Ambisonic spatial blur

Thibaut Carpentier

Audio Engineering Society

Convention Paper

Presented at the 142nd Convention
2017 May 20–23, Berlin, Germany

This paper was peer-reviewed as a complete manuscript for presentation at this convention.

Ambisonic spatial blur

Thibaut Carpentier[1]

[1] *UMR 9912 STMS IRCAM–CNRS–UPMC, Paris, France*

Correspondence should be addressed to Thibaut Carpentier (thibaut.carpentier@ircam.fr)

ABSTRACT

This paper presents a technique for controlling the spatial resolution of an Ambisonic sound field while preserving its overall energy. The proposed method allows to transform a stream encoded in *N*-order Ambisonic to a lower order resolution. The transformation can be continuously operated, indeed simulating fractional order representation of the Ambisonic stream and varying the "bluriness" of the spatial image.

1 Introduction

When working with surround sound scenes (either synthesized or recorded), it is often desirable to modify the spatial characteristics of the sound field. Sound designers, composers and mixing engineers need tools to conveniently manipulate the spatial image either during the production, re-production or post-production stages.

Ambisonic is a recording and reproduction technique that can be used to create spatial audio for circular or spherical loudspeaker arrangements. The technique is based on the representation of the sound field as a combination of orthogonal basis functions, namely the spherical harmonics. One benefit of such intermediate representation is that it can be flexibly manipulated so as to alter the spatial properties of the sound field. Such modifications are achieved by applying a (time and frequency independent) transformation matrix in the Ambisonic domain [1].
A number of transformation operators have been previously proposed. This includes – but is not limited to – rotation matrix [2][3], mirroring [4], spatial emphasis or directional loudness [5], dominance or angular distorsion [6][4] later extended to high order warping [7], etc.

The work presented in this paper proposes a novel transformation operator that allows to control the spatial resolution of an Ambisonic field while preserving its overall energy. Spatial resolution relates to the notions of localization accuracy versus diffuse impression of sound scenes. This is commonly reported as a most salient attribute for composers and sound designers (see e.g. [8][9]) and it is sometimes presented as a "spread" or "blur" parameter in audio processors.
The technique herein presented allows to transform a stream encoded in *N*-order Ambisonic to a lower order resolution. The transformation can be continuously operated, indeed simulating fractional order representation of the sound field. The perceived "bluriness" – for lack of a better word – of the rendered spatial image can therefore be gradually varied.

This paper is organized as follows: Section 2 reviews some fundamentals of Ambisonic theory, highlighting a few important results that will later be needed. In Section 3 we detail the Ambisonic spatial blur transformation operator. Section 4 assesses the effectiveness of the rendered effect.

2 Basics of Ambisonic processing

There is no universally agreed notation for Ambisonic; this section presents the definitions and conventions that will be needed later in the article. We expect the reader to be familiar with this material, so the results appear without proof (see e.g. [2][10] for further details).

2.1 Notations

$\boldsymbol{\vartheta} \equiv (\theta, \phi)$	angular direction in the spherical coordinate system with $x = \cos\phi \sin\theta$, $y = \sin\phi \sin\theta$, $z = \cos\theta$
$\|\|\mathbf{x}\|\|$	Euclidian norm of $\mathbf{x}$
K	number of loudspeakers
N	the Ambisonic order
χ_N	number of Ambisonic components i.e. $\chi_N = (2N+1)$ in 2-D $\chi_N = (N+1)^2$ in 3-D
$\mathbb{I}_N$	the Ambisonic indices i.e. $\{(n,m) \in \mathbb{N} \times \mathbb{Z} : 0 \leq \|m\| \leq n \leq N\}$

