Shake, Rattle and Roll: An attempt to create a "spatially correct" Ambisonic mixdown of a multi-miked organ concert

Jörn Nettingsmeier
Freelance audio and event engineer
Lortzingstr. 11
45128 Essen, Germany
<nettings@stackingdwarves.net>

Abstract

In Summer 2008, the world's first wave field synthesis (WFS) live transmission (of Olivier Messiaen's *"Livre du Saint Sacrément"*) took place between Cologne Cathedral and the WFS auditorium at Technische Universität Berlin.

The music of three spatially separated organ divisions was captured by multiple microphones in a mixture of spot miking and Hamasaki square technique, i.e. without a dedicated main microphone, as this was deemed desirable for the intended reconstruction on a WFS system.[1]

This paper describes an attempt to create a spatially correct mix from the concert recordings using Ambisonic encoding.

The toolkit used for post-production consists exclusively of free software, centered around JACK, Ardour and the AMB plugin set on a Linux system.

Keywords

Ambisonic panning, ambisonic mixing, spatial reconstruction, spot microphones, Hamasaki square

1 Introduction

1.1 The composition

The *"Livre du Saint Sacrément"*, finished in 1984, is the last and greatest organ work of French composer Olivier Messiaen (1908-1992). It consists of 18 sections, and spans a duration of 90-100 minutes. A deeply religious work, it expresses the Christian creed's hope of salvation in a series of movements that either depict stations in the life of Jesus Christ, or present-day religious rituals (*"sacraments"*). It is highly programmatic: each section has a descriptive title and follows a written "storyboard", whose components are represented by motifs or sonic textures and developed according to the passing of events according to Christian belief. To aid the audience in following the work, the composer has chosen excerpts from the Old and New Testament and from quotes attributed to Christian saints to match the program. These quotes are to be displayed to the audience during the performance.

The work features several distinctive elements of Messiaen's musical style: his particular "modes of limited transposition" [2], such as whole-tone and diminished (octatonic) scales, symmetries of time and pitch, a particularly "colourful" use of harmony[2] and the copious use of birdsong as a source of melodic material[3].

1.2 Location, instruments and spatial disposition

Cologne Cathedral is a challenging venue for any organist to perform in, due to its sheer size[4], immense reverberation time of around 13 s and very high ambient noise[5].

The main organ is located in the northern part of the transept, next to the intersection with the central nave. Commonly called the *"Querhaus-orgel"* (transept organ), it was build by Klais, Bonn in 1948, with extensions in 1954 and 2002. It has electric action and consists of 88 stops, 17 of which are placed in a swell enclosure [3]. Its ranks are divided into two facades at an angle of 90 degrees, and a small rear division (*"Rückpositiv"*) at the back of the organ pedestal, facing the choir.

[1] see [1] for a detailed project report

[2] Messiaen described his own perception of harmony as synaesthetic: hearing sounds would inevitably make him imagine colours, which he often includes in his scores as hints to the performer.

[3] see for example Mvt. 15, « La joie de la grâce »

[4] over 400,000m³ of interior volume

[5] caused by the vicinity of several subway lines, the central train station and a number of roads plus, during daytime, a steady stream of visitors

