Projection-based demixing of spatial audio

Derry Fitzgerald, Antoine Liutkus, Roland Badeau

Projection-based demixing of spatial audio

Derry FitzGerald, Antoine Liutkus, *Member, IEEE,* Roland Badeau, *Senior Member, IEEE*

***Abstract*—We propose a method to unmix multichannel audio signals into their different constitutive spatial objects. To achieve this, we characterize an audio object through both a spatial and a spectro-temporal modelling. The particularity of the spatial model we pick is that it neither assumes an object has only one underlying source point, nor does it attempt to model the complex room acoustics. Instead, it focuses on a listener perspective, and takes each object as the superposition of many contributions with different incoming directions and inter-channel delays. Our spectro-temporal probabilistic model is based on the recently proposed α-harmonisable processes, which are adequate for signals with large dynamics, such as audio. Then, the main originality of this work is to provide a new way to estimate and exploit inter-channel dependences of an object for the purpose of demixing. In the Gaussian $\alpha = 2$ case, previous research focused on covariance structures. This approach is no longer valid for $\alpha < 2$ where covariances are not defined. Instead, we show how simple linear combinations of the mixture channels can be used to learn the model parameters, and the method we propose consists in pooling the estimates based on many projections to correctly account for the original multichannel audio. Intuitively, each such downmix of the mixture provides a new perspective where some objects are cancelled or enhanced. Finally, we also explain how to recover the different spatial audio objects when all parameters have been computed. Performance of the method is illustrated on the separation of stereophonic music signals. *Index Terms*—source separation, probabilistic models, non-negative matrix factorization, musical source separation**

I. Introduction

The past decade has seen an explosion in the use of non-negative matrix factorisation (NMF, [20]) related techniques to tackle the underdetermined sound demixing problem, where the number of audio objects to recover is greater than the number of signals —or *mixtures*— available [38], [44]. This is due to its ability to give an additive parts-based decomposition of audio spectrograms, which facilitates interpretation of the returned frequency and time basis functions. These typically correspond to repeating parts in the audio signal such as repeating notes or chords, or drum hits for example.

It was quickly realised that the task of grouping the basis functions to sound objects was a very difficult task, and while progress has been made in tackling this problem [18], [40], the majority of NMF-based separation research has concentrated on the incorporation of additional constraints such as shift invariance in time and/or frequency [13], or on the incorporation of harmonicity constraints [11], [15], as well as on more principled statistical models for audio time-frequency (TF) representations [3], [8]. Nakano et al utilised a Bayesian nonparametric fusion of NMF and hidden Markov models to cluster basis functions associated with the same note together [28]. Other techniques utilised to overcome the clustering problem included training separation models for specific instruments/sources [1], the incorporation of user assistance into the separation framework [10], [6], [34], and the use of additional side information [25], such as the score of a piece of music [7], [17], [37]. This in turn led to the concept of informed sound source separation, where the original isolated objects are available for analysis along with the original mixture, and the information required to separate the mixtures are transmitted as side-information along with the mixture signal [26], [24], [46].

Aside from the above mentioned methods, there has been much work on the use of spatial cues as a means of enabling audio demixing. These are of particular interest in the context of this paper, where spatial information is used to perform separation. Initial attempts to incorporate spatial cues utilised non-negative tensor factorisation (NTF) and then grouped the recovered basis functions according to their spatial position [12]. Extensions to NTF-based spatial methods include the weighting of TF bins based on an interaural intensity difference function [27], and replacing the channel gain axis of the tensor factorisation with a direction of arrival axis [41]. These have been shown to give improvements over the standard NTF approach. Other approaches to utilise spatial information into the separation process include using the spatial covariance matrix [4], [32]. This approach was generalised to deal with reverberant environments in [5], and an alternative approach to estimate these models was presented in [36]. A beamspace approach to utilising spatial cues in an NMF framework was presented in [21]. However, this approach requires knowledge of the distances between the microphones of the array to be implemented. More recently, an approach that combines direction of arrival estimation with spatial covariance matrices was presented in [30]. Again, this requires knowledge of the microphone array geometry in order for the technique to work.

In this paper, we present a novel approach to separating multichannel audio signals into their different constitutive spatial objects, in the case where the microphone array geometry is unknown, and assuming an anechoic mixing model. Rather than analyse the observed multichannel mixture directly, we instead chose to analyze projections of it on many spatial directions. As we show, this idea permits to advantageously combine the computational effectiveness of NTF methods [12] with the adequacy of probabilistically models for spatial audio [5], [33], [30], to yield a method that is able to demix even diffuse spatial objects in a reasonable time.

The remainder of the paper is organised as follows: Section II details both the spatial and spectro-temporal models

D. FitzGerald is with the Cork School of Music, Cork Institute of Technology, Ireland. A. Liutkus is with Inria, Nancy Grand-Est, Multispeech team, LORIA UMR 7503, Université de Lorraine, France. R. Badeau is with LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France.

This work was partly supported by the research programmes EDiSon3D (ANR-13-CORD-0008-01) and KAMoulox (ANR-15-CE38-0003-01) funded by ANR, the French State agency for research.

utilised in the paper, as well as the proposed spatial projection method and techniques for separation based on the spatial projection method. Section III then details how the parameters of the spatial and spectro-temporal models are estimated, while Section IV details the evaluation of the proposed method. Finally, Section V contains some concluding remarks and identifies areas for future work.

II. Notations and model

We assume we observe an I channel audio signal, called *mixture*. A typical case is when the mixture is stereophonic ($I = 2$) in music processing. Its Short Term Fourier Transform (STFT) is written x and is a $F \times T \times I$ tensor, where F and T respectively stand for the number of frequency bins[1] and time frames. $x\left(f,t\right)$ denotes a $I \times 1$ vector, which gives the value of the complex spectrum of each channel of the mixture at TF bin $\left(f,t\right)$. Here we assume that the mixture is the sum of J multichannel signals y_j, that are called the *object images*:

$$\forall\left(f,t\right), x\left(f,t\right) = \sum_j y_j\left(f,t\right). \tag{1}$$

Each object image STFT y_j is hence also a $F \times T \times I$ tensor. To understand this, consider the case of three instruments such as vocals, piano and guitar, mixed down into a stereo mixture ($I = 2$). Each object, such as the vocals, appears as stereophonic *within* the mix. For convenience, the i^{th} channel of an image y_j is noted y_{ij}. Of course, all such channels are not equal. In general, the *mixing process*, i.e. the way a sound engineer produces the mixture, is such that each image will appear as mostly located along one particular *direction*. For instance, in the case of stereophonic music, a guitar may be located mostly on the left while a piano is on the right, with centered vocals. Ambient sound objects are characterized by the fact that they perceptually sound as coming from many directions. In this study, we propose a probabilistic model to account for such spatial audio. First, we describe the simple punctual model in Section II-A and then the general diffuse model in Section II-B.

A. Punctual anechoic model

For each object image y_j, we assume for now there is one single underlying monophonic signal, called the object *source*. The STFT of this signal, written s_j, is an $F \times T$ matrix.

