OPEN ACCESS Freely available online

Audio Engineering Society

Convention Paper 9942

Presented at the 144th Convention
2018 May 23 – 26, Milan, Italy

This convention paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This convention paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib), 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.

Advanced B-Format Analysis

Mihailo Kolundžija[1] and Christof Faller[1,2]

[1]*École Polytechnique Fédérale de Lausanne (EPFL), Switzerland*
[2]*Illusonic GmbH, Bahnstrasse 23, 8610 Uster, Switzerland*

Correspondence should be addressed to Mihailo Kolundžija (mihailo.kolundzija@epfl.ch)

ABSTRACT

Spatial sound rendering methods that use B-format have moved from static to signal-dependent, making B-format signal analysis a crucial part of B-format decoders. In the established B-format signal analysis methods, the acquired sound field is commonly modeled in terms of a single plane wave and diffuse sound, or in terms of two plane waves. We present a B-format analysis method that models the sound field with two direct sounds and diffuse sound, and computes the three components' powers and direct sound directions as a function of time and frequency. We show the effectiveness of the proposed method with experiments using artificial and realistic signals.

1 Introduction

The four-channel B-format signal [1, 2] captures a sound field beyond its single-point sound pressure signal, and is used for multi-channel decoding and flexible, yet accurate spatial sound rendering [3].

With the advent of time-frequency processing schemes and during the course of the last decade, we have seen a number of adaptive multichannel rendering schemes that use B-format. These usually work by determining the parameters of a modeled sound field in time-frequency tiles, and generating multi-channel surround mixes adaptively based on the estimated parameters. Directional Audio Coding (DirAC) [4] models the B-format signal as a single plane wave and diffuse sound in each time-frequency tile, and renders them using amplitude panning and decorrelation. High Angular Resolution Planewave Expansion (HARPEX) [5] models B-format in each time-frequency tile with two plane waves, which are rendered with amplitude panning.

The approach presented in this paper deals only with B-format analysis, while we leave the rendering aspect for another paper. In terms of the complexity of the used sound field model, our analysis framework can be seen as going a step further by combining the models of DirAC and HARPEX; namely, we consider three-dimensional sound fields that consist of two far-field sources ("direct sounds") and diffuse sound in each time-frequency tile.

As the main tool, we use eigendecomposition of the B-format covariance matrix, which we combine with algebraic and geometrical observations related to B-format linear mixes and their directional responses. By doing so, we arrive at simple methods for computing directions of arrival of direct sounds, and the powers of the constituent sounds, both direct and diffuse.

A careful reader will notice similarities between some aspects of the approach presented in this paper and the MUSIC algorithm [6]. We do not claim novelty over or superiority to MUSIC in any way; we rather see our approach as a combination of a model choice and signal transformations that favor the eigendecomposition-based solution also featured in MUSIC, while we also present a geometrical analysis that gives rise to closed-form solutions to the problem of source localization on one, and source signals' power computation on the other hand.

The paper is organized as follows. Section 2 gives an account of the used notation and sound field model. Section 3 describes the B-format signal representation and its relation to the analyzed sound field model. Section 4 gives an overview of ways to obtain various first-order directional responses by combining the B-format signal components. In Section 5, we present a B-format signal analysis that gives rise to a procedure for localizing two concurrent far-field sources, while in Section 6 we give a procedure for estimating the powers of the far-field sound sources and the diffuse sound component. For the sake of completeness, Section 7 adapts the presented approach to a two-dimensional sound field model. Section 8 presents the results from a number of experiments, while the conclusions are in Section 9.

2 Notation and Sound Field Model

In this paper, sound source signals and microphone signals are represented by either their sampled, discrete-time representation, such as $w(n)$, or their short time-frequency representation $W(n,f)$, e.g. short-time Fourier transform (STFT), where n and f denote time and frequency indices, respectively. The presented analysis equally applies to both representations, but for convenience we adopt the latter, uppercase representation, and omit the time and frequency indices for brevity.

We also make use of the notion of *signal vectors*, whose elements are sound source signals, microphone signals, or combinations thereof. For example, a signal vector S may consist of N sound source signals $S_0, S_1, \ldots, S_{N-1}$, and we denote it by

$$\mathsf{S} = \begin{pmatrix} S_0 & S_1 & \cdots & S_{N-1} \end{pmatrix}^\top. \tag{1}$$

For simplicity and without loss of generality, the microphone—and thus our analysis—is placed at the origin. The sound fields we consider consist of multiple far-field sources (or direct sounds) in addition to a perfectly diffuse sound field. We denote by N the number of far-field sources, but we mostly focus on the case where $N = 2$.

Any far-field source s_i is described by its sound pressure at the origin, S_i, and a unit-norm vector $\mathbf{n}_i = \begin{pmatrix} n_{i,0} & n_{i,1} & n_{i,2} \end{pmatrix}^\top$ defining its direction of arrival (DOA) along the three Euclidean axes. The sound pressure of the diffuse sound field component at the origin is denoted by S_d.

We further assume that all sources, including the diffuse sound, can be modeled as zero-mean, mutually uncorrelated random processes. The power of S_i is given by $P_i = \mathbb{E}[S_i S_i^H]$, where $\mathbb{E}[\cdot]$ denotes the expectation operator and S_i^H denotes the conjugate-transpose or Hermitian of S_i. In practice, the expectation $\mathbb{E}[\cdot]$ corresponds to a short-time energy estimate.

