

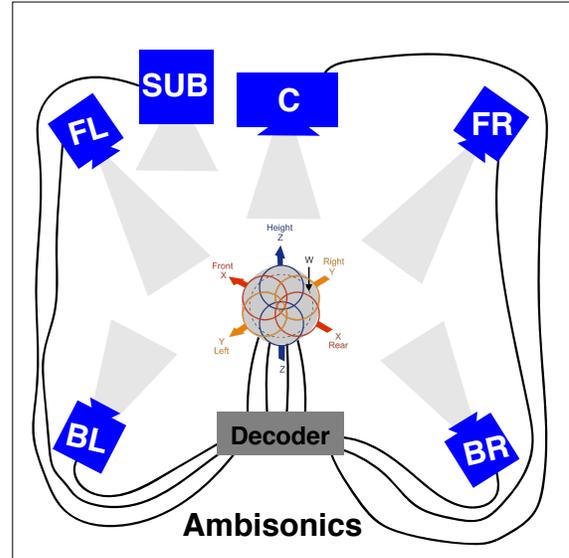
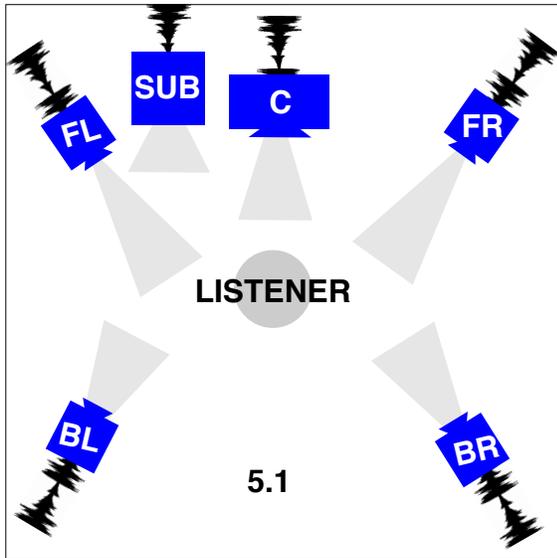
Luís Zanforlin  
May 5, 2018  
Music Acoustics  
Berklee College of Music

## **Synthetic Reverberation Spatial Ambisonic Sound for 360 Video, SFX and Music**

This paper explores existing methods of creating spatial ambisonic synthetic reverberation for 360 video and music using a DAW. This document also proposes a new method of generating empirically-driven spatialized reverberation that can be used in both 360 video as well as agency based experiences such as virtual reality games.

### **Ambisonics:**

Spatialized sound has existed since 1931 through the use of stereophonics, also known as stereo, a technique which involves creating two recordings corresponding to the left and right fields which are reproduced from speakers positioned in the corresponding sides of the listener. Since then the movie industry introduced standards for surround sound which involves the addition of many speakers such as 5.1 and 7.1 in order to surround the listener's sound field creating a more immersive and cinematic experience. The techniques used in surround sound works with the same principle as stereophonics, in order to generate a 5.1 surround delivery, it is necessary to generate six distinct recordings. The reproduction of 5.1 surround sound depends on a play room in which six speakers (five spatial speakers and one subwoofer) are positioned according to the specifications of the recordings. A disadvantage of this technique is that, if you have a 7.1 room and wish to reproduce a 5.1 recording, you are limited to listening in 5.1 unless you are able to generate two extra recordings representing the side speakers.



Ambisonic audio is a 1975 spatialized sound technique which involves generating audio recordings representing the spherical sound field from the perspective of the listener which can later be decoded into corresponding speakers positioned in a room. A different way to make sense of ambisonics is to characterize stereo and surround sound as generating and placing sounds where their sources would be relative to the listener, ambisonics consists in generating sounds as the listener should hear them and extrapolating (decoding) their position by measuring the differences between each field (Left, Right, Up, Down) in order to generate the sounds that need to come out of positioned speakers.

The advantage of ambisonics is that it is possible to reproduce the project in any speaker configuration with the same four or eight (depending on the ambisonic order) audio tracks, taking full advantage of their spatial position. In order to reproduce ambisonic sound it is necessary for the reproduction room to have a decoder containing the position of the room's speakers or headphones orientation. Although this technique requires a limited number of audio channels compared to surround sound, it requires more computational power.

