Audio Engineering Society

Convention e-Brief 111

Presented at the Conference on
Immersive and Interactive Audio
2019 March 27 – 29, York, UK

SPARTA & COMPASS: Real-time implementations of linear and parametric spatial audio reproduction and processing methods

Leo McCormack[1] and Archontis Politis[1]

[1]*Aalto University, School of Electrical Engineering, Department of Signal Processing and Acoustics, Espoo, Finland*

Correspondence should be addressed to Leo McCormack (leo.mccormack@aalto.fi)

ABSTRACT

This paper summarises recent developments at the Acoustics Lab, Aalto University, Finland, regarding real-time implementations of both fundamental and advanced methods for spatialisation, production, visualisation, and manipulation of spatial sound scenes. The implementations can be roughly categorised into panning tools, for both binaural and arbitrary loudspeaker setups, and linear processing tools based on the Ambisonics framework; the latter of which includes: decoders for loudspeakers or headphones and sound scene activity visualisers, which are based on either non-parametric beamforming or parametric high-resolution methods. Additionally, more advanced reproduction and spatial editing tools are detailed, which are based on the parametric COMPASS framework.

Introduction

Recording, production, and reproduction of spatial sound scenes began with the invention of stereophony [1] and has been evolving ever since. The primary aims of spatial sound capture, panning, and playback, has been to offer high perceptual quality and engagement. In the case of recording, the objective is to preserve the spatial characteristics of the original scene as much as possible. To achieve these aims more recent approaches employ sophisticated digital signal processing techniques, which provide increased flexibility during both the recording and reproduction stages. This work presents and discusses a series of real-time implementations of spatial sound tools developed at the Acoustics Lab, Aalto University, Finland, which have been made publicly available as an open-source VST audio plug-in suite, under the name, Spatial Audio Real-time Applications (SPARTA)[1] [2].

Many of the plug-ins are implemented in the time-frequency domain, which allows spectral processing, an uncommon feature compared to existing spatialisation tools, which are often implemented solely in the time-domain. Each plug-in focuses on a certain application. The most fundamental operation is spatialisation of monophonic sound source signals, which has been implemented for loudspeaker and headphone playback by employing amplitude panning or binaural filters, respectively. With regard to recording spa-

[1]The SPARTA & COMPASS plug-in suites can be downloaded from here: http://research.spa.aalto.fi/projects/sparta_vsts/

[2]The SPARTA source code (GPLv3 license) can be found here: https://github.com/leomccormack/SPARTA

tial sound scenes, spherical microphone arrays are often employed, followed by suitable spatial encoding, which converts the microphone signals into spherical harmonic (SH) signals (also referred to as Ambisonic or B-format signals) [2, 3]. An implementation of such an encoder is demonstrated, which integrates a wide range of options, including control over: the microphone sensor directions, array size, baffle type, microphone directivity, and the encoding strategy [4, 5, 6, 7, 8]. Additionally, the plug-in can visualise the encoding performance per frequency, based on a few established metrics [4, 6].

Ambisonics offers a unified framework for linear and non-parametric storage, manipulation and reproduction of spatial sound scenes, formulated in the spherical harmonic domain (SHD). Basic ambisonic operations, such as encoding a monophonic sound source signal at a certain direction or rotating the sound scene [9], are also demonstrated. Due to the generality of SH encoding, decoding ambisonic material may be easily adapted to either: arbitrary loudspeaker setups via time-invariant real-valued gains, or matched to a set of binaural filters for generic or personalised headphone listening. Both loudspeaker and binaural decoders, based on multiple decoding strategies [10, 11, 12], have been implemented. Furthermore, due to the frequency-dependent nature of the plug-ins, different decoders may be assigned at different frequency ranges, providing the ability to independently optimise the low- and high-frequency localisation.

More recently, a set of non-linear and signal-dependent spatial audio processing methods have been developed. These methods aim to achieve higher perceptual accuracy in reproduction and also permit the manipulation of the spatial properties of a recording, in a manner not possible with the traditional linear Ambisonics framework. In this case, the approach relies on the estimation of spatial parameters that correspond to the sound scene. These methods are often referred to as being *parametric*, and were traditionally associated with coding and compression of spatial sound. However, parametric methods have also been applied to SH signals for the purpose of reproduction [13, 14, 15, 16, 17], and have been proven to offer high perceptual quality, even with low spatial resolution material. A recent framework for such parametric spatial audio processing is the: Coding and Multi-directional Decomposition of Ambisonic Sound Scenes (COMPASS) framework [18]. Real-time implementations of loudspeaker and headphone decoders for ambisonic material, employing the COMPASS framework, are also described. Unlike traditional linear Ambisonic decoders, these parametric alternatives feature enhancement controls and advanced parametric processing options; such as re-balancing of the dry and reverberant components in the spatial recording. Furthermore, the framework has also been employed for the up-mixing of lower-order ambisonic material to higher-order material, and suppressing or amplifying sounds from various regions in the sound scene.

