# An Introduction to Higher-Order Ambisonic

**by Florian Hollerweger, April 2005**

## 1 First-Order Ambisonic (B-Format)

### 1.1 Encoding

The roots of Ambisonic date back to the 1970s, when *Michael Gerzon* from the University of Oxford introduced First-Order Ambisonic in form of the so-called *B-Format*, which encodes the directional information of a given three-dimensional soundfield to four channels called W, X, Y, Z:

| | |
|---|---|
| $W = s[\frac{1}{\sqrt{(2)}}]$ | omnidirectional information |
| $X = s[\cos\phi\cos\theta]$ | x-directional information |
| $Y = s[\sin\phi\cos\theta]$ | y-directional information |
| $Z = s[\sin\theta]$ | z-directional information |

where Φ (phi) is the horizontal angle (azimuth), and Θ (theta) the vertical angle (elevation).[1] The equations show that an Ambisonic soundfield can be synthesized by multiplying an audio signal s with the value of a certain function in the desired direction. This is what those functions look like for the different Ambisonic channels (W to Z from left to right):

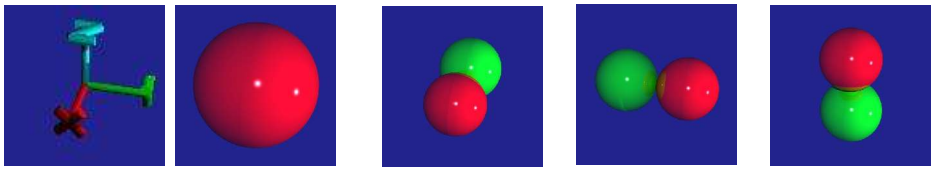

Fig. 1: First Order Ambisonic encoding functions

It is intuitively clear that these signals can be captured by the means of one omnidirectional microphone (the W channel) and three figure-of-eight microphones (channels X, Y, Z), which allows for First-Order Ambisonic recordings of real soundfields. Ideally, those microphones should all be located in the origin of the coordinate system. The *Soundfield microphone* (Fig. 2) has been built for this purpose. It is made up of four cardioid capsules arranged in a tetrahedron, which can be combined as needed to provide the desired polar patterns.

Fig. 2: soundfield microphone

[1] The coordinate system assumed here has the x axis showing into the 0° direction for both, azimuth and elevation. If the listener faces the positive x direction, he finds the positive y direction to his left, and the positive z axis above him. The azimuth goes counterclockwise from 0° to 360°, the elevation is positive for values above, and negative for values below the xy plane.

## 1.2 Decoding

It is important to understand, that the Ambisonic-encoded signals are not feeding any speakers themselves, but carry the directional information of an entire soundfield. That means, that they are completely independent from the loudspeaker layout chosen for *decoding* the soundfield. An Ambisonic decoder is therefore always designed for a specific speaker layout, and an Ambisonic-encoded soundfield can be reproduced on any Ambisonic decoding system. Ambisonic does not imply a certain number of loudspeakers used for reproduction. The only limit it puts on the number of speakers is, that the minimum number of loudspeakers L is equal to the number of Ambisonic channels N:

$$L \geq N$$

In case of the B-Format, this means four speakers for a periphonic (i.e. three-dimensional) loudspeaker setup. In case of a horizontal-only setup, where the Z channel can be neglected (since it's the only channel extending in the z-direction, see Fig. 1), three speakers are sufficient. However, it is fine and even desirable to use more speakers than the number of Ambisonic channels, since this can increase the overall quality of sound localization.[2]

An Ambisonic encoder has however put certain demands towards the layout of the loudspeaker array: it is supposed to be as regular as possible. The more regular it is, the better the results in terms of localization of audio sources will be, just like in Vector Base Amplitude Panning (VBAP) systems. In other words, the decoder will do its job as good as it can with the speaker layout offered to him. There are two major decoding strategies, which will be introduced in the chapters 1.2.1 and 1.2.2:

### 1.2.1 Decoding through Projection

So how are the actual speaker signals derived from the Ambisonic signals? A very basic Ambisonic decoder works like this: each speaker receives its own weighted sum of *all* Ambisonic channels. For each speaker, the weight of an Ambisonic channel equals the value of the according spherical harmonic for the position of that speaker.[3] In other words, the spherical harmonics are spatially sampled by the speakers.

