Audio Engineering Society

Convention Paper 8920

Presented at the 134th Convention
2013 May 4–7 Rome, Italy

This paper was peer-reviewed as a complete manuscript for presentation at this Convention. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Parametric spatial audio reproduction with higher-order B-format microphone input

Ville Pulkki[1], Archontis Politis[1], Giovanni Del Galdo[2], Achim Kuntz[3],

[1] *Aalto University, ELEC, Dept. Signal Processing and Acoustics, Espoo, Finland*

[2] *Ilmenau University of Technology, Digital Broadcasting Research Laboratory, Ilmenau, Germany*

[3] *Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany*

Correspondence should be addressed to Ville Pulkki (`Ville.Pulkki@aalto.fi`)

ABSTRACT

A time-frequency-domain non-linear parametric method for spatial audio processing is presented here, which can utilize microphone input having directional patterns of any order. The method is based on dividing the sound field into overlapping or non-overlapping sectors. Local pressure and velocity signals are measured within each sector, and an individual Directional Audio Coding (DirAC) processing is performed for each sector. It is shown, that in certain acoustically complex conditions the sector-based processing enhances the quality compared to traditional first-order DirAC processing.

1. INTRODUCTION

The B-format signal set consists of signals obtained from microphones with directional patterns following spherical (for the 3-dimensional case) or circular/cylindrical (for the 2-dimensional case) harmonics up to any order. The most common B-format microphones deliver spherical harmonics only up to first order [1]. With such microphones, the output consists of the zeroth-order omnidirectional W signal, corresponding to the pressure signal, and first-order dipole signals X, Y, and Z in an orthogonal arrangement, corresponding to the three components of the particle velocity vector. First-order Ambisonics [2] is based on the use of such a microphone, where for each loudspeaker in a multichannel setup a signal with first-order directional pattern is computed from the B-format by means of a least-squares inversion. Unfortunately, such techniques have issues with blurred spatial images and coloration, since first-order directional patterns are too broad for setups having more than three or four loudspeakers, for the 2D and 3D case, respectively [3].

These issues can be mitigated by time-frequency processing of spatial audio by exploiting the limitations of human spatial hearing mechanisms in auditory

bands. In Directional Audio Coding (DirAC) [4], it is assumed that at one instant and in one critical band, the human auditory system is limited to decoding one cue for direction and another for inter-aural coherence. It is further assumed that if the direction and the diffuseness of the sound field are measured and reproduced correctly, a human listener will perceive the directional and the coherence cues correctly. The time- and frequency-dependent direction and diffuseness values are treated as metadata which is associated to the recorded audio signals. The metadata can be used in the reproduction to enhance the quality of Ambisonics-type of decoding [5]. In the most basic DirAC rendering method, the omnidirectional signal W is divided into a non-diffuse and a diffuse stream for each frequency channel. The non-diffuse stream is steered to the direction obtained from the metadata while the diffuse stream is delivered to all directions after decorrelation.

Although DirAC enhances the audio quality for multichannel listening, some small differences to the reference are still perceived in specific cases [5]. Quite often such issues arise due to an overestimation of the diffuseness parameter in cases where the sound field is not diffuse, which results in artifacts due to decorrelation of dry sound signals. This may be caused from acoustical conditions having multiple spectrally overlapping source signals from different directions, or due to strong early reflections near the boundaries of the room.

Higher-order B-format microphones have been suggested for spatial sound capture to solve the issues with first-order recording and reproduction [6]. In such microphones, signals following higher-order spherical harmonics are computed by utilizing a larger number of capsules. In subsequent higher-order Ambisonics reproduction, the loudspeaker signals correspond to narrower directional patterns than with first-order systems, which enhances the quality of reproduction [7]. Increasing the order of the microphone increases also the number of microphones, and building an N^{th}-order microphone requires at least $2N + 1$ and $(N + 1)^2$ microphones, for the 2D and 3D cases, respectively [6]. In practice, this limits the use of higher-order microphones in many cases. There exist also unsolved technical challenges related to low-frequency noise, and deformation of directional patterns at high frequencies [8], which sets upper limits for the microphone order that can be achieved with real devices with tolerable quality.

The existence of higher-order microphones poses the question whether they could be utilized in parametric spatial audio processing similarly as first-order microphones are utilized in DirAC, with the assumption that the utilization of higher-order signals could mitigate the quality issues in DirAC in acoustically challenging conditions. Although it is intuitively clear that additional directional information for the sound field should help in the process, the solution is not self-evident. It is clear, that more parameters can be analyzed from higher-order input signals, however, the DirAC method itself does not suggest a trivial generalization to higher orders.

This article presents a method to utilize a higher-order (N) microphone input with DirAC. The basic idea is to divide the sound field into $M \geq N$ sectors, all of which capture the pressure and velocity signals with certain direction-dependent weighting factors. The sum of the weighted pressure and velocity signals equals always the pressure and velocity signals of the original sound field. The pressure signals inside each sector are then rendered in a reproduction system available using DirAC. The sector-based processing makes the acoustical conditions less challenging for the DirAC processing, mitigating the quality issues due to e.g. a too high degree of decorrelation in the output.

