ECHO

a journal of music, thought and technology

ORPHEUS
INSTITUUT

Title:	**IrSOb – The Iridescent Sound Object**
Subtitle:	A Model and Architecture for 3D Sound Object Creation and Binaural Reproduction with Standalone Hardware and in VR Spaces.
Author(s):	Philippe Druez
Issue:	<write here>
Publication date:	
Review status:	
License:	CC BY-NC-ND 4.0
Article DOI:	

On The Poetics of Space…

'The poetic image is a sudden salience on the surface of the psyche' and has to be a quest based not oncausality, but in 'reverberation' and 'in this reverberation, the poetic image will have a sonority of being'.

Gaston Bachelar

Abstract

IrSOb (Iridescent Sound Object) proposes a theoretical, object-oriented model and modus operandi for capturing real-world 3D sounds and constructing synthetic 3D virtual-reality sound objects. A proof-of-concept implementation enables participants to experience these sound objects either within a wavefield spatial-audio laboratory or at any location using a dedicated hardware system for binaural reproduction over headphones.

Due to its distinctive characteristics, this new method offers added value compared to current techniques in 3D gaming, sound art, performance, theatre, and scientific research involving embodiment. It also provides a valuable tool for the realistic recreation and archival preservation of musical instruments and a wide range of sounding objects. Thanks to its simplicity and coherent workflow, the method opens new creative possibilities for sound artists and composers working with experimental music.

The Iridescent Sound Object is part of my research at the eXtended Reality and Human Interaction Labs at IPEM (Institute for Psychoacoustics and Electronic Music), within the Systemic Musicology Section of the Department of Art, Music and Theatre Studies at Ghent University (UGent), and at KASK School of Arts and Conservatory Ghent.

Introduction

The use of spatial sound in the gaming and the movie industry is already well established. Major entertainment players made important progress in delivering spatial audio for cinema or home theatres and into the headphones of the consumer or even the museum visitor. The result delivered by these immersive experience systems can be quite impressive and are available from a wide range of manufacturers[1].

Figure 1. Experiencing the The Heidelberg Press Sound Project at Orpheus Institute.

This specific part of my PhD resulted in a method for acquiring faithful 3D sound captures or create non-existing VR sound imagery; a method for recording the 3D sound object; a Max for Live device for reproducing the 3D sound object in a 3D lab environment using Qualisys MoCap[2] (XRHIL in this case) and *Earis*, a hardware system specifically developed and built for this purpose, using binaural headphone technology and designed to be used anywhere.

This article covers the theoretical concept and method of IrSOb and its relation to Ambisonics; an overview of related 3D recording technologies; the architecture of the Earis hardware and its intrinsic iRose spatial calibration system and briefly, *The Heidelberg Press Sound Project* (discussed further) as use case as demo at Orpheus (Figure 1).

IrSOb, the iridescent sound object's methodology is not to be confused with the *object-based* concept where audio designers place a sound as a "point" in 3D space (instead of *channel-based* by assigning a sound to a specific speaker)[3], The Iridescent Sound Object is a VR 3D sound object in itself that has a spatial volume with “tangible” boundaries and can be placed in a VR 3D sound space.

Unlike omnidirectional point sounds (some treated with equalising for better orientation perception), IrSOb offers you how you hear a sounding object in real life as VR experience. These sound objects sound different when listened to from different angles. Additionally it will reflect different sounds to its “environment” and thus create a much more accurate and lively reflecting and reverb ambient experience (future extension). The development started at the XRHIL wavefield lab of UGent (IPEM) but is technology independent. As a proof of concept I developed a basic device to experience the IrSOb everywhere.

---

[1] Dolby Atmos surround sound, XPERI’s DTS Headphone:X, Creative Labs ’Super X-Fi, Windows 10 Sonic, Apple Spatial Audio, Sony 360 Reality Audio (WalkMix) and Waves Nx (combining room emulation and motion tracking)

[2] Other providers of motion capture devices and software like OptiTrack, Rokoko and Move are equally usable.

[3] DAWs (also called SAWs “spatial audio workstations") like Steinberg Nuendo / Cubase and DaVinci Resolve / Fairlight and others…

How Did I Get Here…

When visiting the Immersive Lab at AP Hogeschool Antwerpen I saw a volumetric capture device of the photogrammetry studio and was thinking of replacing the cameras with microphones to capture a "real" 3D sound image of an object inside it. This could be a person singing, a violin playing, robins in a cage or a Bernina

Figure 2. L'image de Mon Père. The Sound Frame producing a 2D sound image.
Insert: The four-mic recording setup.

sewing machine, but nothing larger. Besides the complex and not so portable structure, how to define the number and placement of (and what kind of) microphones in what could be called a "sonogrammetry" device and make it scalable. Capturing a real 3D sound image of a violin performance looked feasible to me but that of ants at work in their nest or a Formula One Ferrari seemed as significant to me and needed a different approach.

