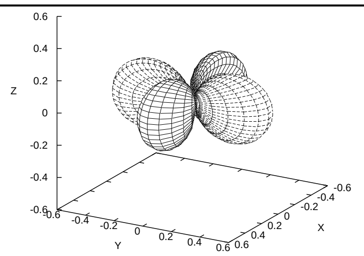
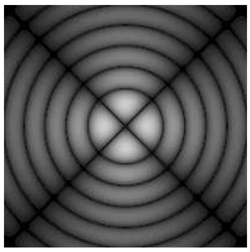
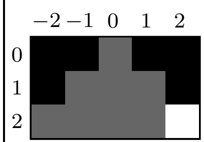
Type / Coef. #	Directivity	Field	Coefficients
$l = 0$, $m = 0$			
$l = 1$, $m = 1$			
$l = 2$, $m = 2$			

Fig. 3: Various representations of acoustic fields.

coefficients being null. This corresponds to the standard representation of spherical harmonics.

- 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.

1.2 Spatial sampling

The recording of an acoustic field is achieved using a set of elementary microphones. The configuration of the set is rigid, that is the relative position of each elementary microphone does not change. In order to distinguish between a single elementary microphone and a set of microphones forming a microphone array, we call an elementary microphone a capsule.

We do not impose any a priori constraint on the capsules forming the microphone array. The capsules can have any characteristic, position and orientation in space. They can be omnidirectional, bidirectional ("figure-of-eight"), or cardioid, for example. A microphone array can contain capsules of differing types.

Spatial sampling consists of measuring some of the information in an acoustic field. The input of a spatial sampler is thus an acoustic field, and the output is a set of signals representing some information measured from the input field. This is a very wide definition and any capsule or microphone array realizes a sampling of the field in which it is immersed. For example, a 0 order or omnidirectional capsule measures the sound pressure at the point in space where it is placed, which is a piece of information from the overall acoustic field. A figure-of-eight capsule measures the velocity of the acoustic field where it is placed which is proportional to the first order derivative of the pressure field along the capsule's orientation. The piece of information measured by a higher order capsule is more abstract and difficult to describe.

A representation of the sampling process is given in figure 4.

Spatial sampling is the spatial equivalent of temporal sampling. A temporal sample is a piece of information from an entire signal the same way a temporal signal coming from a capsule is a piece of information from the entire three dimensional acoustic field.

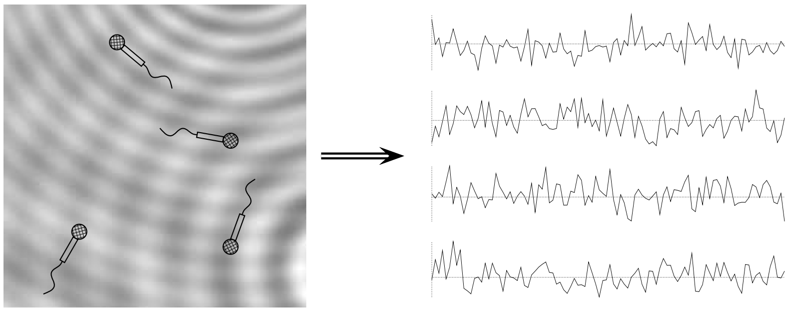
The goal of this paper is not to deal with a comprehensive spatial sampling theory but to describe the main results of its consideration.

The spatial equivalent to Shannon's theorem is a rather complicated issue because we deal with a priori irregular layouts. To limit spectrum aliasing, the Fourier-Bessel decomposition of the estimated acoustic field is truncated at some order, L. This order mainly depends on the number, N, of capsules. In general, since the number of Fourier-Bessel coefficients of order L is $(L+1)^2$, L is taken such that $(L+1)^2 \leq N$.

The goal of the remainder of this section is to obtain the sampling relation, that is the relation that makes it possible to theoretically calculate the signals provided by the microphone array from a known acoustic field. This relation depends on the microphone array characteristics (position, orientation and type of the capsules).

For example, if the capsules are omnidirectional, the sampling they realize is simply measuring the sound pressure at their position. In this case, if the acoustic field is known, it is known at the particular points where the capsules are located, so that it is possible to determine the theoretical signals that the capsules should deliver.

The acoustic field can be known in different ways, including using the previously described model, so that it is possible to determine the signals that the capsules will deliver if the Fourier-Bessel coefficients are known. With common capsules (more precisely, linear capsules), there is a linear matrix relation be-



Acoustic field Sampled acoustic field

Fig. 4: Sampling process.

tween the capsule's signals and the Fourier-Bessel coefficients in the frequency domain. If the capsule's signals are contained in a vector $\boldsymbol{c}$ and if the Fourier-Bessel coefficients are contained in a vector $\boldsymbol{p}$, the relation is given by

$$\boldsymbol{c} = B\boldsymbol{p} \tag{2}$$

The matrix B is called sampling matrix and it depends only on the microphone array characteristics. Examples of such matrices are given in appendix B for the cases of omnidirectional and cardioid capsules.

This matrix represents the way in which the microphone array samples and extracts information from the acoustic field.

1.3 Spatial encoding

Spatial encoding aims to estimate the three dimensional acoustic field using the signals provided by a few capsules. The goal is to calculate the Fourier-Bessel coefficients from the capsule's signals by inverting the sampling relation (2). The inversion process is affected a number of problems:

- The sampling matrix B is generally not square, so B^{-1} does generally not exist.
- The capsules are never exactly positioned where they should be; a real microphone array always has some positioning error.
- The modeling is not perfect, for example, omnidirectional capsules are not perfectly omnidirectional; they measure the pressure on a small surface and not at a point; they do not all have the same sensitivity. Generally, actual cardioid capsules present a significant departure from the ideal model.
- Last but not least, capsules, even the best ones, are noisy.

The first problem can be solved by generalized inversion which allows the inversion of non-square matrices. Ignoring the other points, however, could lead to unsatisfactory results. For example, at low frequencies, the wavelength reaches several meters, so that with a relatively small array, the signals from the capsules resemble each other very much. Since spatial information is obtained from the differences between the capsule's signals, noise on these capsules will become more and more important as the frequency decreases.