2. BACKGROUND

In DirAC the microphone signals are first transformed into the time-frequency domain, after which the spatial parameters of the captured sound field are analyzed using energetic analysis as described in the next section. The results of the analysis are used in the reproduction, which is reviewed in Sec. 2.2.

2.1. Energetic analysis of sound field

The method is presented for the 2D case only, however an extension to the 3D case is straightforward. Let us model the sound field in one point as the sum of an infinite number of plane waves with frequency f, each arriving from the direction of arrival (DOA) φ and carrying the complex signal $S(\varphi)$. Let further $b_A(\varphi)$ denote the pick-up pattern of a generalized microphone, such that the complex signal A

received by the microphone at frequency f is

$$A = \int_{2\pi} S(\varphi) \cdot b_A(\varphi)\ \mathrm{d}\varphi. \tag{1}$$

The DirAC analysis requires the W, X, and Y microphone signals, defined as

$$\begin{aligned} W &= \int_{2\pi} S(\varphi) \cdot b_W(\varphi)\ \mathrm{d}\varphi \\ X &= \int_{2\pi} S(\varphi) \cdot b_X(\varphi)\ \mathrm{d}\varphi \\ Y &= \int_{2\pi} S(\varphi) \cdot b_Y(\varphi)\ \mathrm{d}\varphi, \end{aligned} \tag{2}$$

where the pickup patterns are

$$\begin{aligned} b_W(\varphi) &= 1 \\ b_X(\varphi) &= \cos(\varphi) \\ b_Y(\varphi) &= \sin(\varphi). \end{aligned} \tag{3}$$

Note that these signals correspond to the traditional B-format, with the exception of a $\sqrt{2}$ scaling factor for the velocity components. Subsequent scaling factors that have been proposed in literature to equalize the diffuse field power between higher-order components are omitted and the unnormalized harmonics are used in this study.

The active intensity vector $\boldsymbol{I}_\mathrm{a}$ and energy density E according to the definitions given above are

$$\boldsymbol{I}_\mathrm{a} = -\frac{1}{2\rho_0 c} \Re \left\{ W^* \cdot \begin{bmatrix} X \\ Y \end{bmatrix} \right\} \tag{4}$$

$$E = \frac{1}{4\rho_0 c^2} \left(|W|^2 + |X|^2 + |Y|^2 \right), \tag{5}$$

where ρ_0 is the air density and c is the sound velocity. The estimated DOA of sound θ is simply defined as the opposite direction of $\boldsymbol{I}_\mathrm{a}$.

The diffuseness parameter is then computed as

$$\Psi = 1 - \frac{\| \mathrm{E}\left[\boldsymbol{I}_\mathrm{a}\right] \|}{cE}, \tag{6}$$

which characterizes the field with a real-valued measure ranging from zero to one [9]. A zero value is obtained with a single plane wave in free field while a value of one is obtained in a fully diffuse or fully reactive field, such as one of many uncorrelated uniformly distributed plane waves or one comprised only of standing waves respectively.

2.2. Directional audio coding utilizing one sound signal in synthesis

There exist multiple versions of DirAC for different applications. For simplicity, in this work we consider only the version called mono-DirAC and shown in Fig. 1. In the presented implementation the short-time Fourier transform (STFT) is used, though other transforms can be employed. The directional metadata is computed using energetic analysis from the pressure and velocity signals, and only the pressure signal is utilized in the synthesis of sound. The metadata is transmitted to the reproduction side along with the audio signal W.

In the reproduction, the signal is divided into non-diffuse and diffuse streams using Ψ, as shown in the figure. The non-diffuse stream is reproduced by the loudspeakers using vector base amplitude panning (VBAP) [10], and the diffuse stream is distributed to all loudspeakers after decorrelation. The method is well motivated, since the energy of the signal W is preserved in reproduction for any acoustic scenario. However, it has been shown that in multi-source cases and with reverberation there are some issues [5, 11] with mono-DirAC.

3. DIRAC WITH HIGHER-ORDER MICROPHONE INPUT

The first-order mono-DirAC can be seen as a system, where the sound signal captured with a zeroth-order microphone is steered to loudspeakers using combined information from all microphones. The driving idea here is to formulate DirAC in an analogous way, where the sound signal captured using microphones up to (N-1):th order are steered with parametric information computed from microphone signals of up to Nth order. In the proposed solution, the directional pattern of (N-1):th order microphone input defines a "sector" wherein the analysis and synthesis is performed. With the analysis and the synthesis conducted within each sector, the total sound field is supposed to be recreated when the output signals of all sectors are combined.

3.1. Directional analysis within sectors

We define the DirAC analysis for the i-th sector as follows: Let $c_i(\varphi)$ be an angle dependent gain function which modifies the output of each microphone