I have the unfortunate tendency of transmogrifying any tiny idea into octopus-monsters with dozens of spin-off tentacles and deep holistic and futuristic visions. But the lack of a large studio, a team working for me and the requisite funding reduced this project to a workable field resulting in what became the IrSOb project. That said, the project itself is a bit of a spinoff of my initial PhD research and even more, the hardware developed is a side dish as proof of concept of this spin-off. So began my groundwork on how to capture any object outside a lab environment (portable) and with a minimum of construction hassle. This resulted in *the iridescent sound object* model on which I will throw light in this paper.

Before I started my research on 3D sound in 2019, I created a sound-art work with the title *l'Image de Mon Père (2017).* The context is not so relevant here but the work is what I call a sound frame. It's a metal frame the size of a large painting hanging 30cm from the wall. The rear side contains speakers, reproducing a two-dimensional auditory image (in loop) of the process of drawing on paper (Figure 2). I wasn't really aware of spatial sound capturing at that moment and empirically designed this system, which functioned as desired.

This artwork was the beginning of my journey into spatial sound exploration, resulting in installations (*Radison – Kindergarten, Neuron – Choral Flow*, …) to Ambisonics concerts (*Ethereal Threads, Nanotonal Ascension for*

24 Voices, ...) and VR compositions (*Polytox, Boredom,* ...), concepts like that of the *iPixel* and the idea to produce *AcouStasis*, a wavefield wallpaper containing arrays with hundreds of electrostatic speakers to create instant acoustical emulations, eliminate noise or play the spatial sound of a rainforest in any room (Figure 3).

A Walk Through Sound Fields

With a background in 3D modelling and animation I soon realised the unexplored possibilities of the 3D sound experiments and compositions I heard at the Wavefield lab at IPEM. I saw opportunities in adopting and extrapolating 3D modelling and animation techniques from visual practice to the auditory domain. I quickly got stuck on the basics: what is a 3D sound object —in VR terms—? How big can it be, what shape can it have (Figure 4), does it has a surface, what's inside it…? And even further: besides being volumetric, could it be fog-like, a particle system or NURBS-surface and what about reflection and absorption of multiple sound objects (including walls and people) in a space, (trans)-morphing sound objects and more…

Taming the octopus was key and I started by studying 3D recording techniques. Almost all of them are space-oriented (covered in the next chapter) and one of them caught my attention: the Soundfield microphone invented in 1978 by Michael Gerzon and Peter Craven. Their theoretical design was applied into a microphone system by Calrec Audio Limited, who launched the Soundfield microphone. It has four closely spaced cardioid microphone capsules arranged in a tetrahedron, facing outward. A full-sphere surround sound format was later defined by a group of researchers in the UK and named *Ambisonics*[4].

Figure 4. Possible shapes of 3D sound objects.

The number and types of microphones are defined for different "resolutions", known as first order Ambisonics (FOA) when four microphones are used or higher order Ambisonics (HOA: 2nd, 3rd, …) when additional microphones are added. This brought me back to the *sonogrammetry* device. Why not use the tetrahedron model of a Soundfield microphone and invert it, changing the POV (point of listening) on the "other side" of the wavefield. Space-oriented becomes object-oriented (Figure 8). The maths was already there. The notation of this "inverted" FOA wavefield capture becomes ~FOA.

[4] There is a comprehensive work by Franz Zotter and Matthias Frank on Ambisonics (see references) for those who want to dig deeper into the matter.

A Brief Overview of Research on 3D Sound Recording

Common more complex (i.e. not mono or stereo) techniques for 3D recording using multiple microphones directed to the sound source are mostly orchestral or multi-instrument (ensemble) setups. They may look similar to IrSOb but are space oriented recordings.

The *22.2 multichannel microphone setup* (Will Howie et al. 2016) is an example of such a very complex setup with discrete channels and different flavours of omni- and cardioid microphones. A wide range of complex setups can be found for surround recordings among them the popular *ORTF-3D* using 4 cardioid and 4 super-cardioid mics, the *Hamasaki Cube* with four figure-of-eight microphones, the OCT-3D 2L-Cube and the *Decca Cuboid to name some.*

The 42-channel spherical microphone array is a human sized icosahedron for measuring acoustic violin properties (National Institute of Technology, Japan — Figure 5).

