Aalto University

Aaltodoc
OPEN ACCESS

This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Author(s): Archontis Politis, Tapani Pihlajamäki, Ville Pulkki

Title: Parametric Spatial Audio Effects

Year: 2012

Version: Final published version

Please cite the original version:

Archontis Politis, Tapani Pihlajamäki, Ville Pulkki. Parametric Spatial Audio Effects. In 15th International Conference on Digital Audio Effects (DAFx-12), York, UK, September 2012.

Rights: © 2012 Authors. Reprinted with permission.

This publication is included in the electronic version of the article dissertation:
Politis, Archontis. Microphone array processing for parametric spatial audio techniques.
Aalto University publication series DOCTORAL DISSERTATIONS, 195/2016.

All material supplied via Aaltodoc is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

PARAMETRIC SPATIAL AUDIO EFFECTS

Archontis Politis, Tapani Pihlajamäki, Ville Pulkki

Department of Signal Processing and Acoustics
Aalto University School of Electrical Engineering
Espoo, Finland
archontis.politis@aalto.fi,
tapani.pihlajamaki@aalto.fi,
ville.pulkki@aalto.fi

ABSTRACT

Parametric spatial audio coding methods aim to represent efficiently spatial information of recordings with psychoacoustically relevant parameters. In this study, it is presented how these parameters can be manipulated in various ways to achieve a series of spatial audio effects that modify the spatial distribution of a captured or synthesised sound scene, or alter the relation of its diffuse and directional content. Furthermore, it is discussed how the same representation can be used for spatial synthesis of complex sound sources and scenes. Finally, it is argued that the parametric description provides an efficient and natural way for designing spatial effects.

1. INTRODUCTION

In the audio engineering and tonmeister disciplines, certain audio effects are used that aim to modify the directionality, spaciousness, reverberance or any other characteristics of a recording related to the spatial impression of the sound scene presented to the listener. We refer to this family of manipulations as spatial audio effects. A rough categorisation of them can be made as: a) manipulation of the directional information retained in the recording by means of the stererophonic or multichannel microphone technique used to capture it, b) adding reverberation in a recording or modifying the reverberant sound captured in it, and c) adding spatial cues to a monophonic sound to give it a spatial presence, meaning a perceived direction and spatial extent.

The first category includes directivity control in common stereophonic techniques, such as the stereo-image width control using mid-side stereophony, and generally all methods based on manipulation of directional patterns in order to modify the perceived spatial distribution of the recorded sound scene. A more comprehensive approach has been demonstrated by the ambisonic theory of recording and reproduction [1]. The local sound field information is captured using the B-format recording, or encoding, specification. The most well known ambisonic effects are rotation of the sound field around an arbitrary axis, and the dominance, an effect which amplifies the sound in a specific direction and attenuates gradually the sounds off-axis [2]. Another trivial operation is to orientate a first-order beam created from the B-format signals towards a specific direction, performing basic spatial filtering on the recorded sound scene. These transformations are realised with direct matrixing of the B-format signals.

The second category mentioned above includes all common reverberator designs, based either on physical or perceptual principles, and realised with FIR filters or simple recursive filter structures simulating the diffusion of sound in a reverberant space. The relation between the perceptual impression of reverberation and decorrelation of the source signal has been pointed out more than 40 years ago [3], [4]. Decorrelator designs suitable for reverberation have been presented by [5]. A similar approach are the feedback delay networks presented by [6], [7]. An approach to adapting artificial reverberation in the ambisonic domain has been presented by [8].

Finally, the third category is about modifying the perceived spatial extent of a sound, or apparent source width, which has been traditionally achieved by distributing the sound in different channels and then applying frequency dependent delays on it. In stereophonic reproduction this effect has been termed pseudostereo [9]. Another related technique is decorrelation applied to different channels and the source signal, in a controlled manner [5], [10], [11]. A more physical approach that can achieve a similar effect is to encode the source directivity in some kind of expansion functions, such as spherical harmonics. These functions translate into directional filters which, when applied to the loudspeakers, recreate properly the source extent at the listening spot [12], [13].

Stereophonic and multichannel microphone techniques retain directional information of the captured sound scene, as well as information related to the reverberation or coherence. This information can be manipulated in various ways in order to achieve spatial effects. Parametric spatial audio techniques such as Directional Audio Coding (DirAC) [14], encode this spatial information inherent in a recording or a multichannel mix into a set of parameters. This study presents various spatial effects that can be achieved by manipulation of the recording in the parametric domain. Using the physical analysis and synthesis of DirAC, it is demonstrated that by working in the parametric domain, a greater variety of effects can be achieved than by simple matrixing or direct time-domain methods. At the end of this paper, a link to a companion web page is provided where sound samples for the presented effects are provided.

2. DIRECTIONAL AUDIO CODING

Directional Audio Coding is a parametric method for a perceptually motivated representation, transmission, and reproduction of spatial sound. The model of DirAC assumes that with a time-frequency representation similar or finer to the resolution of the auditory system, it is adequate to encode and decode the local sound field with a reduced set of audio streams, compared to the output channels, and two parameters for each time-frequency tile. These are the direction of arrival (DOA) of incident sound energy and the diffuseness.