2.2 Ambisonic encoding

A plane wave of incidence $\boldsymbol{\vartheta}$ conveying a signal S is encoded in the Ambisonic domain by means of the spherical harmonic decomposition, leading to the following expression of the Ambisonic components:

$$\forall (n,m) \in \mathbb{I}_N, \quad B_n^m(\boldsymbol{\vartheta}) = S\,\mathrm{Y}_n^m(\boldsymbol{\vartheta}) \tag{1}$$

where $\mathrm{Y}_n^m(\boldsymbol{\vartheta})$ are a real-valued set of spherical harmonics defined $\forall (n,m) \in \mathbb{I}_N$ by

$$\mathrm{Y}_n^m(\boldsymbol{\vartheta}) = \mathscr{A}_n^{|m|} P_n^{|m|}(\cos\theta) \begin{cases} \cos(m\phi), & \text{for } m \geq 0 \\ \sin(|m|\phi), & \text{for } m < 0 \end{cases} \tag{2}$$

In this equation, $\mathscr{A}_n^{|m|}$ is a scalar normalization constant and $P_n^m(x)$ represent the associated Legendre functions defined $\forall x \in \mathbb{R} \cap [-1,1]$, $\left\{(n,m) \in \mathbb{N}^2 : m \leq n\right\}$ by

$$P_n^m(x) = (-1)^m \left(1-x^2\right)^{\frac{m}{2}} \frac{d^m}{dx^m} P_n(x). \tag{3}$$

$P_n(x)$ are the Legendre polynomials and they may be expressed using Rodrigues' formula $\forall x \in \mathbb{R} \cap [-1,1]$, $\forall n \in \mathbb{N}$

$$P_n(x) = \frac{1}{2^n\, n!} \frac{d^n}{dx^n} \left(x^2-1\right)^n. \tag{4}$$

Various normalization conventions have been proposed for the scalar constant $\mathscr{A}_n^{|m|}$ (see e.g. [2]); without loss of generality, this paper opts for the full-normalization scheme, noted N3D and N2D in the 3-D and 2-D case respectively, and given for $\left\{(n,m) \in \mathbb{N}^2 : m \leq n\right\}$ by

$$\left(\mathscr{A}_n^m\right)^{\mathrm{N3D}} = \begin{cases} \sqrt{2n+1}, & \text{if } m = 0 \\ (-1)^m \sqrt{2\,(2n+1)\,\frac{(n-m)!}{(n+m)!}}, & \text{if } m \neq 0 \end{cases} \tag{5}$$

and

$$\left(\mathscr{A}_n^n\right)^{\mathrm{N2D}} = \begin{cases} 1, & \text{if } n = 0 \\ (-1)^n \sqrt{\frac{2^{2n+1}\, n!^2}{(2n)!^2}}, & \text{if } n \neq 0 \end{cases} \tag{6}$$

2.3 Ambisonic decoding

Let's note $\boldsymbol{c}$ the vector of real gains involved in the encoding equation (1):

$$\boldsymbol{c}(\boldsymbol{\vartheta}) = \begin{bmatrix} \mathrm{Y}_0^0(\boldsymbol{\vartheta}) \\ \mathrm{Y}_1^{-1}(\boldsymbol{\vartheta}) \\ \mathrm{Y}_1^0(\boldsymbol{\vartheta}) \\ \mathrm{Y}_1^1(\boldsymbol{\vartheta}) \\ \vdots \\ \mathrm{Y}_n^m(\boldsymbol{\vartheta}) \\ \vdots \end{bmatrix} \tag{7}$$

In practice, the encoding is limited to a finite order N, involving χ_N components in vector $\boldsymbol{c}$. Note that the components in equation (7) are presented with the ACN ordering, but other arrangement can be employed as well.