And then there is Markus Noisternig, Franz Zotter, et al. with their surrounding 64-channel spherical microphone array. It's set up at IEM for acoustic research in a semi-anechoic chamber (Figure 6).

Antonio Canclini et al. at the Politecnico di Milano describe a method for the estimation of the three-dimensional radiation pattern of violins, during the performance of a musician. A plenacoustic microphone array captures the energy radiated by the violin in different directions using beamforming based on sub-arrays (Figure 7).

All methods mentioned above (there are many more) have their own specific advantages and uses. The more complex ones are only possible in lab or equipped stage environments. Though the ideal IrSOb recording would also be in an anechoic room for accurate archiving purposes, it offers proficient results for the reproduction in a 3D VR experience in any environment (also outside).

Figure 5. An isocahedron setup in an in an anechoic chamber.

Figure 6. Surrounding 64-channel spherical microphone array.

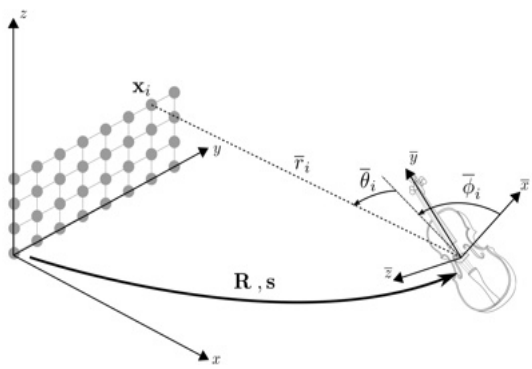

Figure 7. Plenacoustic analysis.

IrSOb — The Iridescent Sound Object

When recording with a SoundField microphone, the direct output from the four cardioid microphone capsules is called the A-format (figure 8, left)[5]. The Ambisonics B-format is the standard audio format produced from the A-format after the Soundfield calculation, consisting of W (pressure signal), X (front-back information), Y (left-right) and Z (up-down information). The capsules are closely positioned in a tetrahedron pointing outward and individually identified as FLU (front left up), FRD (front right down), BLD (back left down) and BRU (back right up). In itself, the signals are unusable since they only contain phase-related information instead of channel based information.

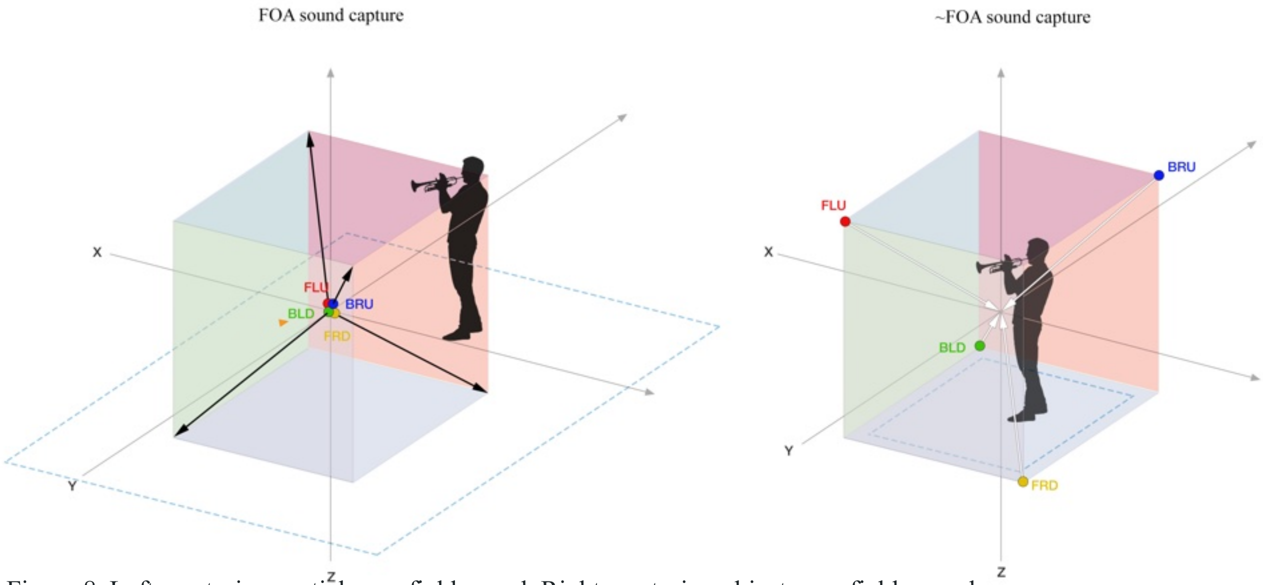

Figure 8. Left: capturing spatial wavefield sound. Right: capturing object wavefield sound.