We denote by $\langle \mathbf{n}, \mathbf{m} \rangle$ the inner product of vectors $\mathbf{n}$ and $\mathbf{m}$. The l^2-norm of a vector $\mathbf{n}$ is denoted by $\|\mathbf{n}\|_2$ and is related to the inner product via $\|\mathbf{n}\|_2 = \sqrt{\langle \mathbf{n}, \mathbf{n} \rangle}$. We also use $\mathrm{diag}(\mathbf{n})$ to denote a diagonal matrix whose main diagonal contains the entries of $\mathbf{n}$.

The notions of *direct sound* and *far-field source* are used interchangeably.

3 B-Format Signal

The standard[1] B-format signal representation [1] uses a four-signal vector

$$\mathsf{B}_s = \begin{pmatrix} X_s & Y_s & Z_s & W_s \end{pmatrix}^\top \tag{2}$$

and corresponds to a coincident sound pickup of the following four signal components:

- sound pressure W_s, with an omnidirectional directional characteristic $d_{W_s}(\theta,\phi) = 1$
- sound pressure gradient (or particle velocity) X_s along the x-axis, with a figure-of-eight or dipole directional characteristic $d_{X_s}(\theta,\phi) = \sqrt{2}\cos\theta\cos\phi$

[1] We use the attribute *standard* and subscript s to denote the B-format signal widely known in the literature [1, 3] and distinguish it from a slightly modified, but equivalent representation that we adopt in this paper.

- sound pressure gradient (or particle velocity) Y_s along the y-axis, with a figure-of-eight or dipole directional characteristic $d_{Y_s}(\theta,\phi) = \sqrt{2}\sin\theta\cos\phi$
- sound pressure gradient (or particle velocity) Z_s along the z-axis, with a figure-of-eight or dipole directional characteristic $d_{Z_s}(\theta,\phi) = \sqrt{2}\sin\phi$.

Angles θ and ϕ are azimuth and elevation, respectively. The scaling by $\sqrt{2}$ of X_s, Y_s, and Z_s is used to balance the average power between the omni and the three dipole channels [3].

Fig. 1 shows the directional responses of the four B-format signal components.

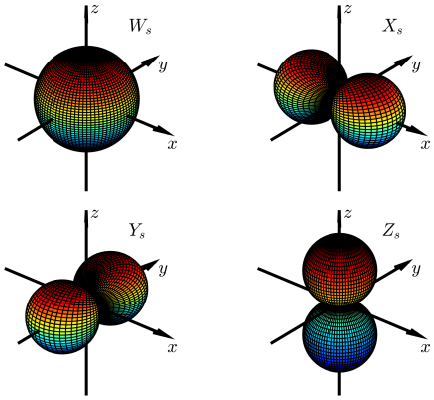

Fig. 1: B-format signal's directional responses.

For reasons elaborated later in the text, we introduce a reversible change of the standard B-format signal vector B_s. This change consists of removing the scale factor $\sqrt{2}$ from the three dipole channels, while introducing the scale factor of $\frac{1}{\sqrt{3}}$ in the omni channel. We denote the modified B-format vector as

$$\mathsf{B} = \begin{pmatrix} X & Y & Z & W \end{pmatrix}^\top. \tag{3}$$

If we define the vector $\mathbf{c}$,

$$\mathbf{c} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{pmatrix}^\top,$$

to be the vector of scale factors modifying the standard B-format vector B_s, then the relation between $\mathbf{B}$ and $\mathbf{B}_s$ is given via

$$\mathsf{B} = \mathrm{diag}(\mathbf{c})\,\mathsf{B}_s\,. \tag{4}$$

The modified B-format signal vector defined in this way corresponds to a set of virtual microphones with the following directional responses:

$$\begin{aligned} d_X(\theta,\phi) &= \cos\theta\cos\phi \\ d_Y(\theta,\phi) &= \sin\theta\cos\phi \\ d_Z(\theta,\phi) &= \sin\phi \\ d_W(\theta,\phi) &= \frac{1}{\sqrt{3}}\,. \end{aligned}$$

Taking the definition of the norm of $d(\theta,\phi)$,

$$\|d\|_2 = \left(\frac{1}{4\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\int_0^{2\pi} |d(\theta,\phi)|^2\cos\phi\,\mathrm{d}\theta\,\mathrm{d}\phi\right)^{\frac{1}{2}}, \tag{5}$$

it is easy to show that the norms of the four modified directional responses, d_X, d_Y, d_Z, and d_W are all equal to $\frac{1}{\sqrt{3}}$.

3.1 B-format of far-field sources

As mentioned before, far-field sources come from directions represented by unit vectors $\mathbf{n}_i$, each pointing towards the DOA of the corresponding source. The direction-dependent gains of any source s_i in the B-format signal vector are given by a four-element vector

$$\mathbf{g}_i = \begin{pmatrix} n_{i,0} & n_{i,1} & n_{i,2} & \frac{1}{\sqrt{3}} \end{pmatrix}^\top. \tag{6}$$

We call vectors $\mathbf{g}_i$ signals' *B-format gain vectors*.

