Author(s): Archontis Politis, Ville Pulkki

Title: Broadband Analysis and Synthesis for Directional Audio Coding using A-format Input Signals

Year: 2011

Version: Final published version

Please cite the original version:

Archontis Politis, Ville Pulkki. Broadband Analysis and Synthesis for Directional Audio Coding using A-format Input Signals. In 131st Convention of the Audio Engineering Society, New York, NY, USA, October 2011

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.

Audio Engineering Society

Convention Paper

Presented at the 131st Convention
2011 October 20–23 New York, USA

Broadband analysis and synthesis for Directional Audio Coding using A-format input signals

Archontis Politis[1], Ville Pulkki[1]

[1] *Aalto University, School of Electrical Engineering, Department of Signal Processing and Acoustics, FI-02150 Espoo, Finland*

Correspondence should be addressed to Archontis Politis (archontis.politis@aalto.fi)

ABSTRACT

Directional Audio Coding (DirAC) is a parametric non-linear technique for spatial sound recording and reproduction, with flexibility in terms of loudspeaker reproduction setups. In the general 3-dimensional case, DirAC utilizes as input B-format signals, traditionally derived from the signals of a regular tetrahedral first-order microphone array, termed A-format. For high-quality rendering, the B-format signals are also exploited in the synthesis stage. In this paper we propose an alternative formulation of the analysis and synthesis, which avoids the effect of non-ideal B-format signals on both stages, and achieves improved broadband estimation of the DirAC parameters. Furthermore, a scheme for the synthesis stage is presented that utilizes directly the A-format signals without conversion to B-format.

1. INTRODUCTION

Directional Audio Coding (DirAC) is a parametric method for a perceptually motivated analysis and re-synthesis of a sound field [1]. The parameters estimated are direction of arrival (DOA) and diffuseness. These parameters are used to recreate the captured sound scene in an efficient manner. In the low-bit rate version of DirAC, intended for applications where encoding, transmission and decoding efficiency are crucial, such as teleconferencing, only a pressure signal and the estimated parameters are included in the DirAC stream. This stream can then be sent to the decoder and reproduced in arbitrary loudspeaker layouts. In the high-bit-rate version, intended for applications where higher-fidelity is desired such as music recording and reproduction, all the audio channels used for the analysis are stored in the DirAC stream.

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 and three orthogonal pressure-gradient signals, known in literature as the B-format signal set. B-format is usually obtained by appropriate matrixing of the signals captured from a microphone array. For 3-dimensional music recording and reproduction, the most widely used microphone setup is a regular tetrahedral array of four cardioid or subcardioid capsules, such as the Soundfield microphone [2], which offers a good compromise between effective coincidence of the derived B-format signals and fidelity of the recorded material.

The conversion from the signals captured from the tetrahedral array, known in literature as A-format, to the B-format set can be achieved by a straightforward linear combination of the signals. In this case, the result matches the ideal frequency-independent B-format patterns under the following assumptions: a) the microphones are coincident, b) the directional patterns of the capsules are ideal first-order ones, c) the directional patterns are similar between capsules and d) the directional patterns are invariant with respect to frequency. From these assumptions, deviations from (a) and (d) have the most prominent effect in the frequency and directional response of the B-format components. Furthermore, when the distance between capsules becomes comparable with the wavelength, spatial aliasing occurs and the omnidirectional and bidirectional components get contaminated with higher order directional components. The effect of non-coincidence is alleviated partly with use of correction filters based on the on-axis theoretical or measured responses of each component [3]. These filters achieve equalization of the responses up to the spatial aliasing limit. Beyond that limit the frequency responses become strongly direction-dependent.

The effect of spatial aliasing and other non-idealities in real arrays, such as position or capsule mismatch, results in a direction-dependent error in the directional analysis with increasing frequency. Consequently, diffuseness is overestimated at high frequencies. An analysis of this effect for a 2-dimensional case has been presented in [4], where the authors achieve broadband estimation using a dual radius array. Furthermore, in the synthesis stage, the high-quality version of DirAC distributes the audio to the loudspeakers by means of virtual microphones produced from the B-format signals as in [5], which are also affected by the non-coincidence of the capsules and the spatial aliasing in a direction-dependent manner.

In this paper we propose an alternative formulation of DirAC analysis and synthesis, performed directly on the A-format signals, which achieves improved broadband estimation of direction and diffuseness compared to the B-format analysis. Moreover, based on the knowledge of the array geometry, we present a method to generate efficiently virtual microphone signals without intermediate conversion to B-format, having an on-axis frequency response closer to the ideal.

