Introduction to Ambisonics

Technical Report · June 2015

1 author:

Daniel Arteaga
Dolby Laboratories, Inc.

23 PUBLICATIONS   **79** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   IDHOA: decoding Ambisonics to irregular layouts View project

Lecture notes – v0.3

Introduction to Ambisonics

Daniel Arteaga

June 2015

Audio 3D – Grau en Enginyeria de Sistemes Audiovisuals
Universitat Pompeu Fabra

Preface

This is a *preliminary version* of some lecture notes on Ambisonics. Many things are missing or can bee improved, like a better list of references, historical remarks, a better exposition of concepts, a grammar and typoc check. Additionally, it will probably contain errors. Use it under your own risk!

License

Text and equations (not the figures) are under a Creative Commons Attribution-ShareAlike 4.0 International License.[1] Figures have been reproduced from various sources under several licenses; please check the caption for information.

Indications on the text

> Important point, or summary of the section.

> ( ) Advanced or lateral note going beyond the scope of the course (some of the footnotes also contain advanced or lateral material).

Contact

For any question or remark, please contact me at daniel (dot) arteaga (at) upf (dot) edu. Any indication of typos, errors, and suggestions for improvement are welcome.

Daniel Arteaga

---

[1] http://creativecommons.org/licenses/by-sa/4.0/

Contents

1 Introduction

1.1 What is Ambisonics?

Ambisonics is a method of codifying a sound field taking into account its directional properties. In traditional multichannel audio (e.g., stereo, 5.1 and 7.1 surround) each channel has the signal corresponding to a given loudspeaker. Instead, in Ambisonics each channel has information about certain physical properties of the acoustic field, such as the pressure or the acoustic velocity.

Ambisonics is a perturbative theory.

0. At *zeroth order*, Ambisonics has information about the *pressure field* at the origin (recording of an omnidirectional microphone at the origin). The channel for the pressure field is conventionally called W.

1. At *first order*, Ambisonics adds information about the *acoustic velocity* at the origin (recording of three figure-of-eight microphones at the origin, along each one of the axis). These channels are called X, Y, Z. Following the Euler equation, the velocity vector is proportional (up to some equalization) to the gradient of the pressure field along each one of the axis.

2. At second and higher orders, Ambisonics adds information about higher order derivatives of the pressure field.

In this course we will limit ourselves mostly to first order Ambisonics, which we shall call simply Ambisonics from here, although we will also do a brief introduction to higher order Ambisonics (HOA).

1.2 Historical remarks

Ambisonics was developed by the pioneering British engineer Michael Gerzon in the 70s [4]. Although hardware Ambisonic systems were soon developed, they were never a commercial success. However, Ambisonics has many nice features and has attracted the interests of researchers in spatial audio since the early beginning. In the 90s, the theory for higher Order Ambisonics was developed [1]. Currently, although there are several examples of commercial application of Ambisonics here and there, still Ambisonics as a complete theory has not found widespread commercial success. It is still nowadays a topic of research.

1.3 Features of Ambisonics

Differently to most approaches, like stereo, 5.1, Ambisonics is:

- Based on *physical principles* of the acoustic field.

- It is *not restricted to single plane waves*, but it can account for any general sound field.

- Completely *layout-independent*, in the sense that the Ambisonics production and postproduction workflow does not depend on the exhibition layout. Stereo, 5.1, etc. all assume that there are loudspeakers in fixed positions, and all the recording, production and exhibition methods are based on this assumption. In contrast, Ambisonics does not depend on a fixed distribution of loudspeakers.

- Even if it is layout-independent, Ambisonics it is *not necessarily object-based*. First order Ambisonics has 4 fixed channels. This is in contrast to other techniques such as VBAP or Wave Field Synthesis, which depend on an object-based method: every audio object is characterized by a mono track and a set of metadata indicating the location and other properties. While Ambisonics can be adapted to an object-based representation, this is not necessary. The fact that Ambisonics has a fixed number of channels it is important to save memory and processing requirements with hundreds or thousands of audio objects, as it is the case in complicated movie productions.

- As we will see now, it is a *complete theory*, covering encoding, recording, postproduction, transmission and reproduction. The possibilities of Ambisonics are showcased in fig. 1.1.
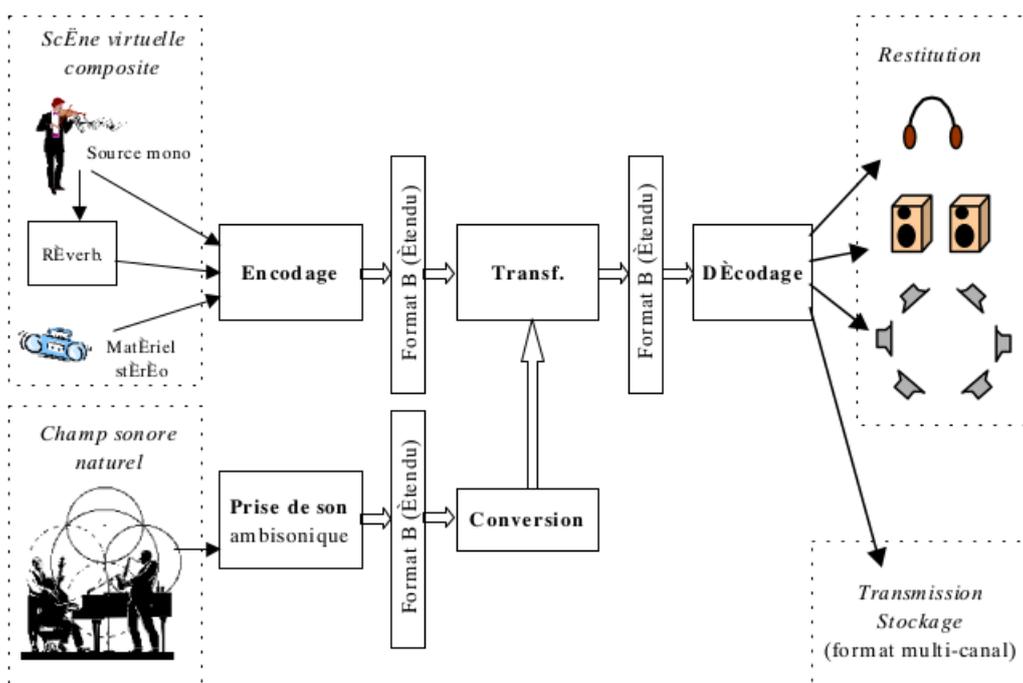

Figure 1.1: Global scheme of an Ambisonics system. Reproduced from [1].