In the *punctual anechoic* model, each channel of the object image is then obtained by a simple delay and gain applied to this unique source. For instance, in the stereophonic case $I = 2$, using a pan-pot linear instantaneous mixing model, we define a *panning* vector $\phi = [\cos\theta, \sin\theta]$, where $\theta \in [0, \pi/2]$ is defined as the *panning* angle, and generalise it to an anechoic case through the incorporation of a delay vector τ of size $I \times 1$ and write:

$$\forall\left(f,t\right), y_j\left(f,t\right) = h\left(\phi,\tau \mid f\right) s_j\left(f,t\right), \tag{2}$$

with[2]:

$$\left[h\left(\phi,\tau \mid f\right)\right]_i = \phi_i \exp\left(-\sqrt{-1}\frac{2\pi f}{2F-1}\tau_i\right) \tag{3}$$

where f is an index for the frequency band considered (from 0 to $F-1$) and τ_i is a delay in samples. This model guarantees that the energy of y_j is the same as that of s_j, while giving a particular spatial position and delay to object j. By convention, the delays will be taken relative to the first channel and so with $\tau_1 = 0$. In the general case of an arbitrary number I of channels for the mixture, we define[3]

$$\mathbb{L} \triangleq \underbrace{\mathcal{C}_I}_{\text{pannings}} \times \underbrace{\mathbb{R}^{I-1}}_{\text{delays}}, \tag{4}$$

as the *pan-delay set*, which comprises two parts. The panning part is simply the unitary sphere $\mathcal{C}_I$ in $\mathbb{R}^I$ and generalizes the stereophonic case. The delays part consists of $I \times 1$ vectors giving the delays of the different channels with respect to the first one, measured in samples. As the first entry of τ is always 0, this leaves $I-1$ degrees of freedom. A *panning-delay* $(\phi,\tau) \in \mathbb{L}$ is thus a couple comprising a $I \times 1$ vector of gains with unit norm $\|\phi\| = 1$ and a $I \times 1$ vector of delays $\tau \in \mathbb{R}^I$. As can be seen, the punctual anechoic model (2) simply generalizes classical stereophonic anechoic modelling to an arbitrary number of mixture channels. When listening to the resulting image y_j, we have the clear sensation that the object comes from a particular point in space, hence the name *punctual* for this model.

B. Diffuse object model

In many cases, the punctual anechoic model is not sufficient to account for real audio signals. Indeed, a typical audio object does not come from only one point in space. The size of the vibrating structure or reverberation in the environment causes the signal to be more accurately modelled as a superposition of many such —virtual— point sources. Inspired by recent research on this topic [30], we generalize the simple model (2) to the case where the image y_j is a weighted sum of anechoic contributions from all panning-delays in $\mathbb{L}$:

$$y_j\left(f,t\right) = \int_{(\phi,\tau)\in\mathbb{L}} h\left(\phi,\tau \mid f\right) q_j\left(\phi,\tau\right) s_j\left(f,t \mid (\phi,\tau)\right) d\left(\phi,\tau\right), \tag{5}$$

where all $\left\{s_j\left(f,t \mid (\phi,\tau)\right)\right\}_{(\phi,\tau)}$ are *object sources*, each one corresponding to the part of the object that originates from location (ϕ,τ). Then, $q_j\left(\phi,\tau\right) \geq 0$ is a *pan-delay gain* that indicates the strength of this location in the object. This model is close to that proposed in [30], except the sum was there taken over all possible source positions in the 3D space, where h was then understood as modelling the acoustic transportation of each source to the microphones. As demonstrated in [30], this model has the important advantage of binding all frequencies for the purpose of estimating the parameters, thanks to the acoustic modeling brought in by the mixing filter h. However, it requires prior knowledge about the microphone array to correctly initialize the mixing filters.

In this study, our model (5) takes another route and amounts to completely drop the physical acoustic modeling undertaken in [30] and rather adopt a *receiver* point of view: each

[1] We assume the redundant "negative" frequencies have been discarded.

[2] The $2F$ in (3) is the window size of the STFT, which is assumed even.

[3] $\triangleq$ denotes a definition

object image $y_j(f,t)$ is now modelled as the superposition of contributions with various panning directions ϕ and delays patterns τ, from the point of view of the *listener*, weighted by a gain $q_j(\phi,\tau)$. For the sake of simplicity, (5) is further simplified as in [30] by approximating the integral over $\mathbb{L}$ by a discrete sum over a fixed set $\mathcal{L}$ of L panning-delays couples from $\mathbb{L}$:

$$y_j(f,t) = \sum_{l=1}^{L} h_l(f)\, q_j(l)\, s_j(f,t \mid l), \tag{6}$$

where $h_l(f)$ and $s_j(f,t \mid l)$ are short-hand notations for $h(\phi_l,\tau_l \mid f)$ and $s_j(f,t \mid \phi_l,\tau_l)$, respectively, and:

$$\mathcal{L} = (\phi_1,\tau_1),\ldots,(\phi_L,\tau_L) \in \mathbb{L}^L \tag{7}$$

is called the *sources locations set* and samples $\mathbb{L}$, typically in a regularly-spaced way.

One originality of our model will then be to assume all object sources $\{s_j(f,t \mid l)\}_l$ are independent, although sharing the same underlying energy. This comes as an alternative to the classical convolutive model, which basically amounts to have all the sources $\{s_j(f,t \mid l)\}_l$ of object j being deterministically related one to the other, rather than independent as in here. This relaxation permits us to model diffuse audio objects, for which the different punctual sources are not necessarily coherent. On the contrary, the probabilistic model we pick for them precisely boils down to assuming they all share the same "energy", while having their own random phase. More precisely, inspired by the recent study [22], we assume that all TF bins (f,t) are independent and choose all $\{s_j(f,t \mid l)\}_l$ as independent and all distributed with respect to a complex isotropic α-stable distribution (denoted by $S\alpha S_c$) [35], which is a generalization of the Gaussian case ($\alpha = 2$), that notably handles large dynamic ranges:

$$s_j(f,t \mid l) \sim S\alpha S_c\left(P_j^{\alpha}(f,t)\right). \tag{8}$$

$P_j^{\alpha}(f,t) \geq 0$ is a nonnegative scalar called the fractional Power Spectral Density (α-PSD) of object j at TF bin (f,t) [42]. It can basically be understood as the energy of object j at TF bin (f,t), shared for this object by the sources at all panning-delays. Given one of these pan-delays, we showed, e.g. in [23] for the Gaussian case and in [22] for $\alpha \in]0\,2]$, that model (8) is equivalent to assuming that the underlying object source waveform is both locally stationary and α-stable, which better fits the large dynamic ranges found in audio signals than the Gaussian assumption. See [29], [42], [16] for previous applications of α-stable distributions to audio processing. In the following, we will often take $\alpha \approx 1$, which has long proved to be an adequate choice for music signals (see, e.g. [22], [19] and references therein for a discussion).