So the signal feeding the i-th loudspeaker is:

$$p_i = \frac{1}{N}\left[W\left(\frac{1}{\sqrt{(2)}}\right) + X\left(\cos\phi_i \cos\theta_i\right) + Y\left(\sin\phi_i \cos\theta_i\right) + Z\left(\sin\phi_i\right)\right]$$

---

[2] Based on these considerations, Michael Gerzon has suggested that *quadrophonic systems* should be driven rather by only *three* (Ambisonic-encoded) channels (W, X, Y) than by four discrete channels (one for each speaker).

[3] This is refered to as re-encoding the loudspeaker positions, since the same encoding functions are applied.

with ($\Phi_i$, $\theta_i$) being the position of the i-th speaker, and N the number of Ambisonic channels.[4]

This decoding strategy assumes regular speaker layouts, like 8 speakers along a circle, separated by equal angles. If the layout is not regular, the decoder will just act as if this was the case, which is called *projection* (of the Ambisonic signals onto the loudspeaker array). For 3D layouts, there is unfortunately only a very limited number (five) regular layouts, the so-called *platonic solids*[5], which means that we are also very limited in the number of loudspeakers used.

### 1.2.2 Decoding through Pseudoinverse

A notation of the decoding equation in matrix form can be useful. Assume, that B is the column vector of Ambisonic channels (B = [W X Y Z]$^T$), p the column vector of loudspeaker signals, and C the *re-encoding matrix*. The entries of C are the values of the spherical harmonics for the loudspeaker positions, with N rows for the different spherical harmonics, and L columns for the speakers. We can express the decoding function as:

$$B = C * p$$

thus:

$$p = C^{-1} * B$$

$C^{-1}$ is the inverse of C (also called *decoding matrix*). To invert C, the matrix needs to be square, which is only the case if L=N (number of speakers = number of Ambisonic channels). Since generally, L>N, the inversion can only be done by the means of the *pseudoinverse*:

$$pinv(C) = C^T (C * C^T)^{-1}$$
$$p = pinv(C) * B = C^T (C * C^T)^{-1} * B$$

The pseudoinverse will only provide useful values if the condition number of C is small. This can be achieved by minimizing the maximal distance and angle between two speakers, i.e. making the layout again as regular as possible. For perfectly regular speaker layouts, decoding by the means of projection is therefore equivalent to decoding with the pseudoinverse.

## 1.3 Soundfield Operations

Since Ambisonic encodes the soundfield created by multiple sound sources rather than

---

[4] Since each speaker receives all Ambisonic channels, the speaker signals have to be normalized by 1/N to avoid clipping.

[5] cube, tetrahedron, octahedron, icosahedron, dodecahedron

the sources themselves, it is straightforward to apply operations on this soundfield as an entity:

### 1.3.1 Rotation, Tilt, Tumble

By the means of simple *rotation matrices*, it is possible to rotate an Ambisonic soundfield around all three axes of an xyz coordinate system. *Rotation* (or *yaw*) refers to the z-axis, *tilt* (or *roll*) to the x-, and *tumble* (or *pitch*) to the y-axis.

This property of Ambisonic soundfields is exploited in binaural 3D audio reproduction techniques including head tracking: in natural acoustic environments, we constantly make small head movements, in order to maximize interaural level and time differences and thus improve our auditory localization. Using headphones, you usually can't make head movements without moving the soundfield along with you. Head tracking devices coupled to a Ambisonic rotation matrix can achieve this with relatively little CPU power.

### 1.3.2 Mirroring

It is possible to easily mirror an Ambisonic encoded soundfield, i.e. create movements of all sources in the soundfield to their diametrically opposed directions.

### 1.3.3 Zoom

Zoom-like operations (also known as focus, dominance, or acoustical lense) can be applied to an Ambisonic soundfield by the means of Lorentzian transformations or filters.

