

# AMBISONICS TUTORIAL

## INTRODUCTION

Ambisonic has several advantages compared to other surroundsound- techniques:

- It supports periphery (meaning the inclusion of the height component of sound)
- The image is stable and precise, independent from the position of the virtual sound (coming from the speaker or from a point in-between the speakers). This means that the sound does not change, as it moves around; thus the sound gets liberated from the speakers.
- The position of the listener is not that important for getting a true localisation as it is in most other surroundsound-techniques. Even listeners far off the sweet spot still perceive a realistic image.
- Once the sound is spatially encoded it can be decoded for performance to any desired speaker-setup, as long it is symmetric.
- Ambisonics is free and efficient.
- It can be combined with the distance -clues of the virtual sound source, resulting to sounds, which are perceived as being closer or more distant to the listener.

## THE AMBISONIC FORMAT

The spatial reproduction with Ambisonics splits up into two basic parts:

1. Encoding  
and
2. Decoding.

## ENCODING

In the encoding part a soundsource is either recorded with a soundfield microphone, or angle and elevation co-ordinates are assigned via the encoding equations to any arbitrary mono source (soundfile/ speech/ instrument/ synthesized/ microfoned) with a computer. The result are 4 Audio Tracks (W, X, Y, Z) in First Order Ambisonics or 9 Audio Tracks (W, X, Y, Z, R, S, T, U, V) in the Second Order Format.

W is omnidirectional/ scalar and does contain simply the pressure gradient of the sound, whereas the others are vectors and correspond to axial directions of the cartesian space. This means that W contains all the sound information, but directionless, as if the whole piece was just mono. X for example however contains just the amount of sound that propagates in X direction. It will contain the full signal (Amplitude =1) for a source right in front or behind (Amplitude =-1) of the listener (lying directly on the X-axis), but nothing (Amplitude =0) of any signal to the left or to the right of the listener lying directly on the Y-axis. For sounds inbetween the directions of the axes their directional energy is split up in proportion according the position. Thus the spatial directions of the sound are

encoded. Some beautiful pictures of the directional lobes here:  
[http://members.tripod.com/martin\\_leese/Ambisonic/harmonic.html](http://members.tripod.com/martin_leese/Ambisonic/harmonic.html)

Here are the encoding equations for 2<sup>nd</sup> order by Richard Furse and Dave Malham (FMH - set of encoding equations). The input signal is multiplied with them to get the information encoded. Angle (A) and elevation (E) have to be assigned as desired. This projects the sounds somewhere at the unit- sphere. For sounds off the unit sphere the distance- clues have to be added.

FMH -set of encoding equations:

| Label |                | Polar Representation     | Cartesian Representation |
|-------|----------------|--------------------------|--------------------------|
| W=    | input signal * | 0.707107                 | 0.707107                 |
| X=    | input signal * | $\cos(A)\cos(E)$         | x                        |
| Y=    | input signal * | $\sin(A)\cos(E)$         | y                        |
| Z=    | input signal * | $\sin(E)$                | z                        |
| R=    | input signal * | $1.5\sin(E)\sin(E)-0.5$  | $1.5zz-0.5$              |
| S=    | input signal * | $\cos(A)\sin(2E)$        | 2zx                      |
| T=    | input signal * | $\sin(A)\sin(2E)$        | 2yz                      |
| U=    | input signal * | $\cos(2A)\cos(E)\cos(E)$ | xx-yy                    |
| V=    | input signal * | $\sin(2A)\cos(E)\cos(E)$ | 2xy                      |

Recommended Reading:

[http://www.york.ac.uk/inst/mustech/3d\\_audio/ambis2.htm](http://www.york.ac.uk/inst/mustech/3d_audio/ambis2.htm) Spatial Hearing Mechanisms and Sound Reproduction by D.G. Malham, University of York, England  
[http://www.york.ac.uk/inst/mustech/3d\\_audio/ambitips.html](http://www.york.ac.uk/inst/mustech/3d_audio/ambitips.html) Ambisonics hints and tips page  
<http://www.muse.demon.co.uk/3daudio.html> Ambisonics by Richard Furse  
[http://members.tripod.com/martin\\_leese/Ambisonic](http://members.tripod.com/martin_leese/Ambisonic) Ambisonics Info-Page by Martin Leese

## DECODING

In the decoding process however these encoded files are assigned to the speakers in proportion to the chosen speaker layout: For each speaker a precise proportion of each spatial encoded direction is assigned, according to its position in the soundfield. Any symmetrical speaker layout may be chosen from 1 to N speakers, whereas the clearness of the spatial reproduction improves, the more speakers are available. Also the vector -channels are combined at a certain ratio with the omnidirectional W.