For the IrSOb model, this design is inverted. The four capsules of the capturing apparatus aren't at the centre anymore but have been moved to a sphere around the point of interest (object K), now at the centre. The cardioid microphones pointing inward are still arranged as a precise tetrahedron to avoid phase shift and to eliminate time of arrival calculation. The virtual listener (participant P) can be anywhere around (or in) the virtual 3D object. The ~A-format (Figure 8, right) allows for the calculation/reconstruction of the 3D sound of the object in the VR space.

The IrSOb model has three zones (Figure 9), each having a set of four sound channels maximum:

- the O-zone (object zone): 4 channels to contain the direct object sound (the ~A-format);
- the A-zone (ambient zone): 1, 2 or 4 channels for the reflected space sound (reverb);
- the I-zone (inside-zone): 1, 2 or 4 channels for the object's internal sound.

[5] Some figures may show inverted as a macron diacritic above the letters instead of prepended by a tilde but they both denote "inverted".

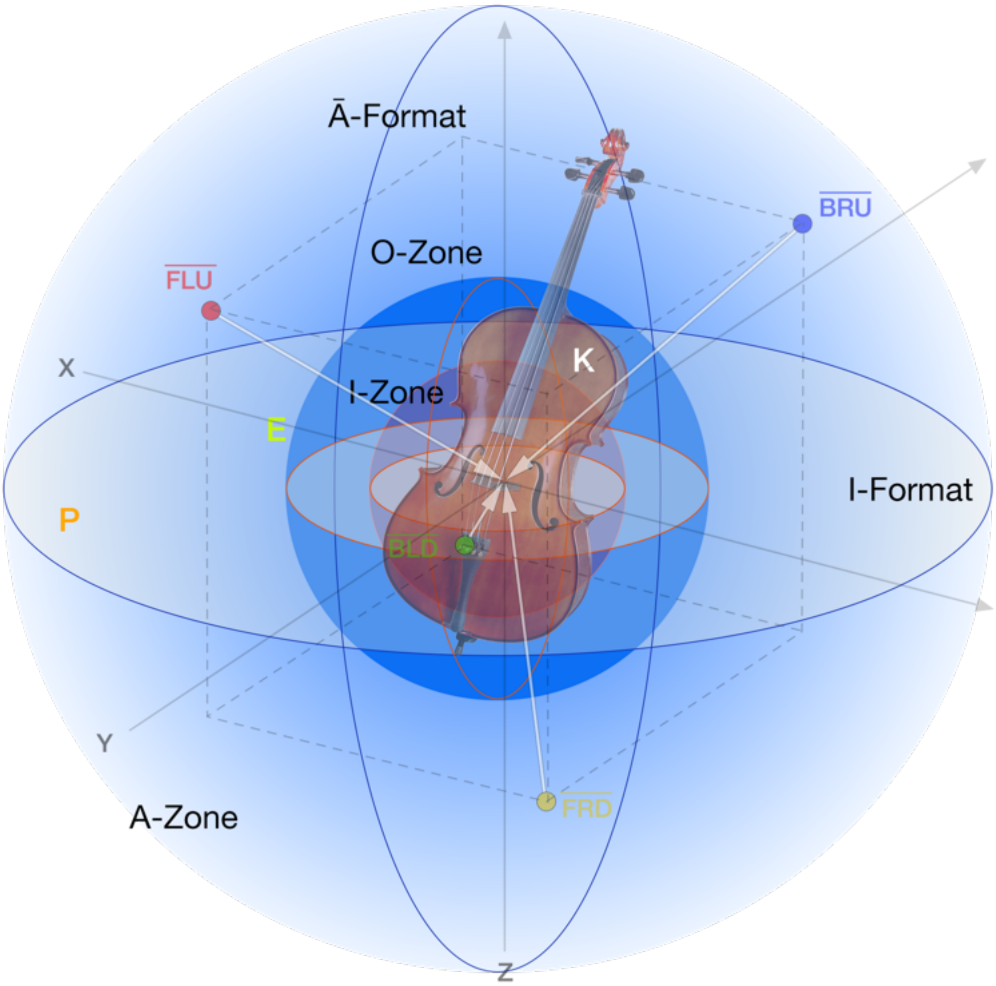

Figure 9. Representation of the inverted tetrahedron setup and the three zones of the IrSOb concept. K: sound object, P: participant, E: Event Horizon.