2 Ambisonics encoding and recording

As a first step in the Ambisonics chain, a sound field needs to be encoded (ie, created synthetically from a mono recording) or recorded (using a special microphone).

2.1 Codification of a plane wave in first order Ambisonics

Any plane wave can be characterized by a sound signal $s(t)$ and a direction of arrival of the sound[1] $\hat{k}$. The unit vector $\hat{k}$ indicate the direction of arrival of the plane wave (the origin of the sound source). This vector can be decomposed in spherical coordinates as follows: $\hat{k} = (\cos\theta\cos\delta, \sin\theta\cos\delta, \sin\delta)$, with θ being the azimuth and δ the elevation of the sound source.

The sound pressure can be constructed in terms of $s(t)$ and $\hat{k}$:

$$p(t,\vec{r}) = s(t + \hat{k}\cdot\vec{r}/c), \tag{2.1}$$

For example, if we have a plane wave of angular frequency ω, then $s(t) = A\sin(\omega t)$ and the above expression transforms into:

$$p(t,\vec{r}) = A\sin(\omega t + \omega\,\hat{k}\cdot\vec{r}/c) = A\sin(\omega t + \vec{k}\cdot\vec{r}), \tag{2.2}$$

where we have used that the wave vector $\vec{k}$ can be expressed as $\vec{k} = (\omega/c)\,\hat{k}$.

A plane wave also has an acoustic velocity vector $\vec{v}$,

$$\vec{v}(t,\vec{r}) = (1/Z)\,s(t + \vec{k}\cdot\vec{r}/c)\,\hat{k}, \tag{2.3}$$

where $Z = \rho_0 c$ is the acoustic impedance, with ρ_0 being the density of air.. In components:

$$v_x(t,\vec{r}) = (1/Z)\,s(t + \vec{k}\cdot\vec{r}/c)\cos\theta\cos\delta, \tag{2.4a}$$

$$v_y(t,\vec{r}) = (1/Z)\,s(t + \vec{k}\cdot\vec{r}/c)\sin\theta\cos\delta, \tag{2.4b}$$

$$v_z(t,\vec{r}) = (1/Z)\,s(t + \vec{k}\cdot\vec{r}/c)\sin\delta, \tag{2.4c}$$

Remind that the acoustic velocity is proportional to the gradient of the pressure.

Both the pressure and the 3 components of the acoustic velocity are measurable quantities. Actually we do have microphones which directly measure the pressure and the acoustic velocity:

- The pressure p is be proportional to the recording of an *omnidirectional* microphone.

---

[1]The direction of arrival is the opposite to the direction of propagation.

Figure 2.1: 3D polar patterns of an omnidirectional microphone (top) and three figure-of-eight directional microphones in each one of the directions (bottom). Adapted from figure in [2].

- The acoustic velocity components v_x, v_y and v_z are proportional to the recording of three *figure-of-eight microphones* placed along the components of the x, y and z axis respectively.

The idea of Ambisonics is to encode in four different channels the pressure and velocity at the origin, in the following way:

$$W(t) = s(t)/\sqrt{2} \tag{2.5a}$$
$$X(t) = s(t)\cos\theta\cos\delta, \tag{2.5b}$$
$$Y(t) = s(t)\sin\theta\cos\delta, \tag{2.5c}$$
$$Z(t) = s(t)\sin\delta, \tag{2.5d}$$

where $s(t)$ can be any mono sound signal.

The channel $W(t)$ is proportional to the pressure field at the origin, and the channels $X(t)$, $Y(t)$ and $Z(t)$ are proportional to the acoustic velocity in each one of the axis. The set of four audio channels (W, X, Y, Z) is called *B-format* and constitutes the basis for first order Ambisonics.

The factor of $\sqrt{2}$ next to $W(t)$ is there for historical reasons (Furse-Malham normalization). Ambisonics is also valid in 2D. In this case, $\delta = 0$ and $Z(t) = 0$.

2.2 Codification of a general sound field in Ambisonics

For a general set of multiple localized sources (multiple plane waves) with signals s_i coming from direction (θ_i, δ_i), the Ambisonics channels can be computed as:

$$W(t) = \sum_i s_i(t)/\sqrt{2} \tag{2.6a}$$

$$X(t) = \sum_i s_i(t) \cos \theta_i \cos \delta_i, \tag{2.6b}$$

$$Y(t) = \sum_i s_i(t) \sin \theta_i \cos \delta_i, \tag{2.6c}$$

$$Z(t) = \sum_i s_i(t) \sin \delta_i. \tag{2.6d}$$

Again, $W(t)$ amounts to the recording of a virtual omnidirectional microphone, and X, Y, Z to the recording of three figure-of-eight microphones in the x, y, z axis.

In Ambisonics it is also possible to codify other kind of sources different to plane waves. However, the codification is more difficult and will not be analyzed here. In any case, working with plane waves is normally enough for the following two reasons

1. Any source (e.g., an extended source) can be codified in terms of multiple plane waves.

2. A spherical wave, which is another important type of source, at a distance from the origin of several wavelengths can be approximated with good precision in terms of plane waves.

( ) The codification of a source distribution Ambisonics is analogous to the decomposition of any function in the circle in terms of Fourier series. Actually, any distribution of sound sources in the solid angle $S(\theta, \delta)$ can be approximated in Ambisonics as:

$$S(\theta, \delta) = \sqrt{2}\, W(t) + \frac{1}{3}[X(t) \cos \theta \cos \delta + Y(t) \sin \theta \cos \delta + Z(t) \sin \delta] + \ldots \tag{2.7}$$

Furthermore, Ambisonics can also be viewed in terms of a perturbative decomposition of the sound field around the origin. By looking for a solution of the wave equation in spherical coordinates one can show that the sound field around the origin is:

$$p(t, \vec{r}) = \frac{\sin r}{r}\, \sqrt{2}\, W(t) + \frac{i}{3}\left[\frac{\sin r}{r^2} - \frac{\cos r}{r}\right][X(t) \cos \theta \cos \delta + Y(t) \sin \theta \cos \delta + Z(t) \sin \delta] + \ldots$$
$$\tag{2.8}$$

when $r = 0$, $p(t, \vec{r}) = \sqrt{2}\, W(t)$ as expected.