Decoding can be applied on the fly right after encoding, no encoded files need to be written then. The more usual way however is to write 9 encoded Files and decode them to derive the files to feed the speakers. Ofcourse, with sufficient CPU-power, it is possible to decode some prefabricated encoded files also in real-time.

The First and Second Order Ambisonic Decoding Equations by Dave Malham and Richard Furse are here: <http://www.muse.demon.co.uk/ref/speakers.html>.

# **SPATIALISATION WITH SET OF THE CSOUND SPATIALISING INSTRUMENTS**

This set of Csound instruments spatialises 20 soundsources independent of each other in 3d space by using 2<sup>nd</sup> order Ambisonics and combining this with distance clues. There are three main facilities in this collection of instruments:

1. Creation of a trajectory for the movement of each of the 20 sources in 3d space
2. Combination of the soundfile with distance clues and creation of a sonic environment
3. Spatialisation of the sound, the reverberation and the early reflections using the Ambisonic equations

Important to know here is that Ambisonics is not capable in creating distance information. Ambisonics is only about the direction of a sound. All sounds are projected onto the surface unit-sphere, the sphere enclosed by the speaker rig, if not enhanced by distance information.

Fortunately distance clues can be well combined with Ambisonics. For some distance clues like e.g. the pattern of early reflections it is even essential, that they get combined with a method like Ambisonics, so they get spread out in different perifonic directions.

For directional information: Ambisonics

For distance information: Distance clues (partly ambisonically distributed)

This separation of tasks has to be pointed out very clearly.

## **TRAJECTORIES**

Task is to create a set of trajectories each containing the angle (A), the elevation (E) and the distance (D) of a sound at any particular moment. That way a position of the Sound at any moment in full 3D space is defined. It does not matter if for this description Polar co-ordinates or Cartesian co-ordinates are taken, they may even be converted into each other. Even a subsequent combination of both is possible: parts of the whole set of trajectories may be described with polar co-ordinates, other parts with Cartesian ones. What never should happen though is, that one trajectory ends at some location while the next trajectory, describing the path of the sound, starts somewhere else: smooth transitions have to be kept, otherwise clicks and pops may occur.

Its starting and ending point and the time it shall take to go from one point to the next define a single trajectory. All these starting and ending- points are linked with a function with sufficient resolution to provide smooth transition. Functions may be straight lines or exponentials as well as any other function desired. This function may be afterwards combined with randomisation or a modulation of one or more oscillations, so that complex ways of movement are generated.

## ENCODING DEPTH

These are the most common distance clues:

1. Attenuation
2. Global Reverb
3. Local Reverb
4. Filtering
5. Early Reflections

### Attenuation

Most important is the fact that the amplitude of a source is in relation to its distance. So distant sounds should get attenuated with the factor  $1/D$ <sup>1</sup> (D= distance in Units, 1 Unit is the distance from the zero-point of the coordinate system to the location of the speaker) Closer sounds increase in amplitude. A sound right on the unit sphere gets the attenuation- factor of one, sounds on other locations get multiplied by  $1/D$ . The opcode "locsigs" in Csound does this by default. Note that the sounds become nearly infinitely loud as they get close to the origin of the coordinate- system: either sounds have to be prevented of getting that close (limitation of D) or the resulting amplitude has to be limited. Note that the sound is related to distance in such a way that there is no longer an absolute amplitude of a sound: The composition changes with the position of the sound.