If a direct inversion is employed, noise will have a very important impact on the derived signals. To solve this problem, we introduce a parameter μ which specifies the desired compromise between spatial resolution and noise amplification. This parameter can vary between 0 and 1. A value of 1 means that we do not want to take noise into account at all. The smaller the value of μ, the less is noise amplified, but at the expense of lowering the spatial resolution.

The encoding process is determined via an encoding matrix, E, which gives the estimated field as a function of the capsule's signals. If the vector $\boldsymbol{c}$ contains the signals and the vector $\hat{\boldsymbol{p}}$ contains the

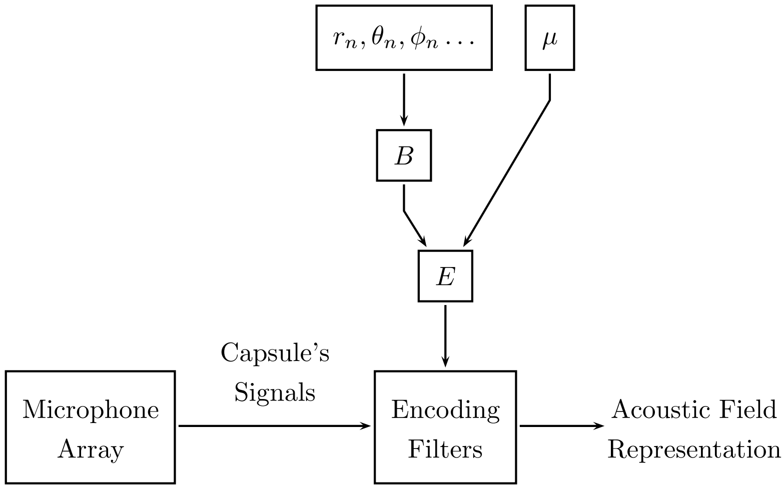


Fig. 5: Encoding scheme.

desired Fourier-Bessel coefficients of the estimated field, the relation is given by

$$\hat{\boldsymbol{p}} = E\boldsymbol{c} \tag{3}$$

The matrix E is calculated from B, the sampling matrix, and μ by

$$E = \mu B^\top \left(\mu B B^\top + (1-\mu)I\right)^{-1}$$

where I is the identity matrix and $X^\top$ denotes the transpose conjugate of matrix X.

Encoding a field with a microphone array requires specifying the sampling matrix B and, from it, calculating the encoding matrix E. Then each coefficient of E, being a function of frequency, represents a filter which can be parameterized. For an array with N capsules and an encoding of order L, there are $N \times (L+1)^2$ filters. The encoding itself is made by applying these filters to the signals provided by the capsules. Figure 5 represents this process.

As Ambisonics is a simplification of Fourier-Bessel theory, and is also based on spherical harmonics, the signals we calculate representing the acoustic field are compatible with generalized Ambisonics, and correspond to order L Ambisonic signals.

2 RESULTS

The preceding paragraphs describe a method for extracting as much spatial information as possible from a microphone array and for providing this information in a format suitable for generalized Ambisonics. However, they do not tell anything about optimal arrays. This section will try to give some hints to find out optimal arrays. Performances of common arrays will then be presented. Finally, practical results obtained with a prototype will be presented.

2.1 Microphones performances

In this paper, two types of microphone array performances are considered:

- **Spatial performances**, which give the three dimensional spatial fidelity of the estimated acoustic field in comparision with the original acoustic field. These include spatial distorsion and virtual sources position errors. Such distorsions can cause deformations on the sound environment, or bad envelopment feeling, for example.
- **Temporal performances**, which give audio quality of the calculated acoustic field. These can be thought of as a comparision between the signal provided by a virtual ideal capsule placed at a point in space and the signal calculated at that point from the estimated field. These include noise amplification and signal distorsion.

Temporal performances are tightly linked to capsules performances. As the encoding process is a linear

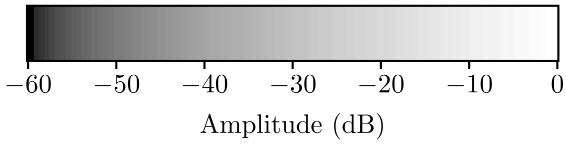


Fig. 6: Color scale used for matrix B visualisation.

matrix calculation, it does not add any distorsion to the signals provided by the capsules. However, it can amplify or weaken noise added by the capsules, depending on the value of μ and on the array layout itself. In the examples given below, μ has the value of 0.9, unless otherwise noticed. Temporal performances are evaluated using the classical signal to noise ratio.

Spatial performances depend mainly on the sampling matrix B. Indeed, this matrix determines how the microphone array takes information from the acoustic field. If an array does not measure any information on a Fourier-Bessel coefficient, there is no chance to be able to obtain any information about this coefficient using this array. For example, figure 8 represents the matrix B calculated until order 3 at the frequency of 1 kHz for a line of 24 capsules along the x axis between -10 cm and 10 cm. The microphone array layout is represented on figure 7. The color represents the amplitude of the matrix coefficient: the darker the coefficient is, the smaller is the modulus of the coefficient. The used scale is represented on figure 6. The amplitudes are relative to the largest coefficient.

This image contains many black columns. This means that this linear array does not take any information on the corresponding coefficients of the field. For example, the column for the coefficient $(1, -1)$ is null, which means that the array does not take any information on the y component of the direction of arrival of a plane wave. So it is impossible to know the projection on the y axis of the direction of arrival of a plane wave recorded by such an array. In other words, the array is unable to distinguish left from right in the sound environment.

Figure 10 represents the matrix B calculated until order 3 at the frequency of 1 kHz for a circular array of 24 capsules in the horizontal plane and which radius is 10 cm. This layout is represented on figure 9.