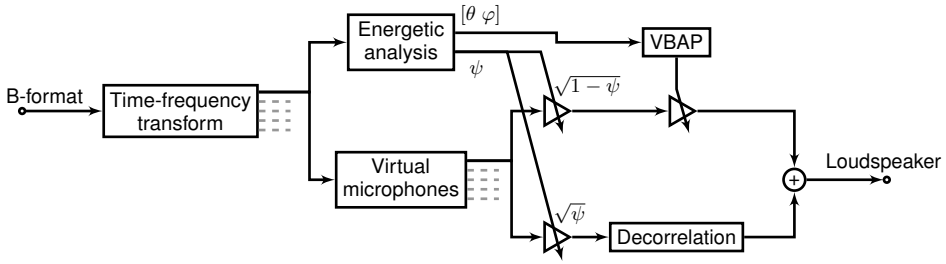

Figure 1: Block diagram of DirAC processing for a single sub-band and a single channel output. Dashed lines correspond to the rest of the sub-bands in the first stage and the rest of the speaker outputs in the second stage.

Direction-of-arrival is assumed to relate to the directional cues of localisation, while diffuseness relates to the sense of reverberation or sound source extent represented by interaural coherence. These parameters are used in the synthesis stage to recreate perceptually the captured sound scene.

2.1. DirAC analysis

The DirAC parameters are extracted with an energetic analysis based on the sound pressure $p(t)$ and particle velocity $\vec{u}(t)$ at the recording position. In the STFT domain, used in this implementation, the respective complex transformed quantities are $P(k,n)$ and $\vec{U}(k,n)$, where k,n are the frequency and time indices of the transform. In principle, DirAC can be used with any type of input permitting analysis of direction and diffuseness. However, the most common 3-dimensional input consists of one pressure, W, and three orthogonal pressure-gradient signals, X, Y, and Z, known in literature as the B-format signal set. Since an omnidirectional transducer captures a signal proportional to the sound pressure and an equalized pressure-gradient transducer captures a signal proportional to sound velocity, these physical quantities are related to the B-format signal set by the following relations

$$P(k,n) = W(k,n) \tag{1}$$

$$\vec{U}(k,n) = -\frac{1}{\sqrt{2}Z_0}\vec{X}'(k,n), \tag{2}$$

where $\vec{X}'(k,n) = [X(k,n)\ Y(k,n)\ Z(k,n)]^T$ is the vector of the B-format pressure-gradient signals and $Z_0 = c\rho_0$ is the characteristic impedance of air. The scaling of $\sqrt{2}$ in (2) is applied due to B-format convention.

For each time-frequency frame the sound field is assumed to be stationary and composed from a plane wave and a perfectly diffuse field. Then an estimate of the direction of the plane wave is given by the net energy flow, expressed by the active intensity vector [15]

$$\vec{I}_a(k,n) = \frac{1}{2}\Re\left\{P(k,n)\cdot\vec{U}(k,n)^*\right\}. \tag{3}$$

Using (1) and (2), $\vec{I}_a$ can be expressed in terms of the B-format signals as

$$\vec{I}_a(k,n) = -\frac{1}{2\sqrt{2}Z_0}\Re\left\{W(k,n)\cdot\vec{X}'(k,n)^*\right\}. \tag{4}$$

The direction of incidence of the plane wave is estimated as the opposite of the active intensity vector

$$\vec{u}_{\mathrm{DOA}}(k,n) = -\frac{\vec{I}_a(k,n)}{\|\vec{I}_a(k,n)\|} \tag{5}$$

or in terms of the B-format signals

$$\begin{aligned}\vec{u}_{\mathrm{DOA}}(k,n) &= \frac{\Re\left\{W(k,n)\cdot\vec{X}'(k,n)^*\right\}}{\left\|\Re\left\{W(k,n)\cdot\vec{X}'(k,n)^*\right\}\right\|} \\ &= \begin{bmatrix}\cos(\theta)\cos(\varphi)\\ \sin(\theta)\cos(\varphi)\\ \sin(\varphi)\end{bmatrix},\end{aligned} \tag{6}$$

where $\theta(k,n), \varphi(k,n)$ are the estimated azimuth and elevation of incidence respectively and $\|\cdot\|$ is the l^2-norm.

The energy density of the sound field at the same point is defined as [15]

$$E(k,n) = \frac{\rho_0}{4}\|\vec{U}(k,n)\|^2 + \frac{1}{4\rho_0 c^2}|P(k,n)|^2 \tag{7}$$

and in terms of the B-format signals

$$E(k,n) = \frac{1}{4\rho_0 c^2}\left[\frac{\|\vec{X}'(k,n)\|^2}{2} + |W(k,n)|^2\right]. \tag{8}$$