### Global Reverb

Another important clue is the ratio of the reverberated part of the sound energy and the dry sound itself. As a sound moves into the distance, the reverb should not get that attenuated as the direct sound. The result is an almost dry sound close to the listener and an increasingly reverberant as it moves away. The model of Chowning may be applied, setting the ratio reverberant sound to  $1/\sqrt{D}$  as the direct sound decreases with  $1/D$ .<sup>2</sup>

### Local Reverb

In the model of John Chowning used here the reverberant energy again is split up into global and local reverb. Whereas the global energy comes from several directions (several reverb-units with slightly different parameters), the local reverb is coupled with the source's position. Their ratio is expressed as  $1/D$  for global reverb and  $1-1/D$  for local reverb. This leads to a more directional amount of reverb from the sources position for distant sounds.

### Filtering

Onto more distant sounds a lowpass filter should be applied. This simulates the loss of high- frequency due to long travelling times through the air. The parameters of the filter should be coupled dynamically to distance (D).

### Pattern of Early Reflections

The perception of the spatial image is increased with the appliance of a dynamic pattern of early reflections. Four virtual walls, ceiling and floor are calculated at a certain distance where the sound energy is reflected. These early reflections approach the listener slightly delayed to the original signal and from different directions. These directions change according to the movement of the source. Also the amplitude changes

---

<sup>1</sup> Charles Dodge, Thomas A. Jerse: Computer Music (Second Edition), chapter 10.2C, p. 314

<sup>2</sup> Charles Dodge, Thomas A. Jerse: Computer Music (Second Edition), chapter 10.2F, p. 319

in respect to the distance, the virtual reflected sound has travelled. According to David Griesinger<sup>3</sup> the parameters should be chosen in that way that the reflections approach in a time- window of 20-50 ms<sup>4</sup> after the source. Two kinds of reflections are calculated here: Specular and diffuse reflections, as the laws of reflection are different. Ofcourse the early reflections are also spatially distributed to simulate their natural direction. The result is a lot more transparency and an increased image of depth aswell an enhanced perception of distance.

### ENCODING MOVEMENT

In the Csound- spatializing .orc the Doppler-shift of frequency has been coded using the formula  $f' = f(c/c-v)$ . Whereas  $f'$  is the resulting frequency,  $c$  is the speed of sound propagation through the air (345m/s) and  $v$  is the speed of the source related to the listener. Note that the sound is related to velocity in such a way that there is no longer an absolute pitch of a sound: the composition changes with the movement of the sound.

### GENERAL (ZAK PATCH SYSTEM)

Please note that the processing of the sound in terms of spatialisation, depth modelling etc. is spread among several instruments within the Csound orchestra. The sound is passed among the instrument as an a-rate- variable, whereas the A (angle), E (elevation) and D (distance) positions are passed a k-rate variables, both via Zak-Pach- System.

### DECODING

The decoding is done in a separate .orc. There are some speaker setups already coded, but they may be completed due to personal needs. Usually there has to be done a test-run with the amplitude factor girescale set to 1. In the output window the maximum amplitudes of every generated soundfile will appear after finishing. The appropriate amplitude factor can be calculated then by dividing 32768 by the maximum output. This factor will differ everytime regarding the used input and the selected speaker setup. If the "soundout"-opcode had been chosen for output, the 8-bit output files however have to be converted into 16 bit files and a header has to be written. This can be done for example with the programme Soundhack (freeware, but mac only: <http://www.soundhack.com/> ). The Files then have to be fed for example into an editing programme to be able to play them simultaneously.

Copyright by Jan Jacob Hofmann For any comments, hints, feedback, bug-reports and questions please contact [jjh@sonicarchitecture.de](mailto:jjh@sonicarchitecture.de) .

[www.sonicarchitecture.de](http://www.sonicarchitecture.de)

---

<sup>3</sup> David Griesinger: „The Psychoacoustics of Listening Area, Depth, and Envelopment in Surround Recordings and their relationship to Microphone Technique“, Session- Paper of the 19 AES Conference, June 2001

<sup>4</sup> „The Theory and Practice of Perceptual Modeling - How to use Electronic Reverberation to Add Depth and Envelopment Without Reducing Clarity“ downloadable at D. Griesingers Pages <http://www.world.std.com/~griesngr>