1 General implementation details

Both the SPARTA and COMPASS plug-in suites employ the Spatial_Audio_Framework[3] and JUCE, for the internal processing and graphical user interfaces (GUI), respectively. For the plug-ins which utilise SH signal input and/or output, the Ambisonic Channel Number (ACN) ordering convention is used. In addition both orthonormalised (N3D) and semi-normalised (SN3D) schemes are supported; note that ACN/SN3D is more popularly referred to as the *AmbiX* format. The maximum transform order for the relevant SPARTA plug-ins is 7th order, whereas for COMPASS plug-ins, the maximum input order is 3rd order. Furthermore, all plug-ins that feature a binaural output option, permit the user to import their own head-related impulse responses (HRIRs); in order to attain personalised headphone listening; via the SOFA (Spatially Oriented Format for Acoustics) standard. Additionally, plug-ins that allow the user to specify the directions for the loudspeakers, sources or sensors, support the saving and loading of the directions via JSON (JavaScript Object Notation) configuration files. Source directions may also be controlled via suitable host automation. The plug-ins also display warning messages, should some aspects of the current configuration not meet the internal processing requirements.

For the time-frequency transform, the alias-free short-time Fourier transform (afSTFT)[4] implementation by Juha Vilkamo (which is optimised to be robust to temporal artefacts during aggressive time frequency domain manipulations) is employed [19]. Furthermore, due to the heavy use of linear algebra operations, the code

[3]The Spatial_Audio_Framework (ISC license) can be found here: https://github.com/leomccormack/Spatial_Audio_Framework

[4]The alias-free STFT (MIT license) can be found here: https://github.com/jvilkamo/afSTFT

also conforms to the relevant BLAS [20] and LAPACK [21] standards, for which Intel's MKL implementations are employed for both Windows (64-bit) and Mac OSX (10.10 or above) versions.

2 The SPARTA Plug-in suite

The SPARTA plug-in suite offers an array of fundamental tools for spatialisation, production and visualisation of spatial sound scenes. It includes elementary binaural and loudspeaker spatialisers, and processing tools based on the Ambisonics framework. Regarding the latter, plug-ins are included to address sound-field reproduction, rotation, spatial encoding, visualisation, and frequency-dependent dynamic range compression.

2.1 SPARTA | AmbiBIN

The AmbiBIN plug-in is a binaural ambisonic decoder for headphone playback of SH signals (see Fig. 1). It employs a virtual loudspeaker approach [22] and offers a max_rE weighting option [11]; which concentrates the beamformer signal energy towards the look-direction of a given virtual loudspeaker. This weighting option also serves to reduce the side-lobes of the directional spread inherent in Ambisonics, for a source emanating from any direction, and concentrates the directional energy in an optimal manner. Furthermore, the number of virtual loudspeakers increases with the transform order and are uniformly distributed. The plug-in also features a built-in rotator and head-tracking support, which may be realised by sending appropriate OSC messages to a user defined OSC port. The rotation angles and matrices [9] are updated after the time-frequency transform, which allows for reduced latency compared to its loudspeaker counterpart (AmbiDEC) paired with the Rotator plug-in.

2.2 SPARTA | AmbiDEC

AmbiDEC is a frequency-dependent ambisonic decoder for loudspeaker playback (see Fig. 1). The loudspeaker directions can be user-specified for up to 64 channels, or alternatively, presets for popular 2D and 3D set-ups can be selected. The plug-in also features a built-in binauraliser. This allows the user to audition the loudspeaker output via convolution of the individual loudspeaker signals with interpolated head-related transfer functions (HRTFs) which correspond to each loudspeaker direction.

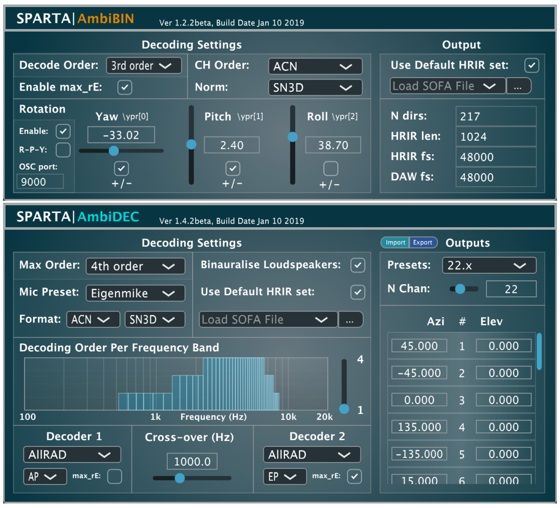


Fig. 1: Linear ambisonic decoders for headphones (top) and loudspeakers (bottom).

The plug-in employs a dual decoding approach, whereby different decoder settings may be selected for the low and high frequencies; the cross-over frequency may also be dictated by the user. Several ambisonic decoders have been integrated, including recent perceptually-motivated approaches, such as the All-Round Ambisonic Decoder (AllRAD) [11] and Energy-Preserving Ambisonic Decoder (EPAD) [12]. The max_rE weighting [11] may also be enabled for either decoder. Furthermore, in the case of non-ideal SH signals as input (i.e. those that are derived from real microphone arrays), the decoding order may be specified for the appropriate frequency ranges. In these cases, energy-preserving (EP) or amplitude-preserving (AP) normalisation may be selected to help maintain consistent loudness between decoding orders.

2.3 SPARTA | AmbiDRC

The AmbiDRC plug-in is a frequency-dependent SHD dynamic range compressor, based on the design presented in [23]. The gain factors are derived by analysing the omni-directional (zeroth order) component for each frequency band, which are then applied to all of the input SH components. The spatial properties of the original signals remain unchanged. The implementation also stores the frequency-dependent gain factors, derived from the omni-directional component, over time and plots them on the GUI to provide the user with visual feedback (see Fig. 2).