At the decoding stage, the so-called "re-encoding principle" [11] can be employed: considering K loudspeakers (supposed to be far enough from the sweet area so that the plane waves assumption holds) with directions $\{\boldsymbol{\varphi}_1, ..., \boldsymbol{\varphi}_K\}$, the re-encoding matrix has χ_N rows and K columns and it writes:

$$\boldsymbol{C} = \left[\boldsymbol{c}(\boldsymbol{\varphi}_1)\ \boldsymbol{c}(\boldsymbol{\varphi}_2)\ ...\ \boldsymbol{c}(\boldsymbol{\varphi}_K)\right]. \tag{8}$$

Provided that there are enough loudspeakers (i.e. $K \geq \chi_N$), the decoding matrix $\boldsymbol{D}$ is given by the pseudo-inverse of $\boldsymbol{C}$:

$$\boldsymbol{D} = \mathrm{pinv}(\boldsymbol{C}) = \boldsymbol{C}^T \left(\boldsymbol{C}\boldsymbol{C}^T\right)^{-1}. \tag{9}$$

Daniel et al. [12][2] have shown that, for regular loudspeaker layouts and under the full-normalization, the previous expression simplifies to

$$\boldsymbol{D} = \frac{1}{K}\boldsymbol{C}^T \tag{10}$$

which is commonly refers to as the "sampling Ambisonic decoder" (SAD).
The decoder may further be altered by applying a set of χ_N corrective gains $\{g_0, g_1, \cdots, g_N\}$ to the encoded stream. The decoding matrix thus becomes:

$$\boldsymbol{D} = \frac{1}{K}\boldsymbol{C}^T \,\mathrm{Diag}\Big\{ g_0\; g_1\; \cdots\; g_N \Big\} \tag{11}$$

2.4 Equivalent panning function

Combining the encoding equation (1) and the decoding matrix (11), Daniel et al. [12][2] have derived an equivalent panning function i.e. the k^{th} loudspeaker is fed with a signal $S_k = S\,G_k$ such as $\forall k \in [1, K]$

$$K\,G_k = \begin{cases} g_0 + 2\sum\limits_{n=1}^{N} g_n \cos(n\,\gamma_k) & \text{in 2-D} \\ \sum\limits_{n=0}^{N} (2n+1)\, g_n\, P_n(\cos\gamma_k) & \text{in 3-D} \end{cases} \tag{12}$$

where γ_k denotes the angle between the virtual source and the k^{th} loudspeaker direction. This equation can be interpreted as an equivalent panning law or alternatively as an equivalent directivity pattern (see also [13][14]).

Assuming a regular loudspeaker layout, Daniel et al. have further proved that the energy Σ_e resulting from the decoding stage simplifies to:

$$\Sigma_e = K\sum_{k=1}^{K} G_k^2 = \begin{cases} g_0^2 + 2\sum\limits_{n=1}^{N} g_n^2 & \text{in 2-D} \\ \sum\limits_{n=0}^{N} (2n+1)\, g_n^2 & \text{in 3-D} \end{cases} \tag{13}$$

3 Energy-preserving spatial blur

The spatial resolution of the Ambisonic stream can be adjusted by controlling the gains $\{g_0, g_1, ..., g_N\}$. Noting $\alpha \in \mathbb{R} \cap [0, 100]$ the spatial blur factor, one can for instance use a weighting function such as $\forall n \in [0, N]$

$$\alpha \;\rightarrow\; g_n(\alpha) = 1 - \frac{1}{1 + e^{-\tau\left(\alpha - 100\frac{N-n+1}{N+1}\right)}} \tag{14}$$

where $\tau \in \mathbb{R}_+^*$ is an arbitrary constant factor that characterizes the slope of a S-shape curve. Such weighting gains allow to continuously transition from order N to lower order representation (see Figure 1).
Note that any other function $g_n(\alpha)$ might have been used and the following results hold without loss of generality.