2. ENERGETIC ANALYSIS OF THE SOUND FIELD AND RELATION TO B-FORMAT SIGNALS

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 sound field with one or more audio streams and two parameters for each time-frequency tile, namely the direction of arrival and the diffuseness. These parameters are extracted from an energetic analysis based on the sound pressure $p(t)$ and particle velocity $\mathbf{u}(t)$ at a point in the sound field. In the STFT domain, used in this implementation, the respective complex transformed quantities are $P(l,n)$ and $\mathbf{U}(l,n)$, where l,n are the frequency and time indices of the transform. 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

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

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

with $\mathbf{X}'(l,n) = [X(l,n)\; Y(l,n)\; Z(l,n)]^T$ the vector of the B-format pressure-gradient signals and $Z_0 = c\rho_0$ 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

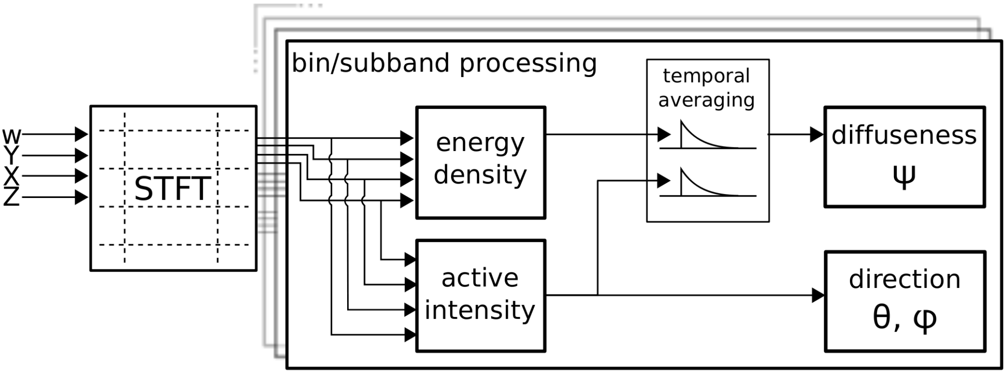


Fig. 1: Flow diagram of the DirAC analysis with B-format input signal.

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 [6]

$$\mathbf{I}_a(l,n) = \frac{1}{2}\Re\left\{P(l,n)\cdot\overline{\mathbf{U}(l,n)}\right\}. \tag{3}$$

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

$$\mathbf{I}_a(l,n) = -\frac{1}{2\sqrt{2}Z_0}\Re\left\{W(l,n)\cdot\overline{\mathbf{X}'(l,n)}\right\}. \tag{4}$$

Hence, the direction of incidence of the plane wave can be estimated from the active intensity vector as

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

or in terms of the B-format signals

$$\begin{aligned}\mathbf{u}_{\mathrm{DOA}}(l,n) &= \frac{\Re\left\{W(l,n)\cdot\overline{\mathbf{X}'(l,n)}\right\}}{\|\Re\left\{W(l,n)\cdot\overline{\mathbf{X}'(l,n)}\right\}\|} \\ &= \begin{bmatrix}\cos(\theta)\cos(\phi)\\ \sin(\theta)\cos(\phi)\\ \sin(\phi)\end{bmatrix}\end{aligned} \tag{6}$$

where $\theta(l,n), \phi(l,n)$ are the estimated azimuth and elevation of incidence respectively.

The energy density of the sound field at the same point is defined as [6]

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

and in terms of the B-format signals

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

Finally, the diffuseness is defined as

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

and in terms of the B-format signals

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

where $\langle\cdot\rangle$ denotes time averaging. 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 energy transport is zero.

The analysis procedure is presented schematically in Figure 1. The temporal averaging for the estimation of diffuseness is realised in the current implementation with first-order recursive filters.

3. BROADBAND ESTIMATION OF DIRAC PARAMETERS USING A-FORMAT SIGNALS

As mentioned above, errors in the parameter estimation occur as the frequency and directional response of the B-format components begin to deviate from the ideal frequency-independent omnidirectional and bidirectional ones, which is an inevitable effect mainly of the non-coincidence of the capsules. An alternative parameter estimation for a 2-dimensional case that is robust at mid-high frequencies has been demonstrated in [7], termed energy-gradient analysis, which exploits the effective energy attenuation between the capsules of the array, due to either inherent directivity or because of shadowing due to some rigid body between them. A similar type of analysis is reformulated in the present study in the case of a regular tetrahedral microphone array.