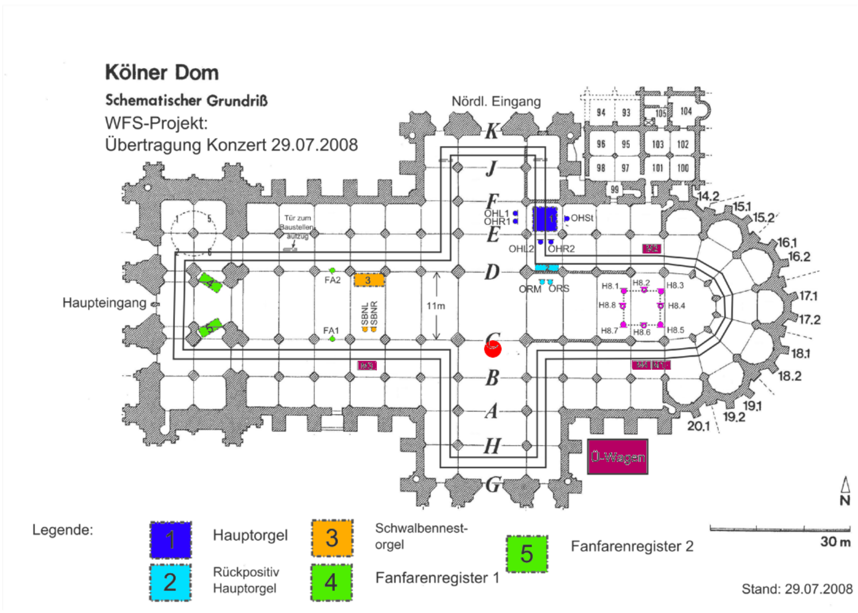

Illustration 1: The organs of Cologne Cathedral. Main organ and microphones in dark blue, rear division in cyan. Schwalbennestorgel *in orange, Microphones on the balcony across the nave. Tuba ranks in green. To the right, the Hamasaki square. The red dot is the virtual listening position, facing north (=upward).*

In 1998, a second instrument was added to the cathedral, again built by Klais. It is suspended from the roof on four steel cables at an acoustically favourable location in the main nave, which lead to its nickname *"Schwalbennestorgel"* (swallow's nest organ). The new organ has mechanic action and 57 registers, with 14 set in a swell [4].
An additional electric action allows it to be remote-controlled from the main console; this setup was used for the performance of the *"Livre"*.

Finally, in 2006, an additional remote division was added to the main organ. Placed atop the main entrance, two high-pressure ranks (at 700 mmAq) housing the *tuba capitularis* and *tuba episcopalis* stops in *"en chamade"* configuration (trumpet-style, facing outward) now enrich the spatial and sonic possibilities of the cathedral organs. These stops are reserved for special occasions and were used sparingly but to great effect during the concert.

The correct spatial arrangement of sounds is an important aspect in the reproduction of organ music. Pipe organs have several layers of spatial separation. Within each stop, sounds obviously move with pitch. This can add a subtle or even dramatic amount of motion to the music, especially if the pipes are arranged in two ranks to the left and right and alternating every half tone (most common on *principal* ranks where this layout is preferred for visual balance). This effect quickly diminishes with distance.
The location of stops within an enclosure are not usually heard outside, since they are commonly mounted one behind the other. However, stops on different facades of the same organ will be localised distinctly.
In Cologne Cathedral, the main organ alone has three clearly separable sections: the two faces to the central nave and the transept, and the rear division towards the choir.

Finally, the two entirely separate organs and the remote *tuba* ranks make up the most dramatic and obvious layer of spatial distribution.

The organist used the latter to great effect, for example by distributing birdsong "dialogues" between two organs as if each bird was sitting on their own tree. Often, voices in a polyphonic setting were physically set apart to increase their independence; at other times, couplers were used to play the same notes in unison from multiple locations.

Illustration 1 shows that the spatial layout of the cathedral organs is somewhat unconventional, and that a listener sitting in the usual place in the central nave facing eastward would experience a strangely lopsided acoustic image, with all sources but one to the left. In reality, this does not diminish the experience much, since the available visual cues support the auditory localisation. But when listening to a reproduction, such a source configuration is sub-optimal, because it neither fully exploits the angular width of the reproduction system nor provides the usual left-to-right balance that is expected in the absence of visual information.
For this reason, the WFS reproduction in Berlin took artistic licence in spreading the sources to

obtain a well-balanced artificial image that does not exist in real life.

To avoid this compromise in the Ambisonic mixdown, it was decided to move the virtual listening position to a place in the southern transept, close to the intersection, facing north. Such a position should obtain a pleasurable left-right balance without extreme rear cues, and allow for later verification of the recreated image with actual organ concerts, ideally using a reference recording from the same spot using a first-order soundfield microphone. Conditions permitting, this will be the subject of an updated version of this paper.