Finally, the diffuseness is defined as

$$\psi(k,n) = 1 - \frac{\|\langle\vec{I}_a(k,n)\rangle\|}{c\langle E(k,n)\rangle} \tag{9}$$

and in terms of the B-format signals

$$\psi(k,n) = 1 - \frac{\sqrt{2}\left\|\left\langle\Re\left\{W(k,n)\cdot\vec{X}'(k,n)^*\right\}\right\rangle\right\|}{\left\langle|W(k,n)|^2 + \|\vec{X}'(k,n)\|^2/2\right\rangle} \tag{10}$$

where $\langle\cdot\rangle$ denotes time averaging and $(^*)$ conjugation. Diffuseness is bounded by $\psi \in [0,1]$ with a value of 0 for a single plane wave, when the net transport of energy corresponds to the total energy density, and a value of 1 for a perfectly diffuse field, where the net

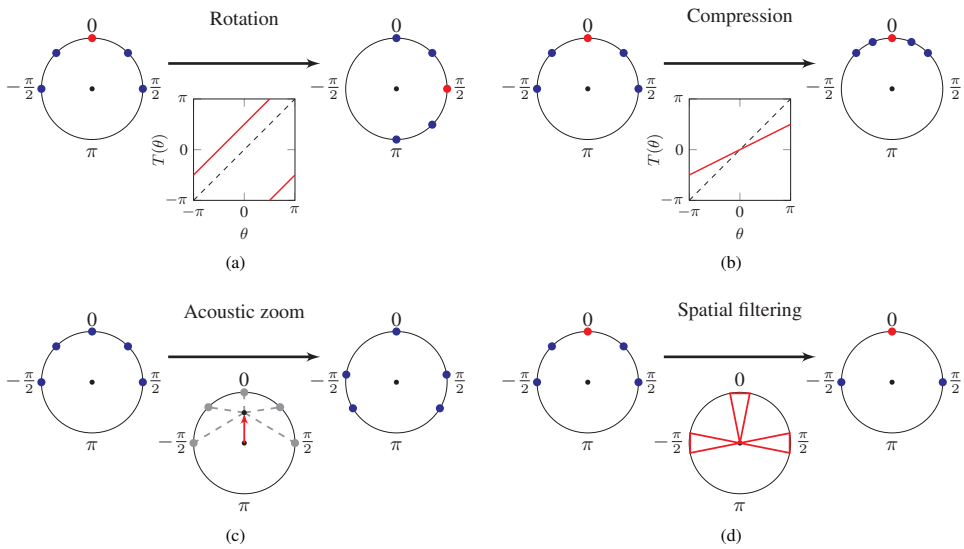

Figure 2: Various effects based on manipulation of the directional metadata.

energy transport is zero. The temporal averaging for the estimation of diffuseness is realised either with a symmetric FIR-filter or with a low-order IIR filter. The averaging is frequency dependent in such a way as to include several periods of the corresponding frequency.

2.2. DirAC synthesis

In the synthesis stage, diffuseness and DOA are used to generate two separate streams, the non-diffuse stream corresponding to the directional part of the recording and the diffuse stream corresponding to the diffuse part of the recording, as in Figure 1. Their respective mixing gain is determined by the diffuseness value. The directional part is then reproduced by means of vector-base amplitude panning [16] operating separately on each time-frequency bin or sub-band. A time-averaging stage is employed on the loudspeaker gains to reduce spurious changes between consecutive frames, which could induce perceivable artifacts in the audio output. The diffuse part is reproduced by means of a separate decorrelation filter per loudspeaker. The final output is then the sum of the two streams for each loudspeaker.

The audio input that is utilised in the synthesis can be either the omnidirectional signal from the B-format, especially in cases where low bit-rate transmission is crucial, or the complete B-format set. In the second case, before the standard DirAC synthesis, the audio is distributed to the loudspeakers by means of virtual microphones, before the VBAP and decorrelation stages. Use of virtual microphones result in better directional separation, natural decorrelation and overall higher audio quality compared to the monophonic case, in expense of increased bandwidth in the case of transmission. Similarly, with regards to parametric spatial processing, certain effects require the full B-format set and others utilise only a monophonic input.

DirAC is primarily a spatial sound coding method which is able to render a recorded sound scene to arbitrary loudspeaker layouts or headphones. However, due to the robustness of its physical basis and its relation to directional auditory perception, it has found many different applications such as upmixing [17], transcoding of different multichannel formats [18], spatial filtering [19] and dereverberation [20], among others.

3. PARAMETRIC SPATIAL AUDIO EFFECTS

This section presents applications of parametric analysis and synthesis, as it is performed in the DirAC framework, to spatial audio effects. As already mentioned, these effects are divided into three broad categories: a) effects that mainly modify the existing directional content of spatially encoded material or a recording, b) effects that modify the reverberation and ambience of a spatial sound recording, c) effects that synthesize spatial cues for monophonic sources, based on user-provided parameters or metadata. Certain effects are based on previous research on applications of DirAC to teleconferencing or virtual reality, such as the 3D projection effects, the spatial filtering and the spatial sound synthesis. Several new proposals are also provided, such as the angular compressions/expansions of the soundfield, the spatial modulation for recordings or impulse responses, the diffuse-field level control and the ambience extraction.