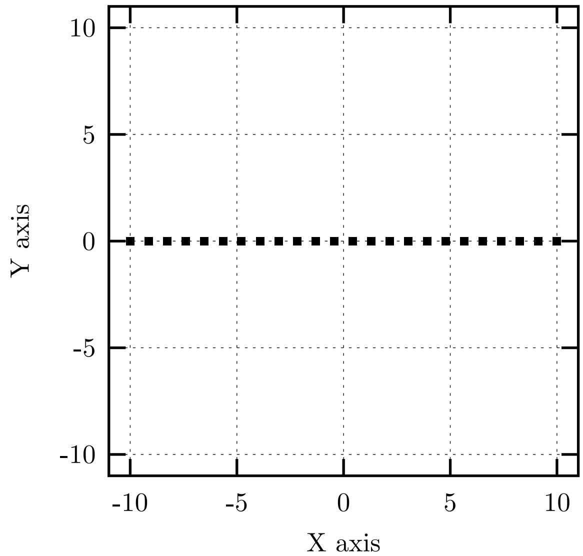


Fig. 7: Layout of the linear microphone array.

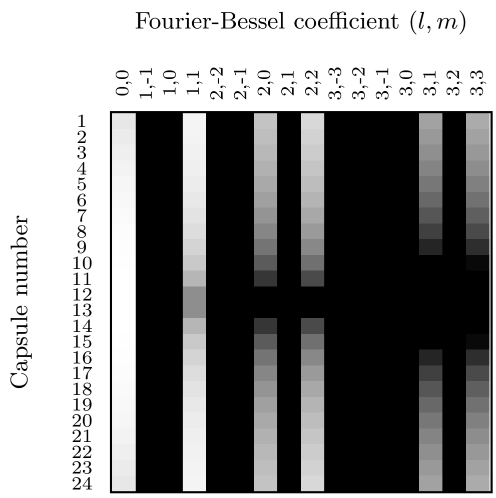


Fig. 8: Matrix B for a linear microphone array.

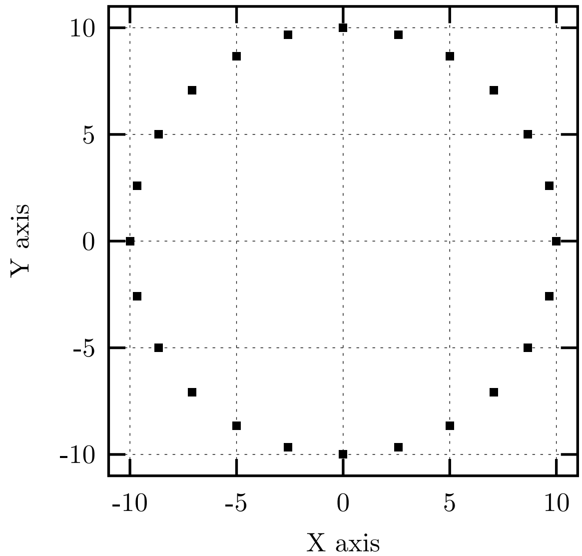


Fig. 9: Layout of the circular microphone array.

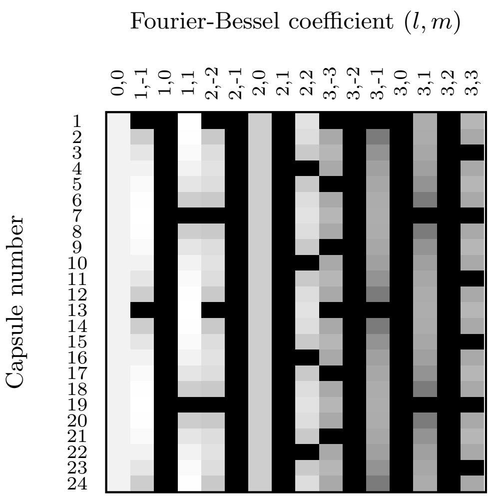


Fig. 10: Matrix B for a circular microphone array.

However, the matrix B is not sufficient to determine the spatial performances of an array. For example, the matrix B of a microphone array consisting of only one capsule can have large coefficients if the capsule is placed at a random position. Obviously, an array consisting of only one capsule has poor spatial performances.

In this article, we introduce a better indicator of the spatial performances of a given array. This indicator associates to each Fourier-Bessel coefficient (l, m) and to each frequency a value that gives the quantity of information on this coefficient that is measured from the acoustic field. We call this value spatial signal to noise ratio of Fourier-Bessel coefficient (l, m). This ratio corresponds to the uncertainty of the estimation of each Fourier-Bessel coefficient of the acoustic field.

Note that spatial signal to noise ratio has nothing to see with classical temporal signal to noise ratio, except that both are ratios between what is considered as an effective signal, and what is considered as noise. The signals and noises are not the same in both cases. In the temporal signal to noise ratio case, the signal is music, voices, etc. while the noise is shuffle or the neighbor's lawn mower, for example. In the spatial signal to noise ratio case, the signal is the part of sound that is correctly positioned in space while the noise part includes all the sounds that are incorrectly positioned. For example, a signal to noise ratio of 30 dB is a very bad temporal signal to noise ratio whereas it is a rather good spatial signal to noise ratio for first order coefficients, since it corresponds to an error of less than 2° on the incoming direction of a plane wave.

The vector ρ_s containing the spatial signal to noise ratios of the Fourier-Bessel coefficients is calculated from the matrices B and E. It is given in dB by the following relation:

$$\rho_{s\ (\mathrm{dB})} = -10 \log_{10} \left|\mathrm{diag}\left((EB - I)^\top (EB - I)\right)\right|$$

where I is the identity matrix and $\mathrm{diag}(X)$ is the diagonal of matrix X. The overall microphone performance is estimated using the mean value of ρ_s.

Figures 12 to 15 give this ratio for linear, circular, spherical and full three dimensional microphone arrays, as well as a representation of the corresponding configuration and the mean spatial signal to noise ra-

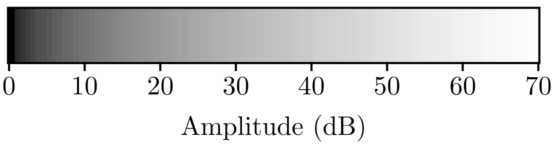


Fig. 11: Color scale used for spatial signal to noise ratio visualisation.

tio $\langle\rho_s\rangle$. The used color scale is shown on figure 11. These figures are shown for arrays having the same bulk, that is the capsules of the linear array are on its x axis from -10 cm to 10 cm, the radius of the circular and spherical arrays is 10 cm, and the capsules of the full three dimensional array lie within a 10 cm-radius ball. Linear and circular layouts are regular, whereas the capsules of the spherical and full three dimensional layouts are randomly placed. Sphere sampling is indeed a difficult issue since regular layouts with 24 capsules on a sphere do not exist. That is why we use random distributions.

