Audio Engineering Society

Convention Paper 6231

Presented at the 117th Convention
2004 October 28–31 San Francisco, CA, USA

Designing High Spatial Resolution Microphones

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

Multichannel recording is one of the most important challenges of today's audio techniques. A good surround recording should provide at the same time good envelopment feeling, accurate localisation, a large sweet spot and respect for tones. Fourier-Bessel theory and advanced signal processing allow to obtain directivities designed from panning laws, which have been designed to optimally drive any multichannel layout. This paper presents the underlying concept of High Spatial Resolution, the spatial equivalent for High Fidelity, and points out why this is a key point to achieve high spatial quality. A very flexible and scalable technology providing High Spatial Resolution, as well as a high-performance 5.0 microphone featuring a compact array of 16 omnidirectional capsules are also presented.

0. INTRODUCTION

Multichannel sound is now widely used in the audio industry and is replacing stereo in more and more applications. However, surround recording techniques are still quite new, and getting a good multichannel recording is not straightforward. Such a recording should indeed provide a good envelopment feeling, a good source localization, a large sweet spot, and respect timbres, of course.

Recently, a large number of developments has been done in order to provide sound engineers with high-performance multichannel recording systems. In particular, a recently described [24] new technology allows to get high selectivity multichannel microphones using arrays of standard capsules and signal processing. An application of this technology using 8 microphones was presented.

The purpose of this paper is to evoke the main points of this technology and to demonstrate its flexibility by bringing to the fore the possibility to choose a compromise between performances and means. It can therefore be interesting to compare the performances of a 16-capsule microphone presented here with the performances of an 8-capsule microphone presented in [24].

1. HIGH SPATIAL RESOLUTION

In a previous article [24], Laborie, Bruno and Montoya presented a new technique for recording multichannel sound. This article introduced the concept of high spatial resolution that allows to design multichannel microphones delivering high spatial quality recordings. We will first present this concept here.

1.1. Possibilities of Current Techniques

Current multichannel recording techniques take advantage of the directivity characteristics of actual microphones to produce a spatial sound environment. Each microphone is connected to a channel, which is in turn connected to a loudspeaker. The position and directivity of each microphone directly determines what the loudspeaker reproduces. These techniques aim at finding adequate positions and directivities of available capsules in order to create intensity differences (ΔI) and/or time differences (ΔT) between channels in order to recreate the correct localisation cues at the listener ears: the interaural time differences (ITD) and/or the interaural level differences (ILD) [4, 5, 25]. Such a system is presented on figure 1.

All multichannel techniques can be classified in 2 families:

- Coincident techniques use directive microphones placed at the same point and generate only intensity differences (ΔI) between channels. The resulting sound image is stable and accurately localised but lacks depth. Moreover, directive microphones such as cardioid ones have some unwanted behavior, such as the well-known proximity effect. The sweet spot is relatively wide. The available systems are Soundfield and Double M-S [9, 38].

- Non-coincident techniques use spaced microphones and generate time differences (ΔT) between channels. The resulting sound image has depth, but is often unstable and the sweet spot is rather small. Furthermore, the delay between signals introduces so-called comb-filters. The available systems are Multi-Microphone Array, INA 5, IRT cross and OCT [38, 33, 20, 35, 36, 34].

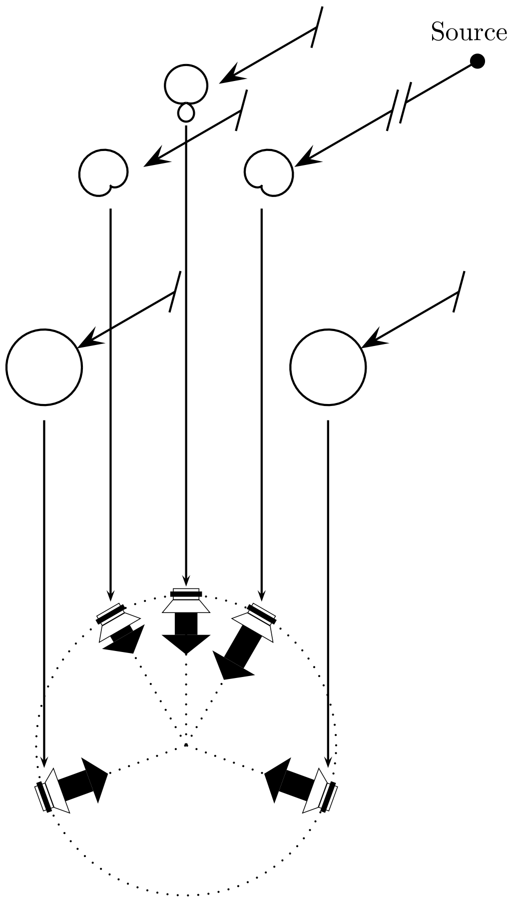


Fig. 1: Traditional 5.0 multichannel microphone.

The psychoacoustic model of multichannel recording is well understood for 2-channel stereo. However, the psychoacoustic models involving simultaneously ΔT and ΔI with more than 2 loudspeakers is complex and has not been fully studied. As a result, the development of current multichannel microphones extensively uses listening tests. Unfortunately, multichannel recording complicates things a lot, because it adds many possibilities concerning microphone placement and properties. This rapidly leads to a huge quantity of different possible multichannel microphones, that should be compared in order to determine the best one.

Moreover, the psychoacoustic model may not be easily simplified. In particular, considering a multichannel system as a set of stereo pairs is not perfectly correct, because each considered stereo pair interacts with one another, resulting in interferences phenomena or parasitic images. Furthermore, it is very difficult to get lateral images in 5.0 because systems are generally not designed to create images between the right or left front speakers and the right or left surround speakers.

1.2. High Spatial Resolution Approach

Current techniques put up with directivities of existing microphones and use them in the best possible way to design multichannel microphones based on psychoacoustic models inherited from stereophony. However, this optimum remains below the possibilities that a 5.0 multichannel system can reach. In order to get a maximum spatialization quality, an other approach should be considered and the multichannel microphone problematic should be considered from a different point of view: what would be the ideal behavior of a multichannel system recording sources over 360°? The answer is quite simple: it should behave in such a way that a recorded source in a given direction will generate a phantom image in the same direction on the multichannel restitution system. This gives, for each incoming direction, the targeted amplitude of each channel, which describes the sensitivity associated to each channel as a panning law. In other words, the recording system does not add any limitation, and each source in the original acoustic field is panned on the restitution system the same way as if it had been panned using a mixing desk.

As a consequence, directivities of panning laws can be interpreted as targeted directivities of a multichannel recording system, because such laws are designed to provide best possible phantom images over 360°. A system with such directivities is represented on figure 2. Examples of panning laws, each corresponding to a possible targeted directivity set for a multichannel microphone, are given on figures 3, 4 and 5 [8, 4, 5, 25].

A multichannel recording system providing such directivities is said to have *High Spatial Resolution* (HSR). This concept is the spatial equivalent for high temporal resolution, otherwise called *High Fidelity* (HiFi). High Fidelity aims at recording sound

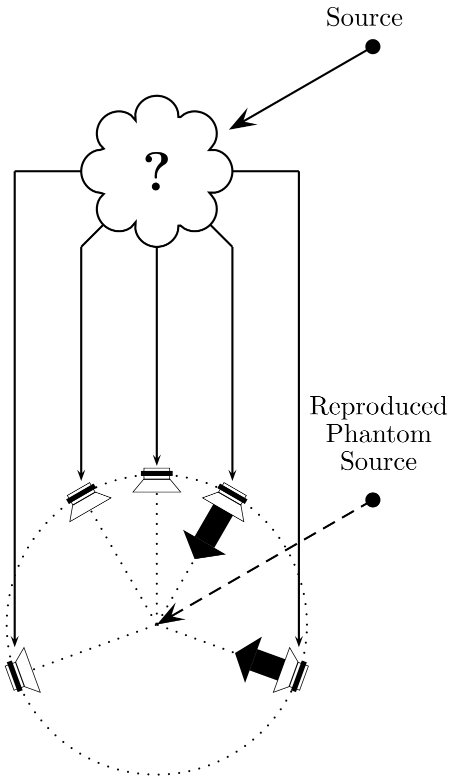


Fig. 2: Ideal 5.0 multichannel microphone.

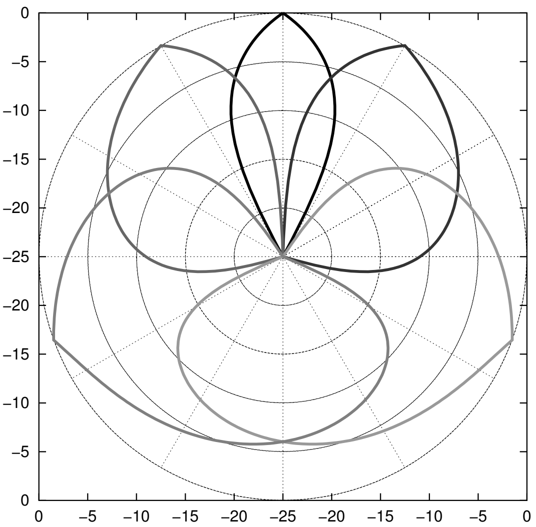


Fig. 3: Intensity panning law.

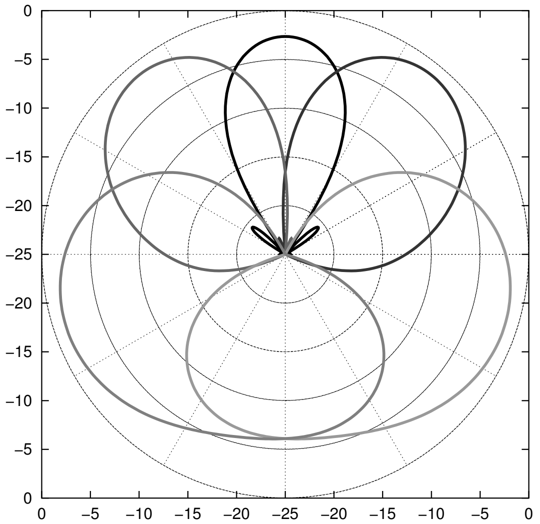


Fig. 4: 5th-order intensity panning law (cf. appendix A).