2.3 Ambisonics recording: the soundfield microphone

In the previous sections we saw how we can artificially generate an Ambisonics signal from a mono recording, by placing that recording in any angular direction through the use of eq. (2.5).

Now we will see that it is also possible to record directly on Ambisonics by using the so-called Soundfield microphone.

We have seen that Ambisonics amounts to:

- An omnidirectional microphone for $W(t)$.

- Three bidirectional (figure-of-eight) microphones for $X(t)$, $Y(t)$ and $Z(t)$.

All three microphones should be ideally coincident in one point of space.

In practice, it is difficult or impossible to fit all three microphones in a sigle point. Instead, the *soundfield microphone* places four cardioid or subcardioid capsules in the vertices of a tetrahedron, as seen in fig. 2.2.

<div align="center">

| Soundfield ST-250 | Soundfield STS-200 | DPA-4 | Core Sound TetraMic |

</div>

Figure 2.2: Several soundfield microphones, without the cover protection. Reproduced from [5].

Since for any microphone it holds that:

$$\text{cardiod mic} = \frac{1}{2}(\text{omni mic} + \text{figure-of-8 mic}),$$

by solving a linear system of 4 unknowns with 4 equations it is always possible to retrieve the Ambisonics B format signals from the raw microphone recordings.

For each soundfield microphone we distinguish:

- The *A format*, the recording of each one of the four microphone capsules.

- The *B format*, the components of the Ambisonics channels (W, X, Y, Z)

A linear system of equations can be used to convert the A format recordings into the B format Ambisonics signals.[2] Additionally, it is necessary to add some filtering to correct for the fact that not all four capsules are coincident, which is relevant at high frequencies.

For some soundfields, the mic preamplifier already does the A to B format conversion in the hardware, so that the preamplifier already has 4 outputs labelled as W, X, Y, Z. For some other soundfields, the mic preamplifier delivers the output of the capsules in A format and the conversion to B format has to be done externally (typically on a computer).

---

[2]For example, for the TetraMic microphone the solution to such system is:

$$W = \text{FLU} + \text{FRD} + \text{BLD} + \text{BRU},$$
$$X = \text{FLU} + \text{FRD} - \text{BLD} - \text{BRU},$$
$$Y = \text{FLU} - \text{FRD} + \text{BLD} - \text{BRU},$$
$$Z = \text{FLU} - \text{FRD} - \text{BLD} + \text{BRU},$$

where FLU, FRD, BLD and BRU are the A-format signals.
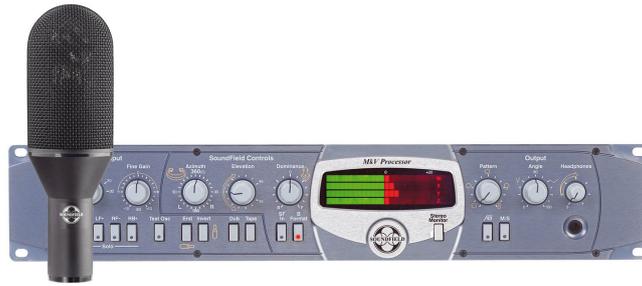
Figure 2.3: SoundField microphone from *TSL* [6], which does the A-format to B-format conversion in the preamplifier electronics.

Figure 2.4: TetraMic from *Core Sound* [7], which delivers signals in A-format, and they have to be converted to B-format externally.

Finally, it is to be noted that by making suitable combinations of the B-format components, the soundfield microphones can be used not only for Ambisonics, but also as directional mono mics (with adjustable directionality), as adjustable pairs of XY stereo microphones, or as 5.1 microphones. Soundfield microphones have been used in professional audio productions.

3 Ambisonics transmission and manipulation

Once Ambisonics has been encoded, or recorded, it needs to be transmitted and manipulated (for instance in postproduction). Let us see how this process can work.

3.1 Ambisonics format for transmission

For transmitting Ambisonics, one only needs to send the 4-channels of the B-format, no matter if the is only one single source, one recorded ambient, reverb, or whatever combination of sound objects.

This is why Ambisonics is a channel-based format (although these channels are based on physical considerations, and so they are very different from the e.g. 6 channels of 5.1 surround). This is to be contrasted to other approaches, such as VBAP or wave field synthesis, which are typically object-based. In object-based approaches one transmits:

1. A single mono audio track

2. Metadata for the position in space (and other characteristics) of the audio track

In Ambisonics, an object-based approach can be used, but it is not mandatory.

( ) However the are other historical formats for Ambisonics transmission. The UHJ format [8], also called C-format, is another way to codify Ambisonics so that it is compatible with normal stereo decoders. It has four channels LRTQ. Only the channel LR are compulsory, T and Q are optional.

   – With a normal mono decoder, the two channels LR are summed and decodified as mono.

   – With a normal stereo decoder, the two channels LR are decodified as stereo.

   – With a special UHJ decoder, the two channels LR can be decoded as an matrixed Ambisonics system on the plane (with spatial information being "matrixed" in the regular stereo)

   – With a special UHJ decoder, the three channels LRT can be decoded as full Ambisonics system on the plane (equivalent to W,X,Y).

   – With a special UHJ decoder, the four channels LRT can be decoded as full Ambisonics system in 3D (equivalent to W,X,Y,Z).

3.2 Ambisonics manipulations

3.2.1 Rotations

The Ambisonics B-format can be easily rotated. Under an arbitrary rotation, characterized by a rotation matrix $\mathbf{R}$, the B-format is transformed as follows:

$$W' = W \tag{3.1a}$$

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \mathbf{R} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{3.1b}$$

In general, any rotation matrix can be decomposed in roll, pitch and yaw For instance, for a rotation around the z axis of an angle θ ('yaw'), the rotation can be expressed as:

$$W' = W, \tag{3.2a}$$

$$X' = X \cos\theta - Y \sin\theta, \tag{3.2b}$$

$$Y' = X \sin\theta + Y \cos\theta, \tag{3.2c}$$

$$Z' = Z. \tag{3.2d}$$

In summary: under a rotation, the channel W behaves as a scalar, and the channels X, Y, Z behave as vectors.