3.1. DOA and diffuseness estimation based on A-format energy gradients

An alternative estimation of the average energy flow can be achieved by appropriate matrixing of the energy of the A-format signals. The unit vectors of the orientation of the four capsules of a regular tetrahedral array, following the common naming of A-format found in literature as in Figure 2, are

$$\begin{aligned}
\mathbf{u}_{\mathrm{LF}} &= (1/\sqrt{3})\left[\begin{array}{ccc} 1 & 1 & 1 \end{array}\right]^T \\
\mathbf{u}_{\mathrm{RB}} &= (1/\sqrt{3})\left[\begin{array}{ccc} -1 & -1 & 1 \end{array}\right]^T \\
\mathbf{u}_{\mathrm{LB}} &= (1/\sqrt{3})\left[\begin{array}{ccc} -1 & 1 & -1 \end{array}\right]^T \\
\mathbf{u}_{\mathrm{RF}} &= (1/\sqrt{3})\left[\begin{array}{ccc} 1 & -1 & -1 \end{array}\right]^T . \qquad (11)
\end{aligned}$$

By weighting the power spectrum of the A-format signals with the vector components of (11) a spatial average of their energy flow can be taken as

$$\mathbf{I}_e(l,n) = [\mathbf{u}_{\mathrm{LF}}\ \mathbf{u}_{\mathrm{RF}}\ \mathbf{u}_{\mathrm{LB}}\ \mathbf{u}_{\mathrm{RB}}] \cdot \left[\begin{array}{c} |P_{\mathrm{LF}}(l,n)|^2 \\ |P_{\mathrm{RF}}(l,n)|^2 \\ |P_{\mathrm{LB}}(l,n)|^2 \\ |P_{\mathrm{RB}}(l,n)|^2 \end{array}\right] . \qquad (12)$$

It can be shown that for a plane wave this vector quantity is codirectional with the intensity vector, $\mathbf{I}_e \parallel \mathbf{I}_a$, and proportional to it, $\|\mathbf{I}_e\| \propto \|\mathbf{I}_a\|$, for an array of 3 pairs of first-order microphones, equidistant from the origin, placed at each cartesian axis with one microphone oriented on the positive and the other on the negative direction. With this regular octahedral arrangement and Equation (12), only two opposite microphones contribute to the estimation of each cartesian component of $\mathbf{I}_e$. Since the A-format microphone does not possess this symmetry, correct estimation occurs only at specific planes separately for azimuth and elevation. Off these planes there is an estimation bias whose magnitude depends solely on the plane wave direction and the orientation of the capsules. Hence, there is an one-to-one mapping of the biased estimate to the true direction, which can be corrected. The analysis is tested in the present study on the horizontal plane only, where no bias occurs for azimuth estimation and the maximum elevation error reaches 15°, while a full 3-dimensional directional bias correction is left for future study. Furthermore, this relation can be used only if effective attenuation exists between the capsules for a specific incidence direction, which in the case of the A-format microphone is provided by the directionality of the capsules.

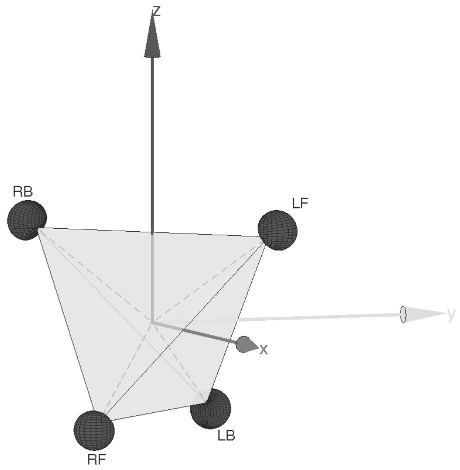


Fig. 2: Capsule arrangement and naming convention of the A-format signals in a regular tetrahedral array.

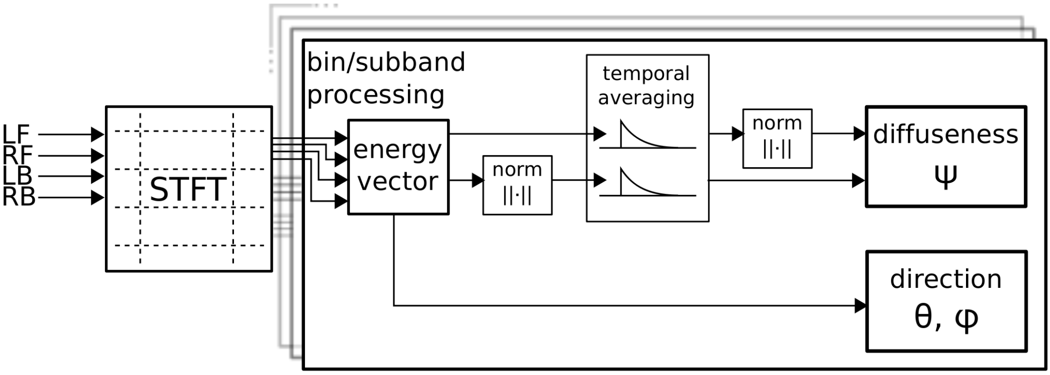


Fig. 3: Flow diagram of the DirAC analysis with A-format input signal.