With N far-field sources, each with the sound pressure signal S_i and B-format gain vector $\mathbf{g}_i$, the B-format signal vector B_f can be expressed as

$$\mathsf{B}_f = \mathbf{G}\,\mathsf{S}\,, \tag{7}$$

where the $4\times N$ B-format signal gain matrix $\mathbf{G}$ has the form

$$\begin{pmatrix} \mathbf{g}_0 & \mathbf{g}_1 & \cdots & \mathbf{g}_{N-1} \end{pmatrix}, \tag{8}$$

and the signal vector S has length N and is given by

$$\mathsf{S} = \begin{pmatrix} S_0 & S_1 & \cdots & S_{N-1} \end{pmatrix}^\top. \tag{9}$$

Using elementary matrix calculus, we can express the B-format covariance matrix $\mathbf{R}_f$ as

$$\mathbf{R}_f = \mathbb{E}\left[\mathsf{B}_f\,\mathsf{B}_f^H\right] = \mathbf{G}\,\mathbb{E}\left[\mathsf{S}\mathsf{S}^H\right]\mathbf{G}^\top = \mathbf{G}\,\mathbf{R}_S\,\mathbf{G}^\top. \tag{10}$$

Since we assumed the far-field sources to be zero-mean and uncorrelated, the signal covariance matrix $\mathbf{R}_S$ is diagonal and of the form $\mathbf{R}_S = \mathrm{diag}\,(P_0, P_1, \ldots, P_{N-1})$, which simplifies the B-format covariance matrix to the following form:

$$\mathbf{R}_f = \sum_{i=0}^{N-1} P_i \mathbf{g}_i \mathbf{g}_i^\top. \tag{11}$$

3.2 B-format of diffuse sound

We denote by B_d the B-format signal vector for perfectly diffuse sound and by X_d, Y_d, Z_d, and W_d its four signal components. Unlike for far-field sources, there is no deterministic relation between the four B-format components; we only know that $W_d = \frac{1}{\sqrt{3}} S_d$. However, by the fact that the four directional characteristics d_X, d_Y, d_Z, and d_W are orthogonal functions, the B-format covariance matrix $\mathbf{R}_d$ has a diagonal form, and its entries are proportional to the squared norms of the four directional characteristics. Since those are made equal by our B-format modification (4), the B-format covariance matrix is proportional to the 4×4 identity matrix $\mathbf{I}$,

$$\mathbf{R}_d = P_d \mathbf{I}, \tag{12}$$

where P_d is the diffuse-sound power in any of the four B-format channels, and is related to the diffuse sound power P_{d_s} in the standard B-format omni channel W_s by $P_d = \frac{1}{3} P_{d_s}$.

3.3 B-format of the analyzed sound field model

Since our sound field model combines far-field sources and diffuse sound, the B-format signal vector is a combination of B_f and B_d,

$$\mathsf{B} = \mathsf{B}_f + \mathsf{B}_d = \mathbf{G}\,\mathsf{S} + \mathsf{B}_d\,. \tag{13}$$

Similarly, the B-format covariance matrix $\mathbf{R}$ is given by

$$\mathbf{R} = \sum_{i=0}^{N-1} P_i \mathbf{g}_i \mathbf{g}_i^\top + P_d \mathbf{I}. \tag{14}$$

4 B-Format Transformations

B-format transformations are simple manipulations of the B-format signal vector or its sub-vectors, such as

$$\mathsf{B}_D = \begin{pmatrix} X & Y & Z \end{pmatrix}^\top \tag{15}$$

or

$$\mathsf{B}_P = \begin{pmatrix} X & Y & W \end{pmatrix}^\top. \tag{16}$$

To avoid any confusion, we denote by B' the signal vector that may take the form of the B-format vector itself or any of its previously mentioned sub-vectors.

A B-format transformation defines a mix of the B-format signal components and takes the form

$$U = \mathbf{m}^\top \mathsf{B}'. \tag{17}$$

The vector $\mathbf{m}$ defines the B-format transformation and is called the B-format *signal transformation vector*. In this paper, we are mainly interested in the B-format transformations whose transformation vectors are of unit norm, namely $\|\mathbf{m}\|_2 = 1$.

4.1 Dipole rotation

Consider any of the three B-format dipole signals, X, Y, or Z. A far-field source coming from the direction defined by a vector $\mathbf{n}$ is acquired with a gain equal to the projection of $\mathbf{n}$ onto the unit-norm vector defining the dipole's axis. This is apparent from the definition (6) of the signal's B-format gain vector, and represents a property that actually holds true for a dipole response aligned along any direction, and not only along the three Euclidean axes.

If one applies the B-format transformation $U = \mathbf{m}^\top \mathsf{B}_D$ using a unit-norm transformation vector $\mathbf{m}$, a far-field source with the DOA defined by $\mathbf{n}$ is acquired with the gain $\langle \mathbf{n}, \mathbf{m} \rangle$. In other words, the gain of the far-field source in U is equal to the projection of $\mathbf{n}$ onto the unit-norm vector $\mathbf{m}$. This in turn means that U is a dipole response along the axis $\mathbf{m}$.