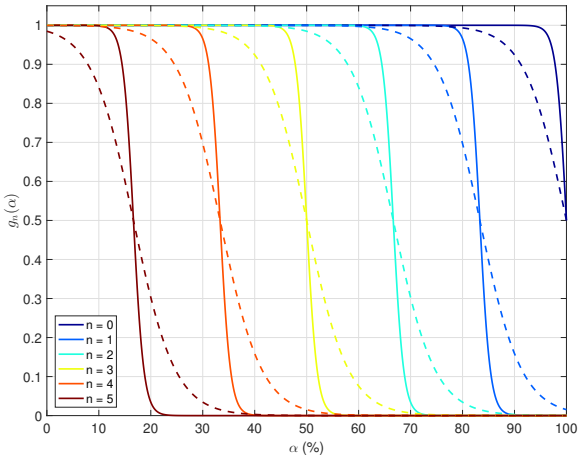

Fig. 1: Ambisonic weighting functions $g_n(\alpha)$ as a function of the blur factor α. Curves are plotted for $N = 5$; solid lines correspond to $\tau = 1$ and dashed lines to $\tau = 0.25$

From now on, it is clear that the loudspeaker gains G_k and the resulting decoder energy Σ_e are dependent on the blur factor α:

$$\Sigma_e(\alpha) = K\sum_{k=1}^{K} G_k^2(\alpha). \tag{15}$$

Choosing $\alpha = 0$ in equation (14) leads to $\forall n \in [0, N]$, $g_n(\alpha) \approx 1$ which means that the Ambisonic stream is unaltered. When α is increased from 0 to 100%, the high order Ambisonic components are progressively "faded out", obviously causing a decrease in the reproduced energy. This energy loss has to be compensated in order to maintain a constant loudness. The energy-preserving spatial blur is achieved by introducing a compensation factor $W(\alpha)$ and modified gains $\{\widetilde{g_0}, \widetilde{g_1}, ..., \widetilde{g_N}\}$ such as $\forall n \in [0, N]$:

$$\widetilde{g_n}(\alpha) = W(\alpha)\, g_n(\alpha). \tag{16}$$

It is clear from equation (13) that the decoder energy with modified gains $\widetilde{\Sigma_e}$ then writes:

$$\widetilde{\Sigma_e}(\alpha) = K\sum_{k=1}^{K} \widetilde{G_k}^2(\alpha) = W(\alpha)^2\, \Sigma_e(\alpha) \tag{17}$$

and this formula holds both for the 2-D and 3-D cases. It is easy to see that

$$W(\alpha) = \sqrt{\frac{\Sigma_e(0)}{\Sigma_e(\alpha)}} \tag{18}$$

ensures energy preservation i.e. $\widetilde{\Sigma}_e(\alpha) = \Sigma_e(0)$ is constant for any value of α. The resulting gains $\{\widetilde{g_n}(\alpha)\}$ are represented in Figure 2.

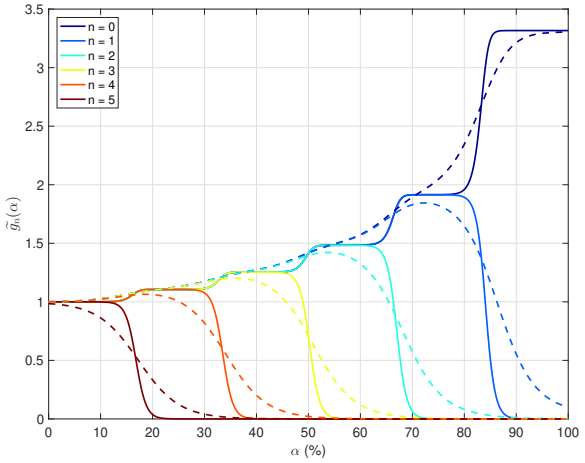

Fig. 2: Energy-preserving Ambisonic weighting functions $\widetilde{g_n}(\alpha)$ for $N = 5$ and $\tau = 1$ (solid lines) and $\tau = 0.25$ (dashed lines).

Note that the proposed approach is compatible with subsequent max-r_E or in-phase weightings [2] that might be applied in the decoder.

4 Results

The spatial blur processor presented in the previous section was assessed through numerical simulations, for various values of the blur factor α. The simulations were conducted for the case study of a single virtual sound source encoded up to the order $N = 5$ and decoded over $K = 11$ regularly arranged loudspeakers.