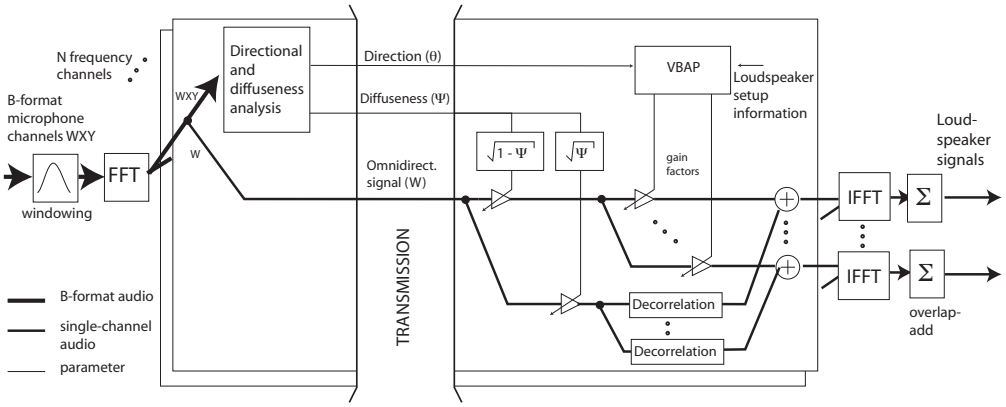


Fig. 1: Block diagram of first-order DirAC with rendering of W signal, also known as mono-DirAC.

by effectively multiplying the pickup pattern as follows

$$\begin{aligned} W_i &= \int_{2\pi} S(\varphi) \cdot b_W(\varphi) \cdot c_i(\varphi)\ \mathrm{d}\varphi \\ X_i &= \int_{2\pi} S(\varphi) \cdot b_X(\varphi) \cdot c_i(\varphi)\ \mathrm{d}\varphi \\ Y_i &= \int_{2\pi} S(\varphi) \cdot b_Y(\varphi) \cdot c_i(\varphi)\ \mathrm{d}\varphi, \end{aligned} \tag{7}$$

such that the equivalent pickup patterns of the modified microphones W_i, X_i, and Y_i are

$$\begin{aligned} b_{W_i}(\varphi) &= c_i(\varphi) \\ b_{X_i}(\varphi) &= \cos(\varphi) \cdot c_i(\varphi) \\ b_{Y_i}(\varphi) &= \sin(\varphi) \cdot c_i(\varphi). \end{aligned} \tag{8}$$

The modified (i.e., valid for the i-th sector) active intensity vector $\boldsymbol{I}_{\mathrm{a_i}}$ and energy density E_i are then simply

$$\boldsymbol{I}_{\mathrm{a_i}} = -\frac{1}{2\rho_0 c} \Re \left\{ W_i^* \cdot \begin{bmatrix} X_i \\ Y_i \end{bmatrix} \right\} \tag{9}$$

$$E_i = \frac{1}{4\rho_0 c^2} \left(|W_i|^2 + |X_i|^2 + |Y_i|^2 \right). \tag{10}$$

The estimated DOA of sound θ_i is again simply defined as the opposite direction of $\boldsymbol{I}_{\mathrm{a_i}}$.

The sector diffuseness variable is computed as

$$\Psi_i = 1 - \frac{\| \mathrm{E}\left[\boldsymbol{I}_{\mathrm{a_i}} \right] \|}{c E_i}. \tag{11}$$

Please note that the sector diffuseness parameter is a local diffuseness specific to the sector.

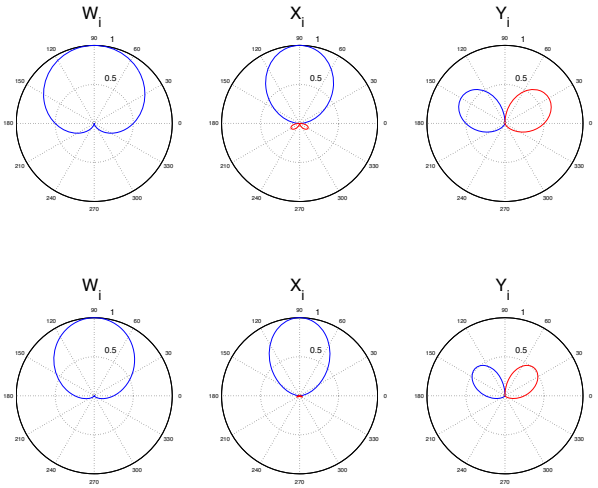
3.2. Selection of the directional pattern defining the sectors

The previous section showed that directional analysis can be conducted inside sectors. It is now of interest how to define $c_i(\varphi)$, which gives the proper form to the sectors. As the target is to reproduce the sound as it was in the original place, it should hold that

$$\sum_{i=1}^{M} c_i(\varphi) = 1. \tag{12}$$

where M is the number of sectors. Additionally, we may also limit $c_i(\varphi) \geq 0$, since allowing negative $c_i(\varphi)$ values would invert the direction of $\boldsymbol{I}_{\mathrm{a_i}}$, which is not desired.

The selection of the directional pattern for $c_i(\varphi)$ is still a free choice, as this equation sets only the scaling of the patterns to be one. Arbitrary patterns can be utilized when the microphone is simulated, such as in virtual realities or in audio coding, however, it is of great interest in this article to consider how to implement such patterns with higher-order microphones in practice. As the higher-order B-format signals are defined to have directional patterns following spherical/circular harmonics, the patterns of the modified microphones in Eq. (8) should be something that can be composed from spherical harmonics. As the direct multiplication between arbitrary harmonics results in another harmonic, the pattern can be selected as a harmonic function. As $c_i(\varphi) \geq 0$, and as we assume that the separation between the sectors should be maximized, the minimum value of $c_i(\varphi)$ is set to zero. According to this, the most suitable pattern that can be used for N^{th}-