1.3 Microphone disposition

For the intended reproduction on the WFS array at Technische Universität Berlin, it was decided early on that close-miked signals of the organ divisions would be most effective. These were to be combined with uncorrelated ambience signals (captured by two intersected Hamasaki squares, one with its lobes in the horizontal plane, the other's in the vertical plane.

The close-up signals were then to be rendered as point sources, whereas the ambience would be added to taste, rendered as plane waves coming from the corners of the listening room.[6] Thus, both the direct sound and some amount of late reverberation were captured, transmitted and recreated separately, but distinct early reflections were not.

Naturally, each microphone picks up its own set of early reflections, but these would obviously provide conflicting spatial cues when mixed. This did not impair the enjoyment of the WFS audience, which was in no position to judge the correctness of the imaging, but the problem regularly surfaced in the studio during the Ambisonic mixdown.

Given the attempted rendering method, it might seem odd that most organ works were captured with narrow A/B stereo pairs. This was done partly to allow some degree of control over the size of the sonic image (by modifying the distance of the two correlated sources in the WFS rendering), and partly out of Tonmeister habit. It certainly did not ease the task of creating a convincing Ambisonic mixdown. However, even the one coincident pair used (the M/S at the Rückpositiv) turned out to be non-trivial. Single, decorrelated sources such as the *tuba* microphones proved the most straightforward.

2 Mixdown

2.1 Target format

Since the author's audio playground contains a hexagonal Ambisonic monitoring system which is capable of horizontal second-order reproduction [7], and full second-order panning was easily achieved by slightly modifying available panners [8], it was decided to go for second-order Ambisonics as the target format, to profit from the greater angular resolution compared to first-order B-Format. This implies the use of a 9-channel master bus, which is easily accomplished using Ardour [9], a free digital audio workstation software. Ardour is exceptionally well-suited to Ambisonic surround production due to its extreme flexibility with multichannel routing: it allows buses, sends and inserts to have an arbitrary number of channels.

During the preparation of this paper, the author did not have access to a full 3D listening rig. It is hoped that this will change until the presentation at LAC 2009, so that mixdown decisions can be fully verified (and corrected if necessary) before the public demonstration.

2.2 Panning considerations

Before starting to work on the mixdown, the azimuth and elevation angles of all spot microphones had to be computed (see illustration 2). These were then applied to a separate panning plugin for each of the captured signals. The standard Ardour panner was bypassed.

It was decided to use both signals of each A/B pair after carefully checking for comb-filtering and colouration. The opening angle between those source pairs was set "to taste", not derived from actual measurements (see Comments section in the illustration).

The positioning of the Hamasaki signals was non-trivial. The Hamasaki squares were suspended at a height of around 23m in the central nave close to the apsis, in order to keep them out of reach of the organs' direct sound fields. Unfortunately, the reverberant field up there has little if any resemblance to what actually happens at the virtual listening spot. Hence, the signals were again used "to taste". A modified panner was used to feed the figure-of-eight signals into the target planes as pure velocity components, without letting them

[6] For an overview of the WFS array, see [5]. Some information about the rendering can be found in [6].