3.1. Modification of directional properties

Complex directional transformations

As each time-frequency bin is associated with an analyzed direction, it is trivial to apply linear directional transformations such as

rotation of the sound scene around some arbitrary axis, by directly modifying the estimated direction of arrival. However, it is more efficient to apply such rotations directly onto the B-format signals by multiplying the velocity components with appropriate standard rotation matrices. Nevertheless, the parametric domain offers more flexibility for non-linear transformations such as companding or expanding directions around some axis. In general, the directional transformation can be defined as a mapping $T(\theta, \varphi)$ from the original DOA estimate (θ, φ) to the modified one $(\theta', \varphi') = T(\theta, \varphi)$. After the transformation has been defined, its operation is applied on the metadata of each frequency bin or sub-band. It is also possible to define frequency-dependent mappings $T(\theta, \varphi, k)$ for more complex directional control.

Graphical control of the transformation can be performed with an angular representation of the source densities or as a transfer function between the input angle and the transformed angle, as in Figures 2(a) and 2(b). Rotation and companding of directions are given as examples.

Sweet-spot translation and zooming

More physical directional transformations can be defined based on specific application scenarios. One of them is the illusion of movement inside the single-point-recorded sound scene. A general strategy of achieving this effect has been presented by one of the authors in [21], where general purpose affine transformations of the perceived sound scene are described by projection of the analysed DOAs onto an arbitrary surface. Translation of the listening point results in a new spatial relation of the listener with respect to the projection surface, and consequently to a new set of DOAs. This procedure is depicted schematically in Figure 2(c). In addition to the directional transformation, a more coherent sense of translation can be achieved by simultaneous modification of the amplitudes of each time-frequency bin, following an inverse distance law from the reference distance of the projection surface. Schultz-Amling et al. [22] further proposed a respective modification of diffuseness, for a teleconferencing application termed acoustical zooming, where the effect is intended to follow automatically zooming of an actual camera in the same place.

The approaches above give the opportunity to the audio engineer to create novel spatial impressions. Applying just the directional transformation results in a perceptual sense similar to the optical analogy of narrowing or widening the field of view inside the recording. If the effect takes into account the modification of amplitude and diffuseness, then a true sense of proximity to sound sources can be created which means that the listening point can be virtually translated inside the recorded scene.

Spatial filtering

Another possible effect in the parametric domain is spatial filtering. Kallinger et al. [19] presented a DirAC-based method in the context of teleconferencing. Here a slightly modified implementation is described, intended as a spatial effect with more control given to the user for the effect parameters. That can be achieved by defining arbitrary beampatterns which result in time-frequency masks applied on each analysis frame. More specifically, in a 2-dimensional formulation, if spatial filtering by a beampattern $D(\theta)$ rotated at angle θ_0 is wanted, then the output S_{sf} of the effect is given as:

$$S_{\text{sf}}(k, n) = W(k, n)G(k, n) \tag{11}$$

where $W(k, n)$ is the omnidirectional signal at each bin or sub-band, and $G(k, n)$ is a real-valued time-frequency mask derived as:

$$G(k, n) = D(\theta_k - \theta_0) \tag{12}$$

where θ_k is the analysed DOA at sub-band k. The above formulation is directly applicable to the 3-dimensional case as well. The beampatterns can be designed based on common analytical formulas such as higher-order gradient patterns or spherical modal beamforming ones, or they can be described freely by the user by means of some suitable graphical interface.

Figure 2(d) depicts an extreme case of isolated sources and spatial-filters designed as rectangular directional windows, separating completely the sources. Even though narrow brick-wall spatial filters are possible, it is advantageous to use continuous patterns to avoid artifacts due to abrupt transitions in the spectral gains. Moreover, a portion of the diffuse sound should be added to the output. The diffuse stream in this case is generated from a first-order virtual microphone from the B-format signal, oriented at the direction of the spatial filter. The directivity of the beam can be altered by the user. To achieve correct power for the diffuse stream, it is scaled according to the power output of the spatial filter and the virtual microphone in a diffuse field, or similarly the ratio of their directivity factors Q_{sf} and Q_{vm}:

$$S_{\text{sf+diff}}(k, n) = W(k, n)G(k, n) + \sqrt{\frac{Q_{vm}}{Q_{sf}}} S_{\text{diff}}(k, n) \tag{13}$$

where S_{diff} is a decorrelated version of the virtual microphone signal. Addition of the diffuse stream helps in masking artifacts which may arise from fast fluctuations of directions inside and outside of the beam.

Spatial Modulation: Morphing

A novel way of modifying the directional (or diffuseness) parameters of one signal is to define a spatial modulation operator. This operator imprints the spatial cues of one signal (control) to the spatial sound of another signal (carrier). Within the parametric domain of DirAC, this is a relatively easy operation as the metadata of the control signal can be directly combined with the carrier signal before synthesis. Furthermore, instead of direct replacement, interpolation of the parameters of the two-signals can be performed to produce a mix between spatial cues. This procedure can achieve an effect similar to spatial morphing. Spatial modulation can be applied in many creative ways. An example is the use of a signal with highly varying but strongly directional content as the control signal and a strongly ambient surrounding sound as the carrier signal. The resulting ambient sound will have the directional properties of the control signal. The process is depicted in the diagram of Figure 3(a), where the block 'combine' refers to applying or mixing the parameters of the control stream to the carrier, before the synthesis.