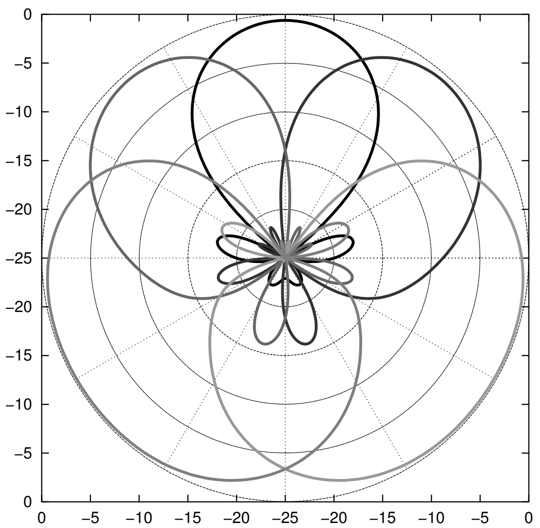


Fig. 5: 4th-order panning law (P. Craven).

precisely, including the least audible details, that is within a wide frequency band. Similarly, HSR aims at recording the acoustic field with a high spatial precision. Spatial resolution corresponds to the directivity. Whereas temporal resolution is characterized by the frequency, spatial resolution is characterized by the spatial frequency, which is also called the *order*. The higher the order is, the higher spatial resolution is. Order 0 corresponds to an omnidirectional directivity and does not select any direction in space. Order 1 corresponds to a figure-of-8, combinations of order 0 and order 1 give all directivities ranging from an omni to a bidirectional pattern, including cardioid and hypercardioid. All these directivities are directly provided by standard capsules. Higher orders give more selective patterns. More information about the notion of order can be found in [24].

Up to now, only azimuthal directivities of a multichannel microphone have been considered. But any sound environment is three-dimensional, and there are always sources outside the horizontal plane, may they be direct or belonging to reverberation. It should then be considered how a multichannel microphone records sources outside the horizontal plane, otherwise its behavior could be undetermined and could lead to unsatisfactory results. In order to achieve that, the full three-dimensional directivities are considered, using spherical harmonics, which are the three-dimensional equivalent for cylindrical harmonics. The targeted directivities of the multichannel microphone should then be specified in three dimensions, and there are several ways to expand the 2D directivity corresponding to a panning law in 3D. One of the possible ways is represented on figure 6.

1.3. Main characteristics of HSR

The High Spatial Resolution approach leads to multichannel directivities that are very different from standard 1st-order directivities. Their main characteristics are:

Selectivity. This point has already been discussed above, and is at the origin of HSR. The main consequence is that a source in a given direction is mainly reproduced by the two closest channels. For example, the frontal channel records sources only in the range $[-30°, 30°]$ (cf. fig. 4).

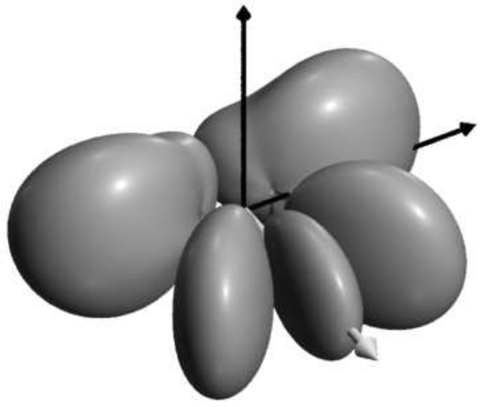


Fig. 6: Possible three-dimensional expansion of a panning law.

It should be noticed that even the rear channels have high spatial selectivity (about 3rd order).

Asymmetry. Except for the frontal channel, directivities are asymmetrical, that is the direction of the associated loudspeaker is not an axis of symmetry of the directivity. This is required if best adaptation to 5.0 irregularity is targeted. If the 5.0 layout were regular, HSR directivities would be symmetrical. Existing microphones do not allow such asymmetry.

Complementarity. The sum of all directivities gives an omnidirectional pattern. This means that, when a source is moving around, its power does not depend on its direction. This property is verified for each frequency, so that the timbre of a source does not vary when the source moves around the microphone.

Such characteristics are not possible using existing microphones. But, using new techniques based on signal processing, it is possible to get such directivities using standard microphones.

2. IMPLEMENTATION

2.1. Principle

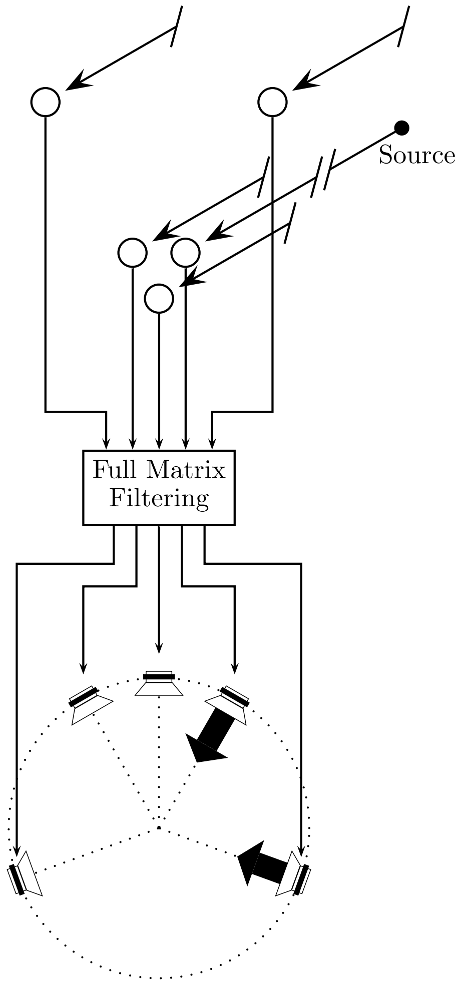


Fig. 7: Full-Matrix-Filtering-based 5.0 multichannel microphone.

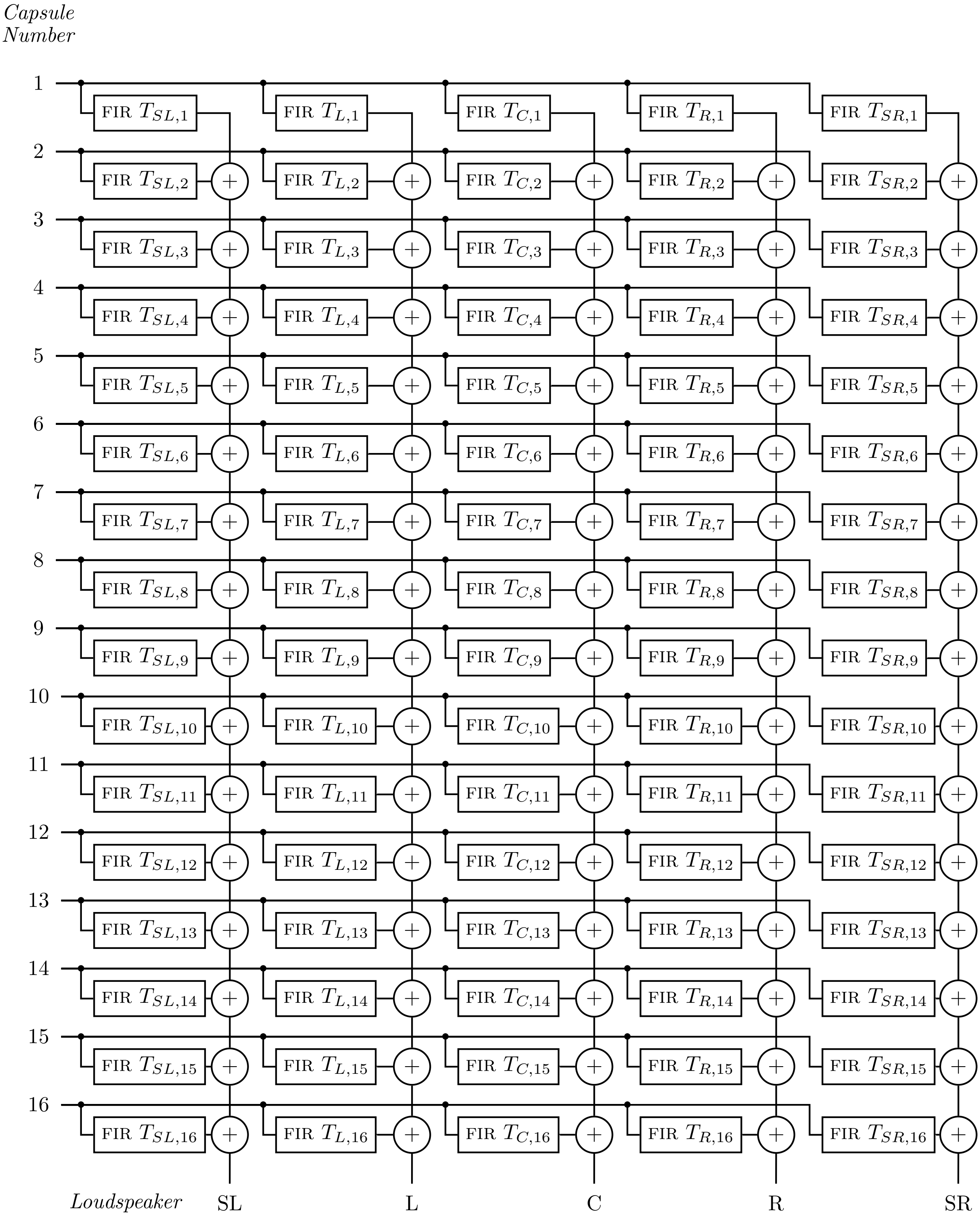


Fig. 8: Full Matrix Filtering.