More generally and in the interest of the following analysis, one can take the signal vector B_D and obtain another dipole signal vector B_D^r along three orthogonal directions $\mathbf{m}_0$, $\mathbf{m}_1$, and $\mathbf{m}_2$, by applying the transform

$$\mathsf{B}_D^r = \begin{pmatrix} \mathbf{m}_0 & \mathbf{m}_1 & \mathbf{m}_2 \end{pmatrix}^\top \mathsf{B}_D = \mathbf{M}^\top \mathsf{B}_D\,, \tag{18}$$

where $\mathbf{M}$ is a 3×3 rotation matrix defined by the three vectors $\mathbf{m}_0$, $\mathbf{m}_1$, and $\mathbf{m}_2$.

4.2 First-order responses

Performing a B-format transformation on the B-format vector B (or its sub-vector B_P), one obtains a signal $U = \mathbf{m}^\top \mathsf{B}$ whose directional response is what is called a general first-order directional response.

4.2.1 First-order response of far-field sources

From (7), we get the B-format signal vector of a single far-field source s,

$$\mathsf{B} = S\mathbf{g}, \tag{19}$$

where S is the sound pressure due to the source at the origin, and $\mathbf{g}$ the source's B-format gain vector related to its DOA $\mathbf{n}$ through (6). After applying the B-format transformation (17) to the B-format signal vector (19), one gets

$$U = S\left[m_0 n_0 + m_1 n_1 + m_2 n_2 + \frac{m_3}{\sqrt{3}}\right], \tag{20}$$

which indicates the directional response obtained through the B-format transformation, i.e., the way the source's direction defined by $\mathbf{n}$ gets mapped to an acquisition gain.

Besides the obvious reduction to a dipole when $m_3 = 0$, we note the well-known *cardioid response* obtained for $m_3 = \frac{\sqrt{3}}{2}$. The cardioid response has a single spatial null in the direction defined by the unit-norm vector

$$\mathbf{n} = \frac{1}{\sqrt{m_0^2 + m_1^2 + m_2^2}} \begin{pmatrix} -m_0 & -m_1 & -m_2 \end{pmatrix}^\top. \tag{21}$$

Fig. 2 illustrates a cardioid response in the xy-plane, whose directional null also lies in the xy-plane.

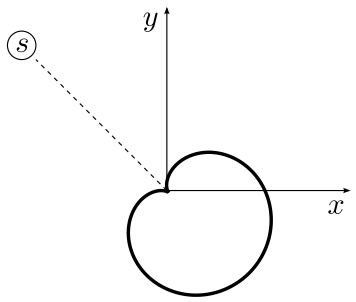

Fig. 2: Cardioid response in the xy-plane.

When m_3 takes on values from the interval $\left[0, \frac{\sqrt{3}}{2}\right)$, the obtained first-order response takes a "tailed cardioid" shape,[2] with multiple directional zeros; these are given by the solutions of the following system with the components of $\mathbf{n}$ as unknowns:

$$m_0 n_0 + m_1 n_1 + m_2 n_2 = -\frac{m_3}{\sqrt{3}} \tag{22}$$

$$n_0^2 + n_1^2 + n_2^2 = 1\,. \tag{23}$$

One can recognize the equation of a plane in (22), while (23) defines the unit sphere centered at the origin; the intersection of the two is a circle, and it exists when $m_3 \in \left[0, \frac{\sqrt{3}}{2}\right]$. The directional nulls of U thus lie on the half-cone (nappe, in mathematical parlance) defined by the unit-norm vectors lying on the intersection circle.[3]

4.2.2 First-order response of perfectly diffuse sound

The B-format signal vector as defined in this paper gives a convenient property of *diffuse sound response invariance under B-format transformations*. By this we mean that the power of diffuse sound is equal across the four B-format channels, and stays the same under any transformation $U = \mathbf{m}^\top \mathsf{B}$ with $\|\mathbf{m}\|_2 = 1$. We show this by first noting that the power of the signal U is given by

$$P_U = \mathbb{E}\left[U\,U^H\right] = \mathbb{E}\left[\mathbf{m}^T \mathsf{B}\,\mathsf{B}^H \mathbf{m}\right] = \mathbf{m}^T \mathbf{R}\,\mathbf{m}\,. \tag{24}$$

Using the expression (12) for the B-format covariance matrix of a diffuse sound field, we get the property of diffuse sound response invariance:

$$P_U = P_d\,\mathbf{m}^T \mathbf{I}\,\mathbf{m} = P_d\,. \tag{25}$$

5 B-Format Source Localization

5.1 Computing the minimum-power B-format transformation

In the following analysis, signal vector B' can be the full B-format vector or any of its sub-vectors, and $\mathbf{R}'$

[2] Tailed cardioid responses include the well-known hyper-cardioid and super-cardioid. Fig. 3 illustrates a tailed cardioid response restricted to the xy-plane.

[3] Note that for the cardioid response, namely when $m_3 = \frac{\sqrt{3}}{2}$, the intersection circle collapses to a single point and the half-cone to a single line.

is the covariance matrix of B'. We consider the sound field model presented in Section 2, so every B-format gain vector $\mathbf{g}_i$ takes the form of its sub-vector $\mathbf{g}'_i$, with the same entry-wise correspondence as that between B and B'. We seek the vector $\mathbf{m}$ and the corresponding B-format transformation $U = \mathbf{m}^\top \mathsf{B}'$, with $\|\mathbf{m}\|_2 = 1$, such that the power of U is minimized.