Since $\mathbf{I}_e$ does not have a magnitude equal to $\mathbf{I}_a$, the intensity-energy density ratio of (9) is not preserved. Therefore, an alternative diffuseness formulation based on the temporal variation of the intensity vector is used, as defined in [8],

$$\psi(l,n) = \sqrt{1 - \frac{\|\langle \mathbf{I}_e(l,n)\rangle\|}{\langle \|\mathbf{I}_e(l,n)\|\rangle}}. \tag{13}$$

Equation 13 is also bounded with $\psi \in [0,1]$ with a value of 1 for a perfectly diffuse field, where the denominator vanishes as the averaged random intensity vectors cancel out, and a value of 0 for a plane wave, where the numerator and denominator are equal. The modified analysis procedure with A-format input signals is presented schematically in Figure 3.

3.2. Simulation of A-format based estimation

The benefit of the A-format analysis compared to the traditional B-format analysis is illustrated in Figure 3, showing the estimated azimuth and respective estimation error for the two methods. Impulse responses (IRs) for an ideal tetrahedral array of cardioid microphones with a radius of 2cm are simulated for plane wave propagation on 72 directions in the horizontal plane. The A-format IRs are also converted to B-format for the respective analysis. The A-to-B-format conversion is performed according to the matrixing scheme of [3]. Further correction filters are employed on the on-axis response of each B-format component. The output to a broadband plane wave is then simulated for each direction by convolving 1 sec of white noise with both IR sets. Furthermore, to introduce internal noise on the analysis signals, additional uncorrelated random noise is added to each of the four A and B components, for each angle, corresponding to a signal-to-noise ratio (SNR) of 20dB with respect to the plane wave. In this manner a more realistic condition is approximated as well as a specific target diffuseness can be defined, as it will be discussed below. The frequency-dependence of internal noise in first-order microphones or the effect of the A-to-B conversion on the SNR is ignored for simplicity. Finally, the A-and-B-format signals are analysed according to the procedure described in section 2 (B-format) and 3.1 (A-format) for each direction and the estimates are averaged across all time frames for each angle.

In Figure 3, the maximum error in the case of the B-format signals is about an order of magnitude larger than the error in the A-format analysis. Moreover, the error is strongly direction dependent, with a minimum at the directions were the plane wave coincides with the axis of the equalized B-format components. The black and white patches in the bottom-left figure, close to 0°azimuth and around 10kHz, are due to estimated values outside the $\theta \in [0\ 2\pi]$ interval, mirrored back into it. It has to be noted that both A and B-format responses are ideal in the sense that only the effect of non-coincidence is simulated. In

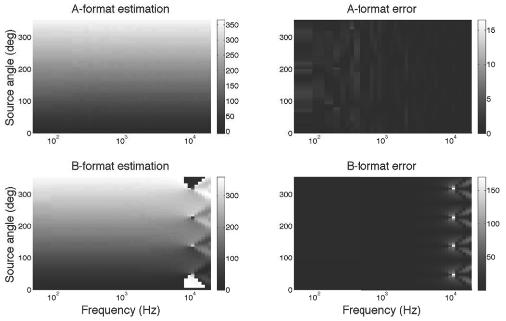


Fig. 3: Azimuth estimation results based on simulated B-format signals and A-format signals for plane wave propagation on the horizontal plane, 2cm array radius and 20dB SNR.

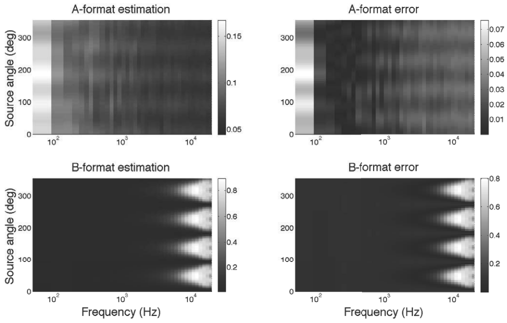


Fig. 4: Diffuseness estimation results based on simulated B-format and A-format for plane wave propagation on the horizontal plane, 2cm array radius and 20dB SNR.

a real-world implementation a less optimal performance is expected due to frequency-dependent directivities or position misalignment of the capsules.