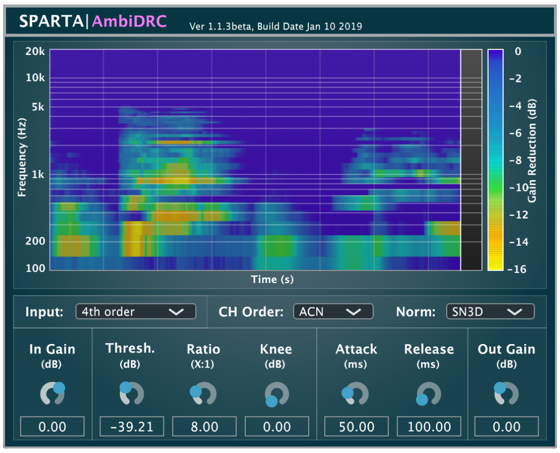


Fig. 2: Frequency-dependent DRC for SH signals.

2.4 SPARTA | AmbiENC

AmbiENC is an elementary ambisonic encoder, which accepts multiple monophonic input signals (up to 64 channels) and spatially encodes them into SH signals at specified directions (see Fig. 3). Essentially, these SH signals describe a synthetic anechoic sound-field, where the spatial resolution of the encoding is determined by the transform order. Several presets have been integrated into the plug-in for convenience (which allow for 22.x etc. audio to be encoded into 1-7th order SH signals, for example). The panning window is also fully mouse driven, and uses an equirectangular representation of the sphere to depict the azimuth and elevation angles of each source.

2.5 SPARTA | Array2SH

The Array2SH plug-in, originally presented in [24], spatially encodes microphone array signals into SH signals (see Fig. 3). The plug-in utilises analytical solutions, which ascertain the frequency and order-dependent influence that the array has on the initial estimate of the SH signals. The plug-in allows the user to specify: the array type (spherical or cylindrical), whether the array has an open or rigid enclosure, the radius of the baffle, the radius of the array (in cases where the sensors protrude out from the baffle), the sensor coordinates (up to 64 channels), sensor directivity (omni-dipole-cardioid), and the speed of sound. The plug-in then determines order-dependent equalisation curves, which need to be imposed onto the initial SH signals estimate, in order to remove the influence

Fig. 3: Monophonic (top) and microphone array (bottom) ambisonic encoders.

of the array itself [25, 26]. However, especially for higher-orders, this often results in large scale amplification of the low frequencies (including the sensor noise at these frequencies that accompanies it); therefore, two popular regularisation approaches have been integrated into the plug-in, which allow the user to make a compromise between noise amplification and transform accuracy [4, 27]. The target and regularised equalisation curves are depicted on the GUI to provide visual feedback. The plug-in also offers a filter design option based on band-passing the individual orders with a linear-phase filterbank, and subsequently shaping their combined encoding directional response with max_rE weights and diffuse-field equalisation, as proposed by Franz Zotter in [28]; this design can be more suitable for reproduction purposes. In addition, for high frequencies that exceed the spatial aliasing limit, where directional encoding fails, all filter designs have the option to impose diffuse-field equalisation on the aliased components, in order to avoid undesirable boosting of high-frequencies.

For convenience, the specifications of several commercially available microphone arrays have been integrated as presets; including: MH Acoustic's Eigenmike, the Zylia array, and various A-format microphone arrays. However, owing to the flexible nature of the plug-in, users may also design their own customised microphone array, while having a convenient means of obtaining the corresponding SH signals. For example, a four capsule open-body hydrophone array was presented in [29], which utilised the plug-in as the first step in visualising and auralising an underwater sound scene in real-time. Furthermore, a 19 sensor modular microphone array was recently developed and presented in [30], which featured MEMS microphones that protruded from the surface of a rigid scatterer via variable length stilts. This allowed control over the sensor radii and, in turn, the ability to more readily optimise the array for a given target application.

The plug-in can also depict two objective metrics, as described in [6, 31], which correspond to the spatial performance of the array in question, namely: the spatial correlation and the diffuse level difference. During this performance analysis, the encoding matrices are first applied to a simulated array, which is described by multi-channel transfer functions of plane waves for 812 points on the surface of the array. Ideally, the resulting encoded array responses should exactly resemble the spherical harmonic functions at each grid point; any deviations from this would suggest a loss in encoding performance. The spatial correlation is a metric that describes how similar the generated patterns are to the ideal SH patterns. A value of 1 indicates that the patterns are perfect, and a value of 0 means that they are uncorrelated. Therefore, the spatial aliasing frequency can be observed for each order, as the point where the spatial correlation tends towards 0. The level difference is then determined as the difference between the mean level over all directions (i.e. the energy sum of the encoded response over all points) and the ideal components. One can observe that higher permitted amplification limits, *Max Gain (dB)*, will result in a wider frequency range of useful spherical harmonic components; however, this will also result in noisier signals.

Note that this ability to balance the noise amplification with the accuracy of the spatial encoding (to better suit a given application) has some important ramifications. For example: the perceived fidelity of ambisonic decoded audio can often be degraded, if the noise amplification is set too high. However, for sound-field visualisation, or for beamformers which employ appropriate post-filtering, this noise amplification is often less problematic; therefore, such applications may benefit from a more aggressive encoding setting.

Fig. 4: Elementary binaural (top) and loudspeaker (bottom) spatialisers.