3.2. Reverberation and ambience modification

Diffuse field level control

The DirAC model of a decomposition of the recording into the directional and diffuse part permits manipulation of their relative power by the user. This can be done in a meaningful way as an adjustment of the direct-to-reverberant ratio (DRR). There is a direct relationship between the diffuseness and the DRR. Generally,

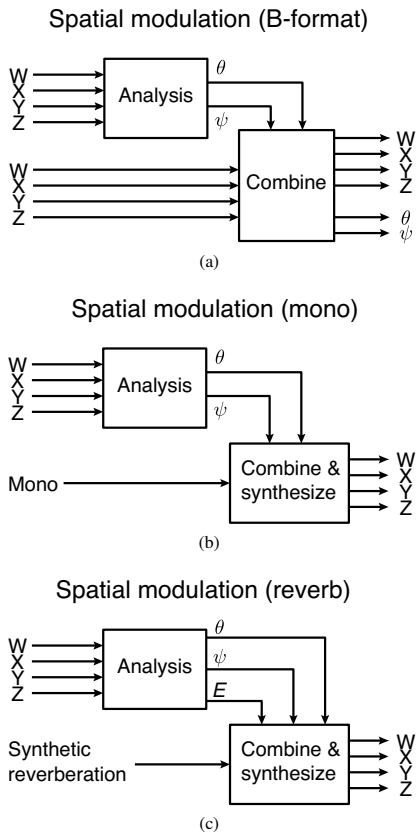

Figure 3: The various flavours of the spatial modulation effect.

since one plane wave and a perfectly diffuse field is assumed in each sub-band, diffuseness expresses the power ratio between the plane wave and the total sound field, while the DRR expresses the power ratio between the plane wave and the diffuse sound. Hence:

$$\psi = \frac{P_{\text{diff}}}{P_{\text{diff}} + P_{\text{pw}}} = \frac{1}{1 + P_{\text{pw}}/P_{\text{diff}}} = \frac{1}{1 + 10^{\Gamma/10}} \tag{14}$$

where P_{diff} is the power of the diffuse field, P_{pw} the power of the plane wave and Γ is the DRR in dB. DirAC analysis gives an estimate $\tilde{\psi}$ of the ideal diffuseness and respectively an estimate $\tilde{\Gamma}$ of the DRR in each time-frequency tile.

Based on this formulation it is straightforward to apply either a global DRR modification $\Delta\Gamma$ on all sub-bands or a modification $\Delta\Gamma(k)$ per sub-band k. The new DRR will be:

$$\Gamma_{\text{mod}}(k,n) = \Gamma(k,n) + \Delta\Gamma(k) \tag{15}$$

and based on Eq. 14 and 15 we can derive the modified diffuseness with respect to $\Delta\Gamma$ as:

$$\psi_{\text{mod}}(k,n) = \frac{\psi(k,n)}{\psi(k,n) + 10^{\Delta\Gamma(k)/10}\,(1 - \psi(k,n))} \tag{16}$$

which can be applied directly in the synthesis stage. One characteristic of the above formulation is that since it results in a diffuseness modification, the overall loudness is preserved.

The specific effect can be directly used in cases where the sound engineer wishes to suppress reverberation in certain frequency bands and to produce a stronger directional impression on the listener. The opposite can be also achieved, with directional sounds being suppressed in favour of the diffuse stream. Control can be given to the user in the style of a parametric equaliser, centered around 0dB of DRR modification, as in Figure 4. For practical use, some upper limits should be imposed on the modification, since the diffuse stream naturally masks artifacts from rapid directional variations in the directional stream.

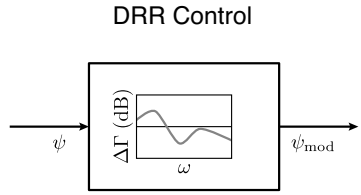

Figure 4: Diffuseness adjustment after modification of the DRR with frequency.

Ambience extraction

The ambience extraction effect attempts to distinguish between foreground sounds having a clear direction, and background sounds or ambience, meaning sounds analysed as mostly diffuse. It differs from the diffuse stream in that instead of scaling the input signals with the diffuseness value and decorrelating the result, it mainly tries to suppress the directional sounds in a direct way. We approach this specific effect with the following steps:

(a) Perform direction-of-arrival and diffuseness analysis in the current windowed frame.

(b) Average the parameters in equivalent rectangular bandwidth (ERB) bands.

(c) For each ERB band k generate a virtual microphone from the B-format components pointing in the opposite direction of the analysed direction of arrival $\vec{u}_k$. The virtual microphone output can be expressed as:

$$S(k,n) = a_k W(k,n) - (1 - a_k)\vec{u}_k^T \cdot \vec{X}'(k,n) \tag{17}$$

where a_k is the directivity coefficient ranging from 0 to 1, adjusting the ratio between the omnidirectional and dipole component, and $\vec{X}'(k,n)$ is the vector of the B-format velocity components.