New approaches have recently been developed in order to improve surround sound recording [37, 2, 21]. They generally use beam-forming techniques to generate directivities of orders more than 1. However, they require considerable means, or new hardware developments. For example, Hulsebos et al. [21] describe a surround microphone with very high performances, but that requires 288 capsules or new hardware developments concerning the microphone membrane. This unfortunately makes these techniques not directly usable by sound engineers today.

Besides, Ambisonics aims at accurately recording the spatial frequencies (spherical harmonics). In its original approach, Ambisonic microphones has long been limited to 1st-order resolution. The most famous implementation is the Soundfield microphone composed of a tetrahedral array of cardioid microphones and a matrix processing [17, 9]. Higher-order microphones were impossible as they required a large number of high-order directivity capsules located at one point. In the early 70s, linear array microphones combined with beam-forming processing have been investigated but led to a narrow frequency range or to a large number of capsules [16, 27].

Recently, a deeper connection between Ambisonics and fundamental acoustics [3, 10, 11, 29, 30] made it possible to overcome the original limitations and to achieve HSR recording. It has been proved that spacing between the recording points is necessary to record high order spatial frequencies using standard zeroest or first order capsules. This approach has been investigated for circular arrays [31], spherical arrays [1, 7, 14, 13, 15, 18, 26] and generalised for any microphone array composed of freely positioned and oriented standard capsules [23, 22]. It has been shown that irregular arrays achieve more efficient spatial recordings over circular and spherical arrays.

In this article [23], an Ambisonic microphone of 3rd order using 24 omnidirectional capsules is described. This microphone uses advanced signal processing in order to obtain the 16 first 3D spatial frequencies corresponding to spherical harmonics up to order 3. The article explains how to determine a matrix of filters that outputs these 16 signals from the 24 input signals coming from the capsules, using a full 3D acoustic field approach.

However, directly applying this technique to design a 5th-order multichannel microphone would require at least 36 omnidirectional capsules (preferably 48), which is not feasible in terms of hardware required and costs for most studios. It must be kept in mind that an Ambisonic microphone of 5th order records sound with the same spatial resolution in every direction. In other words, it is possible to get a 5th-order angular dirac in any direction using a 5th-order Ambisonic microphone, which is not required to get a good 5th order multichannel microphone. A dirac-like directivity is only required for the frontal channel. The main difference between a 5th-order Ambisonic microphone and a 5th-order multichannel microphone is that the listening configuration is known *in advance* in the case of the multichannel microphone. This allows to make additional optimisations that allow to drastically reduce the number of required microphones.

These optimisations do not change the principle, which consists of a Full Matrix Filtering. This means that each input signal coming from a capsule gets filtered, typically using an FFT-based calculus, and contributes to each output channel. Each filter used is proper to each couple (microphone, loudspeaker). This principle is illustrated on figure 8. The resulting system is represented on figure 7.

This approach allows to scale the designed multichannel microphone depending on the performances required and on the hardware resources available. For a 5.0 multichannel microphone, high spatial resolution is possible from an array of 8 omnidirectional capsules [24]. However, it is possible to get even better performances by increasing the number of capsules.

2.2. Optimizations

The main optimization consists in designing filters to get the targeted 5.0 directivities directly, instead of having filters that deliver 5th-order Ambisonic, and composing the targeted directivities out of that.

Determination of the filters is based on the Fourier-Bessel expansion, that will not be detailed here. A previous article [23], as well as fundamental acoustics manuals [6, 32] explain this theory. Fourier-Bessel expansion allows to describe a three-dimensional acoustic field in a region of space as the sum of elementary acoustic fields. The associated weighting coefficients are called Fourier-Bessel co-

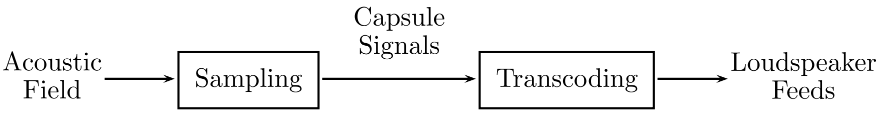


Fig. 9: Transcoding principle diagram.

efficients. More detailed elements can be found in appendix B.

The 3D acoustic field approach described in [23] is based on the acoustic field sampling principle using a microphone array. Acoustic field sampling is the spatial equivalent for time sampling: spatial sampling consists in measuring some information of the acoustic field using standard capsules. The input of a spatial sampler is thus an acoustic field, and the outputs are the signals coming from the capsules, which are the samples of the acoustic field. This principle allows to determine a matrix relation that gives the capsule signals — which are known — from the acoustic field in which these capsules are immersed — which is searched. This relation has a very simple linear relation. If we denote $\boldsymbol{p}$ the vector containing the Fourier-Bessel coefficients of the considered acoustic field, and $\boldsymbol{c}$ the vector containing the capsule signals, the relation is

$$\boldsymbol{c} = B\boldsymbol{p} \tag{1}$$

where B is a matrix depending on the microphone array characteristics, including the capsule directivities and positions. The matrix B is called *sampling matrix*. More information about the vectors and matrices used can be found in appendix B. This relation is in the frequency domain, and each parameter depends on f. This means that $\boldsymbol{c}$ and $\boldsymbol{p}$ are in fact the Fourier transforms of the capsule signals and Fourier-Bessel coefficients. This relation is called *sampling relation*, because it gives the signals $\boldsymbol{c}$, which are samples of the acoustic field, from this acoustic field $\boldsymbol{p}$. Basically, in [23], the relation (1) is inverted using standard matrix inverting methods, and an estimated acoustic field $\hat{\boldsymbol{p}}$ is determined as a function of $\boldsymbol{c}$.

Here, the purpose is not to determine the acoustic field $\boldsymbol{p}$, but loudspeaker feeds $\boldsymbol{v}$ that would produce an acoustic field $\tilde{\boldsymbol{p}}$ close to $\boldsymbol{p}$. The relation between $\boldsymbol{c}$ and $\boldsymbol{v}$ should be linear, so that determining $\boldsymbol{v}$ from $\boldsymbol{c}$ consists in matrix filtering as represented on figure 8. Consequently, what is to be determined is a transcoding matrix T such that

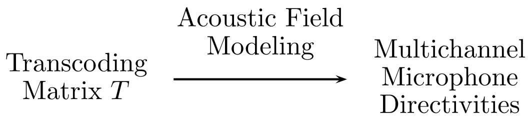
$$\boldsymbol{v} = T\boldsymbol{c} \tag{2}$$

The elements of this transcoding matrix, being a function of frequency, represent the frequency responses of the matrix filters represented on figure 8. The transcoding principle is represented on figure 9.

Equations (1) and (2) allow to know the loudspeaker feeds $\boldsymbol{v}$ from the acoustic field $\boldsymbol{p}$ for a given matrix T:

$$\boldsymbol{v} = TB\boldsymbol{p} \tag{3}$$

If $\boldsymbol{p}$ corresponds to a plane wave in a given direction, $\boldsymbol{v}$ gives the loudspeaker feeds when the multichannel microphone is exposed to a plane wave in that same direction. This corresponds exactly to the definition of the directivities of the multichannel microphone. In other words, the influence of the chosen T over the actual directivities of the multichannel microphone is perfectly known. This is schematically represented below:



As in [23], what we want is exactly the opposite: determining the matrix T from the targeted directivities using the acoustic field modeling. In order to determine T, least squares algorithms are used. Each targeted property of the multichannel microphone directivities is associated with an elementary error that is to be minimized in order to best verify the considered property. Then, a global error $e^2(T)$ is defined as a weighted sum of these elementary errors. The searched transcoding matrix T is then the matrix that minimizes $e^2(T)$. The weighting coefficients allow to change the compromise between the considered properties. The most important property is the proximity of the obtained directivities with the targeted directivities. The elementary error $e_p^2(T)$ is thus introduced and corresponds to the difference between the actual directivities obtained with T and the targeted directivities.

However, all capsules, even the best ones, are noisy and there are always small modeling errors (for example, omnidirectional capsules do not take the pressure at one point, but an average over its membrane). Noise amplification, denoted $e_n^2(T)$, should then be taken into account in the global error, so that transcoding will not amplify that noise.

Furthermore, the directivities should not be taken into account only in the horizontal plane, but also in three dimensions. Otherwise, the elevation behavior would not be controlled, and transcoding matrices could be obtained that amplify the floor or the ceiling too much. A weighting matrix W is consequently introduced in the error $e_p^2(T)$. This matrix corresponds to a spatial weighting to be applied. For example, if all 3D directions are to be weighted the same way, W corresponds to an identity matrix.

The global error $e^2(T)$ is thus given by

$$e^2(T) = e_p^2(T) + \mu\, e_n^2(T)$$

with

$$\begin{aligned} e_p^2(T) &= \operatorname{trace}\big(TB - U_0\big)W\big(TB - U_0\big)^\top \\ e_n^2(T) &= \operatorname{trace} TT^\top \end{aligned}$$

where U_0 is a matrix containing the targeted multichannel microphone 3D directivities, $X^\top$ denotes the transpose conjugate of matrix X, and μ is a parameter that specifies the compromise between spatial resolution and noise amplification. The higher the value of μ, the less is noise amplified, but at the expense of lowering spatial resolution.

The targeted multichannel microphone 3D directivities U_0 are typically determined from panning laws such as the ones represented on figure 3, 4 and 5. However, these directivities, which sum generally are an omni, can be multiplied by a global directivity, so that it is possible to change the overall directivity of the resulting multichannel microphone, and thus its directivity index and distance factor. Figures 10, 11, 12 and 13 represent 4 examples of targeted multichannel microphone 3D directivities. These examples are successively a full 3D omni, an omni with attenuated ceiling and floor, and two directivities with attenuated back. Each figure represents:

- The global directivity in the horizontal plane ($z = 0$). The x axis points to the top, and the y axis points to the left.
- The global directivity in the vertical plane ($y = 0$). The x axis points to the right, and the z axis points to the top.
- A perspective view of the global directivity in three dimensions. The x axis points to the bottom right corner of the image, and the z axis points to the top.
- A perspective view of all directivities in three dimensions. The x axis points to the bottom right corner of the image, and the z axis points to the top.