A similar error analysis is presented for the diffuseness estimation in Figure 4. Here the target diffuseness is related to the SNR of the simulated responses, since it is not possible to distinguish between internal capsule noise and sound due to a perfectly diffuse field. This theoretical value can be given as

$$\psi_{\text{ideal}} = \frac{W_{\text{diff}}}{W_{\text{diff}} + W_{\text{plane}}} = \frac{1}{1+\Gamma} \tag{14}$$

where $W_{\text{diff}}, W_{\text{plane}}$ are the energy of the diffuse field and the plane wave respectively, composing the sound field at the evaluation point, and $\Gamma = W_{\text{plane}}/W_{\text{diff}}$ is the direct-to-reverberant ratio (DRR). In the simulation presented above, SNR replaces DRR resulting in a target diffuseness of $\psi_{\text{ideal}} = 0.09$. As in the direction estimation, the error in the B-format analysis at high frequencies reaches values an order of magnitude above the respective A-format ones, and it follows closely the directional distribution of the azimuth error. This overestimation of diffuseness causes the related directional errors to be partly masked by the diffuse stream. However, it can also result in audible artifacts, especially with transients that are well-localized at low frequencies but sound smeared across multiple directions at mid-high frequencies.

4. DIRAC SYNTHESIS WITH A-FORMAT SIGNALS

In the synthesis part, using either only the W signal or combinations of the B-format signals, a superposition of a plane wave and a diffuse field is synthesized for each time-frequency index. The plane-wave component, or non-diffuse stream, is reproduced by the vector-base amplitude panning technique (VBAP) [9], while the diffuse stream is distributed to all loudspeakers after applying decorrelation. Diffuseness ψ is used to adjust the relative energies between the diffuse and non-diffuse stream.

It has been mentioned above that the high-bit-rate version of DirAC employs all the B-format signals in the synthesis stage. This is achieved by the use of virtual microphones oriented at the directions of the reproduction loudspeakers, that increase separation between channels and reduce the amount of needed decorrelation in the synthesis of the diffuse stream [5]. These virtual microphone signals are created as a linear combination of the B-format components as:

$$S_{\text{vmicB}}(l,n,a,\mathbf{u}_0) = aW(l,n) + \frac{1-a}{\sqrt{2}}\mathbf{u}_0^T \cdot \mathbf{X}'(l,n) \tag{15}$$

where $0 < a < 1$ is a directivity coefficient defining the ratio of the pressure and pressure-gradient component, $\mathbf{u}_0$ is a unit vector defining the orientation

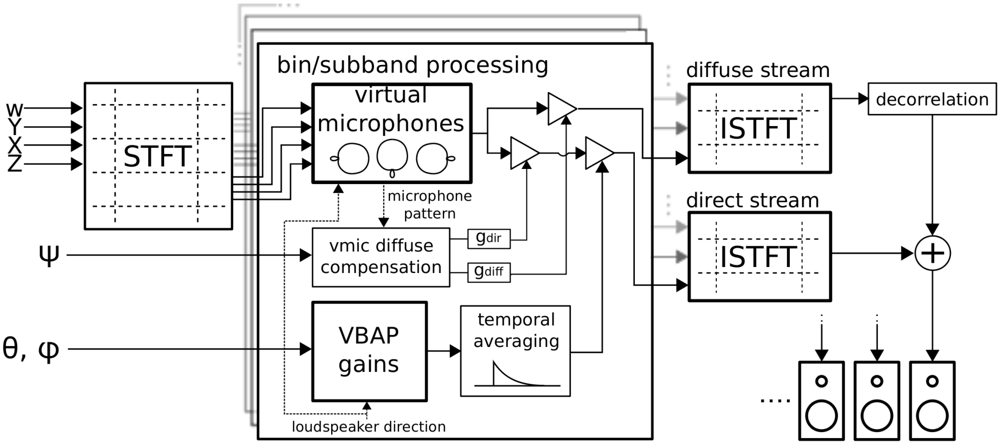


Fig. 5: Flow diagram of the DirAC synthesis with B-format input signal.

of the virtual microphone and $\mathbf{X}'(l,n)$ is the vector of X, Y and Z signals as defined in (2).

The DirAC synthesis procedure is presented more clearly in Figure 5 for a single loudspeaker. The temporal smoothing of the VBAP gains eliminates possible artifacts from sudden directional jumps of a time-frequency tile from one time-frame to the next and has been studied in more detail in [1]. The virtual microphone diffuse-field compensation refers to the gain correction that needs to be applied to equalise the energy of the diffuse and direct stream due to the use of virtual microphones, according to the scheme proposed in [5].

4.1. Direct derivation of virtual microphones from the array geometry

In order to constrain the analysis and synthesis on the use of A-format signals only, without intermediate conversion to B-format, an alternative virtual microphone formulation is employed directly from the arrangement of the microphones of the array. Assuming coincidence, a virtual microphone pointing at $\mathbf{u}_0$ can be derived directly from at maximum 3 transducers with orientations $\mathbf{u}_{\mathrm{m}_1}, \mathbf{u}_{\mathrm{m}_2}, \mathbf{u}_{\mathrm{m}_3}$ as

$$\mathbf{u}_0 = \mathbf{R}\cdot\mathbf{G} \Leftrightarrow \mathbf{G} = \mathbf{R}^{-1}\cdot\mathbf{u}_0 \qquad (16)$$