(d) Since for diffuse sound no suppression should be performed, adapt the directivity of the virtual microphone to range from

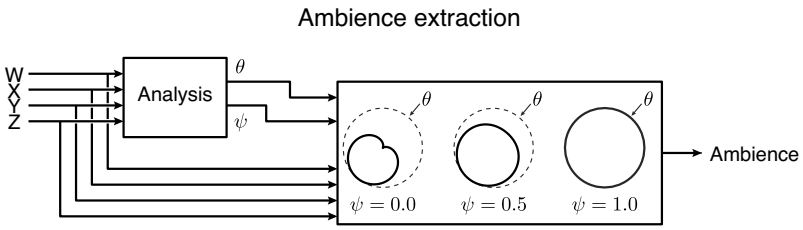

Figure 5: Block diagram of ambience extraction

cardioid for completely non-diffuse sound to omni for diffuse sound. More specifically, the coefficient is adapted as:

$$a_k = \frac{1}{2} + \frac{\psi_k}{2} \tag{18}$$

where ψ_k is the analysed diffuseness at band k.

(e) Perform an IFFT on the processed spectrum and overlap-add to produce the monophonic effect output.

The above procedure is presented schematically in Figure 5. The ambience extraction effect is working effectively for recordings with mild to strong reverberation, where it produces an even reverberant output with strong foreground sounds suppressed. In cases of dry recordings with multiple concurrent sources, artifacts start to appear related to fast fluctuations of the analysed direction and its deviation from the actual source directions. Increasing the temporal analysis resolution with a multi-resolution implementation is expected to improve the sound quality in these cases.

Spatial modulation: Generation of synthetic RIRs

The aforementioned spatial modulation operator can be further applied to parametric reverberation synthesis. In this case, the impulse response of a real space can be parameterised by using for example the Spatial Impulse Response Rendering (SIRR) method, which applies the same analysis and synthesis method as DirAC to room impulse responses [23]. The resulting spatial cues can be imprinted onto a monaural synthetic reverberation. A distinction has to be made between early part synthesis and late part synthesis. The transition time from the early to late reverberation can be determined by the evolution of the diffuseness value, omitting high diffuseness estimates at silent parts in the early reflections region which can be sparse in time. Instead, for the early-to-late reverberation transition, diffuseness should increase monotonically until it reaches an almost constant value, as the response approximates closer the statistical description of the diffuse field. At this point we can assume that the mixing time has been reached.

As it is shown in Figure 3(c), the process resembles the spatial modulation for recordings, however to shape properly the synthetic RIR based on the original one, the estimated instantaneous energy of the original space is included as a third parameter. Then the generated early reflections can be panned to their analysed DOAs and the late part can be rendered with decorrelation. The result can be generated directly for output channels or re-encoded to B-format as a spatial impulse response, as the block 'combine and synthesise' refers to in the picture. It is also possible that the analysed parameters are modified by user-input in some desirable way to affect the spatial properties of the generated reverberation accordingly.

Related parametric approaches for shaping and manipulation of directional room impulse responses have been presented by Melchior et al. [24], starting from a higher-order circular recording of the room response with high-spatial resolution. In that work various acoustical parameters are extracted from the directional recording and an interface is provided to the user for spatial filtering of the room response and shaping of the parameters across time and direction. A similar parametric description for synthesis and manipulation of directional impulse responses can be also realised in the context of SIRR/DirAC, but remains an object for future work.

3.3. Spatial sound synthesis

Spatial sound synthesis from a monophonic source, based on the DirAC parametric decription, has been discussed by Laitinen et al. in [25], in the context of sound scenes for virtual environment applications. However, these same tools can be applied to a music production context with hardly any modifications at all.

The main effect in spatial sound synthesis is the synthesis of persistent spatial cues for a monophonic source. Directional cues are recreated traditionally by panning methods, such as VBAP or ambisonic panning. In virtual-world Directional Audio Coding (VW-DirAC), a monophonic signal is first encoded to B-format for the desired direction and then reproduced with normal DirAC processing. Such a B-format signal processed through DirAC will produce the correct intended DOAs in the analysis stage and zero diffuseness, hence it will be reproduced completely with VBAP in the directional stream.

However, point sources are not very realistic when voluminous real-world objects are considered. More realistic sources should also include synthesis of spatial extent. In VW-DirAC, this effect is achieved by randomising the DOA for each frequency bin of a monophonic source separately. The DOAs can follow a controllable random distribution inside some angular region, which defines the perceived extent of the source. The result can then be synthesized to a B-format, similarly as the single direction case, and reproduced with DirAC. The timbre of the spatially extended source generally does not degrade due to this process, although sounds that are naturally expected to be point-like, such as a singing voice, are perceived to be unnatural. The block diagram of spatial sound synthesis is shown in Figure 6, where the input control parameters are source position and extent. The block 'combine and synthesise' refers to application of parameters to the mono stream and generation of a B-format signal, after panning and decorrelating

the monophonic signal accordingly.

Three-dimensional source positioning and spreading are useful effects in multi-channel mixing. With point-source panning it is possible to place sources in two or three dimensions freely with ease. Furthermore, extent synthesis allows the control of the perceived width of the sources and it can also serve as a distance cue for sound objects that get larger or smaller with distance. These synthesis methods can be controlled dynamically to generate even more effects. The shape of the extent can be controlled either by a single angular parameter, defining an arc around the source position in two dimensions, or a spherical cup in 3 dimensions. Rectangular patches or arbitrary shapes can be also defined.