3.2.2 Other transformations

Other possible transformations of the sound field include:

- Linear transformations that do not depend on the spacial properties (for example, filtering and equalization): they can be done as usual, as long as it is applied equally to all channels.

- Other spatial deformations, such as warping. For instance, one can create the effect of focusing on one direction of the sound field (dominance effect). This will not be studied here.

- Non-linear transformations, such as dymamic range control, are more diffcult to handle and are currently a topic of research.

4 Ambisonics decoding

The objective of the reproduction of Ambisonics is to be able to decode Ambisonics in a set of several loudspeakers distributed around the listener. Given a set of N speakers, the decoding feeds each speaker i, located at a direction $\hat{u}_i = (\theta_i, \delta_i)$, with a signal s_i given by

$$s_i(t) = w_i W(t) + x_i X(t) + y_i Y(t) + z_i Z(t), \tag{4.1}$$

where the set of $4N$ parameters (w_i, x_i, y_i, z_i) define the decoding. To these parameters two basic strategies can be adopted: basic reconstruction of the sound field, or physcoachoustic reconstruction.

4.1 Decoding strategies

The following will be based on Gerzon's localization principles [3].

4.1.1 Basic decoding

In the *basic decoding*, to coefficients are determined under the *assumption of coherence* among signals arriving from the loudspeakers, and the requirement is to accurately reproduce the original pressure and acoustic velocity vector at the origin.

Assuming that all the loudspeakers are at the same distance from the origin, and assuming that all loudspeakers add coherently, the pressure at the origin p is (discarding the geometric decay factor, which is assumed to be constant),

$$p(t) = \sum_{i=1}^{N} s_i(t), \tag{4.2a}$$

and the normalized[1] acoustic velocity $\vec{v}$ is:

$$\vec{v}(t) = \sum_{i=1}^{N} s_i(t) \hat{u}_i, \tag{4.2b}$$

It is useful to define the vector r_v, the *directionality coefficient of the acoustic velocity*, as:

$$r_v = \frac{||\vec{v}||}{|p|} \tag{4.3}$$

---

[1]In general, there would be a factor $1/Z$, where $Z = \rho c$ is the impedance of the air, in front of the acoustic velocity. By "normalized", we mean that we drop this constant factor, as it is usually done in most Ambisonics literature.

For a plane wave (a fully localized source) $r_v = 1$, whereas for a fully de-localized source (e.g. diffuse field) $r_v = 0$. In general, $r_v \in [0, \infty)$ ($r_v = \infty$ e.g. for a standing wave in a pressure node).

Physcoacoustically, this decoding reproduces the impression of the original sound at low frequencies, below 500 Hz approximately, and at distances from the centre not larger than a fraction of the shortest wavelength considered.

4.1.2 Psychoacoustic decodings

In the *psychoacoustic decodings*, an *incoherent sum of the speaker signals* is instead assumed, and it is required that the decoding reproduces the original energy and acoustic intensity at the origin.

Within the *incoherent sum hypothesis*, and assuming that each one of the incoming waves is a plane wave, the normalized[2] signal energy at the origin is:

$$w(t) = \sum_{i=1}^{N} |s_i(t)|^2 \tag{4.4a}$$

and the so-called energy vector[3] $\vec{E}$ is

$$\vec{E}(t) = \sum_{i=1}^{N} |s_i(t)|^2 \tag{4.4b}$$

It is useful to define the quantity r_v, the *directionality coefficient of the energy*, as:

$$r_E = \frac{||\vec{E}||}{w} \tag{4.5}$$

For a plane wave (a fully localized source) $r_E = 1$, whereas for a fully de-localized source (e.g. diffuse field) $r_E = 0$.

For reproduced plane waves, it is physically impossible to fulfil the condition $r_E = 1$ by summing incoherently the signal of several loudspeakers. Instead, the *max-r_E decoding* will try to maximise this value (hence the name). Psychoacoustically, this decoding reproduces the impression of the original sound at high frequencies, above 500 Hz approximately.

The *in-phase* decoding condition is similar to the max-r_E decoding, but with the additional restriction that the different loudspeakers do not emit in the opposite phase simultaneously. This decoding gives a more robust localisation for listeners who are far from the sweet spot.

---

[2]In general, there would be a factor $1/(\rho_0 c^2)$ in front of the expression below. By "normalized", we mean that we drop this constant factor, as it is usually done in most Ambisonics literature.

[3]The energy vector would be a statistical estimator of the acoustic intensity, so that $\vec{I}(t) = \vec{E}(t)/(\rho_0 c)$.

4.2 Physical decoding

For a physical decoding, one imposes that the pressure at the origin should be proportional to W and the velocity to (X, Y, Z), see eq. (4.2). By requering that at the origin the value of the pressure and the velocity is the one that corresponds to $W(t)$ and $(X(t), Y(t), Z(t))$ respectively, one gets:

$$p(t) = \sum_{i=1}^{N} s_i(t) = W(t) / \sqrt{2}, \qquad (4.6)$$

$$\vec{v}(t) = \sum_{i=1}^{N} s_i(t) \hat{u}_i = (X(t), Y(t), Z(t)) \qquad (4.7)$$

This is an system of linear equations with 4 equations and N unknowns (each one of the loudspeakers). Provided the loudspeakers are in reasonable configurations, this equation has a unique solution with $N = 4$. With $N > 4$ the system becomes overdetermined, and there are many different solutions (the solution with less global energy is picked).

( ) It turns out that the above equation can be rewritten in matrix form as:

$$\begin{pmatrix} W(t)/\sqrt{2} \\ X(t) \\ Y(t) \\ Z(t) \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \cos\theta_1\cos\delta_1 & \cos\theta_1\cos\delta_2 & \cdots & \cos\theta_1\cos\delta_N \\ \sin\theta_1\cos\delta_1 & \sin\theta_1\cos\delta_2 & \cdots & \sin\theta_1\cos\delta_N \\ \sin\delta_1 & \sin\delta_2 & \cdots & \sin\delta_N \end{pmatrix} \begin{pmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_N(t) \end{pmatrix} \qquad (4.8)$$

This equation for $N \geq 4$ can be solved by the pseudoinverse method, where a pseudoinverse of a given matrix is given by:

$$D_{\text{pinv}} = C^T (CC^T)^{-1} \qquad (4.9)$$

For certain situations, in particular when the loudspeakers form a regular setup (e.g. they follow the positions of a platonic solid), the matrix CC^T is diagonal and the solution is very simple (see below). Otherwise the solution cannot be expressed in closed analytic form.