2.6 SPARTA | Binauraliser

The Binauraliser plug-in convolves input audio (up to 64 channels) with interpolated HRTFs, in the time-frequency domain (see Fig. 4). The plug-in first extracts the inter-aural time differences (ITDs) for each of the default or imported HRIRs, via the cross-correlation between the low-pass filtered left and right time-domain responses. The HRTFs are then interpolated at run-time by applying triangular-spherical interpolation on the HRTF measurement grid, which may be equivalently realised through the application of amplitude-normalised VBAP gains [32]. Note that these interpolation gains are applied to the HRTF magnitude responses and ITDs separately, which are then re-combined per frequency band. Presets for popular 2D and 3D formats are included for convenience; however, the directions for up to 64 channels can be independently controlled. Head-tracking is also supported via OSC messages in the same manner as the Rotator and AmbiBIN plug-ins.

2.7 SPARTA | Panner

Panner is a frequency-dependent 2D and 3D panning plug-in, based on the Vector-base Amplitude Panning (VBAP) method [32] or Multiple-Direction Amplitude Panning (MDAP) [33]. The latter is employed when the *spread* parameter is defined to be more than 0 degrees (see Fig. 4). Presets for popular 2D and 3D formats are included for convenience; however, the directions for up to 64 channels can be independently controlled for both inputs and outputs; allowing, for example, 9.x input audio to be panned for a 22.2 setup. The panning is frequency-dependent to accommodate the method described in [34], which allows more consistent loudness of sources, when they are panned in-between the loudspeaker directions, under different playback conditions. One should set the *pValue Coeff* parameter to 0 for a standard room with moderate reverberation, 0.5 for a low-reverberation listening room, and 1 for an anechoic chamber.

2.8 SPARTA | PowerMap

The Powermap plug-in depicts the relative sound-field activity at multiple directions, using a colour-map and an equirectangular representation of the sphere (see Fig. 5). The design is based on the software described in [35]. The colour yellow indicates higher sound energy or a higher likelihood of a source emanating from a particular direction, and blue indicates lower sound energy or likelihood.

The plug-in integrates a variety of different localisation functions, all of which are written to operate in the SHD, including: beamformer-based approaches, such as Plane-Wave Decomposition (PWD) [36] and Minimum-Variance Distortion-less Response (MVDR) [37]. Also catered for are: subspace-based approaches, such as Multiple Signal Classification (MUSIC) [25] and Minimum-Norm algorithms [37]. The Cross-Pattern Coherence (CroPaC) algorithm is also featured [38], which employs an iterative side-lobe suppression operation, as described in [35]. Furthermore, the analysis order per frequency band is user definable. Additionally, presets for commercially available higher-order microphone arrays have been included, which also provide approximate starting values, based on theoretical objective analysis of the arrays [4, 6]. The plug-in employs a 812 point uniformly-distributed spherical grid, which is interpolated to attain a 2D image using amplitude-normalised VBAP gains [32] (i.e. triangular-spherical interpolation).

2.9 SPARTA | Rotator

The Rotator plug-in applies a SHD rotation matrix [9] to the input SH signals (see Fig. 6). The rotation angles can be controlled using a head-tracker and OSC messages. The head-tracker should be configured to send a three-element vector, *ypr*[3], to OSC port 9000 (default); where *ypr*[0], *ypr*[1], *ypr*[2] correspond to the yaw-pitch-roll angles, respectively. The sign of the

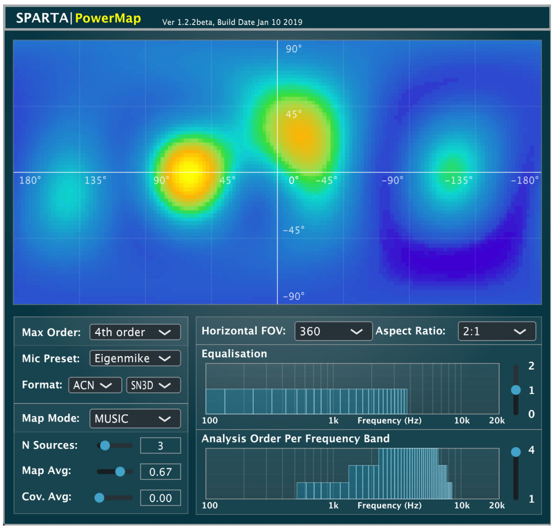


Fig. 5: Ambisonic sound-field visualiser.

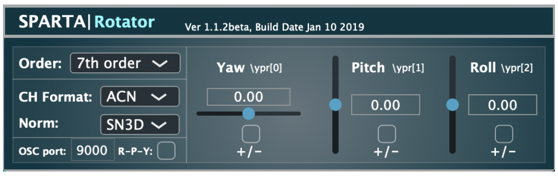


Fig. 6: Ambisonic rotator.

angles may also be flipped, $+/-$, in order to support a wider range of devices. The rotation order [yaw-pitch-roll (default) or roll-pitch-yaw] may also be user specified.

2.10 SPARTA | SLDoA

SLDoA is a spatially-localised direction-of-arrival (DoA) estimator (see Fig. 7); originally presented in [24]. The plug-in first employs VBAP beam patterns (for directions that are uniformly distributed on the surface of a sphere) to approximate spatially-biased zeroth and first-order signals in a least-squares sense. These are subsequently utilised for the active-intensity vector estimation [39]. This allows for DoA estimation in several spatially-localised sectors for each sub-band, whereby estimates made within one sector have reduced susceptibility to interferers present in other sectors.