The spatial modulation presented in Section 3.1 can be directly applied to monophonic sources as well, see Figure 3(b). Instead of modification or replacement of existing spatial cues, the ones from the analyzed recording are directly used as the spatial cues of the monophonic source. A B-format signal can be finally synthesized similarly to the previous techniques.

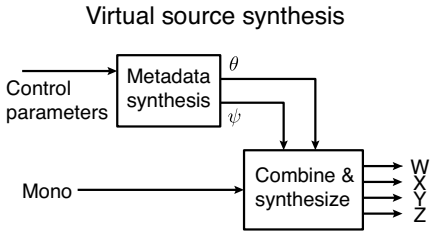

Figure 6: Schematic of spatial source synthesis.

4. DISCUSSION

Various propositions for manipulations of the spatial characteristics of recorded or synthesised sound scenes in the parametric domain have been presented in the previous section. These effects introduce additional tools for creative use by the audio engineer or tonmeister, in addition to established ones such as stereo spreaders or reverberation engines. Apart from ambisonic techniques, which also cover a subset of the manipulations that were presented, there is a lack of effects that work on a recorded sound scene as a whole and permit the tonmeister to focus on the spatial characteristics of the recording and alter them at will. This adds an additional dimension for creative mixing which can be used either to create an enhanced sense of realism from disparate elements or to create non-realistic highly synthetic spatial impressions.

All the spatial effects were accomplished directly in the parametric domain, with appropriate manipulation or combination of parameters. More importantly, the examples converge towards the idea that the parametric domain offers a natural way of realising spatial effects, experimenting and generating new ones. Effects such as ambience extraction or spatial modulation or the complex directional transformations are significantly easier to design and realise as a parametric transformation rather than with more traditional audio signal processing logic. The effect designer is free to choose to manipulate the parameters seen either from a physical perspective or from their perceptual perspective.

All the manipulations discussed above were formulated using the parametric model of DirAC, which can work either with recorded content or content mixed and encoded to some multichannel format. However, the concept and the application are not specific to DirAC. We assume that any parametric spatial audio coding method such as Binaural Cue Coding [26], Spatial Audio Object Coding [27] or the Spatial Audio Scene Coding [28] which are formulated for channel-based content, can be adapted in similar ways, since they all define and extract information related to directional perception and they make a distinction between directional sound events and ambience. More specific to DirAC is the equal treatment of a naturally recorded sound scene and a synthetically created one. After the analysis stage has been performed and the parameters have been extracted, there is no distinction between the recording and the synthetically spatialised or reverberated sound events; the two can be mixed spatially in a coherent way. The parametric manipulations of recorded material can also benefit spatial sound mixing and authoring tools, which are usually object based without facilities to integrate multichannel recordings. An exception seems to be IOSONO's Spatial Audio Workstation [29], which seems able to extract directional components and ambience from stereo recordings and integrate them into the object-based description of the sound scene.

An additional advantage of the presented approach is its computational efficiency. Most of the computational load is needed for the spectral transformations, which is common to all audio effects that work in the spectral domain. After the parameters have been extracted, the spatial manipulations are very cost efficient. Combinations of effects do not increase the computational load since they reduce to a combination and transformation of the parameters only, while the rest of the processing chain remains unchanged. Furthermore, synthetic sound scenes which contain many parametric sound objects can be perceptually compressed into a single sound scene representation and re-encoded in B-format, reducing greatly the number of parameters needed to describe each sound event separately.

5. CONCLUSIONS

This study introduced a series of spatial audio effects for manipulation of directional recordings or synthesis of spatial sound scenes realised in a parametric domain. The effects presented have applications on directional transformations of the sounds, modification of the reverberation and the directional and ambient components and finally on spatial synthesis of complex sound scenes. It has been shown how these effects can be realised efficiently with simple operations directly on the parameters. The extracted parameters allow an intuitive approach for the design of spatial transformations, while unifying the directional content of the recording and the synthetically spatialised sounds.

6. SOUND EXAMPLES

Sound examples for the presented effects are provided online at `http://www.acoustics.hut.fi/go/dafx12-psafx/`.

7. ACKNOWLEDGEMENTS

The Academy of Finland has supported this work. The research leading to these results has received funding from the European

Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n^o [240453].

8. REFERENCES

[1] Michael A. Gerzon, "Ambisonics in Multichannel Broadcasting and Video," *Journal of the Audio Engineering Society*, vol. 33, no. 11, pp. 859–871, 1985.

[2] David G. Malham and Anthony Myatt, "3-D Sound Spatialization Using Ambisonic Techniques," *Computer Music Journal*, vol. 19, no. 4, pp. 58–70, 1995.

[3] Manfred R. Schroeder and Benjamin F. Logan, ""Colorless" Artificial Reverberation," *The Journal of the Acoustical Society of America*, vol. 32, no. 11, pp. 1520, 1960.

[4] James A. Moorer, "About this reverberation business," *Computer Music Journal*, vol. 3, no. 2, pp. 13–28, 1979.

[5] Gary S. Kendall, "The Decorrelation of Audio Signals and Its Impact on Spatial Imagery," *Computer Music Journal*, vol. 19, no. 4, pp. 71–87, 1995.