Figure 12 shows again, that a linear array along the x axis does not give any information on the projection of a plane wave on the y and z axes.

Figures 13 and 14 show that the spatial signal to noise ratio becomes null for all coefficients at some frequencies. This points out the big inconvenient of circular and spherical layouts of radius R: the spherical Bessel functions oscillate and pass through zero periodically. This can be seen on figure 3, where dark areas on the field representations correspond to areas where the acoustic field cancels out. When the spherical Bessel function $j_{l_0}(kR)$ becomes null for a given frequency f_0, the elements $B_{n,l,m}$ of the matrix B cancel out for $l = l_0$ and for all capsules n, since they are all at the same distance from the origin. As a consequence, no information is taken from the acoustic field at this frequency on the Fourier-Bessel coefficients of order l_0. This explains the cancellation of the spatial signal to noise ratio at some frequencies. So the Fourier-Bessel coefficients cannot be acurately determined at these frequencies.

Figures 16 and 17 represent the spatial signal to noise ratio respectively for a circle and a sphere of cardioid capsules, with a representation of the corresponding configuration and the mean spatial signal to noise ratio. The layouts are the same as for figures 13 and 14. These figures show that taking cardioids does not improve performances very much, except for coefficients of order 1 and for low frequencies. In practice, the improvement is affected by the significant departure of real cardioids from the ideal model. In particular, they are generally not cardioid over the full frequency range. As a consequence, the theoretical improvements would be minimized by the performances of capsules themselves. Figure 18 represents spatial signal to noise ratio for a SoundField microphone. The four cardioid capsules are 1 cm away from the center. The SoundField microphone is unable to produce higher order coefficients, which is self-evident as it only has four capsules.

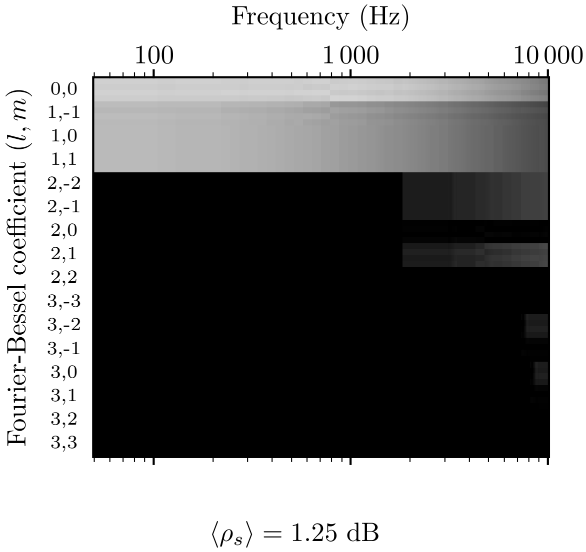


Fig. 18: Spatial signal to noise ratio for a SoundField microphone.

Figure 15 shows that capsules randomly positioned inside a ball give better performances than capsules positioned on a sphere. In particular, the preceding cancellation phenomenon is not observed.

Figures 19 and 20 give respectively the spatial and temporal signal to noise ratios as a function of μ for the 6 preceding layouts consisting of 24 capsules. We suppose here that the signal to noise ratio for the capsule's signals is 80 dB. If it were 90 dB, it would increase all the temporal signal to noise ratios by

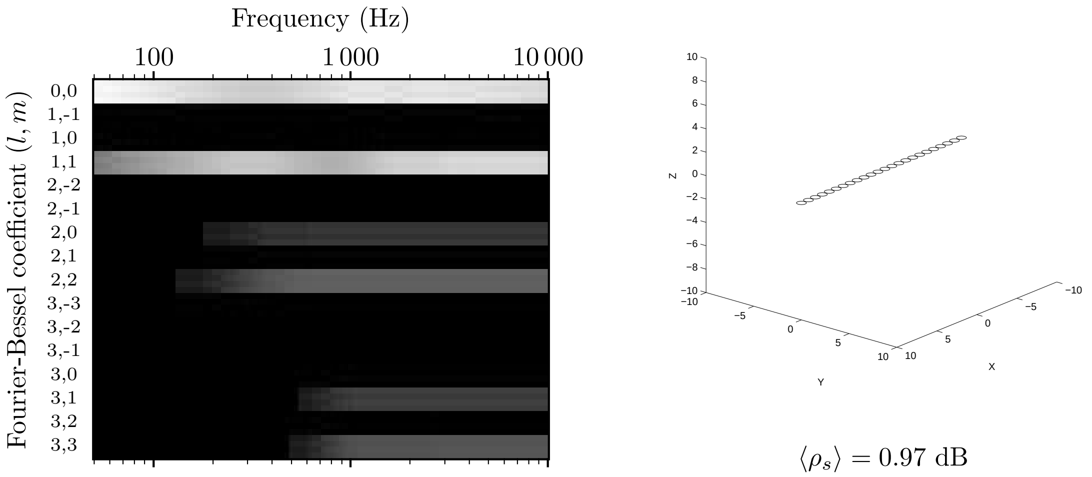


Fig. 12: Spatial signal to noise ratio for a linear microphone array.