4.1 Equivalent panning and directivity pattern

The equivalent panning function and directivity pattern are presented in Figures 3 and 4 respectively. When $\alpha = 0$, all Ambisonic components are being used and they contribute to producing the "finest" angular selectivity (with respect to the simulated setup). When α increases, the contribution of the high order components decreases and therefore the spatial resolution diminishes: the sound field tends towards an omnidirectional pattern.
This is in accordance with the expected behavior of the proposed processor. Note however that $\alpha = 100$ would result in fully correlated signals over all loudspeakers, which is usually to be avoided. Several authors have proposed alternative approaches for controlling the diffuseness of sound fields, typically requiring the use of decorrelation filters (see e.g. [15][16][17][18]).

In Figure 4 it can also be noted that, if the energy compensation factor $W(\alpha)$ is omitted (bottom row in Figure 4), the overall energy – which is proportional to the "area" of the directivity pattern – vanishes towards 0 when α increases.

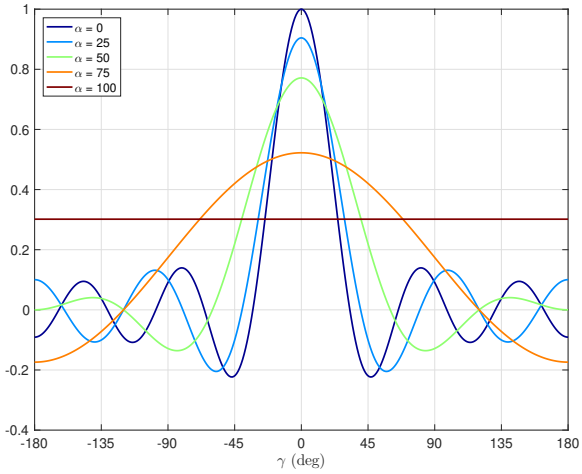

Fig. 3: Equivalent panning law for $N = 5$, $K = 11$ and $\tau = 1$ and various values of the spatial blur α.

4.2 Energy vector

The energy vector $\boldsymbol{E}$ describes the direction and distribution of the energy of the sound field at the listening position [19][20]. It assumes an energetic superposition of the loudspeaker signals S_k and writes:

$$\boldsymbol{E} = \frac{\sum_{k=1}^{K} G_k^2 \boldsymbol{u_k}}{\sum_{k=1}^{K} G_k^2} = r_E \boldsymbol{u_E} \tag{19}$$

where $\boldsymbol{u_k}$ denotes the direction of the k^{th} loudspeaker.