Microphone data – Livre du Saint Sacrement

Source	Microphone	Polar pattern	θ [deg]	d [m]	h [m]	ε [deg]	Δs_{Mic} [m]	Δt_{Mic} [s]	z [m]	$\Delta t_{Mixdown}$ [s]	Comments
H1 horiz.	Sennheiser MKH 800	Fig8	0	30.6	23	0	0	0.0000	38.28	0.1126	location not used
H2 vert.	Sennheiser MKH 800	Fig8	0	34.7	23	90	0	0.0000	41.63	0.1224	“
H3 horiz.	Sennheiser MKH 800	Fig8	0	38	23	0	0	0.0000	44.42	0.1306	“
H4 vert.	Sennheiser MKH 800	Fig8	0	37.3	23	90	0	0.0000	43.82	0.1289	“
H5 horiz.	Sennheiser MKH 800	Fig8	180	36.6	23	0	0	0.0000	43.23	0.1271	“
H6 vert.	Sennheiser MKH 800	Fig8	0	32.7	23	90	0	0.0000	39.98	0.1176	“
H7 horiz.	Sennheiser MKH 800	Fig8	180	28	23	0	0	0.0000	36.24	0.1066	“
H8 vert.	Sennheiser MKH 800	Fig8	0	28.7	23	90	0	0.0000	36.78	0.1082	“
Q1 L/R	2x Schoeps MK 5	Omni	-13	30	12	21.8	5	0.0147	32.31	0.0803	Pair angle 1°
Q2 L/R	2x Schoeps MK 5	Omni	-24	27.9	13	24.98	4	0.0118	30.78	0.0788	Pair angle 5°
Q3 M/S	Schoeps MK 5 / MK 8	Omni / Fig8	-34	20	6	16.7	3	0.0088	20.88	0.0526	S at -124°, no elev., -10dB
Q4	Schoeps MK 5	Omni	-28	32.6	12	20.21	3	0.0088	34.74	0.0933	
S L/R	2x Schoeps MK 21	Sub-Cardioid	62	30	27	41.99	11.3	0.0332	40.36	0.0855	Pair angle 2°
F1	Schoeps CCM41	Hyper-Cardioid	76	61.3	21	18.91	28.7	0.0844	64.8	0.1062	
F2	Schoeps CCM41	Hyper-Cardioid	88	59.3	21	19.5	28.7	0.0844	62.91	0.1006	
Announcer	Sennheiser MD 421	Cardioid	0	10	2	11.31	0.15	0.0004	10.2	0.0296	

θ: Azimuth angle, 0° is due north, positive is counter-clockwise *(measured)*
d: Distance on the floor between virtual listening spot and source *(measured)*
h: Height of source above listening spot *(estimated)*
ε: Elevation angle, 0° is on horizontal plane, 90° is zenith *(atan(h/d))*
$\Delta s_{Mic:}$ Distance from microphone to source *(estimated)*
$\Delta t_{Mic:}$ Delay of sound due to microphone distance from source *(Δs_{Mic} / 340 m/s)*
z: Total distance from listening spot to source (*sqrt(d²+h²)*)
$\Delta t_{Mixdown}$: Additional delay required during mixdown *(z / 340 m/s - Δt_{Mic})*

Illustration 2: The spreadsheet used to compute source angles and delays.

contribute to W, to avoid localisation.[7] The actual microphones of the planar array were oriented towards the walls (i.e. two at 0°, two at 180°), but their signals were spread evenly in the mix (at 45°, 135°, -135° and -45°) to create uniform envelopment without holes to the sides.[8]

The upward array is currently not used due to lack of z-axis reproduction, but it seems likely that a similar approach will be taken here.

As of this writing, no obviously “correct” mixdown approach to the one M/S pair has been found. Currently, the mid component is panned as usual and mixed at 0 dB, while the side signal is fed into the horizontal plane with the positive lobe oriented at 90° to the source direction, using the same approach as for the Hamasaki figures-of-eight. It is mixed at a lower level, again “to taste”, to give some sense of width without blurring the source location too much.

2.3 Delays

In a second step, the timing of the sources was matched to the virtual listening spot using simple delay plugins per channel. The required amount of delay was computed from the source distance z minus the delay introduced by the distance of the microphone(s) from the source.

From the performer's point of view, the extreme latencies from the remote stops to the console are more a nuisance than a feature, but it is the author's belief that they should be reproduced faithfully, since they will have affected the organist's playing (if only to compensate for them) and are thus part of the artistic work.

2.4 Level matching

The relative levels of the sources were matched for musical balance, without any way to determine the correct original ratios. In the end, most signals were used at around 0 dB (which doesn't say much, since the preamps were not gain matched, nor did the capsules have similar sensitivity).

An updated version of this paper will use the original score to obtain more detailed dynamic

[7]The author was made aware of a similar approach being used by reseachers at IEM Graz, but as of this writing, no publications describing this method could be identified. Pointers are welcome.

[8]For azimuth angles, this paper follows the mathematical convention where positive values indicate counter-clockwise rotation. Elevations are positive (=above the floor).

information to revise mixing decisions according to composer's intent.