(b) N = 3

Fig. 2: Directional patterns of sector signals computed from 2nd- and 3rd-order B-format input.

order microphone is $(N-1)^{th}$ order cardioid. Moreover, the prerequisite that the sum of the directional patterns of different sectors has to be unity can be simply fulfilled by utilizing $M \geq N$ equiangular sectors, whose look directions are equal to $i2\pi/M$

$$c_i(\varphi) = G \cdot \left(\frac{1 + \cos(\varphi - \frac{i}{M} 2\pi)}{2} \right)^{N-1}, \qquad (13)$$

where $G = \frac{2^N}{N \cdot M}$ is a gain that normalizes the sum of the sectors to unity. Note, that if $N = 1$ was set, $c_i(\varphi)$ would vanish from the equations, and the directional analysis would be equal to the standard first-order DirAC analysis.

The directional patterns of the microphones for sector i with N^{th}-order B-format microphone input take thus the form of

$$\begin{aligned} b_{W_i}(\varphi) &= c_i(\varphi) \\ b_{X_i}(\varphi) &= \cos(\varphi) \cdot c_i(\varphi) \\ b_{Y_i}(\varphi) &= \sin(\varphi) \cdot c_i(\varphi). \end{aligned} \qquad (14)$$

As an example, the directional patterns for orders $N = 2$ and $N = 3$ are presented in Fig. 2.

The directional pattern of W_i is in each case naturally the cardioid of the corresponding order. The X_i pattern shows a peculiar shape with large a positive lobe, and two minor negative lobes in opposing sides. The Y_i pattern has two lobes as a dipole pattern in general, however, when compared to first-order dipole, both of the lobes are turned towards the look direction of W_i.

3.3. Deriving the signal components for Higher-order DirAC

Since the set of signals we obtain from the order N microphone correspond to spherical/circular harmonics patterns, we can construct any pattern of maximum order N as a linear combination of these patterns. For the 2D case, the sector and analysis patterns are completely determined by the vector of coefficients $\boldsymbol{\alpha}_N^c = [a_0\ a_1\ b_1\ ...\ a_N\ b_N]^T$ that specifies the amplitude of each harmonic component, where a_n and b_n denote the gains for cylindrical components with order n. The derivation that produces the coefficients required for the energetic analysis involves lengthy algebraic and trigonometric operations and is therefore omitted here.

In practice the following steps are performed for the design of the patterns:

(a) Generate the coefficients $\boldsymbol{\alpha}_{N-1}^{c_M} = [a_0\ a_1\ b_1\ ...\ a_{N-1}\ b_{N-1}]^T$ for a non-rotated cardioid

(b) Generate the coefficients of the rotated sectors $\boldsymbol{\alpha}_{N-1}^{c_i}$ from the non-rotated one

(c) Generate the coefficients of the related velocity patterns $\boldsymbol{\alpha}_N^{x_i}$, $\boldsymbol{\alpha}_N^{y_i}$ from the rotated sector ones

(d) Obtain the respective signals by matrixing the B-format channels with the appropriate coefficients from the previous steps

Below in Table 1, we present the coefficients to obtain 2^{nd} and 3^{rd}-order signals.

3.4. Sector response to diffuse field

In the case of a diffuse field, the diffuseness Ψ equals unity measured with a first-order microphone. This is not the case for the local diffuseness Ψ_i of a single sector, which is always $\Psi_i = const < 1$ in global diffuse field. This can be explained by the fact that even if the sound is completely diffuse, the field seen by the sector is not, since the directional distribution

	N	a_0 [1]	a_1 [$\cos\varphi$]	b_1 [$\sin\varphi$]	a_2 [$\cos 2\varphi$]	b_2 [$\sin 2\varphi$]	a_3 [$\cos 3\varphi$]	b_3 [$\sin 3\varphi$]
W_i	2	1/2	$\cos\alpha/2$	$\sin\alpha/2$	0	0	–	–
X_i	2	$\cos\alpha/4$	1/2	0	$\cos\alpha/4$	$\sin\alpha/4$	–	–
Y_i	2	$\sin\alpha/4$	0	1/2	-$\sin\alpha/4$	$\cos\alpha/4$	–	–
W_i	3	3/8	$\cos\alpha/2$	$\sin\alpha/2$	$\cos 2\alpha/8$	$\sin 2\alpha/8$	0	0
X_i	3	$\cos\alpha/4$	$\cos 2\alpha/16 + 3/8$	$\sin 2\alpha/16$	$\cos\alpha/4$	$\sin\alpha/4$	$\cos 2\alpha/16$	$\sin 2\alpha/16$
Y_i	3	$\sin\alpha/4$	$\sin 2\alpha/16$	$3/8 - \cos 2\alpha/16$	$-\sin\alpha/4$	$\cos\alpha/4$	-$\sin 2\alpha/16$	$\cos 2\alpha/16$

Table 1: Coefficients of the harmonic components for the analysis patterns of orders N values 2 & 3, for an ith sector oriented at α. The directional patterns of the harmonic components are given in brackets in the first line of the table.