The title of the figures also indicates the directivity index I_d and the distance factor f_d.

The matrix T is then the least squares solution, that is the matrix that minimizes the error $e^2(T)$. This solution is given by

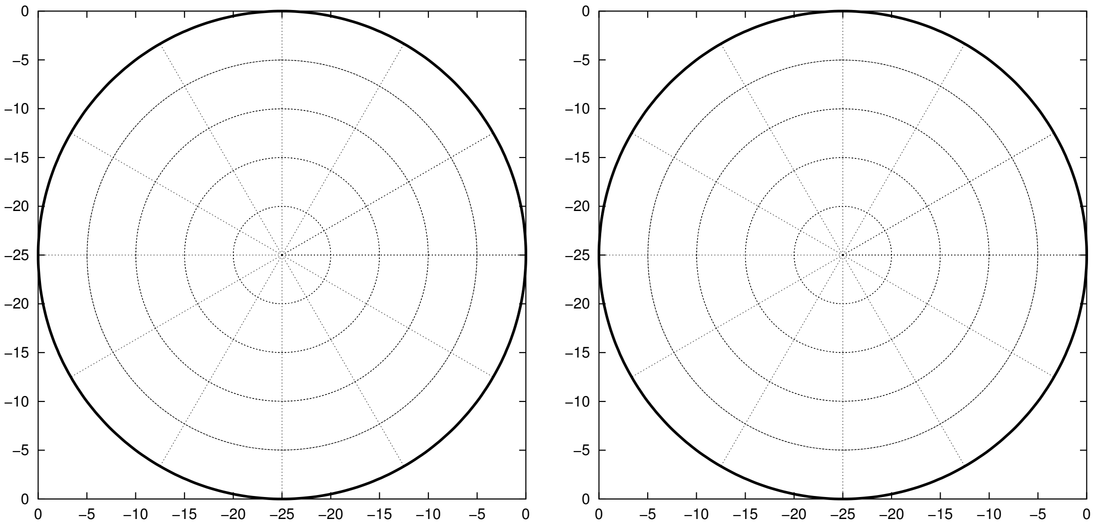
$$T = U_0\, W\, B^\top \big(B\, W\, B^\top + \mu\, I\big)^{-1}$$

where I is the identity matrix.

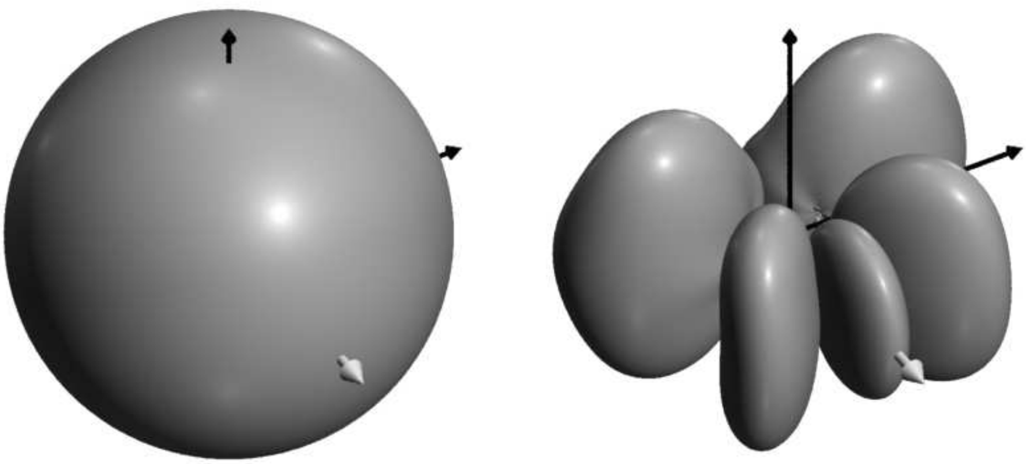
It is also possible to include other elementary errors into account in $e^2(T)$, for example the power of the directivities' first-order derivative in order to reduce secondary lobes.

As mentioned above, the elements of T give the matrix filters that are to be applied on the capsule signals in order to determine the loudspeaker feeds of the multichannel system.

It is widely admitted that ΔT allows to take sound information very well, but that ΔI is more adapted for restitution, transmission, and processing. The matrix T carries out an optimal $\Delta\mathrm{T}\,/\,\Delta\mathrm{I}$ conversion, since the capsule signals are pure ΔT (capsules are omnidirectional), whereas the output multichannel signals are pure ΔI. From the capsules point of

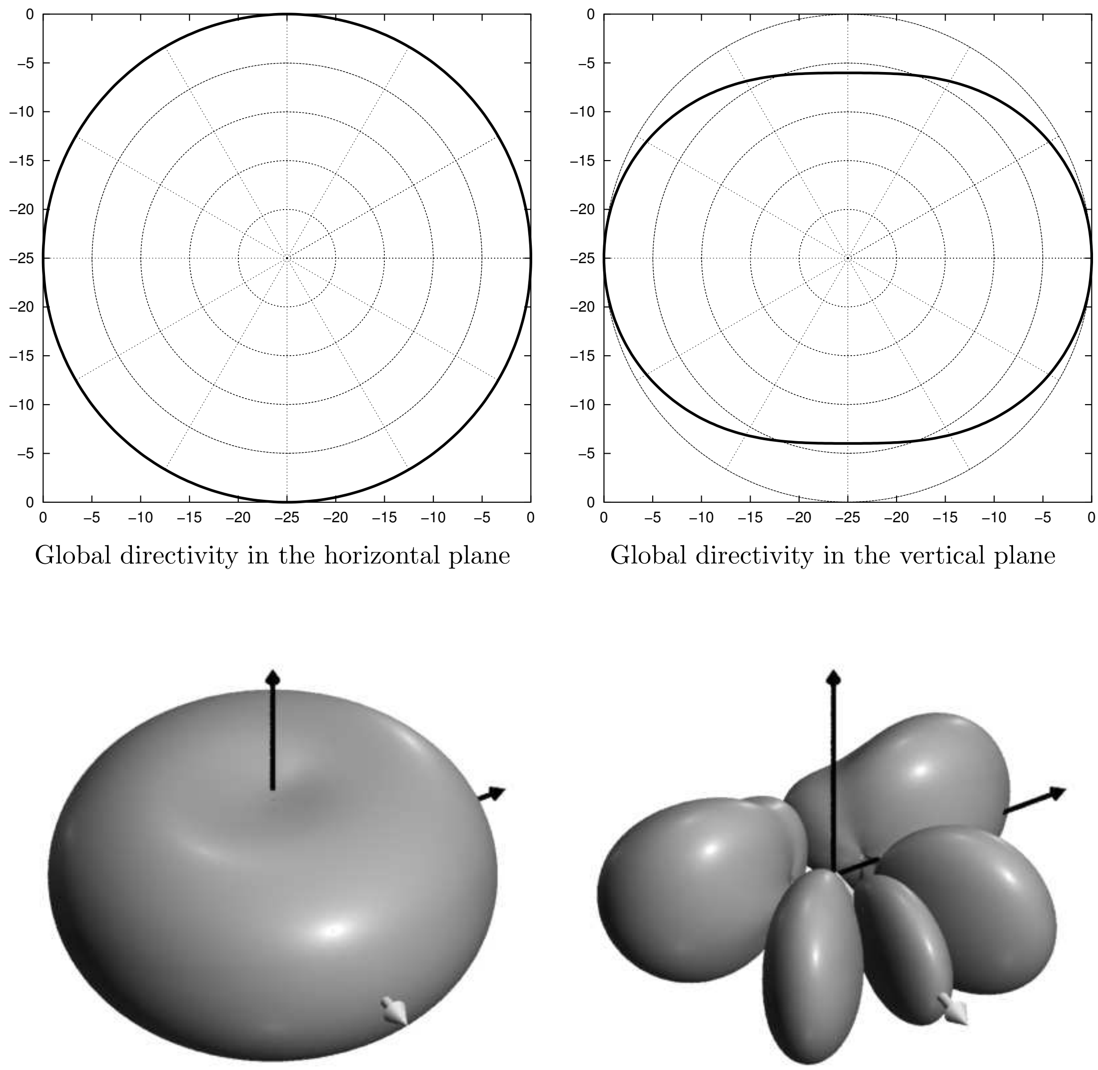


Global directivity in the horizontal plane Global directivity in the vertical plane



Global directivity in 3D Directivities in 3D

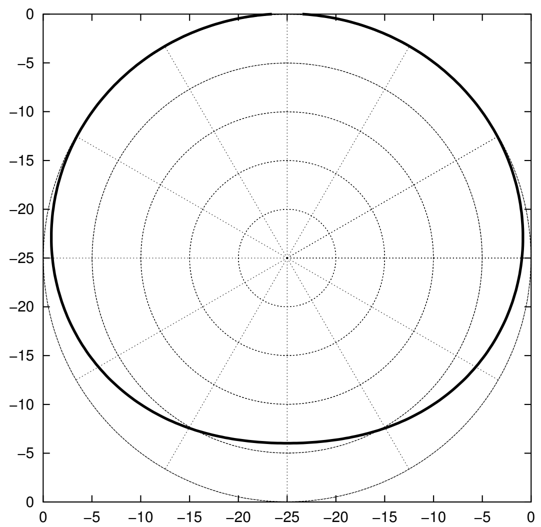
Fig. 10: Targeted 3D directivities: Full Omni ($I_d = 0$ dB, $f_d = 1$)