## B format:

The standard for ambisonic sound reproduction is called **B-format** which includes four channels **W** (omni directional), **X** (front and rear), **Y** (left and right) and **Z** (up and down). When recording ambisonic sound using a sound field microphone it is important to keep the microphone capsules as close to each other as possible in order to prevent phasing, this is accomplished by recording in **A-format** which consists of four capsules positioned as FLU (front left up), FRD (front right down), BLD (back left down), BRU (back right up). **A-format** then is encoded into **B-format** by adding and subtracting the individual channels: **W** (FLU+FRD+BLD+BRU), **X** (FLU+FRD-BLD-BLU), **Y** (FLU-FRD+BLD-BLU) and **Z** (FLU-FRD-BLD+BLU). In order to decode **B-format** to be played in a speaker or headphone system a decoder will use similar calculations to determine what information to send to each individual speaker based on the predetermined speaker locations in a room or the gyroscopic information for the headphone position obtained either from a cellphone, headset or 360 video positions. In the case of headphone reproductions, it is also possible to add HRTF filters in order to enhance the immersive experience.

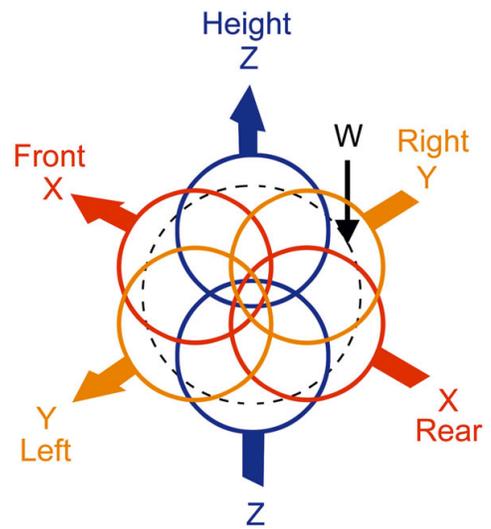


Figure 1 creativefieldrecording.com

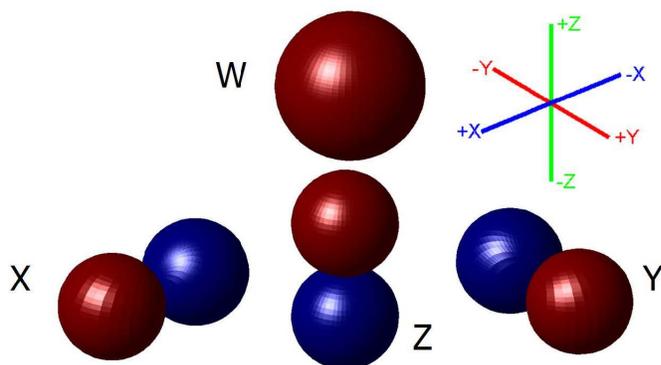


Figure 2 University of Derby

disadvantages. For 360 video applications I will focus on two methods compatible with the practicalities needed for video workflows by being able to be done digitally inside of a DAW;

## Reverb types:

There are many methods of generating artificial reverberation such as Reverberation chambers, plates, springs, algorithmic reverbs, raytraced models and convolution reverbs, however when working with spatial sound these methods have different advantages and

these include algorithmic and convolution reverbs. It is important the DAW supports multichannel tracks going up to eight channels and provides the user the ability to rout individual channels of a track to specified channels of other tracks. Not all DAWs offer multichannel capabilities beyond two channels which is why most 360 audio projects are done in Pro Tools Ultimate or Reason.

### Algorithmic Reverb:

Algorithmic reverb can be added to spatialized ambisonic audio by simply converting **B-format** to **A-Format**, and sending the channel pairs **LB**, **LF** and **RF** and **RB** to stereo aux tracks filtered by identical stereo algorithmic reverb plug-ins which are then combined back into an **A-format** four channel reverb track that can be encoded into **B-format** and added to the **B-format** final mix.

The benefit of this technique is that it can be done using commonly available stereo plug-ins and it requires little processing power. The technique however lacks spatial definition and will always produce the sensation of being the center of a room with evenly spaced walls.

### Convolution Reverb:

Convolution reverbs are derived from impulse responses recorded inside a room in order to capture the room's true reflections. This technique results in an accurate representation of the room and it is the preferred method for creating realistic sounding synthetic reverberation.