The low frequency DoA estimates are depicted with blue icons, mid-frequencies with green, and high-frequencies with red. The size of the icon and its

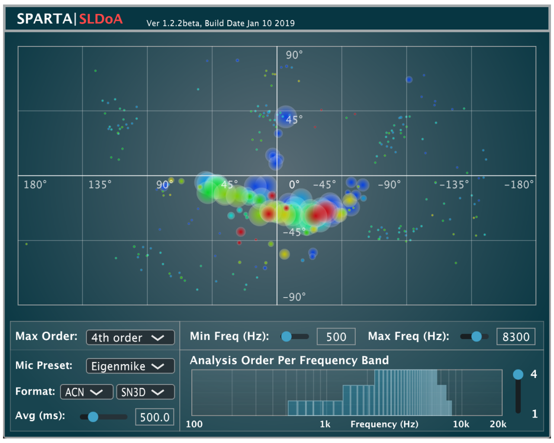


Fig. 7: The SLDoA sound-field visualiser.

opacity correspond to the sector energies, which are normalised for each frequency band. The plug-in employs twice the number of sectors as the analysis order, with the exception of the first-order analysis, which derives the DoA estimate from the traditional active-intensity vector. The analysis order per frequency band, and the frequency range at which to perform the analysis, are also user definable in a similar manner as to the Powermap plug-in. This approach to sound-field visualisation and/or DoA estimation may represent a more computationally efficient alternative, when compared to the algorithms that are integrated into the Powermap plug-in, for example.

3 The COMPASS plug-in suite

Ambisonic decoding is traditionally defined as a linear, time-invariant and signal-independent operation. In the simplest case, reproduction is achieved with a single frequency-independent matrix consisting of real-valued gains. Operations which modify the sound scene may also be conveniently defined in the SHD, in a linear fashion. The operations performed on the SH signals may be simple order-preserving exercises: such as mirroring and rotating. However, more complex operations, may also be carried out, such as directional loudness modifications, which returns signals of a higher-order than that of the original [40].

An alternative to the linear Ambisonics ecosystem is to apply a parametric approach, which is also applicable to the tasks of decoding, modification, and reproduction of multi-channel recordings. Parametric approaches operate in the time-frequency domain and are signal-dependent. They rely on the extraction of spatial parameters from the multi-channel input, which they subsequently utilise to achieve a more informed decoding, or to allow advanced sound scene manipulations.

Parametric methods also have the potential to offer *super-resolution*, in the sense that, even for small compact arrays of a few microphones, they can estimate and reproduce sounds with a spatial resolution that far surpasses the inherent limitations of non-parametric methods. Another advantage of the estimation, is that they perform an "objectification" of the sound scene. As an intermediate representation, instead of signals describing the sound scene as a whole, they produce signals and parameters which describe the scene content; such as directional sounds and their direction, ambient sound, and the energy relations between them. This information provides an intuitive representation of the sound scene and can be used to flexibly manipulate or modify the spatial sound content [41].

However, while parametric approaches may yield enhanced flexibility and spatial resolution, they are also associated with their own limitations. For instance, rather than elementary linear filtering, adaptive spectral processing with dedicated multi-channel analysis and synthesis stages is required. This, therefore, renders them more computationally demanding than their non-parametric counterparts. Furthermore, additional effort must be afforded in the implementation of these methods, in order to minimise spectral processing artefacts and/or musical noise. Finally, the methods must also be designed in such a manner, as to be robust to estimation errors and to provide a perceptually acceptable output; this includes cases where the input sound scene deviates wildly from the assumptions made in the parametric sound-field model. Of the parametric methods operating on SH signals, the most well-known example that has demonstrated robust performance is Directional Audio Coding (DirAC) [13, 17]. Originally intended for reproduction of first-order (FOA) signals, DirAC was later extended for higher-order (HOA) input [16]. Other methods include High-Angular Plane-wave Expansion (HARPEX) [14] for FOA signals, and the ambisonic sparse recovery framework of [15], for both FOA and HOA signals. For a comparison between these published parametric methods, the reader is referred to Fig. 8.

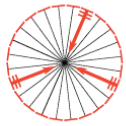
Method	Input	Model	
DirAC (Pulkki, 2006)	FOA (4ch)		1 source component + 1 iso. diffuse component
HARPEX (Berge, 2010)	FOA (4ch)		2 source components
HO-DirAC (Politis et.al., 2015)	HOA (9+ch)		~M sector source + ~M sector diffuse components
Sparse Recovery (Wabnitz, Jin, 2012)	FOA/HOA (4+ch)		≤M/2 source components
COMPASS	FOA/HOA (4+ch)		≤M/2 source components + spatial ambient component

Fig. 8: Comparison of parametric ambisonic processing methods. M in this case refers to the number of ambisonic channels for each order.

The parametric COMPASS framework for SH signals was recently proposed in [18]. It is based on both array processing and spatial filtering techniques, employing a sound-field model that comprises of multiple narrow-band source signals and a residual ambient component. The method is applicable to any input order, and offers more robust estimation and a greater number of extracted source signals with each increase in order. The ambient signal populates all the available SH channels and has its own time-variant spatial distribution, which predominately encapsulates the remaining reverberant and diffuse components. The source and ambient signals may then be spatialised or modified independently. A similar approach to COMPASS, formulated for FOA input only, has also been presented in [42].

3.1 COMPASS | Decoder

The Decoder plug-in[5] is a parametrically enhanced ambisonic decoder for arbitrary loudspeaker setups (see Fig. 9). The GUI for the plug-in offers much of the functionality provided by the AmbiDEC plug-in, but features additional frequency-dependent balance controls between the extracted direct and ambient components, and control between fully parametric decoding and linear ambisonic decoding.