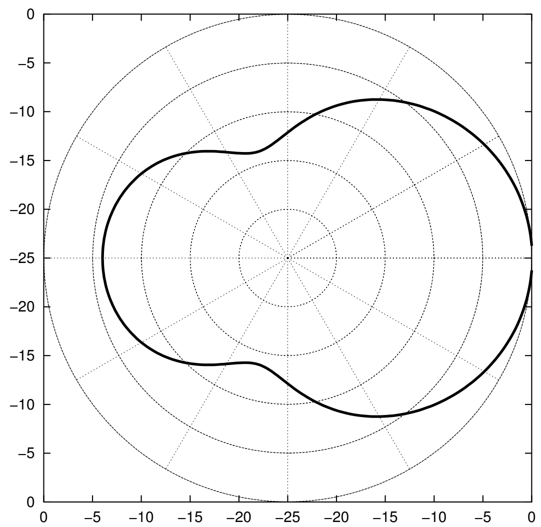
Global directivity in the horizontal plane
Global directivity in the vertical plane

Global directivity in 3D
Directivities in 3D

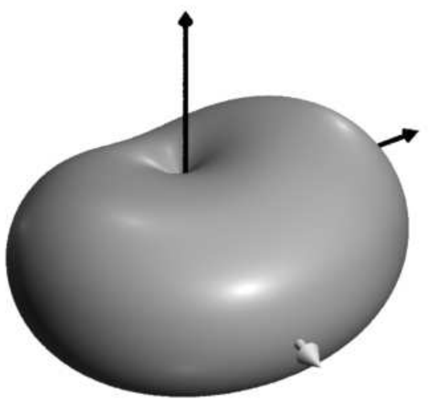
Fig. 11: Targeted 3D directivities: Oblate Omni ($I_d = 2$ dB, $f_d = 1.25$)



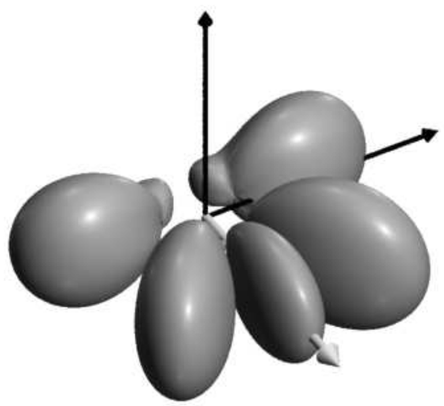
Global directivity in the horizontal plane



Global directivity in the vertical plane

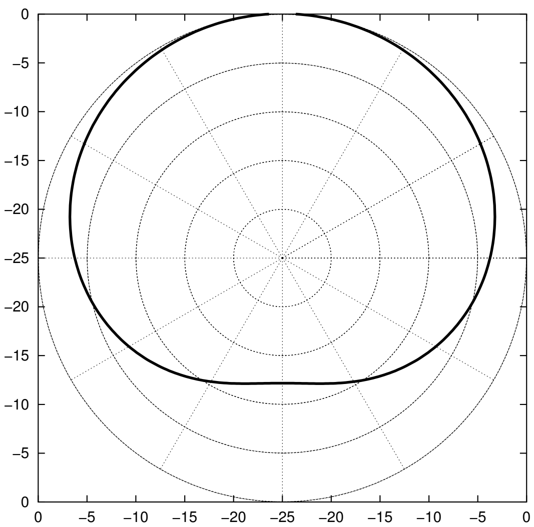


Global directivity in 3D

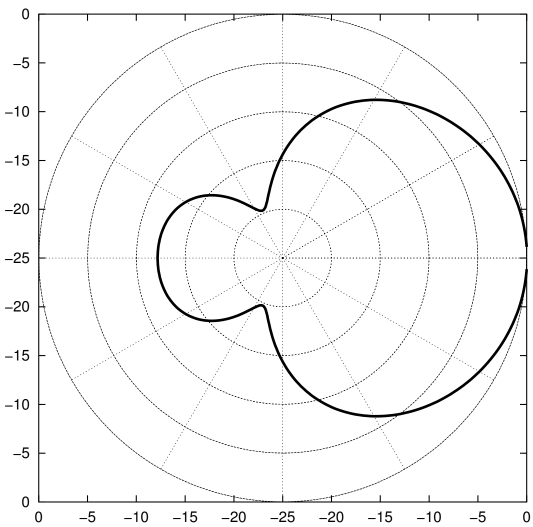


Directivities in 3D

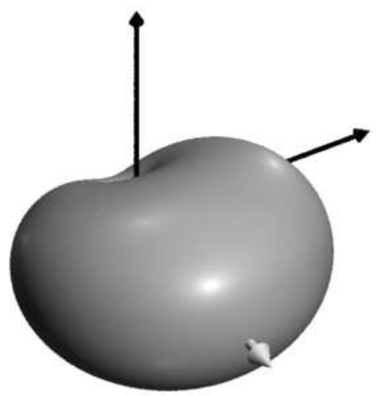
Fig. 12: Targeted 3D directivities: Frontal Emphasis 1 ($I_d = 3$ dB, $f_d = 1.4$)



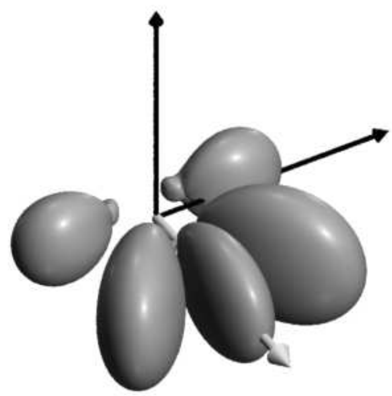
Global directivity in the horizontal plane



Global directivity in the vertical plane



Global directivity in 3D



Directivities in 3D

Fig. 13: Targeted 3D directivities: Frontal Emphasis 2 ($I_d = 5$ dB, $f_d = 1.8$)

view, the acoustic field recording is non-coincident, but from the multichannel signals point of view, the surround recording should be considered as if it were coincident, because there are no time differences between the multichannel signals.

3. HSR IN PRACTICE

High spatial resolution has direct consequences on sound recording, because of its advanced selectivity. This advanced selectivity is represented on figure 14, where HSR directivities and standard resolution directivities are represented. On this figure, 5 cardioid microphones are simply used for standard resolution, which is a little caricatured, but representative of the resulting resolution all the same. This figure brings the fact to the fore, that standard resolution sets 4 of the 5 multichannel signals within a 5 dB range.

3.1. Advantages of an HSR recording

Best channel separation

Channel separation corresponds to the correlation between channels. It should neither be too low nor too high. Insufficient separation implies that all channels have very similar signals, which gives poor spatial performances, in particular concerning the size of the sweet spot. On the contrary, excessive separation deteriorates phantom images, either by creating localization holes between loudspeakers, or by sticking phantom images on the loudspeakers. HSR allows to choose the targeted separation. Figure 14 shows the difference in channel separation between standard resolution and HSR.

Best source punctuality

Source punctuality and precision is an important point in a recording. Using standard resolution, more than two channels generally contribute to each source (cf. fig. 14). One of the resulting effects is that the phantom images are less precise and more diffuse. Advanced separation between channels make phantom images more precise.

Timbre preservation

Each filter of the matrix filtering is determined and optimized for each frequency. Furthermore, the optimization is led so that the sum of the multichannel directivities obtained will give a directivity independent of frequency, generally an omni. The consequence is that the spectrum of a reproduced phantom image does not depend on the direction of the corresponding source in the original sound environment.

Moreover, spatial resolution may vary with frequency without affecting the timbre of the phantom images. The sum of all directivities is indeed, as mentioned before, a fixed directivity. Even if each directivity varies with frequency, their sum does not vary with frequency. Only spatial resolution of each directivity may vary with frequency, not the overall amplitude. In other words, separation between channels may depend on frequency, but the perceived timbre will not be affected.

Phantom images over 360°

The design of the directivities from panning laws implies that any source around the multichannel microphone creates a phantom image on the multichannel reproduction system. If a source regularly moves around the microphone, the phantom image regularly moves around the multichannel listening system too, without being stuck on the loudspeakers nor disappearing, nor changing its amplitude or spectrum.

Good envelopment feeling

The good envelopment feeling of an HSR recording is not guaranteed *a priori*, because only sources are well recorded. However, the microphone is designed in such a way that any source will create a corresponding phantom image in the multichannel sound, and reverberation can be considered as a very large amount of secondary sources all around the microphone. As a consequence, even these secondary sources are panned by the microphone, which results in a good envelopment feeling.

Easy and high quality downmix

As the multichannel signal coming out of an HSR microphone is pure ΔI, directly summing signals does not create comb filters, and keeps sound to the best of the possibilities of the downmix format.