Using (24) and (14), we can express the power P_U as

$$P_U = \sum_{i=0}^{N-1} P_i \, |\langle \mathbf{m}, \mathbf{g}'_i \rangle|^2 + P_d. \tag{26}$$

We know from linear algebra that the unit-norm $\mathbf{m}$ that minimizes P_U in (24) is given by the eigenvector of $\mathbf{R}'$ that corresponds to its smallest eigenvalue. The minimum P_U obtained thereby is equal to the smallest eigenvalue of $\mathbf{R}'$.

Additionally, (26) tells us something more: the power P_U is minimized and equal to the diffuse sound power P_d when the vector $\mathbf{m}$ is orthogonal to all the gain vectors $\mathbf{g}'_i$. This in turn means that the first-order response defined by the optimal $\mathbf{m}$ has directional nulls in all directions $\mathbf{n}_i$ of the far-field sources.

For the latter to be possible in a general case, there is a limit to the number of far-field sources N. If B' refers to the full B-format vector, N can be at most 3; if B' is a sub-vector of B, N cannot be larger than 2.

When $N = 3$, the vector $\mathbf{m}$ minimizing (24) defines a first-order response that has spatial nulls on a half-cone that contains the unknown directions $\mathbf{n}_0$, $\mathbf{n}_1$, and $\mathbf{n}_2$ of the three far-field sources. It however does not allow one to come up with the three unknown directions in a closed form.

A closed-form solution for the far-field directions can be obtained when $N = 2$, and we describe a two-step procedure for obtaining it.

5.2 Localization of two far-field sources in a plane

Without loss of generality, we assume that the sound field consists of two far-field sources, both in the xy-plane, and diffuse sound. We perform the analysis on the B-format signal sub-vector B_P defined in (16), whose covariance matrix is denoted by $\mathbf{R}_P$.

We seek a unit-norm B-format transformation vector $\mathbf{m}_P$ of length 3 that minimizes the power of $U = \mathbf{m}^\top \mathsf{B}_P$. From our previous analysis, we know that the solution is given by the eigenvector of $\mathbf{R}_P$ that corresponds to its smallest eigenvalue. We also know that the obtained first-order response has directional nulls in the directions of the two far-field sources. These two directions in the xy-plane, $\mathbf{n}_0 = \begin{pmatrix} n_{0,0} & n_{0,1} \end{pmatrix}^\top$ and $\mathbf{n}_1 = \begin{pmatrix} n_{1,0} & n_{1,1} \end{pmatrix}^\top$, are the solutions of the following system with two unknowns:

$$\begin{aligned} m_{P,0}\, n_0 + m_{P,1}\, n_1 &= -\frac{m_{P,2}}{\sqrt{3}} \\ n_0^2 + n_1^2 &= 1\,. \end{aligned} \tag{27}$$

The two solutions are easily computed (for instance, by substitution) and we refrain from giving their closed form for the sake of brevity.

Fig. 3 gives a graphical illustration of a first-order response $U = \mathbf{m}_P^\top \mathsf{B}_P$ whose directional nulls point towards the two far-field sources, s_0 and s_1.

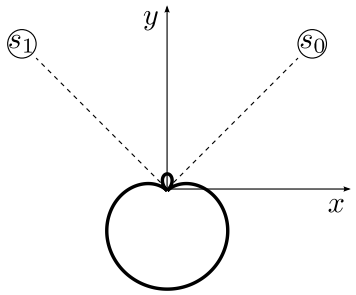

Fig. 3: First-order ("tailed cardioid") response with directional nulls towards s_0 and s_1.

5.3 Localization of two far-field sources in 3D

Here we perform the analysis using the B-format signal sub-vector B_D defined in (15); the covariance matrix of B_D is denoted by $\mathbf{R}_D$.

We start in the same way as before by looking for the unit-norm vector $\mathbf{m}_D$ that minimizes the power of the B-format transformation $U = \mathbf{m}^\top \mathsf{B}_D$. From the previous analysis, we know that U has a dipole response pointing in the direction defined by $\mathbf{m}_D$.

The vector $\mathbf{m}_D$ is orthogonal to both $\mathbf{n}_0$ and $\mathbf{n}_1$. In geometric terms, the signal U has a dipole response whose directional null-plane coincides with the plane defined by $\mathbf{n}_0$ and $\mathbf{n}_1$.

We denote by $\mathbf{V}_D$ the rotation matrix that contains the eigenvectors of $\mathbf{R}_D$ sorted in decreasing order of the corresponding eigenvalues,

$$\mathbf{V}_D = \begin{pmatrix} \mathbf{v}_{D,0} & \mathbf{v}_{D,1} & \mathbf{v}_{D,2} \end{pmatrix}. \tag{28}$$

As mentioned above, $\mathbf{m}_D$ is given by the eigenvector of $\mathbf{R}_D$ that corresponds to its smallest eigenvalue, which in this case means $\mathbf{m}_D = \mathbf{v}_{D,2}$.

The matrix $\mathbf{V}_D$ can be used to rotate the three B-format dipoles in B_D and obtain another vector B_D^r with the following dipole responses:

- $B_{D,2}^r$, along the direction $\mathbf{v}_{D,2}$ orthogonal to $\mathbf{n}_0$ and $\mathbf{n}_1$
- $B_{D,0}^r$ and $B_{D,1}^r$, along orthogonal directions $\mathbf{v}_{D,0}$ and $\mathbf{v}_{D,1}$ which lie in the same plane as $\mathbf{n}_0$ and $\mathbf{n}_1$.