The *Diffuse-to-Direct* control allows the user to lend more prominence to the direct sound components (an effect similar to subtle dereverberation), or to the ambient components (an effect similar to emphasising reverberation in the recording). When set in the middle, the two are balanced. Note that the parametric processing can be quite aggressive, and if the plug-in is set to fully direct rendering in a complex multi-source sound scene with low-order input signals; artefacts may appear. However, with more conservative settings, such artefacts should become imperceptible.

The *Linear-to-Parametric* control allows the user to dictate a mix between standard linear ambisonic decoding and the COMPASS parametric decoding. This control

[5] A video demo of the COMPASS Decoder can be found here: http://research.spa.aalto.fi/projects/compass_vsts/plugins.html

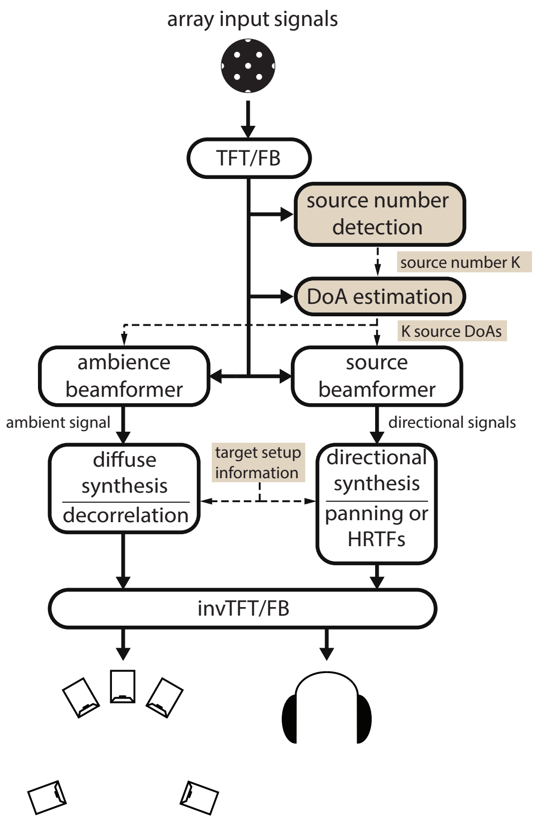


Fig. 9: COMPASS processing flow chart for loudspeaker or headphone reproduction.

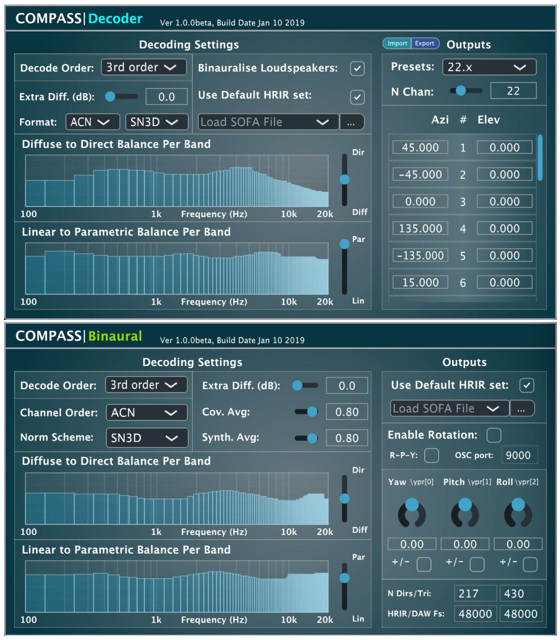


Fig. 10: Parametric Ambisonic decoder plug-ins.

can be used in cases where parametric processing is too aggressive, or if the user prefers some degree of localisation blur offered by linear ambisonic decoding.

The plug-in is considered by the authors as a production tool and, due to its advanced time-frequency processing, requires large audio buffer sizes. Therefore, the plug-in may not be suitable for interactive input with low-latency requirements. For applications such as interactive binaural rendering with head-tracking, the Binaural variant described below is more appropriate.

3.2 COMPASS | Binaural

The Binaural decoder is an optimised version of the COMPASS Decoder for binaural playback, which bypasses loudspeaker rendering and uses HRTFs directly instead. For the plug-in parameters, refer to the description of the COMPASS Decoder above. Additionally, the plug-in can receive OSC rotation angles from a head-tracker at a user specified port, as detailed in the Rotator plug-in description.

This decoder is intended primarily for head-tracked binaural playback of ambisonic content at interactive update rates, usually in conjunction with a head-mounted display (HMD). The plug-in requires an audio buffer size of at least 512 samples (approximately 10msec at 48kHz). The averaging parameters may be employed to influence the responsiveness of the parametric analysis and synthesis, thus providing the user with a means to adjust them optimally for a particular input sound scene.

3.3 COMPASS | Upmixer

The Upmixer plug-in employs COMPASS for the task of up-mixing a lower-order ambisonic recording to a higher-order ambisonic recording. It is intended for users that are already working with a preferred linear ambisonic decoding work-flow of higher-order content, and wish to combine lower-order material with increased spatial resolution. The user may up-mix first, second, or third-order material up to seventh-order material (64 channels).

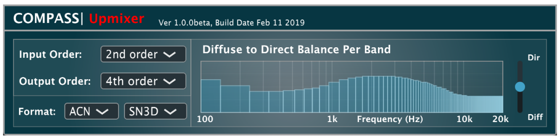


Fig. 11: Parametric Upmixer plug-in.