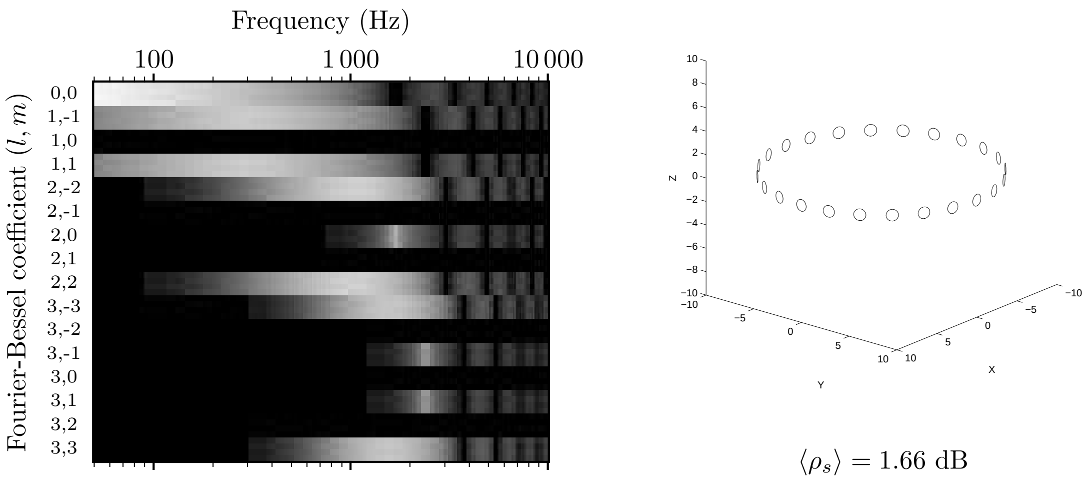


Fig. 13: Spatial signal to noise ratio for a circular microphone array.

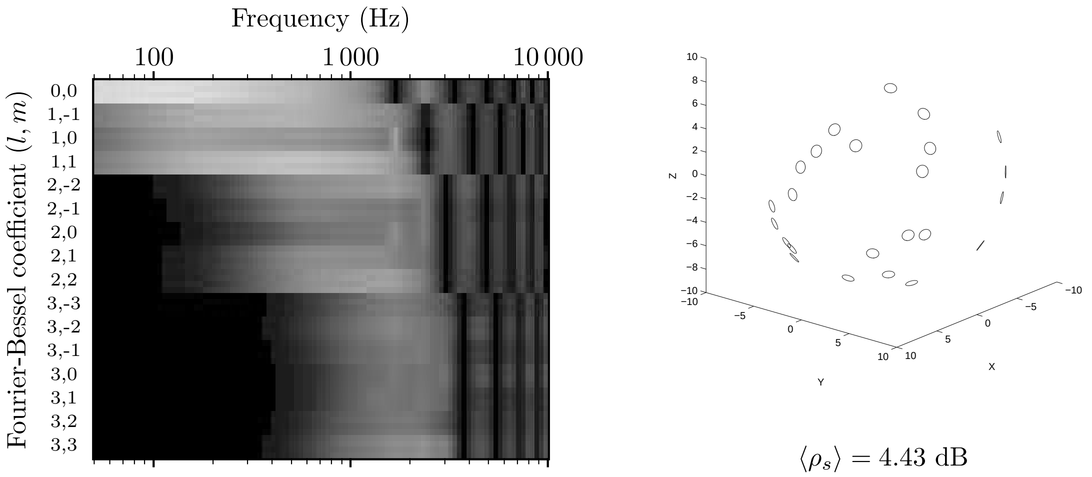


Fig. 14: Spatial signal to noise ratio for a spherical microphone array.

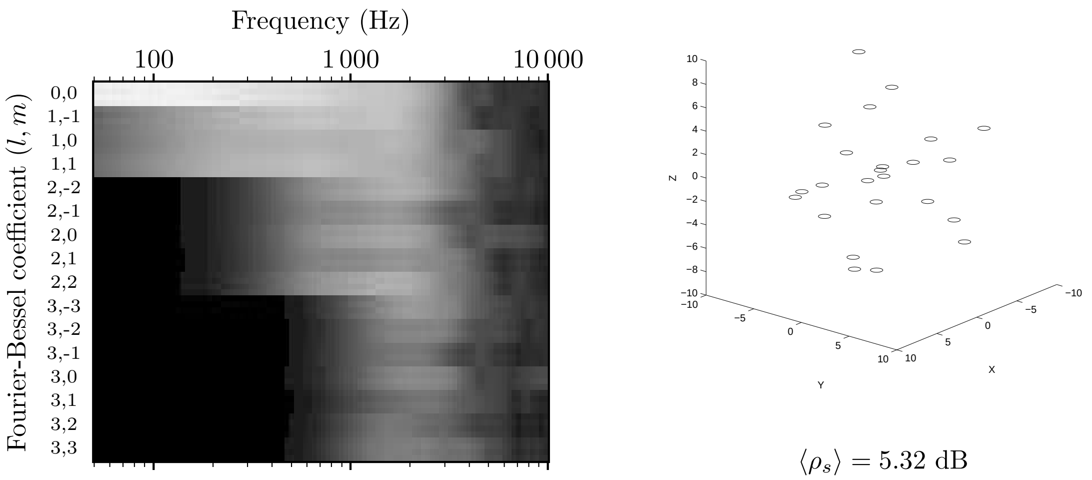


Fig. 15: Spatial signal to noise ratio for a full three dimensional microphone array.

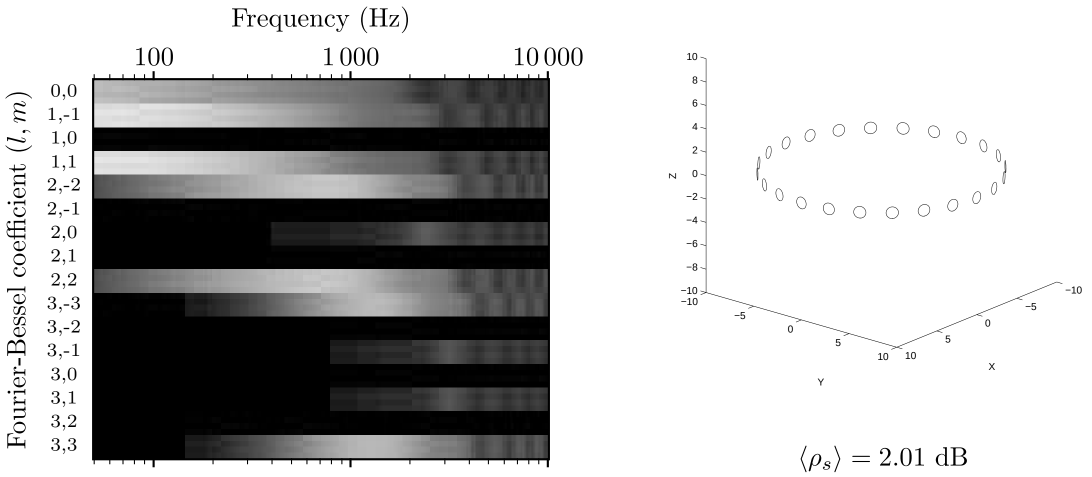


Fig. 16: Spatial signal to noise ratio for a circular microphone array with cardioid capsules.