2.5 A second look at distance coding

The correct reproduction of distance relies on a number of different parameters, only some of which could be matched correctly:

- reduction of sound energy, according to the inverse square law for free-field point sources, a little less for bigger sources within rooms;
  easily obtained by gain reduction
- delay (not an absolute distance cue by itself, it is nonetheless important to reproduce the relative timing of the organ works at the listening spot);
  readily simulated by applying the correct delay times
- air damping, manifesting itself as a loss of high frequencies;
  not currently used, but the next mix revision will include gentle low-pass filters to experiment with air damping effects [10]
- wave front curvature, a familiar aspect of WFS but a comparably new concept in Higher Order Ambisonics;
  has not been investigated further in this paper
- direct-to-reverb ratio;
  quite easy when using artificial reverb, but impossible to control with the Hamasaki setup used for this project

Especially the direct-to-reverb ratio is a thorny subject: neither are the spot mikes totally dry (quite the contrary, in the case of the omnis), nor is the “diffuseness” of the Hamasaki constant for all organ works – there is quite some amount of direct sound from the main instrument, while the swallow's nest organ is far more distant-sounding.

One option would be convolution reverb, but since the impulse responses would have to be captured for each source *at the desired listening spot*, the advantage of being able to select that spot during mixdown is lost.

A minor nuisance were the early reflections picked up by the spot microphones. The swallow's nest organ was almost clean, since the only relevant reflection (from the rear wall) conveniently fell into the least sensitive direction of the sub-cardioids. The main organ's microphones however caught some distinct reflections that were clearly heard as contradicting when the respective channels were solo'ed. In the mix, they did not stand out as badly, but they certainly did nothing to improve the imaging.

3 Results

All in all, the obtained mix sounds convincing and pleasurable on a horizontal rig. It provides a very good degree of envelopment. Source location is very precise. However, the “correctness” of the mix cannot currently be evaluated due to the lack of a co-incident reference recording.

As to the “suspension of disbelief”, the obvious mismatch between the gigantic (albeit not entirely desirable) acoustics of Cologne Cathedral and the minuscule space of the listening room requires that the listener either close her/his eyes or sit in darkness.

A very annoying problem that becomes immediately obvious during short, loud sounds with pauses in between is the occurence of “phantom walls”. Currently, no mixing automation is used[9] and microphones stay open all the time, which means that sound emanating from the main organ will reach the other microphones after a while.

If the original sound is loud, the level at the “wrong” microphones will be non-negligible, and if there is no following sound to mask it, unnatural echoes will appear. **In effect, each microphone creates a false “wall reflection” that should not be there.**

This problem is further emphasized by the use of corrective delays, which spreads the echo incidents further apart and pulls them out of the masking veil of the initial sound.

By far the most obvious false cues came from the *tuba* mircophones. The sound crew had decided to use hypercardioids to reduce the amount of reverberation in this “long shot” setup, but the rear lobes caught so much direct sound and early reflections from the other organs that they distorted the image significantly.

The only way around this appears to be score-based mixing automation, where unused microphones are brought down as much as possible. This is complicated by the fact that the actual choice of stops (and thus the active set of microphones) is for the most part at the discretion of the performing artist. Moreover, even with full automation, the tuba stops' sound would not

[9] with the exception of the announcer's microphone, which is faded out after the opening address

completely mask the false reflections and there would still be a distortion of the image when the tubas are in use.

The lack of distinct early reflections for the virtual listening spot is not immediately obvious, but the acoustic image of the room, while convincing at first, is nowhere near realistic.

4 Conclusion

In retrospect, the chosen recording approach will not be able to provide a totally satisfactory sonic image, mainly due to the creation of false echoes by each source microphone. In addition, no actually correct early reflection signals are available.

On the up side, close miking allows for the creation of different renderings for arbitrary virtual listening spots.

In future undertakings, it might be worthwhile to capture some early reflections at important boundaries using pressure-zone microphones. If a number of such signals were obtained, imaging precision during mixdown could be improved. If the desired virtual listening spot is known in advance, the PZMs can be concentrated around this point to increase efficiency.