C. Spatial projections

Now, our objective will be to estimate the parameters of this multichannel α-stable model. To this purpose, the trick we will be using is to not handle the original multivariate observed mixture $x(f,t) \in \mathbb{C}^I$, but rather many projections of it onto the complex plane. Indeed, directly estimating the parameters of multivariate α-stable distributions is very challenging, but much easier for scalar variables. Our approach then basically consists in pooling the parameters obtained with these scalar projections so as to deduce those of the original multichannel data. This strategy is reminiscent of the pioneering work done in [31], where the same idea was used to estimate the parameters of multivariate α-stable distributions.

Our first step will then be to select a *projection set* $\mathcal{P}$, consisting of M elements from the pan-delay set $\mathbb{L}$:

$$\mathcal{P} = (\phi_1,\tau_1),\ldots,(\phi_M,\tau_M) \in \mathbb{L}^M.$$

The elements of $\mathcal{P}$ are not necessarily the same as those of the sources location set $\mathcal{L}$ defined in (7). We discuss how to choose $\mathcal{L}$ and $\mathcal{P}$ later in Section III-F. For now, consider one of those $(\phi_m,\tau_m) \in \mathcal{P}$ and the corresponding $I \times 1$ complex vector $h(\phi_m,\tau_m \mid f)$ as defined in (3), abbreviated $h_m(f)$. In this study, we will be interested in the scalar $\langle h_m(f), y_j(f,t)\rangle$, which is the dot product $h_m(f)^{\top} y_j(f,t)$ between $h_m(f)$ and $y_j(f,t)$. As a linear combination of $S\alpha S_c$ random variables, it is $S\alpha S_c$ itself [35, Th. 2.1.2 p. 58]. From (6) and (8), it is straightforward to show that we have[4]:

$$\langle h_m(f), y_j(f,t)\rangle \sim S\alpha S_c\left(P_j^{\alpha}(f,t)\sum_{l=1}^{L}|\langle h_m(f), h_l(f)\rangle|^{\alpha} q_j^{\alpha}(l)\right). \tag{9}$$

If we introduce the $L \times 1$ vector $k_m(f)$ as:

$$k_m(f) \triangleq \left[|\langle h_m(f), h_1(f)\rangle|^{\alpha},\ldots,|\langle h_m(f), h_L(f)\rangle|^{\alpha}\right]^{\top}, \tag{10}$$

and the $L \times 1$ vector Q_j as:

$$Q_j \triangleq \left[q_j^{\alpha}(1),\ldots,q_j^{\alpha}(L)\right]^{\top},$$

expression (9) can be written more concisely as:

$$\langle h_m(f), y_j(f,t)\rangle \sim S\alpha S_c\left(P_j^{\alpha}(f,t)\, k_m(f)^{\top} Q_j\right). \tag{11}$$

Exploiting linearity of the dot product and independence of the $S\alpha S_c$ sources, we can easily use result (11) for the dot product between $h_m(f)$ and the mixture and get from (1):

$$\langle h_m(f), x(f,t)\rangle = \sum_j \langle h_m(f), y_j(f,t)\rangle \sim S\alpha S_c\left(\sum_j P_j^{\alpha}(f,t)\, k_m(f)^{\top} Q_j\right). \tag{12}$$

In this study, we gather all $h_m(f)^{\top}$ as the rows of the $M \times I$ so-called *projection matrix* for frequency f, noted $\mathbf{N}_f$:

$$\mathbf{N}_f = \begin{bmatrix} h_1(f)^{\top} \\ \vdots \\ h_M(f)^{\top} \end{bmatrix}. \tag{13}$$

The resulting signals $\langle h_m(f), x(f,t)\rangle$ are gathered into *projection* matrices c_m, each of dimension $F \times T$, as follows:

$$c_m(f,t) \triangleq \langle h_m(f), x(f,t)\rangle. \tag{14}$$

[4]Result (9) can be proved by simply writing the expression for the characteristic function $\mathbb{E}\left[\exp\left(i\Psi\langle h_m(f), y_j(f,t)\rangle\right)\right]$ of $\langle h_m(f), y_j(f,t)\rangle$.

For convenience, $c(f,t)$ will denote the $M\times 1$ vector gathering the different $c_m(f,t)$:

$$c(f,t) \triangleq [c_1(f,t), \ldots, c_M(f,t)]^\top,$$

and the resulting $F \times T \times M$ tensor c is called the *projection tensor*. As can be seen, (14) leads to:

$$c(f,t) = \mathbf{N}_f x(f,t). \tag{15}$$

Now, following (12), the marginal distribution of each entry of the projection tensor is given by:

$$c_m(f,t) \sim S\alpha S_c\left(\sum_j P_j^\alpha(f,t)\, k_m(f)^\top Q_j\right), \tag{16}$$

where $k_m(f)$ is computed only once through (10). As can be seen, the free parameters of this model consist of the $L \times J$ pan-delays gains Q, as well as the $F \times T$ objects α-PSD P_j^α. Concerning the latter, a further possible constraint is to use a Nonnegative Matrix Factorization Model [2], [32], [33]:

$$\forall j, \forall (f,t),\ P_j^\alpha(f,t) = \sum_{r=1}^{R} W_j(f,r)\, H_j(t,r), \tag{17}$$

where $R \in \mathbb{N}^+$ is called the *number of components* and W_j and H_j are $F \times R$ and $T \times R$ nonnegative matrices, respectively. The columns of W_j correspond to spectral patterns and those of H_j correspond to their activations over time. Imposing the parametric model (17) enforces some structure over the α-PSDs and has often proved useful for audio separation [39].

In any case, all parameters of the model are gathered into a parameter set denoted Θ. Depending on whether we adopt the NMF model (17) or not, we can have $\Theta_A = \{W_j, H_j, Q_j\}_j$ or $\Theta_B = \left\{P_j^\alpha, Q_j\right\}_j$.

D. Separation

For now, assume that all parameters have been estimated (we will address the problem of their estimation later in Section III). The question is here: how to perform demixing and separate the original multichannel mixture x given the parameters estimated using the projection tensor c? This is done in the following way: first, the $M \times 1$ projection entries $c(f,t)$ are decomposed into J contributions $y_j^c(f,t)$, which also are of dimension $M \times 1$ and called the *projected images* of the objects, so that:

$$c(f,t) = \sum_j y_j^c(f,t). \tag{18}$$

For this purpose, a simple solution is to discard the dependencies between their different M channels y_{mj}^c and estimate each of them through its marginal expected value given the mixture and parameters [22]:

$$\hat{y}_{mj}^c(f,t) = \frac{P_j^\alpha(f,t)\, k_m(f)^\top Q_j}{\sum_{j'} P_{j'}^\alpha(f,t)\, k_m(f)^\top Q_{j'}} c_m(f,t). \tag{19}$$

As highlighted in [22], procedure (19) generalizes classical Wiener filtering to the case of α-harmonisable processes. Then, given these projected images y_j^c, we aim at recovering the original object images y_j. We achieve this by first noticing that (15) leads to:

$$c(f,t) = \mathbf{N}_f \sum y_j(f,t) = \sum_j \mathbf{N}_f y_j(f,t), \tag{20}$$

and then through (18) that we have:

$$y_j^c(f,t) = \mathbf{N}_f y_j(f,t).$$

Given an estimate $\hat{y}_j^c$ for each projected image y_j^c using (19), a natural solution to estimate the corresponding image y_j is to adopt a least-squares strategy and use:

$$\hat{y}_j(f,t) = \mathbf{N}_f^\dagger \hat{y}_j^c(f,t), \tag{21}$$

where $\cdot^\dagger$ denotes pseudo-inversion. Provided $M \geq I$, this pseudo-inversion will be well behaved. Waveforms of the separated objects images in the time domain are then easily recovered through inverse STFT transforms of the $\hat{y}_j$.