## 1.4 Summary

Ambisonic at its time was pretty much a commercial failure, which is certainly also due to the fact that Gerzon's approach was just way ahead of its time.[6] However, in recent years, it has experienced a comeback in 3D audio applications for loudspeakers and headphones, being extended to Higher Order Ambisonic.

# 2 Higher Order Ambisonic

## 2.1 Basics

In the 1990s, it has been shown that the Ambisonic approach can be extended to higher orders, increasing the size of the sweet spot in which the soundfield is accurately reproduced as well as the overall quality of localization. The price paid for this is that with increasing order, a growing number of additional Ambisonic channels will be introduced (5

[6] It is interesting to note though, that obvious analogies can be found between the Ambisonic approach and the early sterophonic microphone techniques "MS" and "Blumlein pair" (crossed figure-of-eights), if you take a look at their mic layouts including the polar patterns of the microphones.

new channels for second order, 7 new channels for third order, etc.), which also means that the minimum number of required loudspeakers increases (remember L≥N!).
So what do these new Ambisonic signals look like? The mathematical idea behind the extension of the Ambisonic approach to higher orders looks like this:

*1) A soundfield can be regarded as a superposition of plane waves.*

It is intuitively clear that it must therefore be possible to reproduce a soundfield by reproducing the plane waves it is composed of, for example with loudspeakers. The waves emitted by loudspeakers can be assumed plane when the speakers are "far enough away", i.e. when their distance is big compared to the wavelengths of the frequencies they reproduce.

*2) A plane wave can be represented as an infinite series.*

The mathematical expression of this series looks rather complicated. Fortunately, if we limit ourselves to an accurate reproduction of the desired soundfield in the origin of our coordinate system (i.e. the sweet spot), things become much easier, and we can develop the series by the means of *spherical harmonic functions*[7], which consist of simple constant, sine and cosine terms. Spherical harmonic functions come in different orders, with increasing numbers of functions for increasing order. For example, there is only one function of zeroth order. It is the one that is applied in the creation of the W channel of a First-Order Ambisonic system (see Fig. 1)[8]. For the first order, three functions exist, which are applied in the creation of the X, Y, and Z channel. Thus, a First-Order Ambisonic encoded soundfield actually combines the Ambisonic signals of first and zeroth order. Similarly, a third-order system is made up not only by the third-order signals, but also by those of zeroth, first, and second order.
Below are the five spherical harmonics of second order:

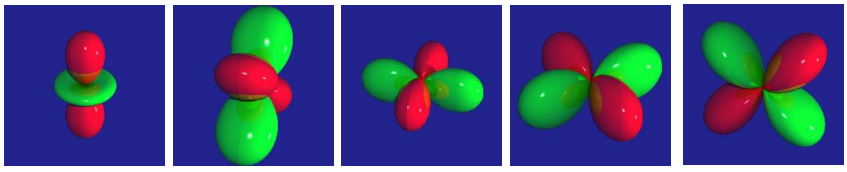

Fig. 3: Second order spherical harmonics

And the according Ambisonic signals:

$$R = s[\frac{1}{2}(3\sin^2\theta - 1)]$$

$$S = s[\cos\phi \sin 2\theta]$$

$$T = s[\sin\phi \sin 2\theta]$$

[7] Spherical harmonic functions describe a function on the surface of a sphere.

[8] So actually most of us got a zeroth order Ambisonic system in their old mono kitchen radio...

$$U = s[\cos 2\phi \cos^2\theta]$$

$$V = s[\sin 2\phi \cos^2\theta]$$

And the seven spherical harmonics of third order:

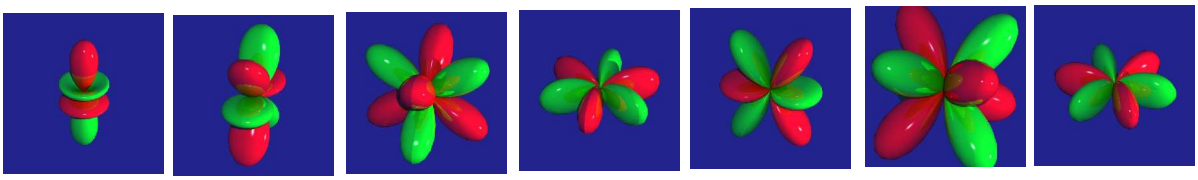

Fig. 4: Third order spherical harmonics

$$K = s[\frac{1}{2}\sin\theta(5\sin^2\theta - 3)]$$

$$L = s[\frac{8}{11}\sin\phi\cos\theta(5\sin^2\theta - 1)]$$

... etc. for channels M, N, O, P, Q ...