The problem of localizing the two far-field sources can now be posed as a two-dimensional localization in the plane defined by the eigenvectors $\mathbf{v}_{D,0}$ and $\mathbf{v}_{D,1}$, using the signal vector $\mathsf{B}_P^r = \begin{pmatrix} B_{D,0}^r & B_{D,1}^r & W \end{pmatrix}^\top$.

Denote by $\mathbf{n}_0^r = \begin{pmatrix} n_{0,0}^r & n_{0,1}^r \end{pmatrix}^\top$ and $\mathbf{n}_1^r = \begin{pmatrix} n_{1,0}^r & n_{1,1}^r \end{pmatrix}^\top$ the solutions to the two-dimensional source localization problem in the plane defined by $\mathbf{v}_{D,0}$ and $\mathbf{v}_{D,1}$. We note that this plane can be viewed as a rotated *xy*-plane, and $\mathbf{n}_0^r$ and $\mathbf{n}_1^r$ obtained by following the procedure described in Section 5.2.

To obtain the three-dimensional directions $\mathbf{n}_0$ and $\mathbf{n}_1$, we need to perform the rotations

$$\mathbf{n}_i = \mathbf{V}_D \begin{pmatrix} \mathbf{n}_i^r \\ 0 \end{pmatrix}, \tag{29}$$

or more simply

$$\mathbf{n}_i = \begin{pmatrix} \mathbf{v}_{D,0} & \mathbf{v}_{D,1} \end{pmatrix} \mathbf{n}_i^r. \tag{30}$$

6 Power of Diffuse Sound and Far-Field Sources

Even when $N = 3$, the solution of (24), given by the eigenvector of $\mathbf{R}$ that corresponds to its smallest eigenvalue, gives the signal whose power P_U equals the smallest eigenvalue of $\mathbf{R}$, but also equals the diffuse sound power P_d. In short, the diffuse sound power P_d is equal to the smallest eigenvalue of the B-format covariance matrix $\mathbf{R}$. The same is obtained when $N = 2$, irrespective of whether $\mathbf{R}$, $\mathbf{R}_D$, or $\mathbf{R}_P$ is used. Note that the diffuse-sound power P_{d_s} of the standard B-format omni channel W_s is obtained via $P_{d_s} = 3P_d$.

The powers P_0 and P_1 of the two far-field sources are slightly more difficult to obtain. One way, presented here, consists of constructing two cardioid responses, d_{C_0} and d_{C_1}, each with its spatial null in the far-field source direction, $\mathbf{n}_0$ and $\mathbf{n}_1$, respectively. Knowing the power P_d of the diffuse sound and the gain of the cardioid in the other far-field source's direction allows computing the power of that source.

Let the vector $\mathbf{m}_i$ define the unit-norm B-format transformation vector giving the cardioid response d_{C_i} described above. It is easy to show that $\mathbf{m}_i$ is given by

$$\mathbf{m}_i = \begin{pmatrix} -\frac{n_{i,0}}{2} & -\frac{n_{i,1}}{2} & -\frac{n_{i,2}}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}^\top. \tag{31}$$

Knowing that $\mathbf{m}_i$ is orthogonal to $\mathbf{g}_i$, we combine (24) and (26) to obtain

$$\mathbf{m}_i^\top \mathbf{R}\, \mathbf{m}_i = \sum_{j \neq i} P_j \left| \langle \mathbf{m}_i, \mathbf{g}_j \rangle \right|^2 + P_d, \tag{32}$$

and from it the power P_j of the far-field source $s_j,\ j \neq i$:

$$P_j = \frac{\mathbf{m}_i^\top \mathbf{R}\, \mathbf{m}_i - P_d}{\left| \langle \mathbf{m}_i, \mathbf{g}_j \rangle \right|^2}. \tag{33}$$

7 Adaption to Two-Dimensional Sound Fields

Our framework can easily be adapted to purely two-dimensional sound fields. With those, far-field sources are located in the *xy*-plane, and so is the diffuse sound.[4]

With the 2D sound field model, the relevant directional responses $d(\theta)$ are functions of azimuth only, and their norm takes the form

$$\|d\|_2 = \left(\frac{1}{2\pi} \int_0^{2\pi} |d(\theta)|^2 \, \mathrm{d}\theta \right)^{\frac{1}{2}}. \tag{34}$$

Contrary to the 3D case, the norms of the three directional responses of the standard 2D B-format vector

[4] By two-dimensional diffuse sound, we mean the sound that has the diffuse sound attributes of homogeneity and directional isotropy in the *xy*-plane.

$\mathsf{B}_s = \begin{pmatrix} X_s & Y_s & W_s \end{pmatrix}^\top$ are already equal. However, to be in line with the previous 3D analysis and have dipole responses of unit gain in their look direction, we modify the standard B-format vector in an analogous fashion, namely

$$\mathsf{B} = \mathbf{s}^\top \mathsf{B}_s \,, \tag{35}$$

where the scaling vector $\mathbf{s}$ now takes the form

$$\mathbf{s} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^\top . \tag{36}$$

Similarly, we denote by $\mathbf{R}$ the covariance matrix of the signal vector B.