of the sound in the sector is heavily biased by $c_i(\varphi)$. The local diffuseness in this case can be expressed as

$$\begin{aligned}\Psi_i &= \frac{\mathrm{E}\left[|W_i|^2\right]}{\mathrm{E}\left[|W|^2\right]} = \frac{S^2_{\mathrm{diff}} \int_{2\pi} |c_i(\varphi)|^2 \mathrm{d}\varphi}{S^2_{\mathrm{diff}} \int_{2\pi} \mathrm{d}\varphi} \\ &= \frac{\int_{2\pi} |c_i(\varphi)|^2 \mathrm{d}\varphi}{2\pi} = \frac{1}{Q_c} \end{aligned} \quad (15)$$

where S^2_{diff} is the power spectral density of the signal of the diffuse field and is independent of direction. The ratio in Eq. (15) corresponds to the power output of the sector in a diffuse field compared to an omnidirectional signal which is equal to the inverse of the directivity factor Q_c of a directional microphone. The formulation in Eq. (15) corresponds to a 2-dimensional diffuse field; for the 3-dimensional case the integrals have to be replaced with their spherical counterparts.

During reproduction, the sound field synthesized by this sector alone would be prominently non-diffuse. However, the target is not that the sound field reproduced by a single sector should be diffuse, but on the contrary, the total sound field superposed from sound field reproductions by *all* sectors should be diffuse.

3.5. Sector response to a mixture of a single plane wave and diffuse field

In the case of a single plane wave with power S^2_{dir} and a diffuse field with power S^2_{diff}, the local diffuseness relates to the global diffuseness as follows. If the direction of the plane wave coincides with the orientation of the sector, so that its power is not affected by the sector pattern, then the local diffuseness is simply

$$\Psi_i = \frac{1}{Q_c}\Psi. \quad (16)$$

However, if the plane wave is coming from any other direction then diffuseness becomes direction-dependent, in contrast to the diffuseness of the original sound field, which is independent of the DOA:

$$\begin{aligned}\Psi_i(\varphi) &= \frac{S^2_{\mathrm{diff}_i}}{S^2_{dir_i} + S^2_{\mathrm{diff}}} = \frac{S^2_{\mathrm{diff}}/Q_c}{c_i^2(\varphi) S^2_{\mathrm{dir}} + S^2_{\mathrm{diff}}} \\ &= \frac{\Psi}{Q_c\left(c_i^2(\varphi) - \Psi c_i^2(\varphi) + \Psi\right)} \end{aligned} \quad (17)$$

where φ is the DOA of the plane wave. In the case that the DOA coincides with a null in the sector response, the local diffuseness takes its maximum value of Eq. (15).

3.6. Analysis results for second-order DirAC with simulated microphone input

The suggested idea was simulated computationally, where single-frequency plane waves from one or two directions were used. The diffuse field was simulated with 360 random-phase sinusoidal plane waves with uniformly distributed frequency within 1/3 octave. Direction and diffuseness are computed for the total sound field (θ and Ψ), and similarly θ_i and Ψ_i are computed for sectors $i = 1$ and $i = 2$, pointing to 0° and 180°, respectively. The values were computed for plane waves from different directions φ, and the results are shown in Fig. 3.

The first example shows the response for the field of a single plane wave. It can be seen, that all analyzed diffuseness values are constantly zero. The analyzed directions θ, θ_1, and θ_2 match also accurately the direction of the plane wave. This shows that sector-based processing does not differ from the traditional first-order DirAC reproduction in simple cases.

In a second case, an interfering antiphasic sinusoid is positioned to the direction of 30°, which is a prob-

lematic case for first-order DirAC, as the diffuseness Ψ is high, and the analyzed direction θ does not correspond to neither signal nor interferer direction. The W signal will also be constantly zero, since the signal and interferer cancel each other. It can be seen that the results for the sectors are affected by the interferer only when the signal and interferer arrive from directions close to each other. When the directions are more separated, the analyzed direction θ_2 corresponds to the direction of the plane wave, and direction θ_1 corresponds to the direction of the interferer. In this case, the signal and the interferer have different levels in W_1 and W_2, which would lead to the reproduction of the signal and interferer in the correct directions with reasonable separation. Note that the diffuseness is analyzed high only when the signal arrives from 330°, i.e., symmetrically from the opposite side of the sector look direction. This happens since both X_i capture signal and interferer with equal level, leading to complete cancellation, and consequently causing the magnitude of $\boldsymbol{I}_{\mathrm{a}i}$ to be zero in both sectors. In this case, W_i would also vanish. This issue could be at least partly solved by using three sectors, as in this case the signal would never vanish in all sectors due to the different symmetry of the setup.

There are also two cases shown in the figure, where the diffuse field is simulated with different direct-to-diffuse energy ratios Γ. It can be seen, that the diffuse sound affects the result of analysis less the closer the signal direction is to the look direction of W_i. Decreasing Γ also causes the diffuseness to be analyzed higher. However, the analyzed value approaches 0.35 in a perfectly diffuse field, which matches with value 0.37 computed from Eq. (15). The analyzed sector diffuseness values are also generally lower than the diffuseness of the total sound field. Assuming that the subsequent reproduction of the sectors produces the correct diffuseness, it can be stated that a lower level of decorrelation is needed there. As the decorrelation process is prone to cause coloration artifacts, this may improve the quality of reproduction in some cases. In addition, the analyzed θ values for the total sound field show large variations in the case of Γ = -10 dB, following the definition of diffuse field where all directions are equally probable. In DirAC reproduction, this might result in artifacts due to fast temporal variations of loudspeaker gain factors. In corresponding directional analysis of sector signals, the analyzed direction values θ_i follow the look direction of the cardioids, which mitigates this issue.