The whole separation procedure, which we call PROJET (PROJection Estimation Technique) is summarized in Algorithm 1. The mixture signal is taken as an input, as well as the different parameters permitting to construct the projection tensor c. Then, the parameters are iteratively estimated and finally used for separation through (19) and (21). In the following, we discuss the parameters estimation method to be used at step 4 of Algorithm 1.

III. Parameter estimation

A natural approach to estimate the parameters Θ of a probabilistic model is to choose Θ so as to maximize the likelihood of the observations. Here, the observations are taken as the entries of the projection tensor c:

$$\hat{\Theta} \leftarrow \underset{\Theta}{\operatorname{argmin}} - \log p(c \mid \Theta). \tag{22}$$

Since all TF bins of the sources are assumed independent in the α-harmonisable model for the object sources, so are those of c. Hence, (22) becomes:

$$\hat{\Theta} \leftarrow \underset{\Theta}{\operatorname{argmin}} - \sum_{f,t} \log p(c(f,t) \mid \Theta) \tag{23}$$

An obvious dependence structure exists between the M different $c_m(f,t)$, due to the way (15) the tensor is constructed. However, instead of taking it into account, we adopt an alternative approach and neglect these dependencies, leading to:

$$\hat{\Theta} \leftarrow \underset{\Theta}{\operatorname{argmin}} - \sum_{f,t,m} \log p(c_m(f,t) \mid \Theta). \tag{24}$$

$$\triangleq \underset{\Theta}{\operatorname{argmin}}\ D_c(\Theta), \tag{25}$$

where $D_c(\Theta)$ is introduced as the global cost function (24) to be minimized. This approximate strategy has for instance been considered for multichannel audio modeling and separation in [32]. It amounts to only fit the parameters using the marginal distribution of the observations c and is called *approximate* maximum likelihood estimation here for this reason. Unfortunately, no analytical expression of the marginal distribution (16) is available in the general case $\alpha \in]0\,2]$, but

Algorithm 1 Overview of PROJET model and method for audio separation through projections. Steps 4a and 4b are achieved using the formulas found in Tables I for β-divergences or II for the Cauchy model.

1) Signal Model (see section II-B)

$$x(f,t) = \sum_j y_j(f,t) \text{ where,}$$

$$y_j(f,t) = \sum_{l=1}^{L} h_l(f) q_j(l)\, s_j(f,t \mid l) \text{ with,}$$

$$h_l(f) = \phi_l \cdot \exp\left(-\sqrt{-1}\frac{2\pi f}{2F-1}\tau_l\right) \in \mathbb{C}^I, \text{ and}$$

$$s_j(f,t \mid l) \in \mathbb{C} \sim S\alpha S_c\left(P_j^\alpha(f,t)\right), \text{ all independent.}$$

2) Input
   - Location set $\mathcal{L} = (\phi_l, \tau_l), \ldots, (\phi_L, \tau_L)$
   - Projection set $\mathcal{P} = (\phi_m, \tau_m), \ldots, (\phi_M, \tau_M)$ (see section III-F)
   - Number of iterations
   - Parameters α and divergence to use
   - Mixture x
3) Initialization
   - Given $\mathcal{P}$, Compute the M $h_m(f)$ with (3)
   - Gather the $h_m(f)$ as the rows of $\mathbf{N}_f$
   - Compute the $c(f,t) = \mathbf{N}_f x(f,t)$
   - Given $\mathcal{L}$, Compute the L $h_l(f)$ with (3)
   - Compute $k_m(f)$ with

   $k_m(f) \triangleq$
   $[|\langle h_m(f), h_1(f)\rangle|^\alpha, \ldots, |\langle h_m(f), h_L(f)\rangle|^\alpha]^\top$
   - Initialize parameters Θ, either $\{W_j, H_j, Q_j\}_j$ or $\left\{P_j^\alpha, Q_j\right\}_j$, randomly.
4) Parameter fitting: for each object j, (see section III)
   a) Update α-PSD: either P_j^α or $\{W_j, H_j\}$
   b) Update pan-delay coefficients Q_j
5) If another iteration is needed, go back to 4.
6) Separation: for each object j:
   a) Estimate the $M \times 1$ projected images $\hat{y}_j^c(f,t)$ through

$$\hat{y}_{mj}^c(f,t) = \frac{P_j^\alpha(f,t)\, k_m(f)^\top Q_j}{\sum_{j'} P_{j'}^\alpha(f,t)\, k_m(f)^\top Q_{j'}} c_m(f,t).$$

   a) Estimate object image, $\hat{y}_j$, through using pseudoinverse of $\mathbf{N}_f$

$$\hat{y}_j(f,t) = \mathbf{N}_f^\dagger \hat{y}_j^c(f,t),$$

   a) Apply inverse STFT to $\hat{y}_j$ to recover waveforms

All updates are computed using the latest available versions of all parameters:

- $P_j^\alpha(f,t) \leftarrow P_j^\alpha(f,t) \cdot \frac{\sum_m k_m(f)^\top Q_j \left[\sigma_m^{\cdot(\beta-2)} \cdot v_m\right]_{ft}}{\sum_m k_m(f)^\top Q_j \left[\sigma_m^{\cdot(\beta-1)}\right]_{ft}}$
- $W_j(f,r) \leftarrow W_j(f,r) \cdot \frac{\sum_m k_m(f)^\top Q_j \left(\sum_t \left[\sigma_m^{\cdot(\beta-2)} \cdot v_m\right]_{ft} H_j(t,r)\right)}{\sum_m k_m(f)^\top Q_j \left(\sum_t \left[\sigma_m^{\cdot(\beta-1)}\right]_{ft} H_j(t,r)\right)}$
- $H_j \leftarrow H_j \cdot \frac{\sum_m \left(\sum_f k_m(f)^\top Q_j \left[\sigma_m^{\cdot(\beta-2)} \cdot v_m\right]_{ft} W_j(f,r)\right)}{\sum_m \left(\sum_f k_m(f)^\top Q_j \left[\sigma_m^{\cdot(\beta-1)}\right]_{ft} W_j(f,r)\right)}$
- $Q_j \leftarrow Q_j \cdot \frac{\sum_{f,t,m} k_m(f) \left[\sigma_m^{\cdot(\beta-2)} \cdot v_m \cdot P_j^\alpha\right]_{ft}}{\sum_{f,t,m} k_m(f) \left[\sigma_m^{\cdot(\beta-1)} \cdot P_j^\alpha\right]_{ft}}$

Table I
MULTIPLICATIVE UPDATES FOR THE β-DIVERGENCE.

only for some particular cases such as $\alpha = 1$ (Cauchy) or $\alpha = 2$ (Gaussian).