For regular layouts, the solution for $N \geq 4$ turns out to be $s_i(t) = w_i W(t) + x_i X(t) + y_i Y(t) + z_i Z(t)$, with

$$w_i = \sqrt{2}/N \qquad (4.10\text{a})$$
$$x_i = 3\cos\phi_i\cos\delta_i/N \qquad (4.10\text{b})$$
$$y_i = 3\sin\phi_i\cos\delta_i/N \qquad (4.10\text{c})$$
$$z_i = 3\sin\delta_i/N \qquad (4.10\text{d})$$

In 2D the equations for decoding would be:

$$w_i = \sqrt{2}/N \tag{4.11a}$$

$$x_i = 2\cos\phi_i/N \tag{4.11b}$$

$$y_i = 2\sin\phi_i/N \tag{4.11c}$$

Remember that $W(t)$ corresponds to the signal recorded by a omnidirectional microphone, and that $X(t), Y(t), Z(t)$ are the signals recorded by three figure-of-eight microphones. The combination of this four microphones is also a microphone, in this case a supercardioid microphone pointing to the direction of the loudspeaker (see fig. 4.1) .

It turns out that in a regular layout the signal emitted by a loudspeaker is the signal that would be recorded by a supercardioid microphone pointing towards that direction. In particular, this means that:

- All loudspeakers emit sound at the same time.
- Given a sound source, loudspeakers in the opposite direction emit in opposite phase.
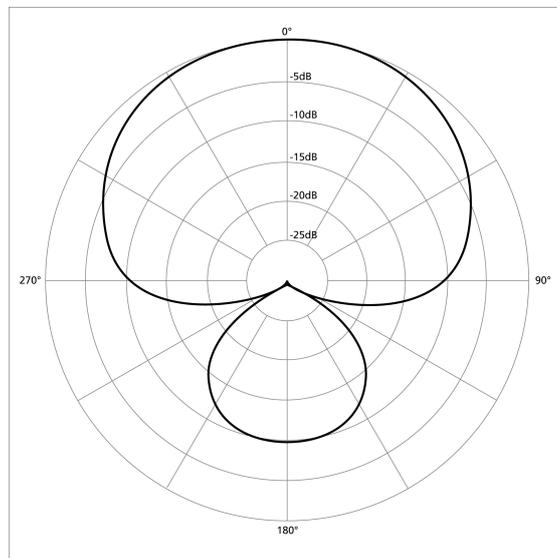

Figure 4.1: Example of a polar pattern of a supercardioid microphone. Figure reproduced from [9].

For the basic decoding, the reconstruction of the pressure and velocity is perfect at the origin, so that $r_v = 1$ for a plane wave. However, the basic decoding only works well for low frequencies (until up 500 Hz approximately) and very close to the sweet spot (exact reproduction center).

4.3 Phsychoacoustic decodings

4.3.1 Max-r_E decoding

In the hypothesis of *max-r_E* decoding, we require reproduction of the energy density and intensity of an incident wave under the assumption of the incoherent summation of all loudspeakers. In particular, the decoding is obtained by requiring that for an incident plane wave the pressure and velocity have the correct magnitude.

However it turns out that it is impossible to fullfill the condition $r_E = 1$; decodings will instead try to maximize this value.

In the case of the *max-r_E* decoding the equations are much more complicate to solve because they are non-linear. However, when the loudspeakers are on a regular layout (platonic solid or similar), a similar solution to the basic decoding can be found. As an example, in 3D, the expressions are:

$$w_i = 1/\sqrt{N} \tag{4.12a}$$

$$x_i = \sqrt{3/2} \cos \phi_i \cos \delta_i / \sqrt{N} \tag{4.12b}$$

$$y_i = \sqrt{3/2} \sin \phi_i \cos \delta_i / \sqrt{N} \tag{4.12c}$$

$$z_i = \sqrt{3/2} \sin \delta_i / \sqrt{N} \tag{4.12d}$$

which again is the equation of a supercardioid microphone pointing towards the directionn of the loudspeaker (this time having a smaller negative lobe). Notice the presence of the factor $1/\sqrt{N}$ instead $1/N$: the incoherent hypothesis makes sound level higher with the basic decoding.

This decoding works well for mid-to-high frequencies, from 500 Hz up, or when the audience is far from sweet spot.

For an incident point source, it is not possible to recover $r_E = 1$: the acoustic intensity will always have a smaller value compared with the energy density. For the max-r_E decoding the value of $r_E = 0.577$ (3D) and $r_E = 0.707$ (2D). This difference is important: the directionality of Ambisonics in 2D is quite acceptable, but this is not the case of 3D.

For non-regular layouts, it is not possible to solve the equations with algebraic or analytic methods. In these cases, one has to use non-linear heuristic search methods.

4.3.2 In-phase decoding

The *in-phase* decoding is similar to the max-r_E decoding, but with the additional restriction that no loudspeakers should be emitting in anti-phase.

The equations for decoding in 3D in regular layouts are:

$$w_i = \sqrt{3/4} \sqrt{2}/\sqrt{N} \tag{4.13a}$$

$$x_i = \sqrt{3/4} \cos \phi_i \cos \delta_i / \sqrt{N} \tag{4.13b}$$

$$y_i = \sqrt{3/4} \sin \phi_i \cos \delta_i / \sqrt{N} \tag{4.13c}$$

$$z_i = \sqrt{3/4} \sin \delta_i / \sqrt{N} \tag{4.13d}$$

This is the equation for a cardioid micropohone pointing towards the direction of the loudspeaker (see fig. 4.2).
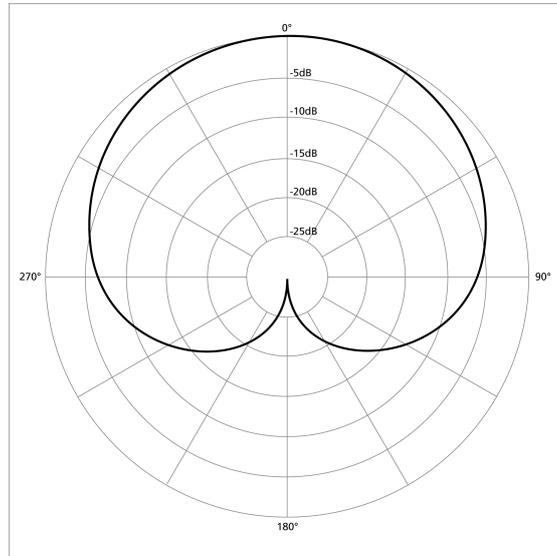

Figure 4.2: Example of a polar pattern of a cardioid microphone. Figure reproduced from [9].