Since with every new order the number of new functions introduced increases by two, the sum over all these functions for a given order is given by

$$N = (M+1)^2 \qquad \text{for 3D reproduction}$$

where N is the number of Ambisonic channels, and M is the order of the system. For horizontal-only reproduction, only the spherical harmonics that are dependent from the z-value are counted, and we end up with:

$$N = 2M + 1 \qquad \text{for 2D reproduction}$$

Obviously, the directional information carried by the spherical harmonics becomes more distinguished with increasing order, giving a more and more accurate localization. At the same time, it is also obvious that recording of Higher Order Ambisonic signals becomes more tricky, since microphones with according polar patterns don't exist. The required polar patterns have to be created by the means of microphone arrays. However, the synthesizing of artificial soundfields is just as easy as it is in a first order system.

It has been mentioned above that the minimum number of loudspeakers required by an Ambisonic system is given by the number of Ambisonic channels ($L \geq N$). It is therefore quite obvious, that the series describing our plane wave has to be truncated at some point (i.e. at a specific order), since the number of available channels and loudspeakers will be finite. This results in an approximation of the original soundfield, being more accurate for higher orders (better localization, bigger sweet spot). It has been shown, that Ambisonic can be as a special case of *holophony*, relating it to the principle of *Wave Field Synthesis*.

## 2.2 Mixed-Order Systems

In designing a *periphonic* (i.e. 3D) audio reproduction system, the psychoacoustic properties of the human hearing can be considered by making use of *mixed-ordered systems*, where the horizontal and vertical parts of a soundfield are encoded separately with different orders. Since the spatial resolution of the human ear is particularly well developed in the horizontal plain, the horizontal part will typically be encoded in a higher order than the vertical part. The number of Ambisonic channels for a mixed-ordered system is given by:

$$N = N_H + N_V = [2M_H + 1] + [(M_V + 1)^2 - (2M_V + 1)]$$

where N is the number of channels, $M_H$ the horizontal, and $M_v$ the vertical order of the system. For example, a system with $M_H = 3$ und $M_v = 1$ consists of N = 8 channels. A full periphonic system of second order requests $N = (M+1)^2 = 9$ channels. Therefore, it is possible to optimize the number of transmitted channels and the quality of the spatial audio reproduction at the same time. Unfortunately, the ability to comfortably rotate the soundfield around the x- and y-axis (*tilt* and *tumble*) is lost within mixed-order systems, only allowing for rotations around the z-axis.

# 3 Extensions to Higher Order Ambisonic Systems

## 3.1 Decoder Flavor

As described above, an Ambisonic decoder attempts to recreate the encoded soundfield in the center of the loudspeaker setup (sweet spot). If we want to enjoy the acoustical pleasures of this recreation, we will unfortunately have to introduce our head into the scene, which causes the following problem: since Ambisonic is sound field oriented rather than sound source oriented, typically *all* loudspeakers in the array contribute to the decoding of the field created by an Ambisonic encoded source. This sounds weird, but the idea is, that in the center of the array, the contributions of the different speakers interact, meaning that for example the signals of loudspeakers opposing each other cancel each other out, since they are 180° out of phase.

However, the introduction of our head introduces reflection and diffraction effects, making such a phase cancellation impossible. This is particularly annoying when a loudspeaker in the opposite direction of an encoded source contributes to the decoding of this source: due to the effects around our head, we will be able to hear this speaker's contribution as a separate source. The effect is also more disturbing for listening positions out of the sweet spot, where the distance to the speakers is not equal any more, and closer speakers will

more likely be heard as separate sources. A solution to this problem is the *in-phase decoder*, which requires that all speakers produce equally phased signals, removing the dependency on cancellation effects in the sweet spot. This is simply achieved by applying gain factors to the basic decoder matrix.