We now consider those particular cases in Sections III-A, III-B and III-C, and then propose a heuristic for the general case in Section III-D. The particular setup of punctual objects is discussed in Section III-E, while the choice of the location and projection sets $\mathcal{L}$ and $\mathcal{P}$ is discussed in Section III-F.

A. Gaussian $\alpha = 2$ case

When $\alpha = 2$, it is known in the dedicated literature that minimizing the negative log-likelihood (24) is equivalent to minimizing the Itakura-Saito divergence between $|c|^2$ and the variance model [8], [23]:

$$\hat{\Theta} \leftarrow \underset{\Theta}{\operatorname{argmin}} \sum_{f,t,m} d_0\left(|c_m(f,t)|^2 \mid \sum_j P_j^\alpha(f,t)\, k_m(f)^\top Q_j\right), \tag{26}$$

where $d_0(a \mid b) = \frac{a}{b} - \log\frac{a}{b} - 1$. The Itakura-Saito (denoted IS) divergence is also a special case of the β-divergence when $\beta = 0$, as are the generalised Kullback-Leibler divergence (denoted KL, $\beta = 1$) and least-squared-error ($\beta = 2$, see e.g. [14], [8], [39]). Adopting the now-classical multiplicative update strategy (see e.g. [9] for a rigorous treatment), the parameters can be updated iteratively using the formulas given in Table I for $\beta = 0$, where:

$$v_m(f,t) = |c_m(f,t)|^\alpha \tag{27}$$

and

$$\sigma_m(f,t) = \sum_j P_j^\alpha(f,t)\, k_m(f)^\top Q_j, \tag{28}$$

while $a \cdot b$ or $\frac{a}{b}$ as well as $a^{\cdot b}$ stand for element-wise operations. In the case of a NMF model, P_j^α is understood as in (17).

B. Cauchy $\alpha = 1$ case, heuristic approach

A common choice for audio modeling is to pick $\alpha = 1$, which boils down to assuming additive STFT magnitudes for additive signals [39]. As shown in [22], the α-harmonisable model provides the probabilistic interpretation for the procedure[5]. Luckily, just as the Gaussian case $\alpha = 2$ above, the

[5]Actually, STFT magnitudes are theoretically additive when α is just slightly greater than 1. However, the Cauchy model is a good approximation in this case. See Section III-D on this point.

case $\alpha = 1$ is amenable to analytical treatment. Indeed, the isotropic 1-stable distribution (16) is identical to the complex isotropic Cauchy distribution, for which the probability density function can be expressed in closed-form [35, ex. 2.5.6 p. 81]. Model (16) then becomes:

$$p\left(c_m\left(f,t\right) \mid \Theta\right) = \frac{\sigma_m\left(f,t\right)}{2\pi\left(v_m\left(f,t\right)^2 + \sigma_m\left(f,t\right)^2\right)^{3/2}}, \quad (29)$$

where v_m and σ_m have been defined in (27) and (28), respectively. To the best of our knowledge, no previous study made the connection between assuming additive magnitude STFT for additive signals and modeling them as locally stationary Cauchy processes as we do here. It is straightforward to show that the global cost function D_c in (25) for maximum likelihood fitting becomes:

$$D_c\left(\Theta\right) \stackrel{c}{=} \sum_{f,t,m} \frac{3}{2}\log\left(v_m\left(f,t\right)^2 + \sigma_m\left(f,t\right)^2\right) - \log\sigma_m\left(f,t\right), \quad (30)$$

where $\stackrel{c}{=}$ denotes equality up to an additive constant independent of Θ. Our objective is now to update one parameter Θ_i (such as an element of some W_j or Q_j), so as to decrease $D_c\left(\Theta\right)$. We can consider several approaches for this purpose. In this study, we mention two of them, both based on multiplicative updates.

A first straightforward but heuristic approach involves the derivative of the global cost function D_c in (30) with respect to the parameter Θ_i to update:

$$\frac{\partial D_c\left(\Theta\right)}{\partial\Theta_i} = \sum_{f,t,m}\left(\frac{3\sigma_m\left(f,t\right)}{v_m\left(f,t\right)^2 + \sigma_m\left(f,t\right)^2} - \frac{1}{\sigma_m\left(f,t\right)}\right)\frac{\partial\sigma_m\left(f,t\right)}{\partial\Theta_i} \quad (31)$$

Since this derivative can be expressed as the difference $G_+\left(\Theta\right) - G_-\left(\Theta\right)$ between two nonnegative terms:

$$\begin{aligned} G_+\left(\Theta_i\right) &= \sum_{f,t,m}\frac{3\sigma_m\left(f,t\right)}{v_m\left(f,t\right)^2 + \sigma_m\left(f,t\right)^2}\frac{\partial\sigma_m\left(f,t\right)}{\partial\Theta_i} \\ G_-\left(\Theta_i\right) &= \sum_{f,t,m}\sigma_m\left(f,t\right)^{-1}\frac{\partial\sigma_m\left(f,t\right)}{\partial\Theta_i}, \end{aligned}$$

we can adopt the now classical multiplicative update procedure pioneered in [20] and update Θ_i through:

$$\Theta_i \leftarrow \Theta_i \cdot \frac{G_-\left(\Theta_i\right)}{G_+\left(\Theta_i\right)},$$

guaranteeing that provided Θ_i has been initialized as nonnegative, it remains so throughout iterations. The whole procedure is summarized in Table II, where z_m is taken as a short-hand notation for:

$$z_m\left(f,t\right) \triangleq \frac{3\sigma_m\left(f,t\right)}{v_m\left(f,t\right)^2 + \sigma_m\left(f,t\right)^2}. \quad (32)$$

All updates are computed using the latest available versions of all parameters:

- $P_j^\alpha\left(f,t\right) \leftarrow P_j^\alpha\left(f,t\right) \cdot \frac{\sum_m k_m(f)^\top Q_j\left[{\sigma_m}^{\cdot-1}\right]_{ft}}{\sum_m k_m(f)^\top Q_j\left[z_m\right]_{ft}}$
- $W_j\left(f,r\right) \leftarrow W_j\left(f,r\right) \cdot \frac{\sum_m k_m(f)^\top Q_j\left(\sum_t\left[{\sigma_m}^{\cdot-1}\right]_{ft}H_j(t,r)\right)}{\sum_m k_m(f)^\top Q_j\left(\sum_t\left[z_m\right]_{ft}H_j(t,r)\right)}$
- $H_j \leftarrow H_j \cdot \frac{\sum_m\left(\sum_f k_m(f)^\top Q_j\left[{\sigma_m}^{\cdot-1}\right]_{ft}W_j(f,r)\right)}{\sum_m\left(\sum_f k_m(f)^\top Q_j\left[z_m\right]_{ft}W_j(f,r)\right)}$
- $Q_j \leftarrow Q_j \cdot \frac{\sum_{f,t,m} k_m(f)\left[{\sigma_m}^{\cdot-1}\cdot P_j^\alpha\right]_{ft}}{\sum_{f,t,m} k_m(f)\left[z_m\cdot P_j^\alpha\right]_{ft}}$

Table II
HEURISTIC MULTIPLICATIVE UPDATES FOR THE CAUCHY $\alpha = 1$ CASE.