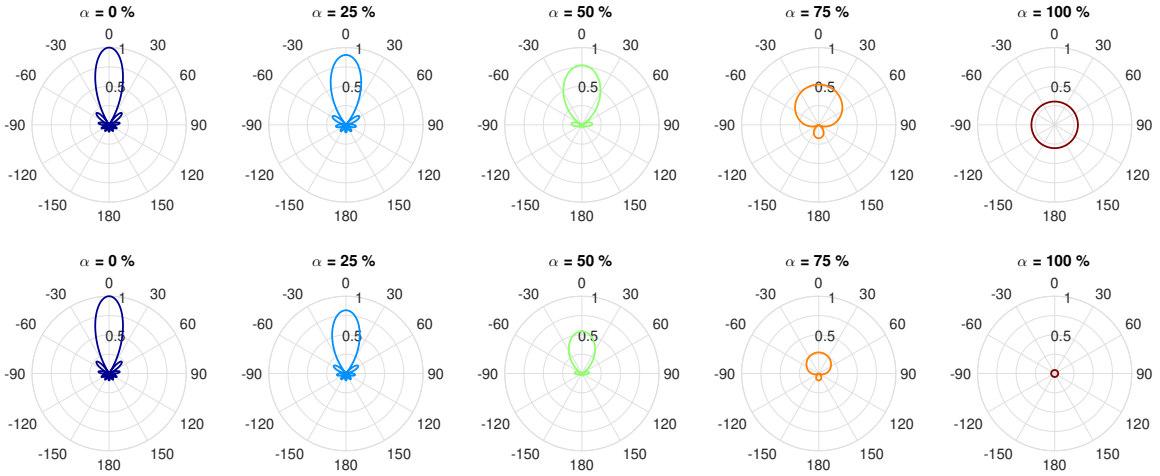

Fig. 4: Equivalent directivity pattern for $N = 5$ and various values of the spatial blur α. With (top) and without (bottom) the energy compensation factor $W(\alpha)$. The figures are plotted for $\tau = 1$. The radial scale is linear.

The direction of the energy vector is typically used as a predictor of the perceived source direction (at high frequencies) while its norm r_E correlates with the perceived source width: the shorter the energy vector, the more widely the energy is distributed across the loudspeakers and the wider the perceived source [21][22].

Daniel et al. [12][2] have shown the following results:

$$r_E = \begin{cases} \dfrac{2\sum_{n=1}^{N} g_n g_{n-1}}{g_0^2 + 2\sum_{n=1}^{N} g_n^2} & \text{in 2-D} \\ \dfrac{2\sum_{n=1}^{N} n\, g_n g_{n-1}}{\sum_{n=0}^{N} (2n+1)\, g_n^2} & \text{in 3-D} \end{cases} \tag{20}$$

Thanks to equation (20) it is possible to plot the norm of the energy vector as a function of the spatial blur α (see Figure 5). As expected, r_E tends towards 0 when the spatial blur is increased.

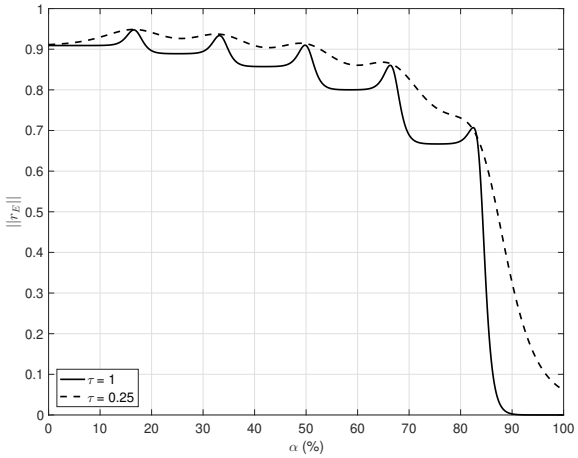

Fig. 5: Norm of the energy vector as a function of the spatial blur α for $N = 5$ and $\tau = 1$ (solid lines) and $\tau = 0.25$ (dashed lines).

4.3 Angular spread

Epain et al. [23] have introduced a mathematical quantity, referred to as the angular spread, which corresponds to the angular width of the area over which the sound energy is distributed. The quantity is defined by

$$\sigma = 2\arccos\left(2r_E - 1\right), \tag{21}$$

and it varies in the range $[0, 2\pi]$. Similarly to r_E, it provides a (rough) predictor of the perceived source width.

The angular spread for the simulated case is plotted in Figure 6, again revealing an increase in the perceived source width as the blur factor α increases.

It is worth noting that the angular spread does not vary linearly with respect to α. This is not really an issue, and informal listening tests confirmed the effectiveness of the proposed spatial blur processor. Nonetheless, it would be possible to design another set of weighting functions $\{g_n(\alpha)\}$ – instead of the S-shape functions used in equation (14) – so as to offer a linear depen-

dency of the angular spread with respect to the blur factor (i.e. $\frac{d\sigma}{d\alpha} = \text{constant}$). The design of such weighting functions is left for future work.