4. HIGHER-ORDER DIRAC SYNTHESIS

In higher-order mono-DirAC synthesis, each sector is seen as an independent first-order mono-DirAC, and the reproduction is performed separately for each of them. The suggested flow diagram for higher-order DirAC in the particular case of second-order microphone input with two sectors is shown in Fig. 4. The figure shows how the sound field is divided into two sectors, both of which are then analyzed and synthesized separately. Note that the outputs from both sectors are connected to all loudspeakers.

In the proposed method, W_i signals are first divided into diffuse and non-diffuse streams. The non-diffuse streams are reproduced to all loudspeakers using, e.g., amplitude panning, thus no changes are needed when compared with first-order mono-DirAC.

When a diffuse field is recorded, the accumulated sound field reproduced by all sectors should be distributed evenly to all directions around the listener. We may approach this by monitoring the directional distribution of the *non-diffuse* streams in second-order DirAC reproduced over a 36-channel horizontal loudspeaker setup. The RMS average of each loudspeaker is shown as a polar plot in the figures. The distribution of a single sector shown in Fig. 5 a) follows a half-dipole pattern centered towards the look-direction of the sector. When combining the energy distribution produced by two sectors with opposing look directions, the resulting distribution of sound is clearly a figure-of-eight, as shown in Fig. 5 c).

This suggests that the diffuse streams should be reproduced such that the total distribution of sound is even with frequency. In the current implementation of HO-DirAC, this has been done by applying the diffuse stream to the loudspeakers using a loudspeaker-individual gain relative to the magnitude of the directional pattern of Y_i in the directions covered by each loudspeaker. The resulting directional distribution of the diffuse stream of one of the sectors shows thus a two-lobe distribution, as shown in Fig. 5 b). The resulting directional distribution of