C. Cauchy $\alpha = 1$ case, majoration-equalization approach

Even if the updates found in Table II are derived straightforwardly using classical non-negative methodology, there is no guarantee that they indeed lead to a decrease of the cost-function $D_c\left(\Theta\right)$ at each step. An alternative way to derive update rules for the parameters is to adopt the Majoration-Equalization (ME) approach presented in [9] to our problem. In essence, the strategy first requires identifying a majoration of the cost-function (30), which is of the form:

$$\forall\left(\Theta,\hat{\Theta}\right), D_c(\hat{\Theta}) \leq g\left(\hat{\Theta},\Theta\right)$$

with $\forall\Theta, D_c\left(\Theta\right) = g\left(\Theta,\Theta\right)$. Then, given current parameters Θ, we look for a value of $\widehat{\Theta}$ different from Θ, such that the right member of the majoration is still equal to $D_c(\Theta)$. This approach guarantees that the cost function will be non-increasing over the iterations, and it is known to provide a faster convergence rate than the "majorise-minimise" approach. Besides, remember that in the case of β-divergences, this strategy leads to the regular NMF multiplicative update rules [9]. Due to space constraints, we simply mention here the majoration we used[6]:

$$\begin{aligned} \forall\left(\Theta,\hat{\Theta}\right),\ D_c(\widehat{\Theta}) &\leq D_c(\Theta) \\ &+ \sum_{f,t,m}\frac{3}{2}\frac{\widehat{\sigma}_m^2\left(f,t\right) - \sigma_m^2\left(f,t\right)}{\sigma_m^2\left(f,t\right) + \left|c_m\left(f,t\right)\right|^2} + \frac{\sigma_m\left(f,t\right)}{\widehat{\sigma}_m\left(f,t\right)} - 1, \end{aligned} \quad (33)$$

as well as the corresponding updates for the model parameters in Table III. We highlight the fact that the Cauchy cost function (30) is guaranteed to be non-increasing over the iterations using these updates.

D. General case $\alpha \in]0\,2]$: lower order moment fitting

In the general case, a maximum likelihood approach is unpractical to achieve (24) due to the lack of an analytical expression for the $S\alpha S_c$ probability density function. However, the model (16) states that the marginal distributions of the entries $c_m\left(f,t\right)$ of the projection tensor are $S\alpha S_c$, with scale parameter $\sum_j P_j^\alpha\left(f,t\right)k_m\left(f\right)^\top Q_j$. For such variables with $\alpha < 2$, the α^{th} order moment $\mathbb{E}\left[\left|c_m\left(f,t\right)\right|^\alpha\right]$ is undefined:

$$\mathbb{E}\left[\left|c_m\left(f,t\right)\right|^\alpha\right] = \infty.$$

[6]In (33), σ denotes the model (28) computed using Θ, while $\hat{\sigma}$ uses $\hat{\Theta}$.

All updates of a matrix Θ_i (some P_j, W_j, H_j, or Q_j) are computed using the latest available versions of all parameters through:

$$\Theta_i \leftarrow \Theta_i \cdot \frac{b\left[\Theta_i\right]}{a\left[\Theta_i\right] + \sqrt{a\left[\Theta_i\right]^{\cdot 2} + 2b\left[\Theta_i\right] \cdot a\left[\Theta_i\right]}}$$

where $a\left[\Theta_i\right]$ and $b\left[\Theta_i\right]$ are of the same size as Θ_i and given by:

Θ_i	$a\left[\Theta_i\right]$
$P_j^{\alpha}\left(f,t\right)$	$\frac{3}{4}\sum_m \frac{\sigma_m(f,t)k_m(f)^{\top}Q_j}{\sigma_m^2(f,t)+\lvert c_m(f,t)\rvert^2}$
$W_j\left(f,r\right)$	$\frac{3}{4}\sum_{t,m} \frac{\sigma_m(f,t)H_j(t,r)k_m(f)^{\top}Q_j}{\sigma_m^2(f,t)+\lvert c_m(f,t)\rvert^2}$
$H_j\left(t,r\right)$	$\frac{3}{4}\sum_{f,m} \frac{\sigma_m(f,t)W_j(f,r)k_m(f)^{\top}Q_j}{\sigma_m^2(f,t)+\lvert c_m(f,t)\rvert^2}$
$Q_j\left(l\right)$	$\frac{3}{4}\sum_{f,t,m} \frac{\sigma_m(f,t)P_j^{\alpha}(f,t)[k_m(f)]_l}{\sigma_m^2(f,t)+\lvert c_m(f,t)\rvert^2}$

and:

Θ_i	$b\left[\Theta_i\right]$
$P_j^{\alpha}\left(f,t\right)$	$\sum_m \frac{k_m(f)^{\top}Q_j}{\sigma_m(f,t)}$
$W_j\left(f,r\right)$	$\sum_{t,m} \frac{H_j(t,r)k_m(f)^{\top}Q_j}{\sigma_m(f,t)}$
$H_j\left(t,r\right)$	$\sum_{f,m} \frac{W_j(f,r)k_m(f)^{\top}Q_j}{\sigma_m(f,t)}$
$Q_j\left(l\right)$	$\sum_{f,t,m} \frac{P_j^{\alpha}(f,t)[k_m(f)]_l}{\sigma_m(f,t)}$

Table III
MAJORATION-EQUALIZATION UPDATES FOR THE CAUCHY $\alpha = 1$ CASE.