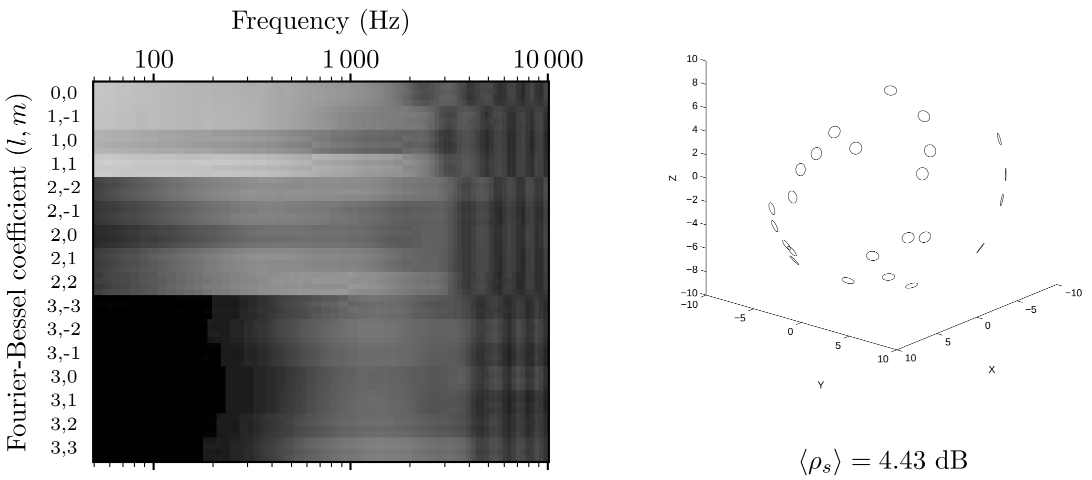


Fig. 17: Spatial signal to noise ratio for a spherical microphone array with cardioid capsules.

10 dB. The ratios are calculated using a mean over all the Fourier-Bessel coefficients and over the frequencies 100 Hz, 200 Hz, 500 Hz, 1 kHz, 2 kHz and 5 kHz. These graphs confirm the fact that temporal performances decrease and spatial performances increase when μ increases. They also point out the fact that irregular layouts give better overall results that regular ones.

So our theory allows to use any capsules layout, but it is also expected that an irregular layout is better than a regular one. This offers very versatile usage of this theory. For example, microphone arrays having good performances for high frequencies are obtained by placing all capsules near to the origin. On the opposite, large microphone arrays have good performances in low frequencies. This surround sound recording technique allows to mix capsules near to and far from the origin, thus giving relatively uniform performances along the frequency spectrum.

2.2 Prototype realization

In order to implement this theory and to verify the previous results, we realized a prototype with 24 omnidirectional capsules leading to third order. The previous section showed that regular layouts are not necessarily the best ones. We so had to find out a good capsules layout for our prototype. We carried out a lot of simulations to find optimal arrays. These simulations consisted in calculating the spatial and temporal signal to noise ratios for a lot of layouts, and in keeping layouts with best performances.

The constraints on this microphone array were that it should not be too big (its radius should not exceed 20 cm), and it should be realizable, that is have capsules not too near to each other. Several millions of layouts were simulated and we chose one among these.

The spatial signal to noise ratio of the chosen microphone is represented on figure 21. This figure uses the same color scale represented on figure 11.

We realized this microphone array using omnidirectional electret capsules and carbon fiber to fix them to each other. The calculations are performed by a standard PC. They consist in:

- Calculating the third order encoding filters using the principle described in section 1.3.

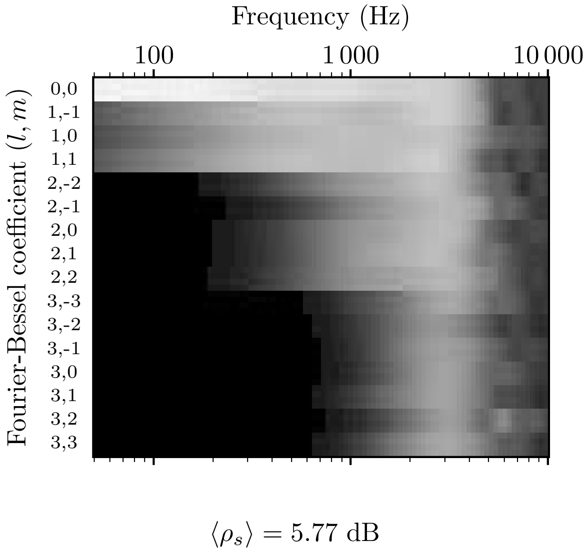


Fig. 21: Spatial signal to noise ratio for our prototype.

- Record the signals provided by the capsules in a 24-channel `wav` file.
- Apply the encoding filters on these recorded capsule's signals.
- Record the third order encoded signals in a 16-channel `wav` file.

Of course, the first step is realized only once for all the recordings we carried out. The encoding by applying filters can be made in real time on a modern PC. We listened to the encoded acoustic field using a 16-loudspeaker irregular full sphere restitution configuration and a decoder we made. The results were quite satisfying and in many cases, it is possible to think that we are into the action.