3.4 COMPASS | SpatEdit

The SpatEdit plug-in is a parametric spatial editor, which allows the user to emphasise or attenuate specific directional regions in the sound scene. This plug-in uses the analysis part of the COMPASS framework, namely the multi-source DoA estimates, and the ambient component; in order to extract or modify sounds which emanate from user-defined spatial targets. These targets are controlled via markers, which may be placed at arbitrary angles on the azimuth-elevation plane. The number of markers is dictated by the transform order. Higher-orders allow the placement of more markers and hence finer spatial control. Additionally, the user may specify a gain value applied to each target, which can amplify, attenuate, or attempt to eliminate sound incident from that direction. In the case of the linear processing mode, a reasonable degree of spatial enhancement or separation can be attained with zero distortion; especially at higher-orders. However, should the user wish to conduct more aggressive manipulations of the sound scene, the plug-in may also operate in the parametric mode. In this case, the target source signals can become more spatially-selective, by taking into account the analysed parameters. For each target, an additional enhancement value can be specified, this essentially determines how spatially-selective the parametric processing may aspire to be. Higher values can separate and modify the sounds coming from the targets to a much greater degree, but with the undesirable possibility of introducing more artefacts. Since artefacts in parametric processing are always scene-dependent, the user is then responsible for ascertaining which settings work most optimally for a given input scene.

The plug-in may be configured to either output the individual beams and the residual; or it can be set to output the manipulated scene in the SHD via suitable re-encoding. This allows the user to process sounds from individual regions separately (e.g. apply equalisation), or receive the modified scene re-packaged conveniently in the input ambisonic format for immediate rendering or further ambisonic processing.

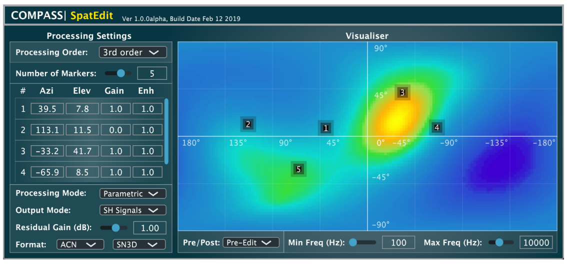


Fig. 12: Parametric spatial editor plug-in.

Acknowledgements

We would like to thank all of the Aalto Acoustics Lab researchers who have contributed, or supported the creation of these plug-in suites, namely Dr. Symeon Delikaris-Manias, Dr. Sakari Tervo, Prof. Ville Pulkki, and Prof. Tapio Lokki.

Our many thanks are also extended to Daniel Rudrich and Dr. Franz Zotter from the Institute for Electronic Music and Acoustics (IEM), Gratz, Austria, for their cross-collaborative efforts. Special thanks are also extended to all those who submitted bug reports during the beta releases of the plug-ins.

References

[1] Blumlein, A. D., “British Patent No. 394325,” *Improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems*, 1931.

[2] Rafaely, B., *Fundamentals of spherical array processing*, Springer, 2015.

[3] Jarrett, D. P., Habets, E. A., and Naylor, P. A., *Theory and applications of spherical microphone array processing*, Springer, 2017.

[4] Moreau, S., Daniel, J., and Bertet, S., “3D sound field recording with higher order ambisonics–Objective measurements and validation of a 4th order spherical microphone,” in *120th Convention of the AES*, pp. 20–23, 2006.

[5] Jin, C. T., Epain, N., and Parthy, A., “Design, optimization and evaluation of a dual-radius spherical microphone array,” *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 22(1), pp. 193–204, 2014.

[6] Politis, A. and Gamper, H., "Comparing modeled and measurement-based spherical harmonic encoding filters for spherical microphone arrays," in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2017 IEEE Workshop on*, pp. 224–228, IEEE, 2017.

[7] Alon, D. L. and Rafaely, B., "Spatial decomposition by spherical array processing," in *Parametric time-frequency domain spatial audio*, pp. 25–48, Wiley Online Library, 2018.

[8] Schörkhuber, C. and Höldrich, R., "Ambisonic microphone encoding with covariance constraint," in *Proceedings of the International Conference on Spatial Audio*, 2017.

[9] Ivanic, J. and Ruedenberg, K., "Rotation Matrices for Real Spherical Harmonics. Direct Determination by Recursion," *The Journal of Physical Chemistry A*, 102(45), pp. 9099–9100, 1998.

[10] Gerzon, M. A., "Periphony: With-height sound reproduction," *Journal of the Audio Engineering Society*, 21(1), pp. 2–10, 1973.

[11] Zotter, F. and Frank, M., "All-round ambisonic panning and decoding," *Journal of the audio engineering society*, 60(10), pp. 807–820, 2012.

[12] Zotter, F., Pomberger, H., and Noisternig, M., "Energy-preserving ambisonic decoding," *Acta Acustica united with Acustica*, 98(1), pp. 37–47, 2012.

[13] Pulkki, V., "Spatial sound reproduction with directional audio coding," *Journal of the Audio Engineering Society*, 55(6), pp. 503–516, 2007.

[14] Berge, S. and Barrett, N., "High angular resolution planewave expansion," in *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics May*, pp. 6–7, 2010.

[15] Wabnitz, A., Epain, N., McEwan, A., and Jin, C., "Upscaling ambisonic sound scenes using compressed sensing techniques," in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*, pp. 1–4, IEEE, 2011.