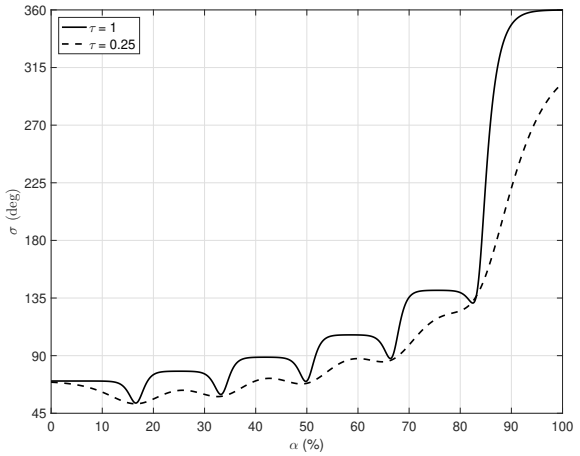

Fig. 6: Angular spread as a function of α for $N = 5$ and $\tau = 1$ (solid lines) and $\tau = 0.25$ (dashed lines).

4.4 Implementation

The spatial blur processor has been implemented as a Max/MSP external objet, part of the Ircam Spat software package [24]. The processor is typically inserted in-between the Ambisonic encoder and the decoder (see Figure 7); the parameters exposed to the end user are the blur factor α and the S-shape slope factor τ. Any change in the parameters (and consequently in the gains $\{g_0, g_1, ..., g_N\}$) is implemented with a short ramping time (a few milliseconds) to avoid audio clicks during realtime rendering.
The CPU load of the blur processor is very low as it only necessitates multiplying the Ambisonic stream with a set of gains, resulting in χ_N multiplications per audio tick.

5 Conclusion

This paper presented a novel technique for controlling the spatial resolution of an Ambisonic sound field while preserving its overall energy. The method allows the user to continuously transform the encoded Ambisonic stream from a high order representation ("finest" angular selectivity) to a low order omnidirectional pattern. The derivation of the transformation gains is easy and the realtime implementation of the spatial blur processor is simple and efficient.

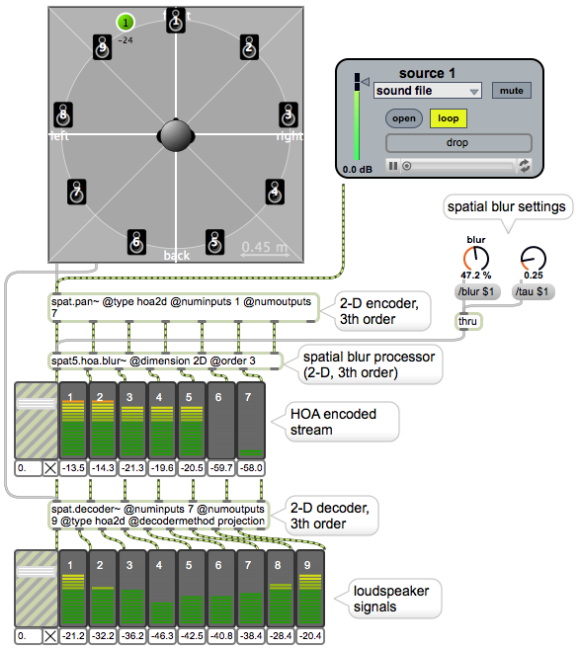

Fig. 7: Workflow within the Max/MSP environment; example with order $N = 3$ and $K = 9$ loudspeakers in 2-D.

References

[1] Chapman, M. and Cotterell, P., "Towards a Comprehensive Account of Valid Ambisonic Transformations," in *Proc. 1st Ambisonics Symposium*, Graz, Austria, 2009.

[2] Daniel, J., *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*, Ph.D. thesis, Université de Paris VI, 2001.

[3] Daniel, J., "Evolving Views on HOA : From Technological To Pragmatic Concerns," in *Proc. of the 1st Ambisonics Symposium*, Graz, Austria, 2009.

[4] Chapman, M., "Symmetries of Spherical Harmonics: Applications to Ambisonics," in *Proc. of the 1st Ambisonics Symposium*, Graz, Austria, 2009.