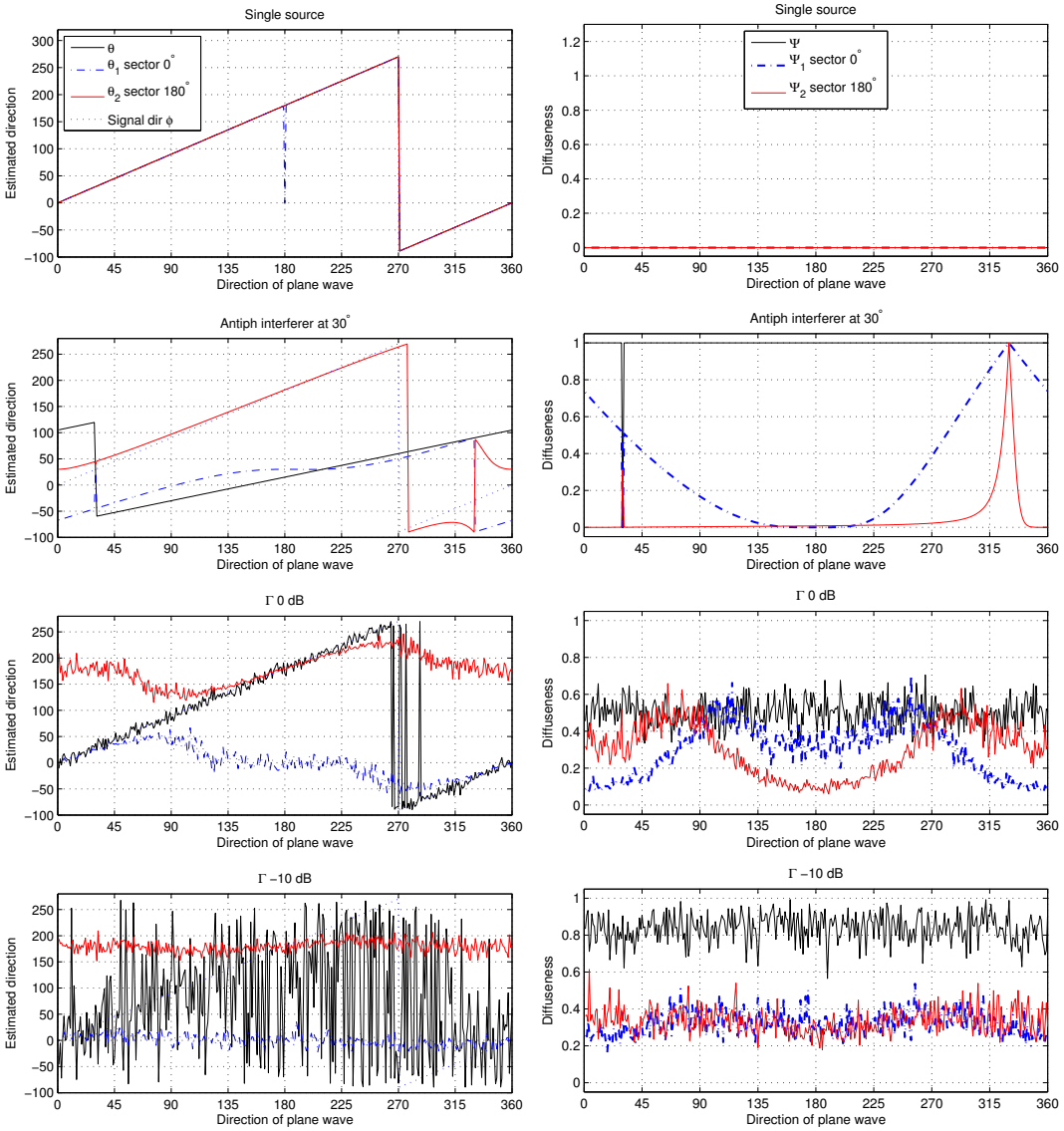


Fig. 3: Analyzed direction and diffuseness values for first-order DirAC (θ, Ψ) and second-order DirAC with two sectors $(\theta_1, \Psi_1$ and $\theta_2, \Psi_2)$ in various scenarios.

summed diffuse streams of the two sectors has directional pattern of figure-of-eight perpendicular to the main directions of the sectors, as shown in in Fig. 5 d). With this method, the sound energy is relatively evenly distributed to all directions when the recorded sound field is diffuse, as shown in Fig. 5 e). Please note that the pattern is not perfectly round. However obtaining a perfectly enveloping distribution does not represent a problem, as the human capability to perceive such spatial distributions of sound is limited.

The method has been tested by using different simulated sound scenarios with multiple sources and simulated room acoustics. Different orders of ideal microphones have been simulated, and a software

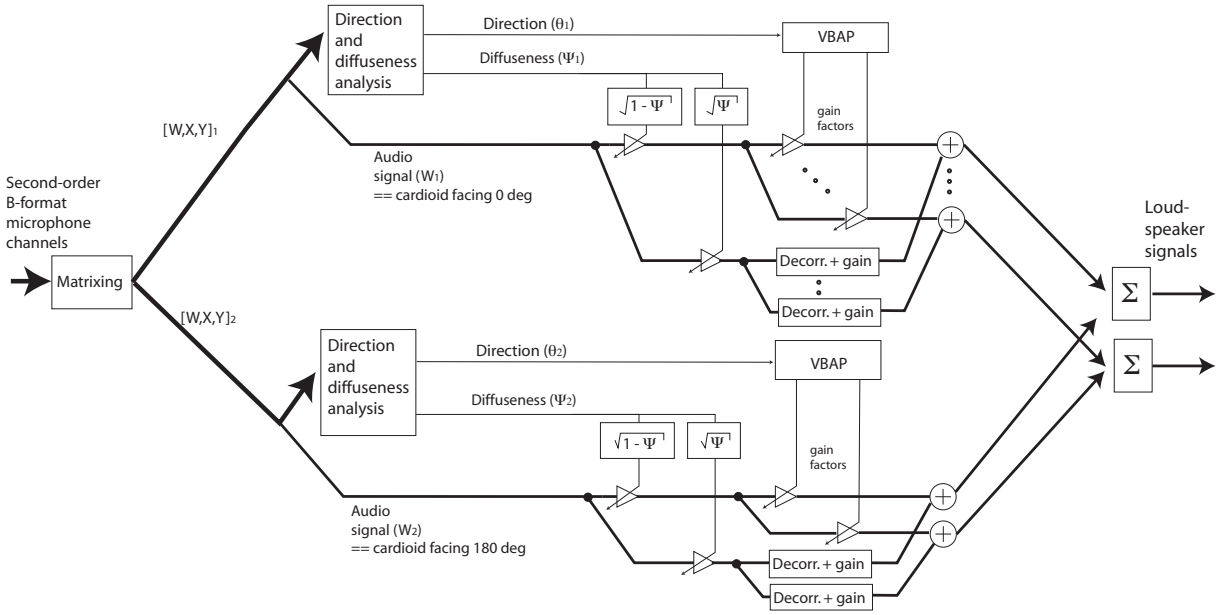


Fig. 4: Implementation of second-order DirAC with two sectors. The input B-format signal is matrixed into two sets of directional signals. A complete directional analysis is performed for these two signal sets, and a mono DirAC decoding is performed for both sets.

capable of decoding any order with any number of cardioid sectors was used to decode the sound into a multichannel loudspeaker setup. Informal listening tests revealed that sector processing indeed mitigated the quality issues in conditions particularly challenging to first-order mono-DirAC reproduction. The processing was not found to cause any new artifacts.

5. DISCUSSION

This article presents the theory of a sector-based higher-order DirAC method. Although so far only informal listening tests were carried out, the potential of the system can already be seen. The problematic cases with traditional DirAC are the cases where spectrally overlapping sound signals with similar levels arrive simultaneously from different directions to the microphone. It is implicitly clear, that when such problematic sound fields are divided into sectors, such overlapping with similar levels occurs less frequently mitigating the issue. This can be seen in Fig. 3 in that the antiphasic interferer affects the result only when it originates from a direction close to the signal direction. In other words, the acoustical situations where the single-plane-wave-with-diffuse-field data model of DirAC is violated are less common with sector-based processing. When the sectors are made narrower, the problematic cases are mitigated further. The required degree of decorrelation is also decreased in the case of diffuse field, which is quantified in Eq. (15), which mitigates some coloration issues due to decorrelation.