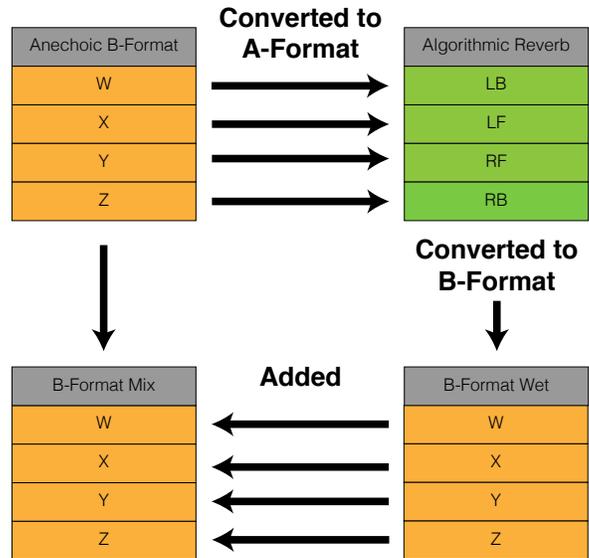
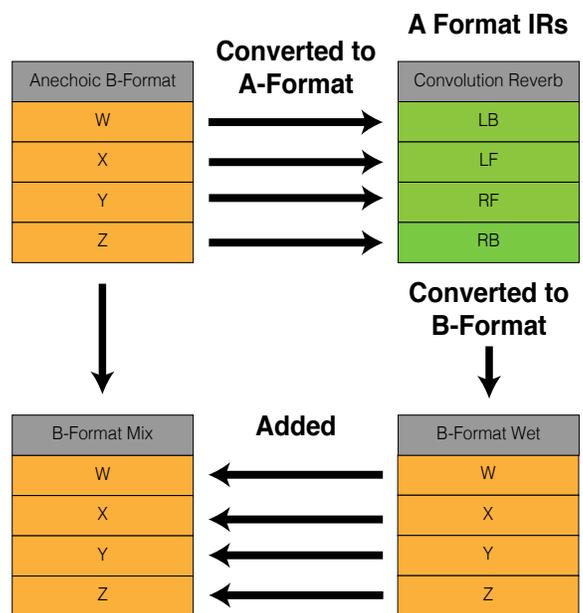


Figure 3 Ambisonic Algorithmic Reverb



Employing convolution reverb in ambisonic spatialized sound can be done the same way as with algorithmic reverbs, however this method has the same downsides as using an algorithmic reverb which defeats the purpose of using convolution reverbs intended to deliver a more realistic representation of a space. A better way to employ convolution reverb to ambisonic sound is to record impulses using a sound field microphone in **A-format**. The anechoic audio is then converted to **A-format**, sent to a four-channel bus where a four-channel convolution reverb plug-in uses the sound field impulse response to deliver a wet **A-format** track. The wet **A-format** track is then converted to **B-format** and added to the **B-format** final mix.

Figure 4 Ambisonic A-format Convolution Reverb

Although this technique delivers an empirically derived spatialized ambisonic reverberation of a space it only accurately simulates the reverberation of a sound source placed in the same exact location where the impulse was generated relative to the listener. This is a problem if we assume the intention behind using spatialized audio is to place sources in multiple locations. In order to combat this issue, without generating an impulse for every source audio location in a particular project, a technique showcased by Dr. Bruce Wiggins at the 2017 *Sounds in Space* seminar can be used.