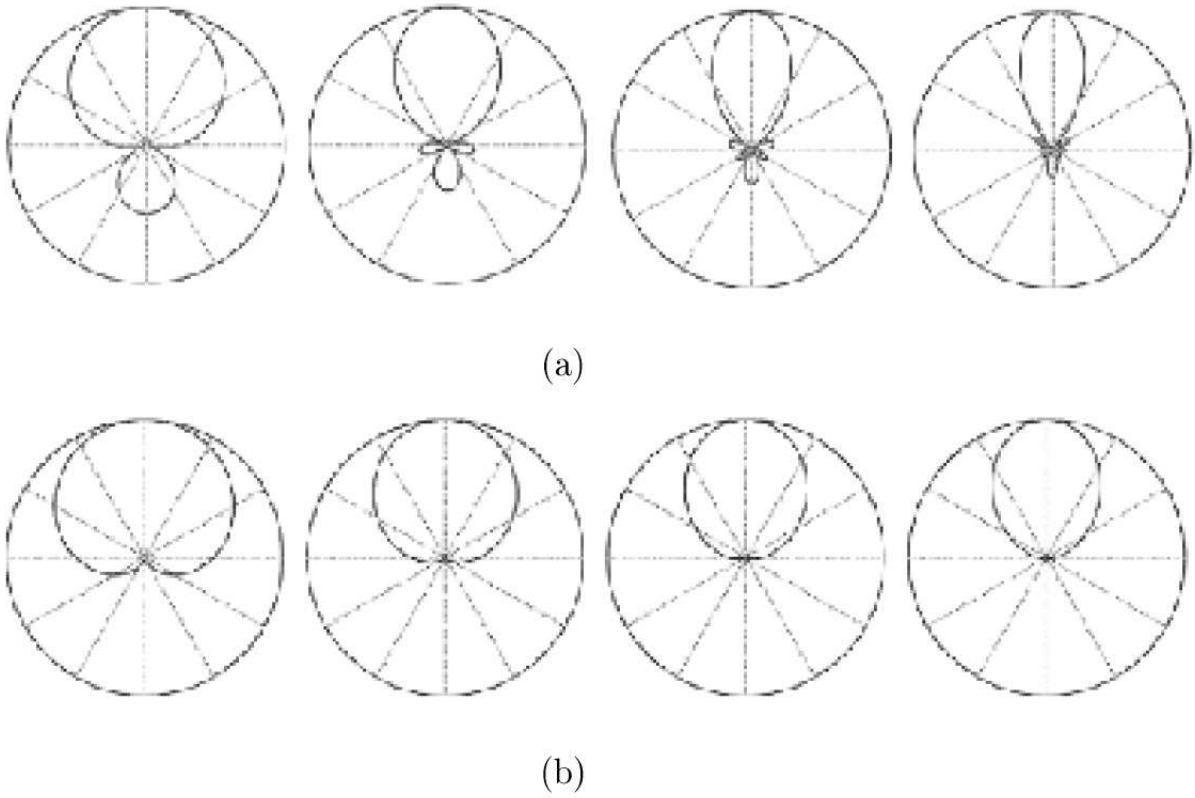


Fig. 5: polar patterns of
(a) a basic decoder
(b) an in-phase decoder
for a source encoded at 0°, and for increasing Ambisonic order (from left to right)

As the figure above shows, removing the contributions from directions opposite to the source is done at the cost of widening the source position. By increasing the order of the system, the lost localization quality can be regained. Fig. 5 can be drawn as an ordinary function showing the amplitude of speakers as a function of angle:

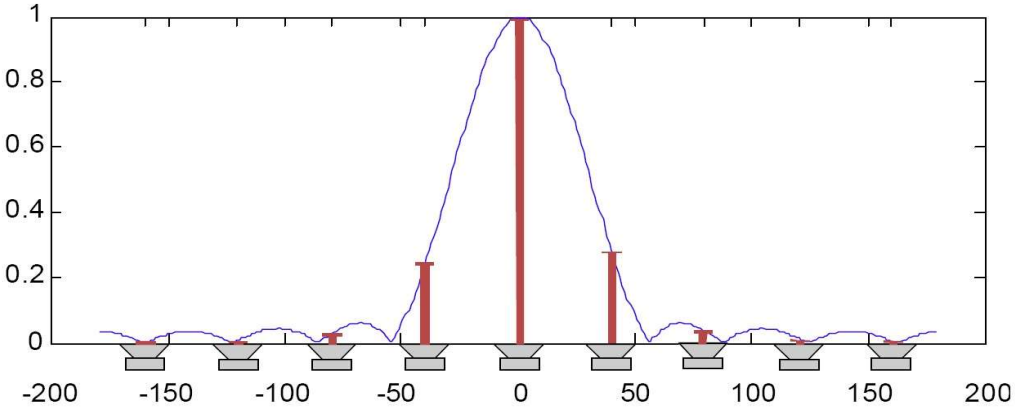


Fig. 6: Window-applied Decoding
x-axis: azimuth in [ °]
y-axis: speaker amplitude

An obvious analogy to the characteristics of windows used in FIR filter design can be found here: an in-phase decoder reduces the amplitude of the side lobes at the cost of widening the main lobe. This way, the principles of window design can be applied in the optimization of decoders for Higher Order Ambisonic systems. This process called *window-applied decoding* can be used to find compromises between the extremes of basic and in-phase decoding, depending on the auditorium the decoder is built for.

## 3.2 Distance Coding

Up to now, we have considered the position of our Ambisonic encoded sound sources to be only dependent on the azimuth and elevation, quietly assuming, that their distance is equal to the radius of the loudspeaker array used for reproduction (an assumption which is familiar from VBAP). However, natural acoustical environments consist of sound sources in varying distances. To reproduce or synthesize natural sounding soundfields, it is therefore necessary to think about how to encode the distance of sources. For sources far away, this can be achieved in a straightforward way by the means of artificial reverb algorithms; again this is similar to how it is done in VBAP. Things become more tricky though if one wishes to locate sources *within* the speaker array. At least two approaches exist for this problem, and both of them try to encode the distance by synthesizing wavefront curvatures: the curvature of a point-like sound source expands spherically, so the further you move away from the source, the less curved the wavefront will be, leaving you with a plane wave at an infinite distance. Remember, that one major idea of Ambisonic is to assume that the loudspeakers emit such plane waves, which removes any dependancy on the distance of a sound source, since the amplitude of a plane wave does *not* decay with distance like a spherical wave does.[9] While the mathematical treatment of plane (non-curved) wavefronts is much easier, we actually loose our distance information this way.

### 3.2.1 Distance Coding by Near Field Compensation

One approach to regain this information bases on considering the fact, that the loudspeakers are not emitting plane waves anyway, if the distance of the speaker array is finite (which is usually the case). This finite distance results in a bass-boost effect (similar to the proximity effect experienced with velocity microphones like the Shure SM58): the closer from infinity I come with my speaker array, the more bass frequencies I introduce. It

[9] The amplitude of a spherical wavefront decays with 1/r. This is the familiar effect of things being more quite if they are far away. The decay is due to the fact, that the energy of the source has to spread over an increasing spherical surface as the wavefront expands from its origin. The energy of plane wavefronts does not have to spread, since it travels in only one direction, which is why their amplitude does not decay with distance.

has therefore been suggested to introduce a new encoding format called *NFC-HOA* (Near-Field Compensated Higher Order Ambisonic), which compensates this bass boost effect already in the *enc*oding stage by the means of filters. This way, distance coding can elegantly be introduced at the same time: if I want a source at the distance of my loudspeaker array, simply don't apply any compensation of the bass-boost. The speakers are going to reproduce the according wavefront curvature anyway! If you want sources outside the speaker array, reduce the bass, since the speaker array will cause a bass-boost because it is "too close". If you want inside sources, even turn up the bass, to simulate the additional bass boost you would get from even closer speakers. The price being paid for this is, that the radius of the loudspeaker array now has to be known at *encoding* time, removing the clear separation between encoding and decoding stage. However, additional filters can compensate for the difference between the radius assumed at encoding time, and the actual radius of a loudspeaker array the soundfield is decoded too.