The directionality properties of in-phase are somewhat worse that for max-r_E. For 3D $r_E = 0.5$, and for 2D $r_E = 0.677$.

The in-phase decoding is designed to work better with extended audiences, where the loudspeakers emitting in anti-phase can be a distraction.

Sometimes a mixed decoding strategy is used, where for low frequencies the basic decoding is used, and for high frequencies an in-phase or max-r_E decoding is employed. The crossover is done via a suitable shelf filter pair.

4.4 Decoding to stereo, 5.1 and other standard formats

2D Ambisonics can be decoded to 5.1, 7.1 surround or similar ayouts using the principles above, but it has to be taken into account that these are irregular decodings (so that the decoding coefficients have been computed by numeric methods or other advanced techniques). Similarlly, 3D Ambisonics can be decoded to 3D layouts such as 9.1, 10.1, 22.2, Auro 3D, Dolby Atmos, etc. Only the correct decoding coefficients need to be computed.

Ambisonics, to some extent, can also be decoded to stereo, taking into account that the directionality properties of the Ambisonics will be partially lost. In this case, the best solution is to simulate a virtual XY pair from the Ambisonics channels, corresponding to two channels L and R. Selecting the aperture and directionality of the virtual XY pair will lead to different stereo decodings.

5 Introduction to higher order Ambisonics (HOA)

5.1 Basic concepts

We have seen that one of the main drawbacks of Ambisonics is the poor directionality. It turns out that first order Ambisonics encodes any sound field in terms of the recording of an omnidirectional microphone (zero-th order microphone) and 3 figure-of-eight microphone (first order microphone). One can think of extending the expansion in terms of higher order microphones. This is what higher Order Ambisonics (HOA) does.

HOA decomposes the sound field in terms of the recording of a set of microphones called *spherical harmonics*. Mathematically, they correspond to the angular portion of the solution to the wave equation. Spherical harmonics have the following notation: $Y_{lm}(\phi, \delta)$, where l is the Ambisonics order, $m = -l, \ldots, l$ indicates the particular coefficient, and where ϕ and δ are azimuth and elevation.
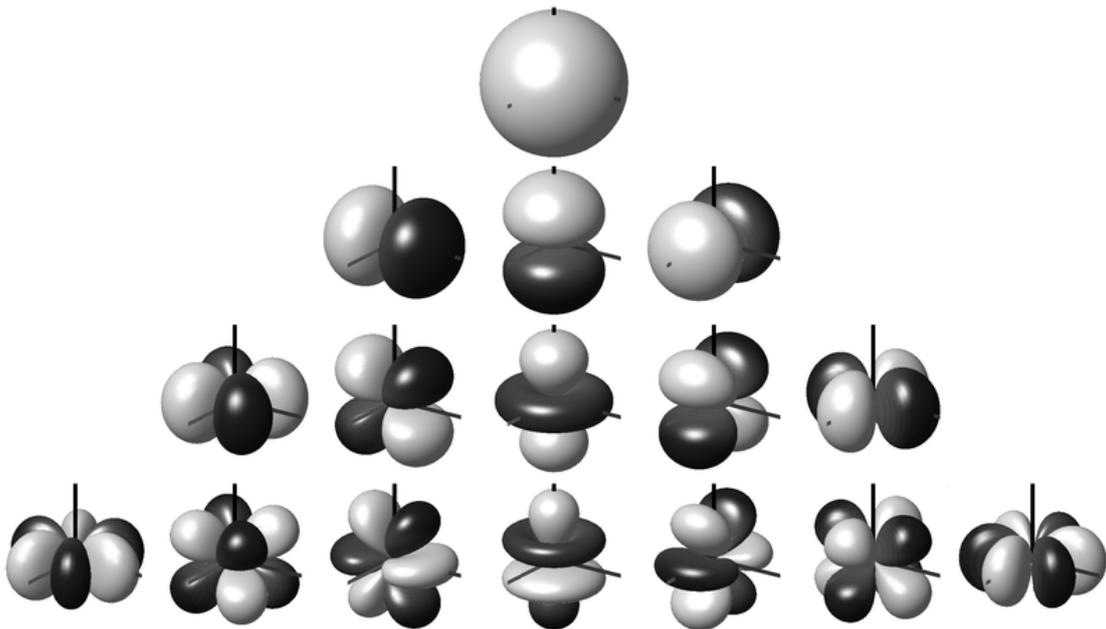
Figure 5.1: 3D polar patterns of a spherical harmonics up to third order. Figure extracted from [2].

ℓ	m	Spherical harmonic in N3D
0	0	1
1	-1	$\sqrt{3}\sin(\theta)\cos(\phi)$
1	0	$\sqrt{3}\sin(\phi)$
1	1	$\sqrt{3}\cos(\theta)\cos(\phi)$
2	-2	$\dfrac{\sqrt{15}}{2}\sin(2\theta)\cos^2(\phi)$
2	-1	$\dfrac{\sqrt{15}}{2}\sin(\theta)\sin(2\phi)$
2	0	$\dfrac{\sqrt{5}}{2}(3\sin^2(\phi)-1)$
2	1	$\dfrac{\sqrt{15}}{2}\cos(\theta)\sin(2\phi)$
2	2	$\dfrac{\sqrt{15}}{2}\cos(2\theta)\cos^2(\phi)$

Figure 5.2: Spherical harmonics according to the fully normalized real representation. From these expressions ϕ is the table is equivalent to δ in the text. Extracted from [10].