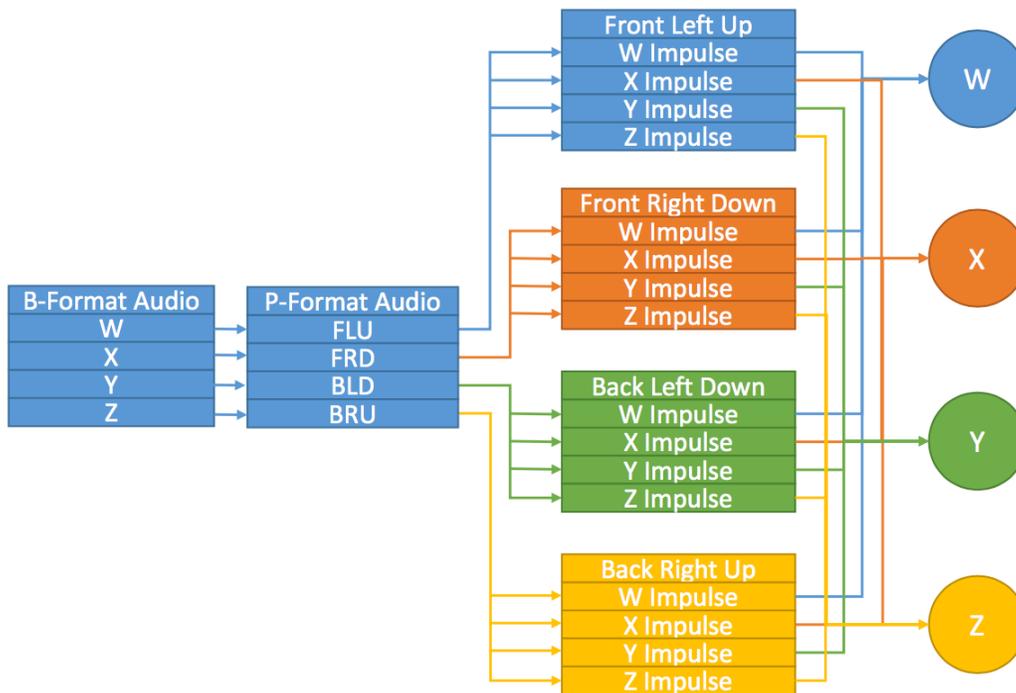


Figure 5 5 Ambisonic P-Format Convolution Reverb (University of Derby)

This technique involves recording impulse responses using a sound field microphone in the middle of a room recording four consecutive impulses from speakers placed on the room's extremities in **P-format** configuration. **P-format** (Spatial PCM Sampling) uses **FLU** (front, left up), **FRD** (front, right down), **BLD** (back, left down) and **BRU** (back, right, up) corresponding to the impulse response's speaker positions. In this technique the anechoic **A-format** sound is encoded to **P format**:  $FLU = (W+X+Y+Z)/2$ ,  $FRD = (W+X-Y-Z)/2$ ,  $BLD = (W-X+Y-Z)/2$  and  $BRU = (W-X-Y+Z)/2$  and each of the four encoded channels are then sent to a four-channel bus with a multi-channel convolution reverb plug-in containing the corresponding sound field impulse response. The four channel buses are then converted back into **B-format** where they are added to the **B-format** final mix.

This technique delivers an empirically derived reverberation track that changes depending on the position of the sound source resulting in realistic and spatialized sounding reverberant mix. Compromises for this technique includes the lack of available four channel convolution reverb plug-ins, the need to record multiple impulse responses generating a total of 16 files per room and the computationally demanding task of processing 16 convolution reverbs. When using this technique accuracy decreases as the order of ambisonic (spatial resolution) increases. As a result, audio sources originating from areas between the original impulse responses will return a combination of the two closest impulse response recordings generating spatially inaccurate early reflections such as the **A-format** convolution reverb method.

### **Proposed Method:**

A more reliable alternative to Bruce's method is to use ray tracing for the early reflections and a single four channel ambisonic convolution reverb for the reverberation tail. Ray tracing is a technique that involves modeling the geometry of the original environment and simulating the reflections bouncing on the walls of the model in order to calculate the angle of incidence of each reflection and the distance traveled. The simulated data can then inform the delay time, volume and imaging needed to generate each reflection. The problem with ray tracing is its high computational cost which prevents real time playback on consumer computers. For that reason, a combination of convolution reverb and ray tracing is an ideal compromise between fidelity and processing demand.

In order to create a plug-in that uses the combined ray trace method, it is necessary to capture ambisonic impulse responses as well as 3D geometric data of the recreated environment (which can be done by imaging or by modeling a floor plan). The ray tracing element of the plug-in is then incorporated into the spatializer plug-in (which converts monophonic sound sources into ambisonic) where it will generate the early reflections. The late reflections are then created by a convolution reverb plug-in placed on the ambisonic track at the end of the signal flow.