where $\mathbf{R} = [\mathbf{u}_{\mathrm{m}_1}\ \mathbf{u}_{\mathrm{m}_2}\ \mathbf{u}_{\mathrm{m}_3}]$ denotes the matrix of the microphones' orientations and $\mathbf{G} = [g_1\ g_2\ g_3]^T$ the vector of gains that should be applied to each microphone to steer the directional response to $\mathbf{u}_0$. These gains should be scaled to normalise the on-axis response of the virtual pattern to unity according to $\mathbf{G}_{\mathrm{norm}} = \mathbf{G}/N_{\mathrm{G}}$ with

$$N_{\mathrm{G}} = \sum_{i=1}^{3} g_i \left(a_{\mathrm{m}} + (1-a_{\mathrm{m}})\mathbf{u}_0^T\cdot\mathbf{u}_{\mathrm{m}_i}\right) \qquad (17)$$

where a_{m} is the directivity coefficient of the array microphones.

Compared to ideal B-format virtual microphones, the ones derived from (16) do not possess a directivity pattern independent of orientation and the directivity coefficient cannot be user-defined. The pattern of the virtual microphone for an array of cardioid capsules varies from subcardioid, for any orientation between the capsules, to cardioid, if the orientation coincides with one of the real capsules. The directivity coefficient for a specific direction can be computed as

$$a_{\mathrm{vmic}} = \frac{1}{2}\left[\sum_{i=1}^{3} g_{\mathrm{norm}_i}\left(a_{\mathrm{m}} - (1-a_{\mathrm{m}})\mathbf{u}_0^T\cdot\mathbf{u}_{\mathrm{m}_i}\right) + 1\right]. \qquad (18)$$

Virtual microphones derived from the A-format signals offer lower separation between the loudspeaker signals used for the synthesis stage, compared to the B-format formulation. However, there are a number of advantages in terms of fidelity and reduced coloration of the decoded material. First, in the cases where the virtual microphone coincides with one capsule or with the line joining two capsules, only the signals from one or two capsules are used respectively. This is optimal in the sense of amplification of internal noise of the capsules, which is always going to be greater in the case of a four-capsule matrixing scheme such as the A-to-B-format conversion. Second, access to A-format permits further processing for improvement of the frequency response of the virtual microphones, using a conventional delay-and-sum beamforming technique.

Since the acoustic delays between the capsules are known for a specific propagation direction, the phase differences from the origin can be compensated. This results in a coincident response for that direction and an on-axis flat frequency response of the derived virtual microphone. In detail, to make the capsule signal coincident with the pressure at the origin, a filter response of

$$H_i(\omega, \mathbf{u}_0) = e^{-jkr\mathbf{u}_0^T \cdot \mathbf{u}_{\mathrm{m}_i}} = e^{-j\omega\Delta_i} \qquad (19)$$

where r is the radius of the array and $k = \omega/c$ is the wavenumber of the plane wave, should be applied on the signal, or equivalently a delay of Δ_i in the time domain. Since this delay can be anti-causal for capsules after the origin in the direction of propagation, a time offset corresponding to the radius can be applied to restore causality. The total delay applied to each microphone signal is thus

$$\Delta_i' = \frac{r(1 + \mathbf{u}_0^T \cdot \mathbf{u}_{\mathrm{m}_i})}{c}. \qquad (20)$$

Finally, the on-axis equalized virtual microphone output in the STFT domain is given by

$$S_{\mathrm{vmicA}_{eq}}(l, n, \mathbf{u}_0) = \sum_{i=1}^{3} S_i(l, n) g_{\mathrm{norm}_i} e^{-j\omega_l \Delta_i'} \qquad (21)$$

where S_i is the i^{th} microphone signal and $\omega_l = 2\pi f_l$ is the angular frequency at index l.

It is also possible to apply the same correction to B-format virtual microphones for a specific direction, by applying the filters of (19) to all A-format signals before conversion. However, for generic B-format material that has been externally converted, it is not possible to correct specific directions without knowledge of the array and the conversion process. In terms of the DirAC synthesis stage, by applying (21), sounds coming from or close to the direction of the speaker are colored as little as possible in both the diffuse and non-diffuse stream. Even though for directions significantly off the loudspeaker axis, both the equalized and non-equalized virtual microphones present similar frequency response deviations, off-axis directions do not contribute significantly to the decoded channel due to the application of the VBAP gains.