The maximum order l at which we perform the expansion constitues the HOA order. Each order has a $2l+1$ channels. In total, Ambisonics of order l has $(l+1)^2$ channels. In 2D, Ambisonics of order l has $2l + 1$ channels (by the way, in 2D the HOA expansion amounts to the standard Fourier series).

With the so-called fully normalized real representation with the, the values of the first spherical harmonics are:

$$Y_{0,0}(\theta,\delta) = 1, \tag{5.1a}$$

$$Y_{1,-1}(\theta,\delta) = \sqrt{3}\sin\theta\cos\delta, \tag{5.1b}$$

$$Y_{1,0}(\theta,\delta) = \sqrt{3}\sin\delta, \tag{5.1c}$$

$$Y_{1,-1}(\theta,\delta) = \sqrt{3}\sin\theta\cos\delta, \tag{5.1d}$$

The spherical harmonics up to second order can be checked in fig. 5.2. Any distribution of sources can be expanded in terms of sound fields as follows:

$$S(t;\phi,\delta) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} B_{lm}(t)Y_{lm}(\theta,\phi) \tag{5.2}$$

$B_{lm}(t)$ correspond to the *HOA channels*, and contain the traditional W,X,Y,Z, as we will see in a minute. The spherical harmonics $Y_{lm}(\theta,\phi)$ are the basis of the expansion.

( ) The perturbative expansion of the sound field around the origin can be written in spherical harmonics as:

$$p(t,\vec{r}) = \sum_{l=0}^{\infty} i^l j_m(kr) \sum_{m=-l}^{l} B_{lm}(t) Y_{lm}(\theta,\phi) \tag{5.3}$$

where $j_m(kr)$ are the spherical Bessel functions. The HOA expansion is a multipolar expansion of the sound field around the origin. The l-th order Ambisonics coefficients are formed up to l-th order derivatives of the sound field around the origin.

5.2 Encoding and recording

Given a plane wave (point source), with signal $s(t)$, coming from θ_s, δ_s, the encoding works as follows:

$$B_{l,m}(t) = s(t) Y_{lm}(\theta_s, \delta_s) \tag{5.4}$$

the first order HOA channels are:

$$B_{0,0}(t) = \sqrt{2}\, W(t) = s(t), \tag{5.5a}$$

$$B_{1,-1}(t) = \sqrt{3}\, Y(t) = s(t)\, \sqrt{3} \sin\theta_s \cos\delta_s, \tag{5.5b}$$

$$B_{1,0}(t) = \sqrt{3}\, Z(t) = s(t)\, \sqrt{3} \sin\delta_s, \tag{5.5c}$$

$$B_{1,-1}(t) = \sqrt{3}\, X(t) = s(t)\, \sqrt{3} \sin\theta_s \cos\delta_s, \tag{5.5d}$$

The fully normalized convention ensures that the coefficients of the expansion have the same exact mathematical expression than the spherical harmonics. Notice the different order with respect to first order Ambisonics. Any distribution of sources can be created with multiple plane waves (point sources)

There are microphones that can record directly in higher order Ambisonics, like the Eigenmike microphone (fig. 5.3, that has 32 capsules, whose ouptut can be converted to HOA.

5.3 Transmission and manipulation

Up to a given order l, HOA has $(l+1)^2$ channels. At 3rd order, 16 channels, regardless of the number of sound sources. These channels can be saved in an ordinary multichannel wave file.

There are several possibilities for channel normalization, and also for channel order. The normalization we have presented is called "fully normalized 3D", and the channel order (first order by increasing l, then by increasing m) is called Ambisonics channel number [10].

Rotation of HOA is considerably more complicated that first order Ambisonics, but can be done. One needs to find the correct representation of the rotation group in the real spherical harmonics basis. Other spatial manipulations, like warping, are currently a topic of research.

Figure 5.3: Eigenmike by mh acoustics. 32-capsule microphone that can record in Ambisonics. Extracted from [11].

5.4 Reproduction

HOA of order l needs at least $(l + 1)^2$ loudspeakers for reproduction. For order 2, at least 9 speakers; for order 3, at least 16 speakers; for order 4, at least 25.

HOA decoding proceeds similarly as first order Ambisonics. There are two possibilities for decoding: either reconstructiong the spherical harmonics at the origin, or trying to deliver a physchoacoustic decoding which maximizes the directionality properties of the energy. Decoding to regular layouts is relatively easy, but decoding to irregular layouts is an active topic of research: one must either apply algebraic methdods (pseudoinverse, basic decoding), or non-linear search methods (psychoacoustic decodings), or combinations of these.

Similarly to first order Ambisonics, the result of the decoding in regular layouts is that every loudspeaker emits the signal of a virtual microphone pointing in the direction of the loudspeaker. For the basic and max-r_E decoding, this virtual microphone is supercardioid; for the in-phase decoding, the virtual microphone is cardioid, as can be seen in fig. 5.4.

The directionality properties of HOA are better than first order Ambisonics (after all, that's the reason for using HOA). Table 5.1 shows the best directionality coefficients of HOA in 3D depending on the decoding.

Table 5.1: Possible values for the directionality coefficient r_E in HOA depending on the order.