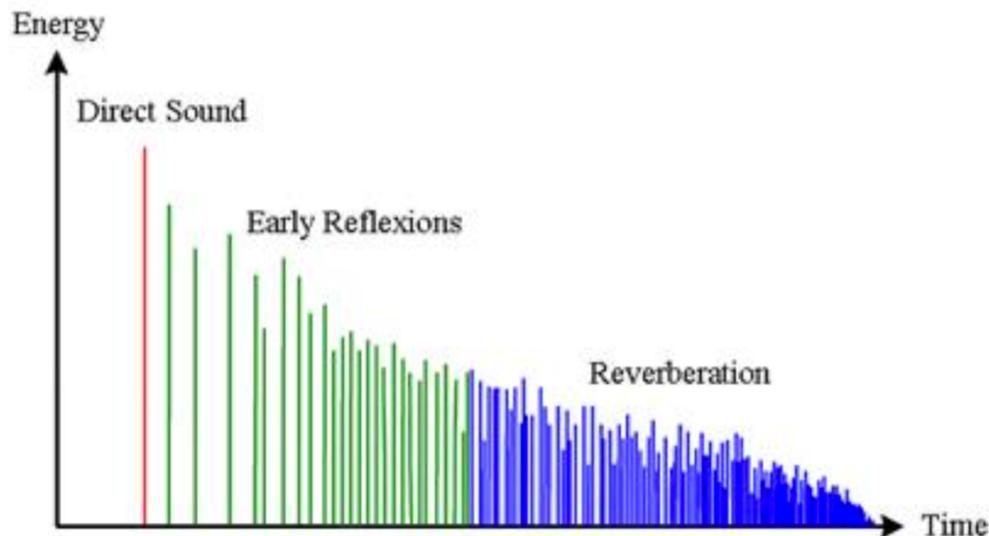


Figure 6 impulse response

Since the early reflections are the main indicators of directional localization used by our ears, there is no need to process the chaotic late reflections that compose the reverberation tail. The number of reflections could be adjusted by delaying the convolution reverb start and trimming the beginning of the impulse response recording. This way the user can improve the fidelity of the reverberation depending on their computer's processing power (helpful in virtual reality applications). The ray tracing/convolution technique is ideal for agency-based experiences such as virtual reality games because the game engine can easily inform the location of each sound source as well as the geometry and materials of all interacting objects and walls. The game engine could also change the number of reflections calculated in order to increase availability of computational power for rendering the game's graphics in real time.

## Conclusion:

There are many ways to generate synthetic reverberation for ambisonic spatialized sound inside of a DAW. In order to choose the right technique, it is essential to evaluate the project's need for realistic reverberation, spatialized reverberation, the availability of **A** and **P-format** sound field impulse responses and the processing power of the computer used to generate the synthetic reverb. **A-format** algorithmic reverbs deliver a computationally lightweight way to add reverb to anechoic sounds with little spatialized fidelity. **A-format** convolution reverb delivers a realistic sounding reverb for a fixed position anechoic source while **P-format** convolution reverb delivers realistic spatialized reverb that depends both on the availability of **P-format** sound field impulse responses and a fast processing unit. A combination of early reflection ray tracing and **A-format** convolution reverb delivers the most realistic synthetic reverberation while requiring 3D data and **A-format** impulse responses, best suited for virtual reality and agency-based applications.

## Citation:

Wiggins, Dr Bruce. "Measured Reverbs for Ambisonics and VR (Convolution Reverbs for Ambisonics)." *Https://Www.brucewiggins.co.uk*, University of Derby, 17 Aug. 2017, <[www.brucewiggins.co.uk/wp-content/uploads/2017/07/01 Wiggins - Ambisonic Convolution Reverbs.pdf](http://www.brucewiggins.co.uk/wp-content/uploads/2017/07/01_Wiggins_-_Ambisonic_Convolution_Reverbs.pdf)> Accessed 5 May 2018

Yeary, Jay. "Ambisonics B-format for immersive audio." *TV Technology*, Dec. 2015, p. 30+. *Business Collection*, <[http://link.galegroup.com/apps/doc/A441492716/ITBC?u=milin\\_b\\_berklee&sid=ITBC&xid=1f0c4fe4](http://link.galegroup.com/apps/doc/A441492716/ITBC?u=milin_b_berklee&sid=ITBC&xid=1f0c4fe4)> Accessed 5 May 2018.

Pavlov, Antonio. "Spatial audio for spherical video." *Videomaker*, Aug. 2017, p. 60+. *General Reference Center GOLD*, <[http://link.galegroup.com/apps/doc/A500260548/GRGM?u=milin\\_b\\_berklee&sid=GRGM&xid=37a14024](http://link.galegroup.com/apps/doc/A500260548/GRGM?u=milin_b_berklee&sid=GRGM&xid=37a14024)> Accessed 5 May 2018.