However, the p-moments of $c_m\left(f,t\right)$ for $p < \alpha$ are defined and we have [35, p. 19]:

$$\lim_{p \uparrow \alpha} \left(\alpha - p\right) \mathbb{E}\left[\left|c_m\left(f,t\right)\right|^p\right] = \alpha \lambda_{\alpha} \sigma_m\left(f,t\right), \qquad (34)$$

where λ_{α} is a constant that only depends on α, and $\sigma_m\left(f,t\right)$ has been defined in (28). Hence, if we pick $p < \alpha$ that is sufficiently close to α, we may assume that:

$$\mathbb{E}\left[\left|c_m\left(f,t\right)\right|^p\right] \approx \lambda_{p\alpha} \sigma_m\left(f,t\right),$$

where $\lambda_{p\alpha}$ is now a constant that only depends on p and α. Consequently, the empirical p-spectrogram $v_m\left(f,t\right) \triangleq \left|c_m\left(f,t\right)\right|^p$ of the observations basically ought to match the scale parameters $\sigma_m\left(f,t\right)$ given by the model just like in variance modeling (26) for the Gaussian case, but up to a multiplicative constant. For any $\alpha \in \left]0\,2\right]$ and $p < \alpha$ close enough to α, this leads to:

$$\begin{aligned} v_m\left(f,t\right) &\triangleq \left|c_m\left(f,t\right)\right|^p \\ &\approx \lambda_{p\alpha} \sum_j P_j^{\alpha}\left(f,t\right) k_m\left(f\right)^{\top} Q_j. \end{aligned} \qquad (35)$$

A reasonable parameter estimation strategy we call *fractional lower order moment fitting* (FLOM) is then to estimate Θ so as to enforce the approximation (35), where the precise value for the $\lambda_{p\alpha}$ constant is of no importance, since it is independent of the model parameters Θ. To this purpose, we can proceed the same way as in (26), but possibly using another cost function than d_0 such as the more general β-divergence d_{β} as a proxy for the relevant $S\alpha S_c$ probability density function. The corresponding updates can be found in Table I.

As can be seen, this FLOM approach requires an estimate of α to be available so as to pick $p < \alpha$ close enough to α for (35) to hold. Previous studies [22], [16] suggest that a value $\alpha \approx 1.1$ is reasonable for audio, giving one interesting justification for the common choice $p = 1$. Choosing the KL divergence ($\beta = 1$) for magnitude spectrogram fitting, as routinely done in PLCA studies for audio (see e.g. [39] and references therein) may thus be interpreted as adopting this FLOM strategy and has long proved efficient. However, our derivations suggest that a maximum likelihood approach rather leads to the Cauchy updates presented above in Section III-B.

E. The case of punctual objects

Modelling an object j as being punctual and anechoic like in Section II-A amounts to assuming that only one element of the $L \times 1$ vector Q_j is nonzero. This constraint can be enforced very easily by taking $J = L$ and imposing $Q = \mathbf{I}_L$, the $L \times L$ identity matrix. For this particular case, there is hence no need to update the pan-delay distributions q_j and some simplifications over the general formulas found in Tables I and II can be used for implementation. Remarkably, when the NMF model (17) is *not* chosen, so that the P_j^{α} are left unconstrained, they are the only parameters to be estimated. In that case, the update for P_j^{α} presented in Table I ($\beta = 0$) and in Table III are guaranteed to identify the global optimal parameters for the model in the Gaussian ($\alpha = 2$) and Cauchy ($\alpha = 1$) cases, respectively.

If a background "ambient" object is to be considered, a natural approach is to adapt this punctual solution and to select the pan-delay directions that exhibit a sufficient energy $\sum_{ft} P_j^{\alpha}\left(f,t\right)$ as punctual objects, while gathering all others as the ambient object in a final grouping stage.

F. Choosing the panning and projection sets

In the above sections, we have assumed that both the location set $\mathcal{L} = \left(\phi_1,\tau_1\right),\ldots,\left(\phi_L,\tau_L\right)$ and the projection set $\mathcal{P} = \left(\phi_1,\tau_1\right),\ldots,\left(\phi_M,\tau_M\right)$ were known and fixed, permitting the computation of both the projection tensor c through the projection matrix $\mathbf{N}_f = \left\{h_1\left(f\right),\ldots,h_M\left(f\right)\right\}^{\top}$ and the elements of the dictionaries $k_m\left(f\right)$ as in (10). However, the choice of these parameters proves to be important for good performance of the proposed methods and is discussed now.

Concerning the location set $\mathcal{L}$ in the stereo $I = 2$ case, our experience suggests that a good strategy is to have $L > J$, i.e. more locations than objects, and to choose $\mathcal{L}$ such that the panning directions equally span the stereo space, while having the delays regularly spanning some $\left[-\tau_{max}\;\tau_{max}\right]$ interval.

The choice of the projection set is also important. Whereas taking $M = L$ and $\left(\phi_m,\tau_m\right) \leftarrow \left(\phi_l,\tau_l\right)$ would seem a reasonable choice, experience shows that separation is better when each projection direction, say $h_m\left(f\right)$, is orthogonal to one of the $h_l\left(f\right)$. This choice guarantees that the energy incoming from location $\left(\phi_l,\tau_l\right)$ will cancel out in $\left\langle h_m\left(f\right), x\left(f,t\right)\right\rangle$, permitting this particular projection to be useful for the fitting of the parameters of all panning-delay locations with $\left(\phi_l,\tau_l\right)$ filtered out. This happens in the stereo case for $\left(\phi_m,\tau_m\right) \leftarrow \left(\phi_l - \pi/2, -\tau_l\right)$. We note that having $M < L$ was sufficient for good performance, while computationally more efficient than larger M.

In any case, for the stereo signals considered in our evaluations, we had $\mathcal{L}$ regularly sampling $[0,\, \pi/2] \times [-\tau_{max},\, \tau_{max}]$ and $\mathcal{P}$ regularly sampling $[-\pi/2,\, 0] \times [-\tau_{max},\, \tau_{max}]$.

IV. Evaluation

In this section, we first describe the test set used to evaluate the proposed PROJET method, followed by the evaluation criteria. Finally, we describe the experiments undertaken. All the audio and test sets are available publicly, as well as MATLAB and Python implementations of the proposed algorithms[7].

A. Test sets

The test set was created from the MSD100 development set used in SiSEC 2015[8]. The development set consists of 50 full length songs created from mixtures of 4 objects, and all have a sample rate of 44.1kHz. The original recordings for these objects are available as part of the development set. To create the test set for this paper, 30 second excerpts were extracted, with the same start and end points relative to the start of the song, for all objects. These mono excerpts were then used to create stereo images using the pan-delay model described in (2). The resulting stereo objects images were then summed to create the mixture signals. In order to evaluate the effects of panning and delay separately, two distinct sets of mixtures were created. In the first set of mixtures, the objects were mixed with an equal angle between them, with no delay between the channels. For example, if the 4 sources were equally spaced over the 90 degrees, they would be positioned at θ_j equal to 0,30,60, and 90 degrees respectively, resulting in a 30 degree angle between the objects in terms of the pan-pot mixing model used. In order to test the robustness of the algorithm with respect to the angle between the objects, this angle between the objects was varied, from 5 degrees to 30 degrees in steps of 5 degrees for the case of oracle/informed separation. This results in a set of 300 test mixtures for the informed/oracle pan-only case, giving a total of 1200 separated sources to evaluate with respect to the chosen separation metrics. For the blind separation pan-only case, the angle between objects was varied from 10 degrees to 30 degrees in steps of 10, for example resulting in a total of 150 test mixtures, with a total of 600 separated sources to be evaluated. Also tested in the pan-only case was the effect of incorporating an NMF model in conjunction with the spatial model, as described in (17). In the second set of mixtures, the angle between the sources was fixed at a value of 20 degrees, and the delay between left and right channels for each of the 4 sources was varied from $-\tau_B$ to $2\tau_B$ in steps of τ_B samples, where τ_B is the baseline delay for that set of mixtures. This was repeated for a number of different baseline delays, with values of $\tau_B = [1, 5, 10, 20, 50]$ being tested. This resulted in 250 mixtures and 1000 separated sources each being tested for both the blind and oracle cases utilising delay.