| Order	max-r_E	in phase
1	0.577	0.500
2	0.775	0.567
3	0.861	0.750
4	0.906	0.800
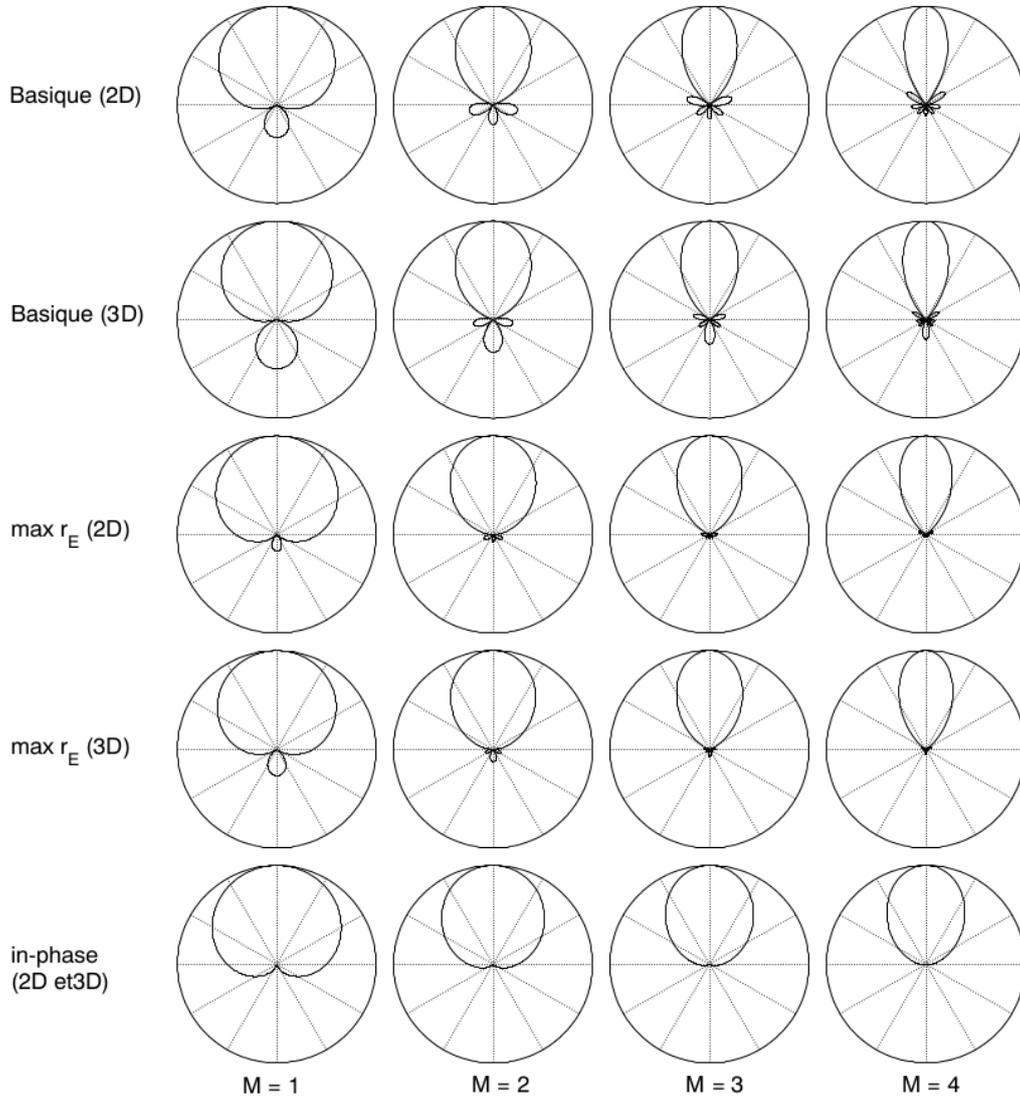
Figure 5.4: Equivalent virtual microphones used for HOA reproduction. Each microphone reproduces the signal of a virtual higher order microphone pointing in the direction of that loudspeaker. Reproduced from [1].

6 Conclusions: advantages and drawbacks of Ambisonics

6.1 Advantages

- Ambisonics is a *complete theory*, covering encoding, recording, postpro duction, transmission and reproduction, based on physical principles of the acoustic field.

- Ambisonics is completely *independent of the exhibition format*. The Ambisonics channels do not have nothing to do with the traditional channels of a stereo, or 5.1, and are instead based on physical principles of the acoustic field.

- Ambisonics has a *fixed number of channels*, in contrast to object-based methods (e.g. VBAP, WFS). It can be adapted to an object-based method if required, though.

- It provides a *smooth listening experience*: pannings are smooth and do not suffer from "jumping" artifacts of e.g. VBAP. Also, the sensation of immersion in the sound field is good.

- It requires only a *moderate number of loudspeakers*, in contrast to WFS that requires hundreds or thousands of loudspeakers.

6.2 Drawbacks

- *Poor directionality properties*, worst than e.g. VBAP, specially at first order and in 3D. 1st order 3D is close to being unacceptable. At higher orders (2,3,4) the situation is much beter.

- *Small sweet spot.* The sweet spot is smaller than in VBAP, but much larger than in transaural. It can be improved by going to higher order and using psychoacoustic decodings.

- *Technically challenging.* Specially, the decoding to non-regular layouts is complicated. Also, higher order Ambisonics is non-trivial.

- Less control over the final exhibited result, as compared to traditional surround methods.

- Non-coincident microphone array recordings are difficult to codify in Ambisonics.

7 References

Basic references

[1] Jérôme Daniel. "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia". PhD thesis. Université Paris 6, 2000. URL: `http://gyronymo.free.fr/audio3D/download_Thesis_PwPt.html`.

[2] *Ambisonics — Wikipedia, The Free Encyclopedia.* 2015. URL: `http://en.wikipedia.org/w/index.php?title=Ambisonics&oldid=656474391` (visited on 05/09/2015).

[3] Michael A Gerzon. "General metatheory of auditory localisation". In: *Audio Engineering Society Convention 92.* Audio Engineering Society. 1992.

Complementary references

[4] Michael A Gerzon. "Periphony: With-height sound reproduction". In: *Journal of the Audio Engineering Society* 21.1 (1973), pp. 2–10.

[5] Angelo Farina. *A-format to B-format conversion.* URL: `http://pcfarina.eng.unipr.it/Public/B-format/A2B-conversion/A2B.htm` (visited on 05/09/2015).

[6] *SoundField Digital Surround Sound Microphone Systems.* URL: `http://www.tslproducts.com/soundfield-type/soundfield-microphones/` (visited on 05/09/2015).

[7] *Core Sound.* URL: `http://core-sound.com` (visited on 05/09/2015).

[8] *Ambisonic UHJ format — Wikipedia, The Free Encyclopedia.* 2014. URL: `http://en.wikipedia.org/w/index.php?title=Ambisonic_UHJ_format&oldid=633579838` (visited on 06/09/2015).

[9] *Microphone — Wikipedia, The Free Encyclopedia.* URL: `http://en.wikipedia.org/w/index.php?title=Microphone&oldid=661279706` (visited on 05/09/2015).

[10] *Ambisonic data exchange formats — Wikipedia, The Free Encyclopedia.* 2015. URL: `http://en.wikipedia.org/w/index.php?title=Ambisonic_data_exchange_formats&oldid=644989991` (visited on 05/09/2015).

[11] *mh acoustics.* URL: `http://mhacoustics.com` (visited on 05/09/2015).