Figure 22 is a directional representation of an extract of one of our recordings. The axes represent elevation and azimuth of a direction and the color represents the amplitude of the power of the sound environment in that direction. For example, the center of the image corresponds to the direction $(\theta, \phi) = (90°, 0°)$ which is the frontal direction, the direction $(\theta, \phi) = (90°, 90°)$ corresponds to the left and $(0°, 0°)$ corresponds to the ceiling. The color

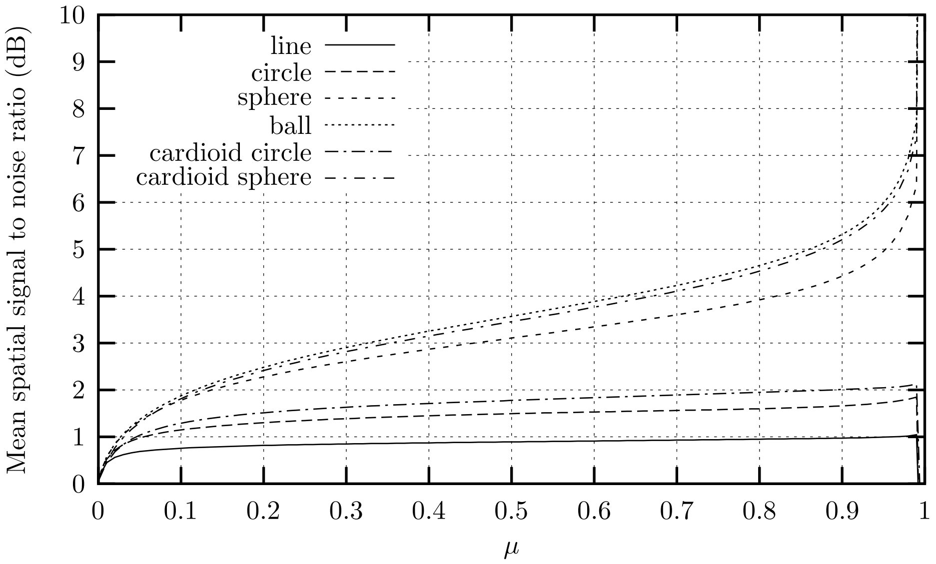


Fig. 19: Mean spatial signal to noise ratio as a function of μ.

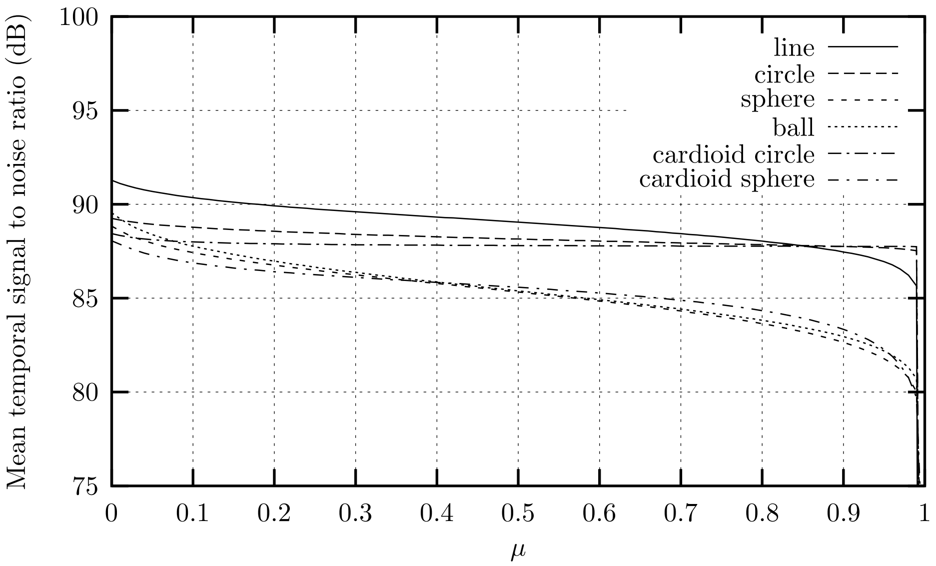


Fig. 20: Mean temporal signal to noise ratio as a function of μ.

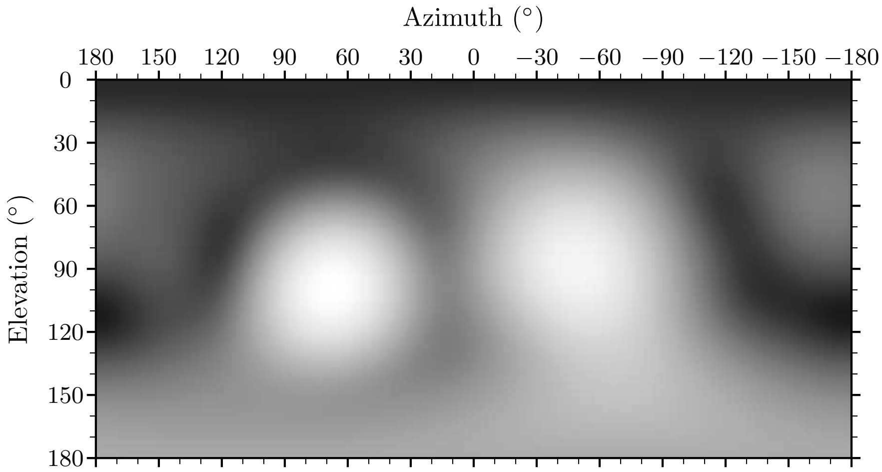


Fig. 22: Visualisation of a sound extract to order 3.

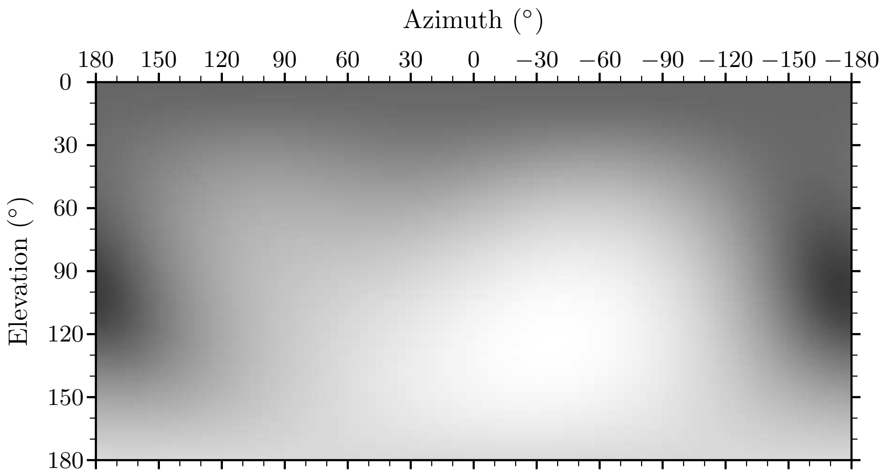


Fig. 23: Visualisation of a sound extract to order 1.