The analysis for DOA estimation is analogous to the 3D case. The B-format transformation $\mathbf{m}$ that nulls the two far-field sources is given by the eigenvector of $\mathbf{R}$ that corresponds to its smallest eigenvalue, and the two DOAs, $\mathbf{n}_0$ and $\mathbf{n}_1$, are obtained as the solutions of the following system:

$$\begin{aligned} m_0 n_0 + m_1 n_1 &= -\frac{m_2}{\sqrt{2}} \\ n_0^2 + n_1^2 &= 1 \,. \end{aligned} \tag{37}$$

The diffuse sound power P_d is given by the smallest eigenvalue of $\mathbf{R}$, and is related to the diffuse sound power P_{d_s} of the standard B-format omni channel W_s via $P_{d_s} = 2P_d$.

Finally, the powers P_0 and P_1 of two far-field sources are obtained by forming two cardioid responses, d_{C_0} and d_{C_1}, with directional nulls defined by $\mathbf{n}_0$ and $\mathbf{n}_1$, respectively.

It is easy to show that those two cardioid responses are given by the B-format transformation vectors

$$\mathbf{m}_i = \begin{pmatrix} -\frac{n_{i,0}}{\sqrt{3}} & -\frac{n_{i,1}}{\sqrt{3}} & \frac{\sqrt{2}}{\sqrt{3}} \end{pmatrix}^\top , \tag{38}$$

while the powers of the two sources are obtained from (33). The B-format gain vectors $\mathbf{g}_i$ for the 2D case have the form

$$\mathbf{g}_i = \begin{pmatrix} n_0 & n_1 & \frac{1}{\sqrt{2}} \end{pmatrix}^\top . \tag{39}$$

8 Experimental Results

8.1 Time-frequency overlapping sources

Time-frequency approaches considering only one direct sound per time and frequency, e.g. [7, 4], fail in correctly identifying direct sound when several sources are overlapping in time and frequency.

Ideally, we expect that consideration of two direct sounds in the model allows to identify two sources, even when they overlap in time and frequency.

For the purpose of testing this scenario, we generated a B-format signal consisting of two white Gaussian signals, S_0 and S_1, as direct sounds, and a third white Gaussian signal S_d as diffuse sound. The two direct sounds have azimuths $\theta_0 = -80^\circ$, $\theta_1 = 40^\circ$, and elevations $\phi_0 = -60^\circ$, $\phi_1 = 20^\circ$.

The power ratio $\frac{P_1}{P_0}$ between the powers of S_1 and S_0 is 6 dB, while the power ratio $\frac{P_1}{P_d}$ between the power of S_1 and the power of diffuse sound S_d is 12 dB. We consider power ratios rather than powers because we process the B-format in time and frequency with non-uniform frequency resolution, which renders the aggregate analysis of the computed power values inadequate.

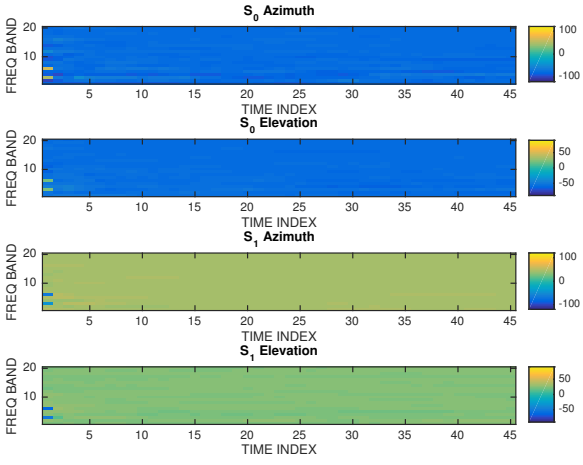

Fig. 4: Estimated direct sound directions across time and frequency. Values are color-coded with ranges indicated by side color bars.

First we present the results of the DOA analysis in the time-frequency domain, shown in Fig. 4. Apart from some variations at the start, where covariance matrix estimates are less accurate, the estimated directions of

the two direct sounds appear to be time- and frequency-invariant. They also seem to be close to the true values, but the visual inspection is insufficient to give an accurate account thereof, which is why we show a more global result overview in the following figures.

Fig. 5 shows the histograms of the two sources' computed DOAs taken over all time-frequency tiles. The estimated DOAs concentrate around and peak at the true values. We also note that the DOA histogram of the second, stronger source, has a smaller spread.

Fig. 6 shows histograms of the estimated power ratios between the two far-field sources on one, and the second far-field source and diffuse sound on the other hand. As with the estimated source directions, we see that both power ratio histograms peak at the true power ratio values.

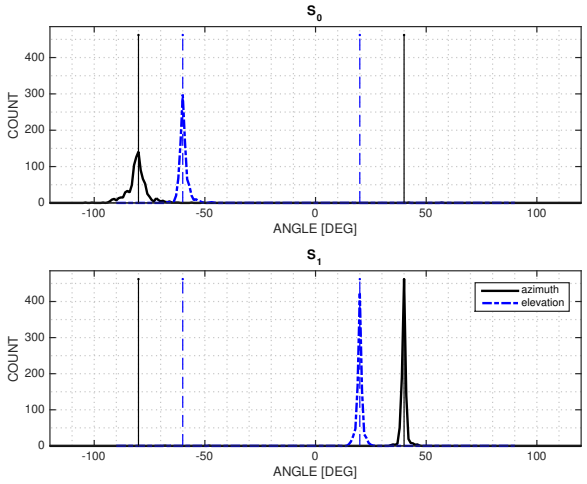

Fig. 5: Histograms of estimated azimuths and elevations.