4.2. Equalization of the energy of diffuse and non-diffuse stream

When virtual microphones are employed in the synthesis, the balance between diffuse and non-diffuse stream is disturbed due to the reduced power capture from a directional microphone in a diffuse field. To counteract that, equalization of the gains between the diffuse and non-diffuse field is applied, based on the pattern of the B-format virtual microphones. A similar scheme is followed for the A-format microphones. The random efficiency (RE) of a first-order directional microphone is given by [10] as

$$RE(a_{\mathrm{vmic}}) = 2a_{\mathrm{vmic}} - 1 + \frac{4}{3}(1 - a_{\mathrm{vmic}})^2. \qquad (22)$$

The gains that should be applied then to the output of the virtual microphones are

$$g_{\mathrm{diff}} = \frac{1}{\sqrt{RE(a_{\mathrm{vmic}})}} \qquad (23)$$

and

$$g_{\mathrm{ndiff}}(l, n) = \frac{1}{\sqrt{1 + \psi(l, n)[RE(a_{\mathrm{vmic}}) - 1]}} \qquad (24)$$

for the diffuse and non-diffuse stream respectively. In the case of B-format virtual microphones, a_{vmic} is constant between the loudspeakers, hence the gains are common between them. For A-format virtual microphones, $g_{\mathrm{diff}}, g_{\mathrm{ndiff}}$ have to be calculated separately for each one of them.

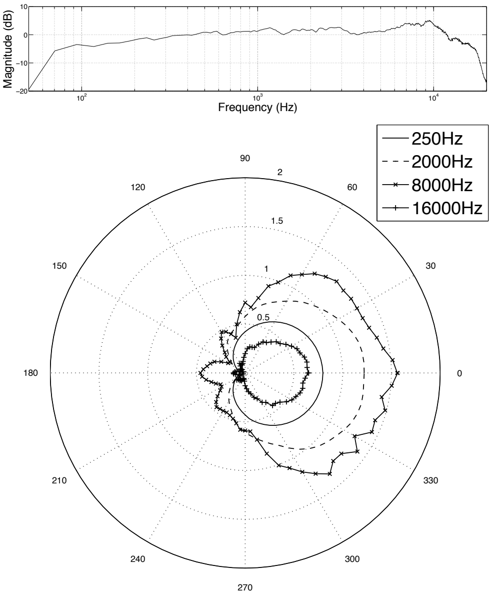
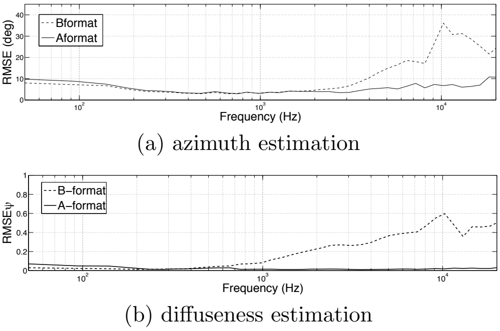


Fig. 6: Response of a single capsule (LF) from the measured A-format microphone. On-axis magnitude response (top) and directivity patterns on horizontal plane for four frequency bands (bottom).

5. MEASUREMENTS AND RESULTS

To validate the performance of the proposed A-format processing in terms of DirAC analysis and synthesis in realistic conditions, measurements were performed in a commercial tetrahedral array. The measured microphone was the Soundfield SPS200, having cardioid capsules and an array radius of approximately 2.5 cm. A-format impulse responses (IRs) were obtained with the exponentially-swept sine method [11] in anechoic conditions in the horizontal plane of the microphone in steps of 5 degrees. Respective B-format IRs were also generated by feeding the A-format IRs through the conversion software that ships with the SPS200 microphone. Apart from the A and B-format responses, a single capsule was also measured to examine its frequency response and the consistency of its directionality with respect to frequency. Figure 6 shows the on-axis normalized magnitude response and the directivity pattern for different frequencies.



(a) azimuth estimation

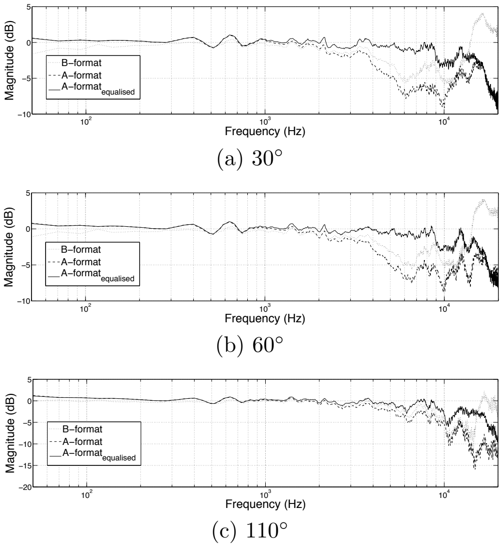
(b) diffuseness estimation

Fig. 7: RMSE for (a) azimuth estimation and (b) diffuseness estimation, for the measured array of 2.5cm radius.