Two Prototypes

I developed two IrSOb proof-of-concept systems. The first at IPEM's ASIL[6] 3D VR sound lab as an E4L[7] and Max for Live device that I wrote and the second was *Earis,* the extra-lab binaural headset device.

[6] ASIL: the Art & Science Interaction Lab offers an acoustically treated black-box with Qualisys Motion Capture and an 80 speaker audio system for spatial & immersive audio.

[7] E4L (Envelop for Live) is an open-source spatial audio production toolkit allowing artists to perform, produce and explore DIY 3D VR sound projects. E4L operates within Ableton Live Suite.

Both systems translate the ~A-format into a two-dimensional horizontal plane (the I-format),resulting in a 360º radial sound image of the object experienced in a headphone.

The 3D sound image of the object is called the ~B format. A two-channel orthogonal raw left/right signal is derived from the I-format and further processed in function of the location and orientation of participant P regarding to the virtual source K. This processing results in the O-zone and includes:

- Sound level and frequency filtering (psychoacoustic adjustment[8]) in function of the angular position for the left and right ear of the participant regarding to the VR sound object (a basic binaural filter);
- the stereo separation (field width) and volume level in function of the distance to the VR object (closer to object results in wider *stereo* field);
- the balance of direct (O-Zone) and ambient (or reflected, reverb – A-zone) signal in function of distance and head rotation.

[8] No real psychoacoustic or phase related adjustments are done for now. A true HRTF filter could be applied in a later stage. A deeper understanding of binaural listening can be read in the work of Jens Blauert's The Technology of Binaural Listening.

The source of the reverb signal can either be simulated (e.g. the Convolution Reverb in Envelop for Live) or recorded by an extra set of microphones (same axis of the direct mics, pointing 180º away from the object) to create the A-zone. Obviously, the use of a Soundfield mic is difficult if not impossible since it should be placed in the exact centre of the tetrahedron where the object is[9]. The use of four extra mics results in a space oriented reverb image, that will equally change in function of position and listening angle. It can be treated very similar as

Figure 10. First test recording setup. CW from left: the sound object is temporarily replaced by
an Ambisonics mic recording the "room tone"; the sound object with distinct front/rear

a B-format.

The Event Horizon E, is the spherical border between the O-Zone and the A-Zone (Figure 9). When crossed, the participant “enters” the object. Since the IrSOb sound object has a (virtual) volume, a third set of mics can capture sound inside the object (the I-zone) and will vary according to the object or desired effect. A single mic in a cello could suffice, contact mics on a skull, a Soundfield mic in a car, a hydrophone in an aquarium… The use of MEMs mics can extend this interior sound experience considerably. The content (as with the whole IrSOb object) of this peculiar zone can be anything from a real recording to a fantasised inner world sound montage (in 3D).

[9] An off-centre recording is acceptable in some cases and/or a mathematical correction is possible.

The IrSOb model is technology independent. The first test started with recording a sound source with distinct spatial sound radiation (Figure 10). Then followed the development at the ASIL lab. I patched a MAX for Live device (Figure 11) that connects to the lab's HybridEngine M4L device in order to get the MoCap data (position in the space and head rotation) of a participant and calculate the binaural signal related to the virtual object positioned in the lab VR space. The processed signal was then transmitted from the central server over analog UHF to the headphone of the participant. The (quick and incomplete) overview (Table 1) below gives a idea of the difference between a lab environment and the developed Earis hardware experience (both using the same IrSOb model).

Figure 11. The IrSOb M4L device is connected to the HybridEngine at IPEM.

One advantage of the IrSOb concept is the independent recording (or synthesising) and re production technique. One can record sound in 3D in a different order than that of the reproduction system, making it flexible and future proof.

Property	Lab	Earis
Number of simultaneous users	At the time of testing: two. Restricted to central processing, MoCap simultaneous users and wireless transmission channels. (Qualisys max. 10)	As stated by the manufacturer: up to 8 beacons and 64 users.
Number of simultaneous IrSOb objects (not tested)	Depends on number of users (of the MoCap and the central processing of the data per user to be streamed per user.	Depends on the participant's processing unit (CHAP) and data communication over the UWB and WiFi.
Latency	Depends on number of users of the MoCap and the central processing of the data per user to be streamed per user.	Less influenced by number of users and beacons. Dependent of CHAP processing unit.
Accuracy	Centimeters	Centimeters (more beacons increase resolution but add to latency)
Location	Constraint to lab	Everywhere

Table. 1 – Comparison of lab and field experience.