8.2 Four far-field sources

We mixed a second B-format signal from a multi-track recording with two simultaneous vocals, a violin, and an acoustic guitar. The azimuths and elevations of the instruments are shown in Table 1. The objects were purposely relatively densely located in the front, within the azimuth range of $[-60°, 60°]$.

The fact that the sources are not densely time-frequency overlapping is not unfavorable for the approaches that can accommodate only one direct sound, such as DirAC. We thus give a comparison between the proposed B-format DOA analysis and that of DirAC.

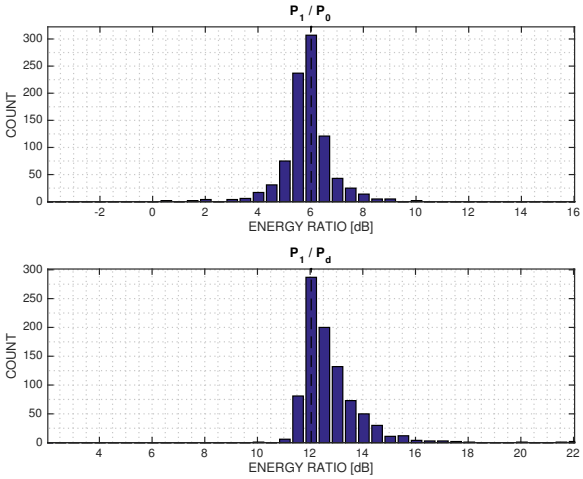

Fig. 6: Histograms of estimated power ratios $\frac{P_1}{P_0}$ and $\frac{P_1}{P_d}$.

Name	*Azimuth*	*Elevation*
Main vocal	$-60°$	$20°$
Background vocal	$60°$	$0°$
Violin	$-30°$	$20°$
Guitar	$30°$	$-30°$

Table 1: Directions of mixed sources.

Fig. 7 compares the results of the DOA analysis of the proposed approach (top two panels) and DirAC (bottom two panels). The first panel gives the histogram of the computed azimuths for both sources taken together, while the second panel gives the analogous histogram with computed elevations; the same holds for the third and fourth panel, which refer to the angles computed with DirAC. Since there are more than two sources, the azimuth and elevation estimations in Fig. 7 vary, depending on the sources' momentary relative strengths. The histograms in the top two panels of Fig. 7 have clear peaks at the source positions specified in Table 1. The peaks in the DirAC DOA analysis, shown in the bottom two panels, are less pronounced, while only the most prominent source, i.e. the guitar, is frequently and correctly localized.

9 Conclusions

We presented an approach for analyzing the B-format signal with a sound field model that considers two direct sounds and diffuse sound in each time-frequency tile. In order to give intuition and show mathematical soundness and rigor, we gave an account of every

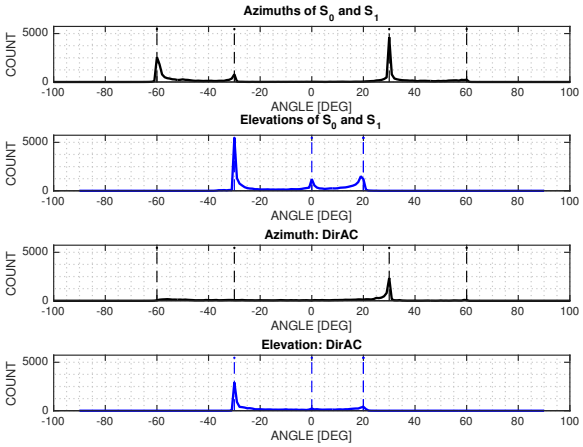

Fig. 7: Histograms of azimuths and elevations estimated with the proposed approach (top two panels) and DirAC (bottom two panels).

step that led to the presented analysis, and we showed its applicability to both three-dimensional and two-dimensional sound fields. Finally, we showed its effectiveness on a few examples.

What still remains is to show how the richer sound field model used in this paper, together with the method for computing its parameters, can serve the principal B-format signal applications, namely high-resolution decoding and spatial sound rendering. We leave this endeavor for a follow-up publication.

References

[1] Farrar, K., "Soundfield Microphone: Design and Development of Microphone and Control Unit," *Wireless World*, pp. 48–50, 1979.

[2] Farrar, K., "Soundfield Microphone - 2: Detailed Functioning of Control Unit," *Wireless World*, pp. 99–103, 1979.

[3] Gerzon, M. A., "Practical Periphony: The Reproduction of full-sphere sound," in *Preprint 65th Conv. Aud. Eng. Soc.*, 1980.

[4] Pulkki, V., "Spatial Sound Reproduction with Directional Audio Coding," *J. Aud. Eng. Soc.*, 55(6), 2007.

[5] Berge, S. and Barrett, N., "High Angular Resolution Planewave Expansion (HARPEX)," in *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics*, pp. 4–9, 2010.

[6] Schmidt, R., "Multiple emitter location and signal parameter estimation," *IEEE Trans. Ant. Prop.*, 34(3), pp. 276–280, 1986.

[7] Faller, C., "Multi-Loudspeaker Playback of Stereo Signals," *J. Aud. Eng. Soc.*, 54(11), pp. 1051–1064, 2006.