Full compatibility with usual panning laws

An HSR recording is spatialized using directivities based on panning laws. The direct consequence is that an HSR recording is fully compatible with mono sounds artificially panned. Spot microphones can thus be added to the signals delivered by the multichannel microphone without any problem. The only

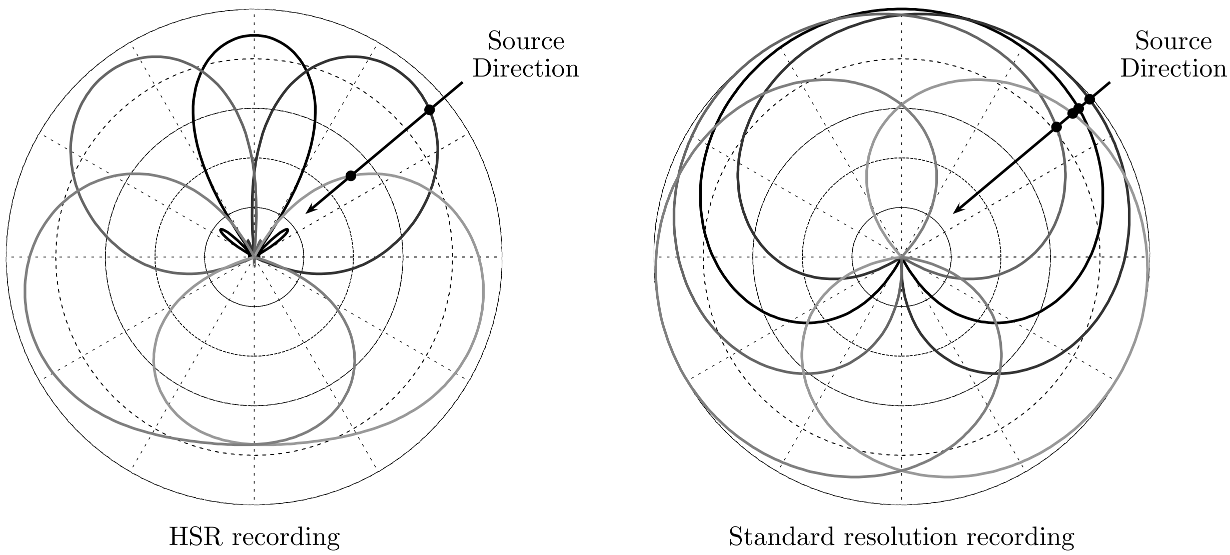


Fig. 14: Comparison between the recording of a source using HSR and using standard resolution.

points that need special care are that the delays between the spot microphones and the multichannel microphone should be compensated, and the panning direction of the spot microphone should be the same as the direction of the corresponding phantom image in the HSR recording.

Spatialization control outside the horizontal plane

The optimizations described in paragraph 2.2 allow to control the directivities outside the horizontal plane. For example, it is possible, using the same microphone array, to reduce the overall levels for sounds coming from the ceiling, the floor or the back. It is also possible to change the way a source is spatialized when it is above or below the horizontal plane. For example, a targeted spatialization for a source just above the microphone could be for the source to be sent on each loudspeaker with the same level.

Optimal distance factor control

As the directivities are controlled in three dimensions, it is possible to change the overall level of sources coming from directions outside the frontal scene. This corresponds to changing the distance factor of the microphone, without having to change the capsules, and even after the recording if the capsule signals have been recorded instead of the multichannel signals. Examples of different possible 3D directivities and distance factors are given on figures 10, 11, 12 and 13.

Recording angle control

As microphone directivities are specifically designed, it is possible to change the recording angle by changing the direction and size of the main lobes of the frontal directivities. Of course, this means that the frontal scene is distorted, but this is precisely the aim of changing the recording angle.

Best treatment robustness

The good separation between channels implies that treatments affecting a channel only modify phantom images close to the considered channel. The rest of the scene is not affected by this treatment. As a consequence, HSR is quite robust to any treatment, such as level change, equalization, or coding.

3.2. Performances obtained

As mentioned before, this multichannel microphone designing technique allows to cope with various constraints, such as the desired performances, the capsule characteristics (omnidirectional, bidirectional, cardioid...) and quality (cheap electret, studio-class), the number of capsules, the overall size of the array, and the targeted directivities. It is thus possible to design multichannel microphones leading to quite different characteristics using this very technique of full matrix filtering, with one or several microphone arrays.

A multichannel microphone consists of a microphone array and a processing platform. We designed two microphones arrays, both of them approximately 20 cm by 20 cm large and using omnidirectional capsules, 8 for the one and 16 for the other. A very important number of simulations was led in order to determine the best positions of the capsules for each multichannel microphone. These simulations consisted in choosing arbitrary positions for the array, determining the matrix T associated with this array using the method described in paragraph 2.2, and evaluating the resulting performances. These performances were evaluated for a very large number of configurations, and the best ones have been kept. A map of the retained 16-capsule array is given on figure 15.

The processing platform computes the full matrix filtering described in paragraph 2.1 and 2.2. It also allows to select the global directivity among one of figures 10, 11, 12 and 13.

Panpot directivities, such as the ones on figures 3, 4, or 5, on which an HSR microphone is based, are only patterns and it is not physically possible to get these directivities at each frequency using a reasonable number of capsules.

The performances of the 8-capsule microphone are not included here, as they were presented in a previous paper [24]. Three-dimensional measures had been made in an anechoic chamber to check that they match the theoretical results. They consisted of a 5°-regular analysis of the sphere covering all azimuths and all elevations from 0° to 135°. For each of these 1873 directions, the frequency response had been measured using log-sine sweep signals and deconvolution method.

The actual directivities of the 16-capsule microphone obtained with the omni filters are given on figure 16 (2D) and 17 (3D). These simulated directivities are calculated using the equation (3) for the determined matrices T: for each incoming direction of a plane wave, the loudspeaker feeds are determined, which gives the resulting directivities when the incoming direction varies. For low frequencies, 5th order is not reached because the corresponding wavelengths are too big in relation to the microphone size. For high frequencies, the performances remain quite high because the number of capsules and their proximity allow to push away the aliasing effect beyond the audible frequencies. In particular, the microphone has the best performances in the range from 700 Hz to 4 kHz, which Griesinger [19] considers as the dominant frequency range for localization. It should be recalled that the variations of each directivity with frequency do not affect the timbre of phantom images, because the sum of these directivities is always the global directivity (cf. § 2.2) throughout the whole frequency range.

A comparison between the performances of this 16-capsule microphone (figure 16) and the performances of the 8-capsule microphone previously mentioned (figure 18 in [24]) brings the following facts to the fore:

- separation is better with 16 microphones above 500 Hz,
- secondary lobes are reduced above 500 Hz,
- directivities are much closer to the targets at 5 kHz or more,
- there are few differences at low frequencies.

The first two points can be explained by the larger number of microphones, which allows to get better performances at a given frequency. The third point is due to the presence of capsules closer to each other, which increases the aliasing frequency. Last, the fourth point is explained by the fact that the performances at low frequencies are tightly linked to the size of the microphone, and both microphones have about the same size.

The figure 18 represents the three-dimensional simulated directivities of a unique 8-capsule microphone array using 8 different filters whose following parameters vary:

The coverage angle which determines the way the microphone narrows or expands the frontal scene. This angle is the direction for which the directivity of the left and right virtual microphones is maximum.

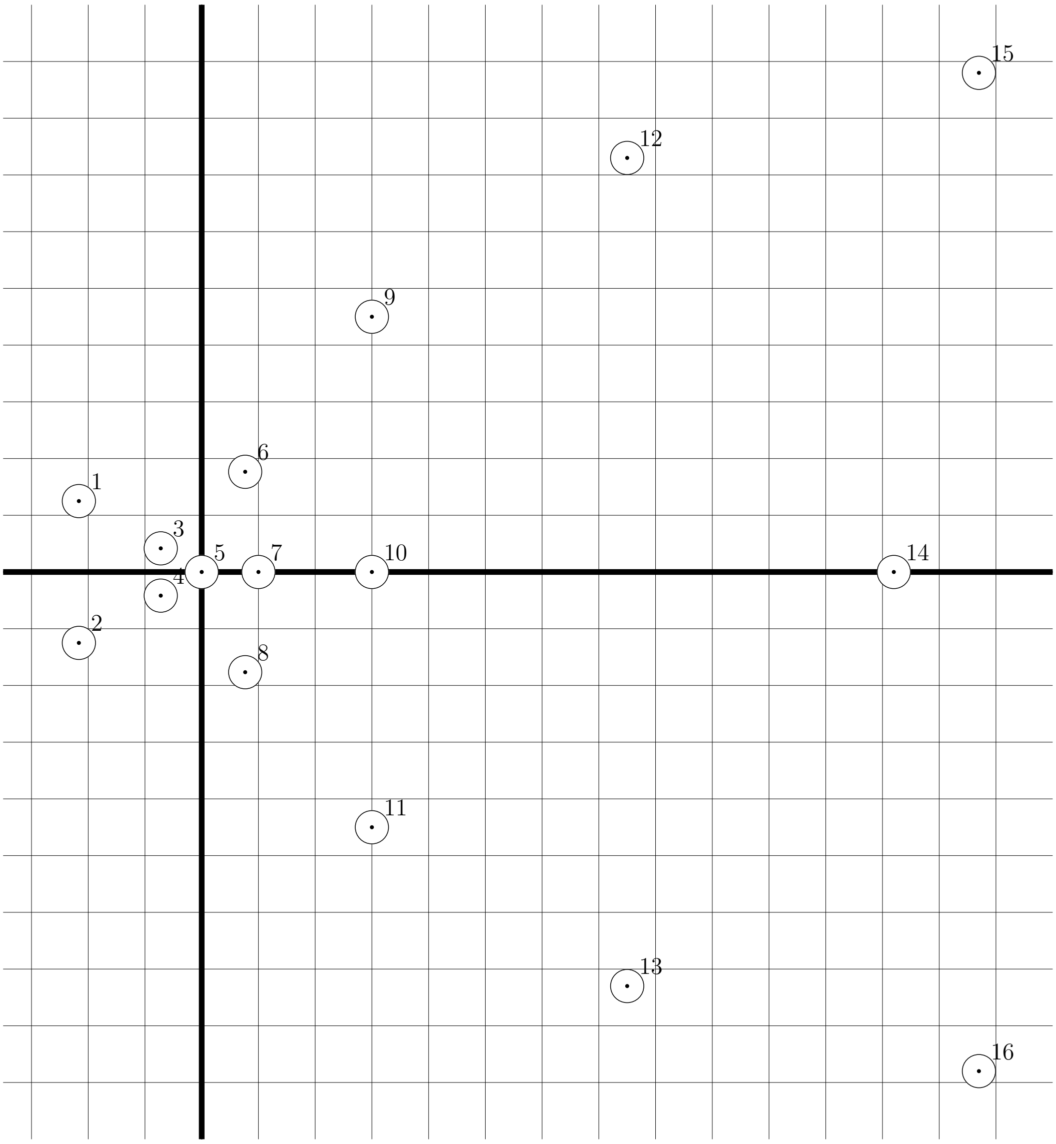


Fig. 15: Map of the chosen microphone array to the scale 3/4.