### 3.2.2 Distance Coding by a Hybrid Holophony / HOA Approach

Another idea is based on a hybrid approach of holophony (i.e. Wave Field Synthesis in 3D) and Higher Order Ambisonic. In the first stage, the curvature of the wavefront (and thus the distance of the source emitting the wavefront) is encoded by the means of holophony to a *virtual loudspeaker array*. In a second stage, the virtual loudspeaker signals are then encoded into an Higher Order Ambisonic soundfield as Ambisonic sources with static position information.

This approach makes it possible to exploit the advantages of holophony (possible distance coding), without the huge amounts of loudspeakers demanded by holophonic systems. Obviously, the number of virtual speakers as well as their layout and spacing are primary design parameters for this approach.

## 3.3 Considering Sound Source Characteristics (O-Format)

Introducing the curvature of spherical wavefronts in distance coding, we have said goodbye to the mathematically useful but rather unrealistic concept of plane wavefronts.[10] However, a spherical wavefront has its origin in a point-like sound source. Quite obvious, real sound sources are unlikely to be point-like (think of a piano!), and even worse, their polar characteristics will be heavily frequency dependent.

It is possible to approximate the surface shape of a sounding object by the means of the

[10] It's not unrealistic in so far as any soundfield can be understood as a superposition of such plane waves. It is just unrealistic to assume that a loudspeaker at a finite distance will emit a plane wave!

by now well-known spherical harmonics.[11] This allows for efficiently modelling the impulse response of the sounding object, which can thus be embedded into an Ambisonic-encoded soundfield including its spatial characteristics. This representation of sounding objects is reffered to as the so-called *O-Format*.

### 3.4 Room Reflection Cancellation

Another quiet assumption we have made is the one of free field conditions, which would mean that only the direct loudspeaker signals will contribute to the restoration of the encoded soundfield in the sweet spot. However, in real life wall reflections disturb the reproduction of the original soundfield. Different algorithms exist for applying room reflection cancellation:
By measuring the directional impulse responses of the reproduction room in the sweet spot, it is possible to interpret the reflexions as virtual sources and encode additional Ambisonic sources which destructively interfere with these reflexions (phase cancellation).

Another approach expolits the analogies of Ambisonic to Holophony: the room reflections are measured by the means of a microphone array (a typical holophonic recording technique). The transfer functions from each speaker to each microphone are measured and compared to the free-field conditions. This leads to a set of filters applied to the speakers, which allows for reflection cancellation, but only within the bounds of the microphone array.

## 4 Conclusion

The possibility of higher-order extension makes the Ambisonic approach very flexible in terms of *scalability* – also due to the *self-compability* of Ambisonic-systems, which is an effect of the complete separation of *encoding* and *decoding* stage: higher-order encoded soundfields can be reproduced on lower-order decoding systems (by simply ignoring the channels exceeding the order of the reproduction system) and vice versa (by happily accepting the fact that you have more speakers available than you actually need).

On the other hand, a speaker layout which provides good properties regarding regularity for a decoder of a certain order, does not necessarily have as good properties for other orders. This is of course a special issue in the design of mixed order systems.

Also, the *soundfield operations* mentioned above (rotations, mirroring, zooming) become increasingly difficult or impossible at all to implement for higher orders.

---

[11] Again, this is a scalable approach, since higher order spherical harmonics will allow for closer approximation.

However, the advantages of Higher Order Ambisonic in terms of efficiency regarding CPU and hardware (number of channels and loudspeakers) make it a very attractive approach for 3D audio reproduction.

Higher Order Ambisonic systems can be improved by including *distance coding*, considering the physical properties of sounding objects (*O-Format*) and applying *room reflection cancellation* algorithms.

## Links

http://audiolab.uwaterloo.ca/~jeffb/info/thesis.html
thesis on 2nd and 3rd order systems

http://members.tripod.com/martin_leese/Ambisonic/faq_latest.html
FAQ

http://www.ambisonic.net/
general forum

http://www.mshparisnord.org/cicm/dl_en.htm
externals for MAX/MSP, PD

http://www.york.ac.uk/inst/mustech/3d_audio/
Ambisonic at University of York, open source Ambisonic VST plugins

florian@create.ucsb.edu