6. CONCLUSIONS

A method to utilize higher-order microphone input in time-frequency-domain parametric spatial audio processing is suggested here, where the recorded sound field is divided into sectors, and modified pressure and velocity signals are measured within each sector. The sector signals are then reproduced using a modified Directional Audio Coding (DirAC) technique. It is shown that the division of sound field into sectors mitigates certain acoustically challenging conditions causing artifacts in traditional first-order DirAC.

7. ACKNOWLEDGMENT

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

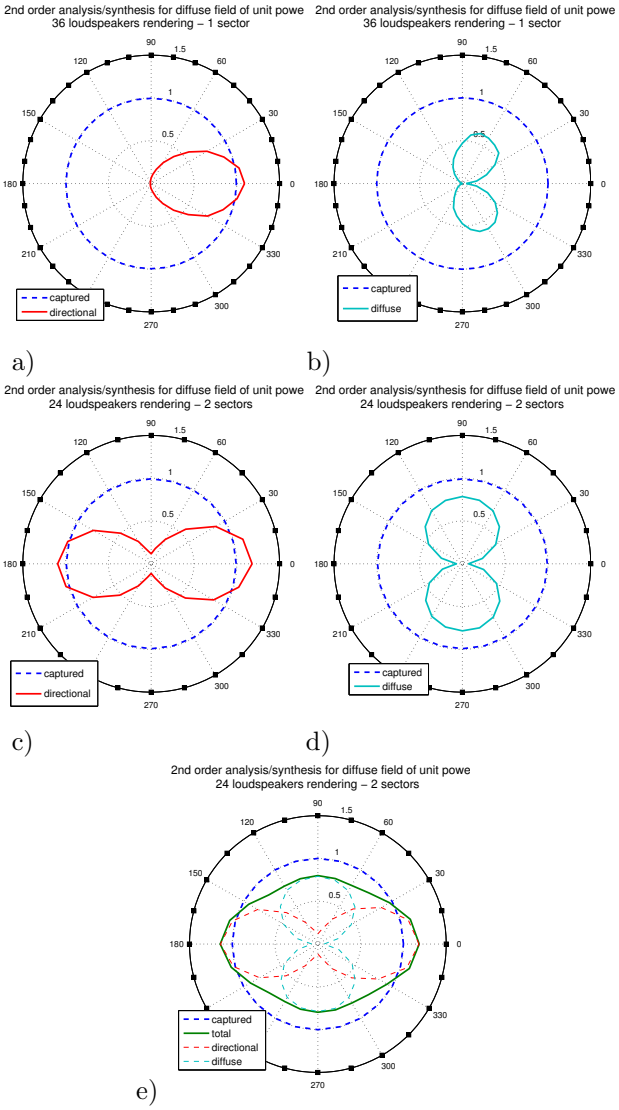


Fig. 5: Directional distribution of diffuse and/or non-diffuse streams of different sectors on reproduction of diffuse field. a) One sector, non-diffuse stream only. b) One sector, diffuse stream. c) Summed non-diffuse streams of two opposite-facing sectors. d) Summed diffuse streams of opposite-facing sectors. e) All four streams summed.

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

8. REFERENCES

[1] Michael A. Gerzon. The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound. In *50th AES Convention*, London, UK, 1975.

[2] Michael A. Gerzon. Periphony : With-Height Sound Reproduction. *Journal of the Audio Engineering Society*, 21(1):1–10, 1973.

[3] Audun Solvang. Spectral Impairment for Two-Dimensional Higher Order Ambisonics. *Journal of the Audio Engineering Society*, 56(4):267–279, 2008.

[4] Ville Pulkki. Spatial sound reproduction with directional audio coding. *Journal of the Audio Engineering Society*, 55(6):503–516, 2007.

[5] Juha Vilkamo, Tapio Lokki, and Ville Pulkki. Directional Audio Coding: Virtual Microphone-Based Synthesis and Subjective Evaluation. *Journal of the Audio Engineering Society*, 57(9):709–724, 2009.

[6] Sébastien Moreau, Jérôme Daniel, and Stéphanie Bertet. 3D Sound Field Recording with Higher Order Ambisonics Objective Measurements and Validation of Spherical Microphone. In *120th AES Convention*, Paris, France, 2006.

[7] Stéphanie Bertet, Jérôme Daniel, Etienne Parizet, Laëtitia Gros, and Olivier Warusfel. Investigation of the perceived spatial resolution of higher order ambisonics sound fields: A subjective evaluation involving virtual and real 3d microphones. In *30th International Conference of AES: Intelligent Audio Environments*, Saariselkä, Finland, 2007.

[8] Boaz Rafaely, Barak Weiss, and Eitan Bachmat. Spatial Aliasing in Spherical Microphone Arrays. *IEEE Transactions on Signal Processing*, 55(3):1003–1010, 2007.

[9] J. Merimaa and V. Pulkki. Spatial Impulse Response Rendering I: Analysis and synthesis. *J. Audio Eng. Soc.*, 53(12):1115–1127, 2005.

[10] Ville Pulkki. Virtual source positioning using vector base amplitude panning. *J. Audio Eng. Soc.*, 45(6):456–466, June 1997.

[11] 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 (TAP)*, 9(2):8, 2012.