The global directivity which has already been explained.

This figure brings to the fore the fact that the full matrix filtering technique brings a great flexibility in the choice of the virtual microphone directivity.

4. CONCLUSION

High spatial resolution allows to consider the problematic of multichannel sound recording from a new point of view. As channel separation is increased and directivities can be controlled in three dimensions, it is possible to adjust quite precisely the way how sound is spatialized.

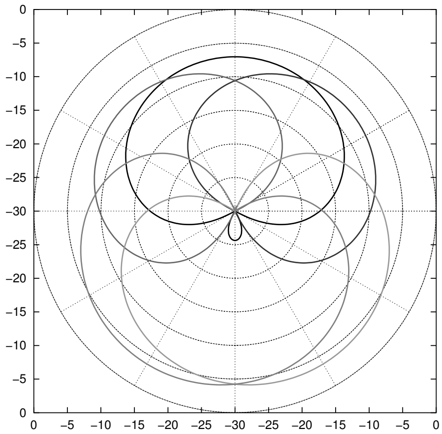
High spatial resolution is achievable using standard 8 omnidirectional capsules arranged in a small array, but it is possible to scale the hardware using the full matrix filtering technique, depending on the targeted characteristics. This microphone array is connected to a processor that computes the high spatial resolution multichannel signals in real time. Directivities up to 5th order are obtained this way.

This technique offers new perspectives as it is possible to get any targeted directivities using microphones of any type. This holds for example for 7.1 or 10.2, and a 7.0 or 10.0 microphone is easily feasible only changing the filter design, and possibly also the microphone array, which is relatively straightforward. Filter design can be optimized for a specific use, for example the recording angle and 3D directivities can be adapted for specific recording conditions.

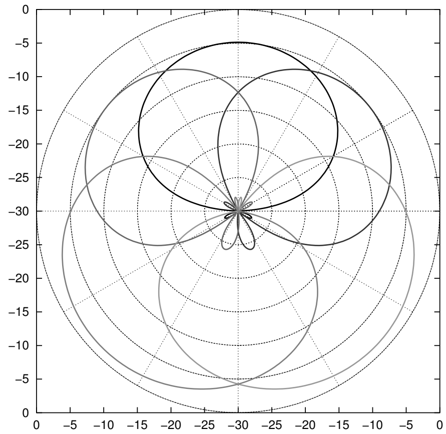
The comparison between performances of an 8-capsule microphone [24] and an equivalent 16-capsule microphone puts into evidence the flexibility of this technology and the ability it offers to choose a compromise between engaged means and obtained performances.

5. ACKNOWLEDGEMENTS

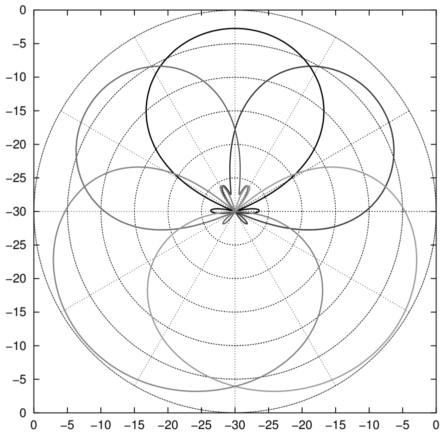
The authors wish to thank Laurent Givernaud, Jean-Michel Cauquy, Bernard Lagnel and Christian Prévot of Radio France for having made it possible to make recordings. They also wish to thank Olivier Warusfel and Guillaume Vandernoot for having given them the possibility to make measures in the IRCAM. They also wish to thank Jörg Wuttke of Schoeps for his interest in surround microphone technologies.



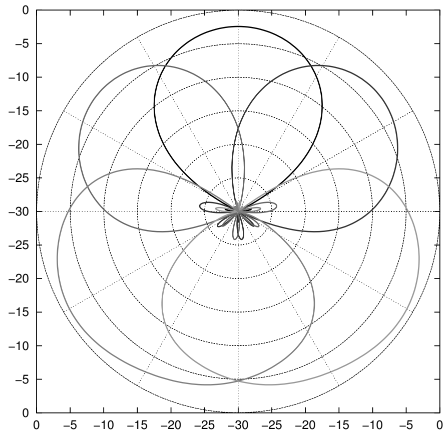
Directivities at 200 Hz



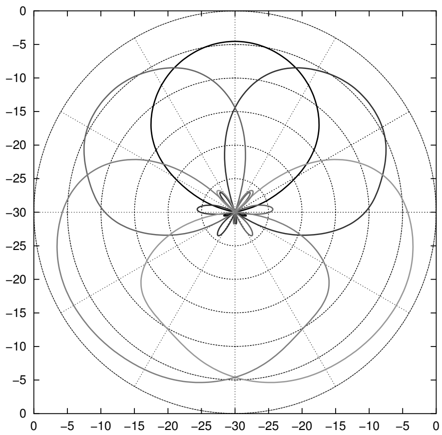
Directivities at 500 Hz



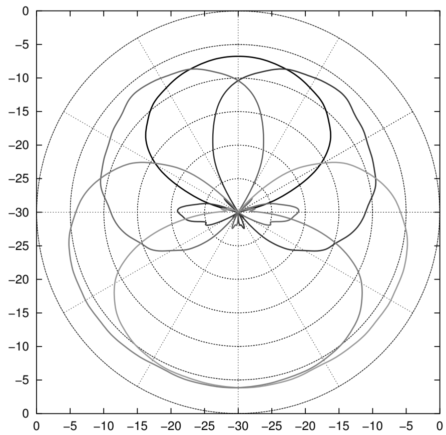
Directivities at 1 kHz



Directivities at 2 kHz

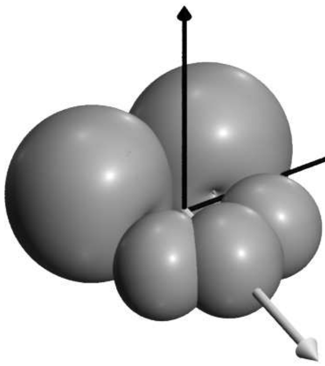


Directivities at 5 kHz

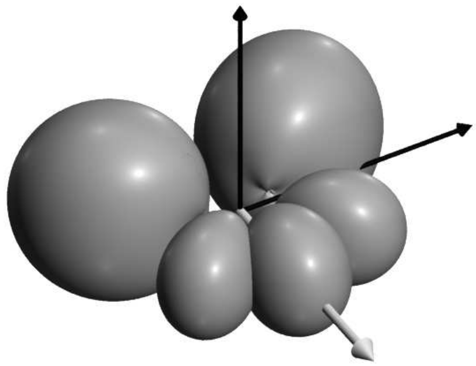


Directivities at 10 kHz

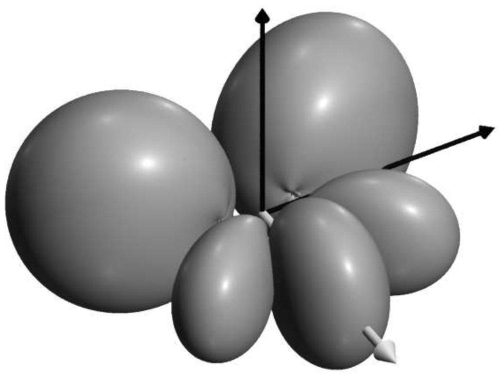
Fig. 16: Simulated directivities with 16 capsules.



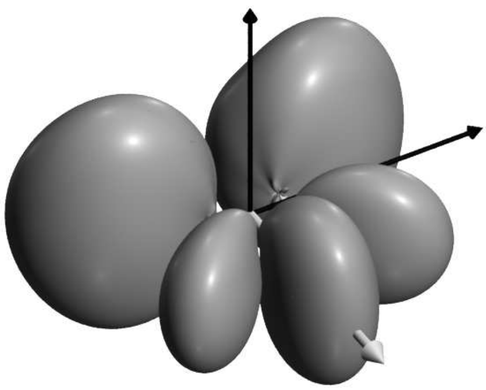
Directivities at 200 Hz



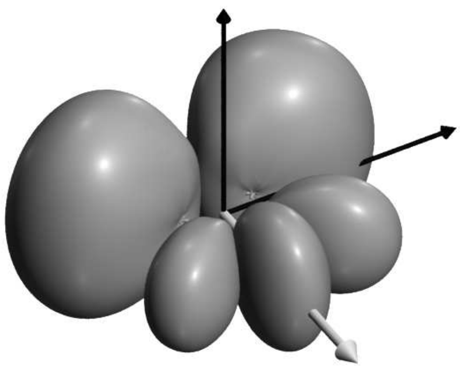
Directivities at 500 Hz



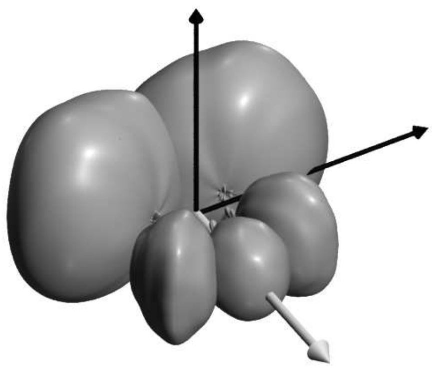
Directivities at 1 kHz



Directivities at 2 kHz

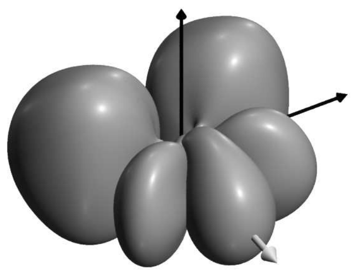


Directivities at 5 kHz

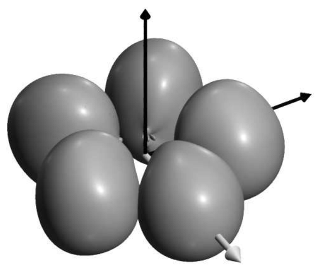


Directivities at 10 kHz

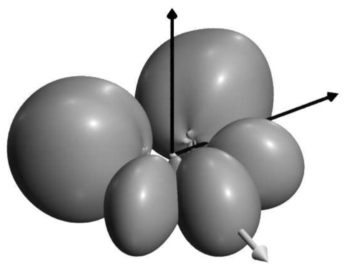
Fig. 17: Simulated 3D directivities with 16 capsules.