[5] Kronlachner, M. and Zotter, F., "Spatial transformations for the enhancement of Ambisonic recordings," in *Proc. of the 2nd International Conference on Spatial Audio (ICSA)*, Erlangen, Germany, 2014.

[6] Gerzon, M. A. and Barton, G. J., "Ambisonic Decoders for HDTV," in *Proc. of the 92nd AES Convention*, Vienna, Austria, 1992.

[7] Pomberger, H. and Zotter, F., "Warping of 3D Ambisonic Recordings," in *Proc. of the 3rd Ambisonics Symposium*, Lexington, KY, USA, 2011.

[8] Peters, N., Marentakis, G., and McAdams, S., "Current Technologies and Compositional Practices for Spatialization: A Qualitative and Quantitative Analysis," *Computer Music Journal*, 35(1), pp. 10 – 27, 2011.

[9] Baalman, M. A., "Spatial Composition Techniques and Sound Spatialisation Technologies," *Organised Sound*, 15(3), pp. 209 – 218, 2010.

[10] Abramowitz, M. and Stegun, I. A., *Handbook of Mathematical Functions*, General Publishing Company, 1970.

[11] Daniel, J., Nicol, R., and Moreau, S., "Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging," in *Proc. of the 114th AES Convention*, Amsterdam, Netherlands, 2003.

[12] Daniel, J., Rault, J.-B., and Polack, J.-D., "Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions," in *Proc. of the 105th AES Convention*, San Francisco, CA, USA, 1998.

[13] Neukom, M. and Schacher, J. C., "Ambisonics Equivalent Panning," in *Proc. of the International Computer Music Conference*, Belfast, Ireland, 2008.

[14] Neukom, M., "Ambisonic Panning," in *Proc. of the 123rd AES Convention*, New York, NY, USA, 2007.

[15] Zotter, F., Frank, M., Kronlachner, M., and Choi, J.-W., "Efficient phantom source widening and diffuseness in ambisonics," in *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*, pp. 69 – 74, Berlin, Germany, 2014.

[16] Potard, G. and Burnett, I., "Decorrelation Techniques for the Rendering of Apparent Sound Source Width in 3D Audio Displays," in *Proc of the 7th Int. Conference on Digital Audio Effects (DAFx'04)*, pp. 280 – 284, Naples, Italy, 2004.

[17] Weger, M., Marentakis, G., and Höldrich, R., "Auditory perception of spatial extent in the horizontal and vertical plane," in *Proc of the 19th International Conference on Digital Audio Effects (DAFx-16)*, Brno, Czech Republic, 2016.

[18] Kendall, G. S., "The Decorrelation of Audio Signals and its Impact on Spatial Imagery," *Computer Music Journal*, 19(4), pp. 71 – 87, 1995.

[19] Makita, Y., "On the directional localization of sound in the stereophonic sound field," Technical report, EBU, Review Part A, 73, 1962.

[20] Gerzon, M. A., "General metatheory of auditory localization," in *Proc. of the 92nd AES Convention*, Vienna, Austria, 1992.

[21] Frank, M., *Phantom Sources using Multiple Loudspeakers in the Horizontal Plane*, Dissertation, Institute of Electronic Music and Acoustics, Graz, Austria, 2013.

[22] Zotter, F. and Frank, M., "All-Round Ambisonic Panning and Decoding," *Journal of the Audio Engineering Society*, 60(10), pp. 807 – 820, 2012.

[23] Epain, N., Jin, C., and Zotter, F., "Ambisonic Decoding With Constant Angular Spread," *Acta Acustica united with Acustica*, 100, pp. 928 – 936, 2014.

[24] Carpentier, T., Noisternig, M., and Warusfel, O., "Twenty Years of Ircam Spat: Looking Back, Looking Forward," in *Proc. of the 41st International Computer Music Conference*, pp. 270 – 277, Denton, TX, USA, 2015.