[6] John Stautner and Miller Puckette, "Designing multi-channel reverberators," *Computer Music Journal*, vol. 3, no. 2, pp. 52–65, 1982.

[7] Jean-Marc Jot and Antoine Chaigne, "Digital delay networks for designing artificial reverberators," in *90th AES Convention*, Paris, France, 1991.

[8] Joseph Anderson and Sean Costello, "Adapting Artificial Reverberation Architectures for B-format Signal Processing," in *Ambisonics Symposium 2009*, Graz, Austria, 2009, pp. 2–6.

[9] Franz Zotter, Matthias Frank, Georgios Marentakis, and Alois Sontacchi, "Phantom Source Widening with Deterministic Frequency Dependent Time Delays," in *14th International Conference on Digital Audio Effects (DAFx-11)*, Paris, France, 2011.

[10] Charles Verron, Mitsuko Aramaki, Richard Kronland-Martinet, and Grégory Pallone, "A 3-D immersive synthesizer for environmental sounds," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1550–1561, 2010.

[11] Guillaume Potard and Ian Burnett, "Decorrelation techniques for the rendering of apparent sound source width in 3D audio displays," in *7th International Conference on Digital Audio Effects (DAFx-4)*, Naples, Italy, October 2004.

[12] Dylan Menzies, "W-Panning and O-Format, Tools For Object Spatialization," in *22nd International Conference of AES: Virtual, Synthetic and Entertainment Audio*, Espoo, Finland, 2002.

[13] Dylan Menzies and Marwan Al-Akaidi, "Ambisonic synthesis of complex sources," *Journal of the Audio Engineering Society*, vol. 55, no. 10, pp. 864–876, 2007.

[14] Ville Pulkki, "Spatial sound reproduction with directional audio coding," *Journal of the Audio Engineering Society*, vol. 55, no. 6, pp. 503–516, 2007.

[15] Frank J. Fahy, *Sound intensity*, Taylor & Francis, London, 2nd edition, 1995.

[16] Ville Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, June 1997.

[17] Ville Pulkki, "Directional audio coding in spatial sound reproduction and stereo upmixing," in *28th International Conference of AES: The Future of Audio Technology-Surround and Beyond*, Pitea, Sweden, 2006.

[18] Mikko-Ville Laitinen and Ville Pulkki, "Converting 5.1 Audio Recordings to B-format for Directional Audio Coding Reproduction," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, 2011, pp. 61–64.

[19] Markus Kallinger, Giovanni Del Galdo, Fabian Kuech, Dirk Mahne, and Richard Schultz-Amling, "Spatial filtering using directional audio coding parameters," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Taipei, Taiwan, 2009, pp. 217–220.

[20] Markus Kallinger, Giovanni Del Galdo, Fabian Kuech, and Oliver Thiergart, "Dereverberation in the spatial audio coding domain," in *130th AES Convention*, London, UK, 2011.

[21] Tapani Pihlajamäki and Ville Pulkki, "Projecting Simulated or Recorded Spatial Sound onto 3D-Surfaces," in *45th International Conference of AES: Applications of Time-Frequency Processing in Audio*, Helsinki, Finland, March 2012.

[22] Richard Schultz-Amling, Fabian Kuech, Oliver Thiergart, and Markus Kallinger, "Acoustical Zooming Based on a Parametric Sound Field Representation," in *128th AES Convention*, London, UK, 2010.

[23] Juha Merimaa and Ville Pulkki, "Spatial impulse response rendering 1: Analysis and synthesis," *Journal of the Audio Engineering Society*, vol. 53, no. 12, pp. 1115–1127, December 2005.

[24] Frank Melchior, Christoph Sladeczek, Andreas Partzsch, and Sandra Brix, "Design and Implementation of an Interactive Room Simulation for Wave Field Synthesis," in *40th International Conference of AES*, Tokyo, Japan, 2010.

[25] Mikko-Ville Laitinen, Tapani Pihlajamäki, Cumhur Erkut, and Ville Pulkki, "Parametric time-frequency representation of spatial sound in virtual worlds," *ACM Transactions on Applied Perception*, vol. 9, no. 2, 2012.

[26] Christof Faller and Frank Baumgarte, "Binaural Cue Coding: A novel and efficient representation of spatial audio," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2002, vol. 2, pp. 1841–1844.

[27] Jeroen Breebaart, Jonas Engdegå rd, Cornelia Falch, Oliver Hellmuth, Johannes Hilpert, Andreas Hoelzer, Jeroen Koppens, Werner Oomen, Barbara Resch, Erik Schuijers, and Leonid Terentiev, "Spatial Audio Object Coding (SAOC) - The Upcoming MPEG Standard on Parametric Object Based Audio Coding," in *124th AES Convention*, Amsterdam, The Netherlands, 2008.

[28] Michael.M. Goodwin and Jean-Marc Jot, "Spatial Audio Scene Coding," in *125th AES Convention*, San Francisco, CA, USA, 2008.

[29] "IOSONO Spatial Audio Workstation," `http://www.iosono-sound.com/spatial-audio-workstation.html`, [accessed 1-July-2012].