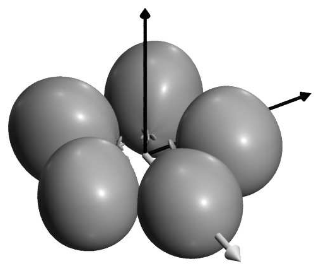
$CA = 40°$ — Omni



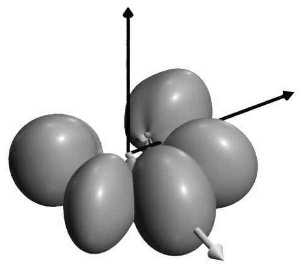
$CA = 72°$ — Omni



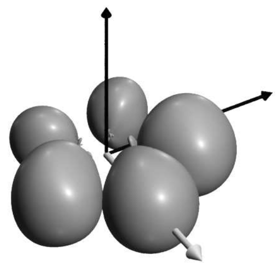
$CA = 40°$ — Oblate



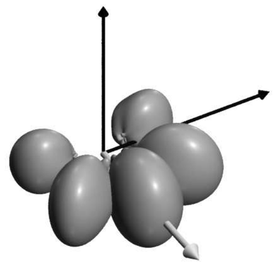
$CA = 72°$ — Oblate



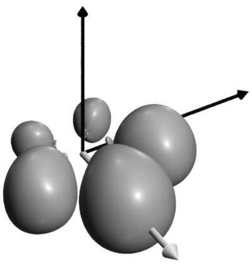
$CA = 40°$ — Frontal Emphasis 1



$CA = 72°$ — Frontal Emphasis 1



$CA = 40°$ — Frontal Emphasis 2



$CA = 72°$ — Frontal Emphasis 2

Fig. 18: Simulated 3D directivities with 8 capsules and various filters at 2 kHz. CA is the coverage angle.

APPENDICES

A. EQUATIONS OF THE 5TH-ORDER PANNING LAWS

Here are the equations that give the 5th-order directivities of the panning laws that are used throughout this article.

$$
\begin{aligned}
\mathit{feed}_C &= 0.08333 + 0.16286\cos\phi + 0.15187\cos 2\phi + 0.13485\cos 3\phi + 0.11360\cos 4\phi + 0.09024\cos 5\phi\\
\mathit{feed}_R &= 0.15278 + 0.19053\cos\phi - 0.20492\sin\phi - 0.00799\cos 2\phi - 0.21249\sin 2\phi\\
&\quad - 0.08529\cos 3\phi - 0.09837\sin 3\phi - 0.05832\cos 4\phi - 0.03253\sin 4\phi - 0.04284\cos 5\phi - 0.02502\sin 5\phi\\
\mathit{feed}_L &= 0.15278 + 0.19053\cos\phi + 0.20492\sin\phi - 0.00799\cos 2\phi + 0.21249\sin 2\phi\\
&\quad - 0.08529\cos 3\phi + 0.09837\sin 3\phi - 0.05832\cos 4\phi + 0.03253\sin 4\phi - 0.04284\cos 5\phi + 0.02502\sin 5\phi\\
\mathit{feed}_{SR} &= 0.30556 - 0.27196\cos\phi - 0.30007\sin\phi - 0.06794\cos 2\phi + 0.06337\sin 2\phi\\
&\quad + 0.01786\cos 3\phi + 0.08685\sin 3\phi + 0.00152\cos 4\phi - 0.02502\sin 4\phi - 0.00228\cos 5\phi + 0.00152\sin 5\phi\\
\mathit{feed}_{SL} &= 0.30556 - 0.27196\cos\phi + 0.30007\sin\phi - 0.06794\cos 2\phi - 0.06337\sin 2\phi\\
&\quad + 0.01786\cos 3\phi - 0.08685\sin 3\phi + 0.00152\cos 4\phi + 0.02502\sin 4\phi - 0.00228\cos 5\phi - 0.00152\sin 5\phi
\end{aligned}
$$

B. 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

$$
\begin{aligned}
j_l(x) &= \sqrt{\frac{\pi}{2x}}\, J_{l+\frac{1}{2}}(x)\\
y_l^m(\theta,\phi) &= \frac{1}{\sqrt{2\pi}}\, P_l^{|m|}(\cos\theta)\ \mathrm{trg}_m\,\phi
\end{aligned}
$$

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.

In this paper, the following notations are used:

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

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

- $\boldsymbol{c}$ denotes a vector containing the time Fourier transform of the capsule signals:

$$\boldsymbol{c} = \begin{pmatrix} C_1(f) & C_2(f) & \ldots & C_{N_{\mathrm{caps}}}(f) \end{pmatrix}^{\mathrm{t}}$$

- $\boldsymbol{v}$ denotes a vector containing the time Fourier transform of the multichannel signals (or loudspeaker feeds):

$$\boldsymbol{v} = \begin{pmatrix} V_1(f) & V_2(f) & \ldots & V_{N_{\mathrm{chan}}}(f) \end{pmatrix}^{\mathrm{t}}$$

- B is the sampling matrix associated with the considered microphone array:

$$B = \begin{pmatrix} B_{1,0,0}(f) & B_{1,1,-1}(f) & B_{1,1,0}(f) & B_{1,1,1}(f) & \ldots & B_{1,L,-L}(f) & \ldots & B_{1,L,L}(f) \\ B_{2,0,0}(f) & B_{2,1,-1}(f) & B_{2,1,0}(f) & B_{2,1,1}(f) & \ldots & B_{2,L,-L}(f) & \ldots & B_{2,L,L}(f) \\ \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ B_{N_{\mathrm{c}},0,0}(f) & B_{N_{\mathrm{c}},1,-1}(f) & B_{N_{\mathrm{c}},1,0}(f) & B_{N_{\mathrm{c}},1,1}(f) & \ldots & B_{N_{\mathrm{c}},L,-L}(f) & \ldots & B_{N_{\mathrm{c}},L,L}(f) \end{pmatrix}$$

  For omnidirectional capsules placed, in spherical coordinates, at (r_n, θ_n, ϕ_n), the elements of B are given by:

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

- T is the transcoding matrix containing the filter parameters of the full matrix filtering:

$$T = \begin{pmatrix} T_{1,1}(f) & T_{1,2}(f) & \ldots & T_{1,N_{\mathrm{caps}}}(f) \\ T_{2,1}(f) & T_{2,2}(f) & \ldots & T_{2,N_{\mathrm{caps}}}(f) \\ \vdots & \vdots & & \vdots \\ T_{N_{\mathrm{chan}},1}(f) & T_{N_{\mathrm{chan}},2}(f) & \ldots & T_{N_{\mathrm{chan}},N_{\mathrm{caps}}}(f) \end{pmatrix}$$

REFERENCES

[1] 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.

[2] J. Backman. Microphone array beamforming for multichannel recording. In *proceedings of the 114th AES Convention*, March 2003.

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

[4] B. Bauer. Phasor analysis of some stereophonic phenomena. *JASA*, 33(11), November 1961.

[5] B. Bernfeld. Simple equations for multichannel stereophonic sound localization. *J. Audio Eng. Soc.*, 23(7), September 1975.

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

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

[8] Peter Graham Craven. Continuous surround panning for 5-speaker reproduction. In *proceedings of the 24th AES International Conference*, 2003.

[9] 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.

[10] 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.

[11] Jérôme Daniel. Spatial sound encoding including near field effect: Introducing distance coding filters and a viable new ambisonic format. In *proceedings of the 23rd AES International Conference*, May 2003.

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

[13] 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.

[14] Gary W. Elko and Jens Meyer. Electroacoustic systems for 3-D audio? A report from the pittsburg meeting. *The newsletter of the JASA*, 12(3), 2002.

[15] 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.

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

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

[18] B. N. Gover, J. G. Ryan, and M. R. Stinson. Designing a spherical microphone array for the directional analysis of reflexions and reverberation. In *proceedings of the AES 115th Convention, preprint 5862*, October 2003.

[19] D. Griesinger. Stereo and surround panning in practice. In *proceedings of the 112th AES Convention*, May 2002.

[20] U. Herrmann and V. Henkels. Main microphone techniques for the 3/2 stereo-standard, March 1997. University of Dusseldorf.

[21] E. Hulsebos, T. Schuurmans, D. de Vries, and R. Boone. Circular microphone array for discrete multichannel audio recording. In *proceedings of the 114th AES Convention*, March 2003.

[22] Arnaud Laborie, Rémy Bruno, and Sébastien Montoya. Immersivesound website, March 2003. URL: `http://www.immersivesound.com`.

[23] 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.

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

[25] Y. Makita. Localisation directionnelle du son dans un champ sonore stéréophonique, June 1962.

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

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

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

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

[30] 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.

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

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

[33] Günter Theile. Multichannel natural music recording based on psychoacoustic principles, May 2001. IRT.

[34] M. Williams. Multichannel microphone array design: Segment coverage analysis above and below the horizontal reference plane. In *proceedings of the 112th AES Convention, preprint 5567*, May 2002.

[35] M. Williams and G. Le Dû. Multichannel microphone array design. In *proceedings of the 108th AES Convention, preprint 5157*, February 2000.

[36] M. Williams and G. Le Dû. The quick reference guide to multichannel microphone arrays – Part I: Using cardioid microphones. In *proceedings of the 110th AES Convention, preprint 5336*, May 2001.

[37] AES Staff Writer. Novel surround sound microphone and panning techniques. *J. Audio Eng. Soc.*, 52:74–80, January/February 2004.

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