5.1. Parameter estimation results

The performance of the direction and diffuseness estimation using both A and B-format input is investigated by performing DirAC analysis on the measured responses, after convolving them with 1sec of white noise. After averaging across all the analysis frames, the root mean-square error (RMSE) between the actual directions and the estimated ones is plotted in Figure 7a. It can be observed that above 2kHz the A-format direction estimation clearly outperforms the B-format one. Hence, in reality, estimation errors appear at a lower frequency than the approximate theoretical spatial aliasing limit given by [2] as $f_{\mathrm{al}} = c/\pi r$, corresponding to about 4.3kHz for an array of 2.5cm. Below 1kHz the B-format estimation performs slightly better with decreasing frequency, due to the effect of less effective level differences and increased internal microphone noise to the energetic average of (12). Where optimal performance is needed, the two approaches can be combined in the analysis stage by converting to B-format using a standard matrixing scheme and estimating direction in two frequency regions.

Figure 7b displays the RMSE of diffuseness, computed with the same procedure. The specific case corresponds to anechoic conditions hence the target ψ_{ideal} is assumed to be zero. The B-format analy-



(a) 30°

(b) 60°

(c) 110°

Fig. 8: On-axis frequency response of B-format (light-dashed), A-format (thick-dashed) and equalized A-format (solid) derived virtual microphones, for on-axis directions of (a) 30°, (b) 60°and (c) 110°. The measured array has a radius of 2.5cm.

sis begins to deviate significantly from the theoretical value already from 1kHz. This is believed to be due to the diffuseness formulation of (9) where both the numerator and denominator are affected independently by discrepancies in the B-format signals. On the contrary, the A-format analysis and the diffuseness formulation of (13) follow closely the theoretical value. The benefit of correct diffuseness estimation with A-format in terms of audible artifacts and broadband localisation, especially for transient sounds, was also verified by informal listening tests.

5.2. Performance of A-format virtual microphones

Concerning the synthesis part, comparisons on the effect of the B-format and the A-format virtual microphones (Vmics) on the spectral content of the decoded audio stream are investigated. To isolate the effect of non-coincidence from the frequency response of the capsule and measurement loudspeaker, the on-axis measured response of Figure 6 is deconvolved from the responses of the generated Vmics. Figure 8 shows the magnitude response of Vmics steered at three source directions, produced from B-format, A-format and the equalized A-format of (21). It is clear that both A and B-format derived Vmics suffer from severe deviations from the ideal response at high frequencies, depending on the direction of the Vmic. By applying the filters of (19) there is a clear improvement of the on-axis responses as can be seen from the solid line of Figure 6. These equalization filters depend only on the loudspeaker directions and need to be calculated only once for a specific layout.

6. CONCLUSIONS

In this contribution we have presented a DirAC analysis and synthesis formulation based on access to the A-format signals of a regular tetrahedral array. We have shown that the new formulation achieves better parameter estimation at high frequencies than the B-format analysis in realistic conditions. The two approaches can be combined with B-format analysis at low frequencies and A-format analysis at mid-high frequencies. Furthermore, the virtual microphones utilized in the synthesis stage of high-quality DirAC, are reformulated for A-format components. By applying equalization based on the knowledge of the microphone array geometry we have shown that, compared to B-format processing, it is possible to achieve a flatter frequency response of the derived virtual microphones on the directions of the loudspeakers, decreasing coloration in the decoded material.

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

[2] M. A. Gerzon, "The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound," in *50th AES Convention*, London, UK, 1975.

[3] C. Faller and M. Kolundzija, "Design and Limitations of Non-Coincidence Correction Filters for Soundfield Microphones," in *126th AES Convention*, Munich, Germany, 2009.

[4] G. Del Galdo, O. Thiergart, and F. Kuech, "Nested microphone array processing for parameter estimation in Directional Audio Coding," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA'09.*, New Paltz, NY, USA, 2009, pp. 273–276.

[5] J. Vilkamo, T. Lokki, and V. Pulkki, "Directional Audio Coding: Virtual Microphone-Based Synthesis and Subjective Evaluation," *Journal of the Audio Engineering Society*, vol. 57, no. 9, pp. 709–724, 2009.

[6] F. J. Fahy, *Sound intensity*, 2nd ed. London: Taylor & Francis, 1995.

[7] J. Ahonen and V. Pulkki, "Broadband Direction Estimation Method Utilizing Combined Pressure And Energy Gradients From Optimized Microphone Array," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2011. ICASSP 2011.*, Prague, Czech Republic, 2011.

[8] ——, "Diffuseness estimation using temporal variation of intensity vectors," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA'09.* New Paltz, NY, USA: IEEE, 2009, pp. 285–288.

[9] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *Journal of the Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, 1997.

[10] J. Eargle, *The microphone book*, 1st ed. Focal Press, 2001.

[11] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique," in *108th AES Convention*, Paris, France, 2000.