However detailed the captured early reflections are, it seems unlikely that they are able to mask the false echoes created by the spot mikes, so a combination with mixing automation seems mandatory. Microphones shouldbe selected to minimize leak from other organ divisions. Closer miking might help as well, which would in turn mandate the use of more microphones to ensure proper coverage, at the risk of comb filtering.

It remains to be seen how a first-order soundfield recording would fare in comparison. Despite the artifacts, there are clear benefits of ambi-panned closed miking: a great deal of clarity and transparency of sound, and the greater localisation precision and larger sweet spot of higher-order ambisonics.

Using a first-order main microphone with discrete spot mikes panned in higher order will be a mixed blessing (pun intended) that will most likely do more harm than good. However, a spherical array capable of second-order recording might be an improvement.

Ultimately, the approach to recording boils down to the decision between a mathematically correct yet unforgiving coincident main microphone and discrete multi-miking and its creative freedom in post-production. The latter will never be “the real thing”, but the question remains whether a convincingly faked reality might not convey the composer's intention just as well or better than a “correct” recording, especially in the extreme acoustics of Cologne Cathedral.

5 Acknowledgements

Thanks to Robert-Schumann-Hochschule für Musik and Technische Universität Berlin, Lehrstuhl für Audiokommunikation, for letting me use the recording of the concert. Special thanks to Hans Schlosser for answering my many questions, and to the RSH student crew (Simon Hildenbrand, Benjamin Simon, Katrin Körber, Jindrich Maßner, Simon Spillner, Josuel Theegarten and Michael Zöllner) for being great guys to work with.

Kudos to Prof. Winfried Bönig, the performing artist, for an inspiring rendition of the “Livre”, and to all the helpful people at the Dombauhütte and at KHM in Köln.

The floor plan of the cathedral was provided by the Dombauhütte; Microphone layout added by Michael Zöllner (with slight changes by the author). Used with permission.

References

[1] Jörn Nettingsmeier: Birds on the wire – Olivier Messiaen's Livre du Saint Sacrément in the world's first wave field synthesis live transmission. http://stackingdwarves.net/public_stuff/event_documentation/wfs_live_transmission_2008, 2009

[2] Wikipedia, “Modes of limited transposition”, http://en.wikipedia.org/wiki/Modes_of_limited_transposition, retrieved Jan 2009

[3] Klais Orgelbau, Stop list of the Querhausorgel, http://www.orgelbau-klais.com/m.php?tx=16 , retrieved Jan 2009

[4] Klais Orgelbau, Stop list of the Schwalbennest-orgel, http://www.orgelbau-klais.com/m.php?tx=15 , retrieved Jan 2009

[5] Marije Baalman, How to control 840 channels. Working with large scale wave field synthesis. In: next_generation 2007: Musik im Raum, hrsg. vom ZKM | Institut für Musik und Akustik, Symposiumsbericht, Karlsruhe, pages 52-55,

http://container.zkm.de/musik/downloads/next_generation_2007.pdf, 2007

[6] Hans Schlosser et al.: "Mer lasse d'r Dom in Kölle" - Weltpremiere einer WFS-Live-Übertragung", workshop presentation at VDT International Convention (Tonmeistertagung) 2008.
http://stackingdwarves.net/public_stuff/linux_audio/tmt08/Workshop_TMT08.pdf, 2008

[7] Fons Adriaensen, AmbDec, an Ambisonic decoder capable of first- and second-order two-band reproduction, including near-field and delay compensation,
http://www.kokkinizita.net/linuxaudio/downloads/index.html, 2008

[8] Fons Adriaensen, AMB plugins, a set of Ambisonic utility LADSPA plugins, 0.3.0
http://www.kokkinizita.net/linuxaudio/downloads/index.html, 2009
As of release 0.4.0, full second-order panners are included.

[9] Paul Davis et al., the Ardour Digital Audio Workstation, release 2.7.1,
http://ardour.org, 2009

[10] ISO 9613-1, Acoustics - Attenuation of sound during propagation outdoors - Part 1: Calculation of the absorption of sound by the atmosphere, 1993