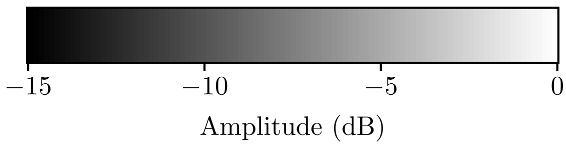


Fig. 24: Color scale used for field visualisation.

scale is given on figure 24. In this extract, two saxophones were playing in a church. The two spots corresponding to each saxophone are very well distinguishable. We can also see the reflections on the floor. Figure 23 is the same representation truncated to order 1. The difference between the two images is self-explanatory about the added value of higher orders.

3 CONCLUSION

This new comprehensive approach of full three dimensional sound recording allows to make a third order Ambisonic microphone using 24 omnidirectional capsules inside a ball of radius less than 20 cm. The encoding can be performed in real time on a modern PC.

This theory overcomes the limitations of Ambisonics due to the lack of good high order capsules and to the mandatory positioning at the center. With our technique, it is possible to use omnidirectional capsules reasonnably spaced without approximations, except for the functioning of the capsules themselves.

We have also realized a multichannel 5.0 microphone using 8 capsules and delivering 5 signals corresponding to 5 virtual microphones of 5th order. The reason why we achieve to have 5th order signals with only 8 capsules is that in the case of multichannel signals, the a priori on loudspeakers position allows to optimize the position of the capsules. Even for a 5.0 configuration, 5th order directivities constitute a great improvement over classical 1st order directivities, or even 3rd order directivities, because of the proximity of the three frontal loudspeakers. The improvements concern source punctuality, sweet spot size and localization precision. However, orders greater than 5 might be useless for a 5.0 configuration.

An application for a pattent has been filed in May 2002 on this technique.

4 ACKNOWLEDGEMENTS

The authors would like to thank Dr. Véronique Larcher of ELSA Productions and Bob Cain of Arcane Methods for their help.

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 in the litterature. A $(-1)^m$ factor is for example sometimes to be found.

B EXAMPLES OF SAMPLING MATRICES

The sampling matrix gives the relation between the acoustic field and the signals provided by a microphone array immersed in this field. We note $c_n(t)$ the signal provided by the capsule n of a microphone array consisting of N omnidirectional capsules, placed at positions (r_n, θ_n, ϕ_n). We place these signals and the Fourier-Bessel coefficients in two vectors as follows:

$$\boldsymbol{c} = \begin{pmatrix} C_1(f) \\ C_2(f) \\ \vdots \\ C_N(f) \end{pmatrix} \qquad \boldsymbol{p} = \begin{pmatrix} P_{0,0}(f) \\ P_{1,-1}(f) \\ P_{1,0}(f) \\ P_{1,1}(f) \\ P_{2,-2}(f) \\ \vdots \\ P_{L,-L}(f) \\ \vdots \\ P_{L,L}(f) \end{pmatrix}$$

where $C_n(f)$ and $P_{l,m}(f)$ represent the Fourier transforms of respectively $c_n(t)$ and $p_{l,m}(t)$. The matrix B consists of coefficients $B_{n,l,m}(f)$. Omnidirectional capsules simply measure the pressure at the point where they are, thus the sampling relation is simply directly derived from the Fourier-Bessel expansion given by equation (1). We have

$$\boldsymbol{c} = B\boldsymbol{p}$$

with

$$B_{n,l,m}(f) = 4\pi\, j^l\, j_l(kr_n)\, y_l^m(\theta_n, \phi_n)$$

In the case of capsules presenting any directivity, it can be proved that a similar relation is obtained, with a different matrix B. We give this matrix for cardioid capsules :

$$B_{n,l,m}(f) = 2\pi\, j^l \Bigg(j_l(kr_n)\, y_l^m(\theta_n, \phi_n) - j\Big(j_l'(kr_n)\, y_l^m(\theta_n, \phi_n)\; u_n^r - \frac{j_l(kr_n)}{kr_n}\; \sin\theta_n\, {P'}_l^{|m|}(\cos\theta_n)\; \mathrm{trg}_m\, \phi_n\; u_n^\theta + \frac{j_l(kr_n)}{kr_n\, \sin\theta_n}\; m\, y_l^{-m}(\theta_n, \phi_n)\; u_n^\phi \Big)\Bigg)$$

with

$$\begin{aligned} u_n^r &= \sin\theta_n\, \sin\theta_n^\alpha\, \cos(\phi_n - \phi_n^\alpha) + \cos\theta_n\, \cos\theta_n^\alpha \\ u_n^\theta &= \cos\theta_n\, \sin\theta_n^\alpha\, \cos(\phi_n - \phi_n^\alpha) - \sin\theta_n\, \cos\theta_n^\alpha \\ u_n^\phi &= \sin\theta_n^\alpha\, \sin(\phi_n^\alpha - \phi_n) \end{aligned}$$

Here, $(\theta_n^\alpha, \phi_n^\alpha)$ is the orientation of the capsule n in the spherical coordinate system. Primes denote first order derivatives.

REFERENCES

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

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

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

[4] 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.

[5] 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.

[6] Jérôme Daniel, Jean-Bernard Rault, and Jean-Dominique Polack. Ambisonic encoding of other audio formats for multiple listening conditions. In *proceedings of the 105th AES Convention, preprint 4795*, September 1998.

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

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

[9] Stanley P. Lipshitz. Stereo microphone techniques... are the purists wrong? *J. Audio Eng. Soc.*, 34:716–744, September 1986.

[10] Dave G. Malham. Spatial hearing mechanisms and sound reproduction. Technical report, University of York, 1998. URL: `http://www2.york.ac.uk/inst/mustech/3d_audio/ambis2.htm`.

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

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

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

[14] James A. Moorer and J. H. Vad. Towards a rational basis for multichannel music recording. In *proceedings of the 104th AES Convention, preprint 4680*, May 1998.

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

[16] 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.

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

[18] Francis Rumsey. *Spatial Audio.* Focal Press, 2001.

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

[20] Jorg Wuttke. General considerations on audio multichannel recording. In *proceedings of the 19th AES International Conference*, 2001.