[16] Politis, A., Vilkamo, J., and Pulkki, V., "Sector-based parametric sound field reproduction in the spherical harmonic domain," *IEEE Journal of Selected Topics in Signal Processing*, 9(5), pp. 852–866, 2015.

[17] Politis, A., McCormack, L., and Pulkki, V., "Enhancement of ambisonic binaural reproduction using directional audio coding with optimal adaptive mixing," in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2017 IEEE Workshop on*, pp. 379–383, IEEE, 2017.

[18] Politis, A., Tervo, S., and Pulkki, V., "COMPASS: Coding and Multidirectional Parameterization of Ambisonic Sound Scenes," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6802–6806, IEEE, 2018.

[19] Vilkamo, J. and Backstrom, T., "Time–Frequency Processing: Methods and Tools," in *Parametric Time-Frequency Domain Spatial Audio*, pp. 3–23, John Wiley & Sons, 2018.

[20] Hanson, R. J., Krogh, F. T., and Lawson, C., "A proposal for standard linear algebra subprograms," 1973.

[21] Anderson, E., Bai, Z., Dongarra, J., Greenbaum, A., McKenney, A., Du Croz, J., Hammarling, S., Demmel, J., Bischof, C., and Sorensen, D., "LAPACK: A portable linear algebra library for high-performance computers," in *Proceedings of the 1990 ACM/IEEE conference on Supercomputing*, pp. 2–11, IEEE Computer Society Press, 1990.

[22] Bernschütz, B., Giner, A. V., Pörschmann, C., and Arend, J., "Binaural reproduction of plane waves with reduced modal order," *Acta Acustica united with Acustica*, 100(5), pp. 972–983, 2014.

[23] McCormack, L. and Välimäki, V., "FFT-based dynamic range compression," in *Sound and Music Computing Conference*, 2017.

[24] McCormack, L., Delikaris-Manias, S., Farina, A., Pinardi, D., and Pulkki, V., "Real-Time Conversion of Sensor Array Signals into Spherical Harmonic Signals with Applications to Spatially Localized Sub-Band Sound-Field Analysis," in *Audio Engineering Society Convention 144*, Audio Engineering Society, 2018.

[25] Teutsch, H., *Modal array signal processing: principles and applications of acoustic wavefield decomposition*, Springer, 2007.

[26] Williams, E. G., *Fourier acoustics: sound radiation and nearfield acoustical holography*, Academic press, 1999.

[27] Bernschütz, B., Pörschmann, C., Spors, S., Weinzierl, S., and der Verstärkung, B., "Soft-limiting der modalen amplitudenverstärkung bei sphärischen mikrofonarrays im plane wave decomposition verfahren," *Proceedings of the 37. Deutsche Jahrestagung für Akustik (DAGA 2011)*, pp. 661–662, 2011.

[28] Zotter, F., "A linear-phase filter-bank approach to process rigid spherical microphone array recordings," in *Proc. IcETRAN*, Palic, Serbia, 2018.

[29] Delikaris-Manias, S., McCormack, L., Huhtakallio, I., and Pulkki, V., "Real-time underwater spatial audio: a feasibility study," in *Audio Engineering Society Convention 144*, Audio Engineering Society, 2018.

[30] González, R., Pearce, J., and Lokki, T., "Modular Design for Spherical Microphone Arrays," in *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*, Audio Engineering Society, 2018.

[31] Bertet, S., Daniel, J., and Moreau, S., "3D sound field recording with higher order ambisonics-objective measurements and validation of spherical microphone," in *Audio Engineering Society Convention 120*, Audio Engineering Society, 2006.

[32] Pulkki, V., "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *Journal of Audio Engineering Society*, 45(6), pp. 456–466, 1997.

[33] Pulkki, V., "Uniform spreading of amplitude panned virtual sources," in *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA'99 (Cat. No. 99TH8452)*, pp. 187–190, IEEE, 1999.

[34] Laitinen, M.-V., Vilkamo, J., Jussila, K., Politis, A., and Pulkki, V., "Gain normalization in amplitude panning as a function of frequency and room reverberance," in *Audio Engineering Society Conference: 55th International Conference: Spatial Audio*, Audio Engineering Society, 2014.

[35] McCormack, L., Delikaris-Manias, S., and Pulkki, V., "Parametric acoustic camera for real-time sound capture, analysis and tracking," in *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17)*, pp. 412–419, 2017.

[36] Jarrett, D. P., Habets, E. A., and Naylor, P. A., "3D source localization in the spherical harmonic domain using a pseudointensity vector." in *EUSIPCO*, 2010.

[37] Van Trees, H. L., *Optimum array processing: Part IV of detection, estimation, and modulation theory*, John Wiley & Sons, 2004.

[38] Delikaris-Manias, S. and Pulkki, V., "Cross pattern coherence algorithm for spatial filtering applications utilizing microphone arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, 21(11), pp. 2356–2367, 2013.

[39] Fahy, F. J. and Salmon, V., "Sound intensity," *The Journal of the Acoustical Society of America*, 88(4), pp. 2044–2045, 1990.

[40] Kronlachner, M., *Spatial transformations for the alteration of ambisonic recordings*, Master's thesis, Institute of Electronic Music and Acoustics, University of Music and Performing Arts, Graz, Graz, Austria, 2014.

[41] Politis, A., Pihlajamäki, T., and Pulkki, V., "Parametric spatial audio effects," *York, UK, September*, 2012.

[42] Kolundzija, M. and Faller, C., "Advanced B-Format Analysis," in *Audio Engineering Society Convention 144*, Audio Engineering Society, 2018.