[7] www.loria.fr/~aliutkus/projet/

[8] https://sisec.inria.fr

B. Separation evaluation criteria

The metrics chosen to evaluate the effectiveness of our proposed PROJET method are Signal to Distortion Ratio (SDR), Signal to Artefacts Ratio (SAR), Signal to Interference Ratio (SIR) and Image to Spatial Distortion Ratio (ISR), as implemented and defined in the BSS Eval Toolbox [43]. We make use of version 3 of the toolbox. While SDR provides a measure of the overall sound quality of the separated object, SAR measures the presence of artefacts and SIR measures the amount of interference or *bleed* from other objects present in the mixture. Finally, ISR measures distortions in the spatial position of the recovered object images.

C. Oracle/Informed Source Separation

In the case of oracle —or *informed*— demixing, the PROJET algorithm was provided with knowledge of the actual angles and delays to which the 4 objects were positioned. The oracle case was tested for two reasons, firstly to determine an upper limit for performance of the technique, and secondly to demonstrate its use for the informed source separation case, where the amount of side information to be transmitted with this technique is remarkably small. The panning/delay location set $\overline{\mathcal{L}}$ then consisted of $L = J = 4$ positions corresponding to the positions of the actual objects. In other words $\overline{\mathcal{L}} = (\phi_1, \tau_1), \ldots, (\phi_J, \tau_J)$, where ϕ_j is the pan position of the jth object and τ_j is the delay position. These angles were then used to define the projection matrices $\mathbf{N}_f$ such that any vector in $\mathbf{N}_f$ is orthogonal to the corresponding vector in the location set $\overline{\mathcal{L}}$, with the number of projections M set equal to the number J of objects. In this punctual case, as noted in Section III-E, the panning coefficients Q_j do not have to be updated, and we estimate the P_j^α via two methods. In the first one, the P_j^α are estimated directly, while in the second case they are constrained via an NMF model as per (17). In any case, the parameters were all initialised randomly. We computed the STFT with a window of size 4096 samples, and a hop size of 1024 samples, i.e. with 75% overlap between frames, using a Hann window.

We first describe the oracle tests done in the angle-only case. As noted previously, here the angle between the objects was varied from 5 degrees to 30 degrees in steps of 5 degrees with the delay fixed at 0 samples for all sources. The oracle angle-only PROJET algorithm was then tested using 3 different sets of update equations. The first of these are the heuristic multiplicative updates for the Cauchy distribution (denoted CH) given in Table II and the majorisation-equalisation (denoted CME) updates given in Table III, both corresponding to $\alpha = 1$. Still with $\alpha = 1$, we then used the generalised Kullback-Leibler divergence (denoted KL) which corresponds to the $\beta = 1$ case (see Table I for update equations) and which also represents an example of the FLOM fitting strategy described in Section III-D. The algorithms all ran for 200 iterations. Also tested was the case where P_j^α was constrained according to the NMF model described in (17). Here, P_j^α was modelled with a fixed number of components R. Two values of R were considered in our experiments, $R = 5$, and $R = 20$. As the results obtained from both these values were broadly similar,

only those for $R = 5$ are presented here. The average results are shown as box plots for each of the angles tested with a circle indicating the mean performance over the separated signals, a line indicating the median of the test results, and the extent of the boxes indicating the range between the 25th and 75th percentiles of the results.

As a baseline, the algorithms are measured against DUET [45], a well known algorithm for separating stereo mixtures, that creates binary masks for the objects based on interchannel level and phase differences. In this case, as the test mixtures are linear instantaneous, the phase cue is not available, and so the level difference between the mixture channels is used as the cue to DUET. Like PROJET, DUET was provided with the correct objects positions, and the time-frequency bins that fell within a given angle on either side of the actual objects positions were associated with the object. Here, the transition point between one object and another is the angle halfways between them.

Figure 1 shows the average results for the BSS_Eval metrics obtained for Cauchy Heuristic, Cauchy Heuristic with NMF, Cauchy ME, Cauchy ME with NMF, KL, KL with NMF and DUET. It can be seen that in the oracle/informed separation case, all 3 sets of update equations perform well with respect to SDR, with a maximum difference of less than 0.3 dB between the best performing algorithm (KL-NMF) and the worst performing algorithm (CME). It is also clear that the proposed approaches all considerably outperform the DUET algorithm. With respect to the incorporation of the NMF model, the overall performance has slightly decreased in all cases except for KL-NMF, where there has been a small improvement in performance. Also, the range between the 25th and 75th percentiles of the results has decreased slightly, suggesting than the NMF constraint helps achieving better physically meaningful estimation. It should be noted that both the Cauchy ME and Cauchy Heuristic update equations have a very similar performance, both with and without the incorporation of the NMF model.

With respect to SIR, the various PROJET update equations again have very similar results and offer good levels of rejection, with or without the NMF model, demonstrating that the spatial projection model is what is mainly driving the separation, regardless of whether the source spectrograms are constrained or not. Concerning DUET, it gives greater rejection in terms of SIR than the methods proposed herein. Regarding SAR, here KL performs best, closely followed by both Cauchy methods. However, all these methods considerably outperform DUET in terms of SAR. This is partly due to the binary nature of the masks used with DUET, which minimise interferences (good SIR scores), at the cost of a much perceptually degraded audio quality and an increase in the level of audible artefacts. The presence of artefacts is observed to remain basically independent of the angle between the objects for the proposed method. It can be seen that there is again a small decrease in SAR induced by using NMF, except for when KL is used as a cost function. Regarding ISR, the performance of both Cauchy methods and KL is again very similar, and exhibit a performance that is basically independent of angle. DUET is seen to perform best with regards to spatialisation of the recovered objects. Finally, the results obtained for ISR show an overall decrease in spatialisation performance when using the NMF model, regardless of cost function, which shows that the use of the NMF model degrades the spatial performance of the separation method. Taken together, these results suggest that there is little need to employ the NMF model, unless one is attempting to deal with the case where multiple sources originate from the same direction, in which case the NMF model —or any spectrogram model— can be used to help separate these sources by introducing spectral prior information. As such, the NMF model was not tested in subsequent tests. Further, the results for both Cauchy updates are so similar that for subsequent tests only the Cauchy Heuristic results are shown.

Figure 2 then shows the results obtained for keeping a set angle of 20 degrees between the sources and varying the baseline delay as described previously. These delays were implemented in the time domain to ensure a more realistic mixing scenario. Again, DUET is used as a baseline to measure performance against, and here the phase as well as amplitude cues are used in estimating the binary masks for DUET. It can be seen that the separation performance, while still good overall, degrades with increasing baseline delay. Further, it can be clearly seen that DUET fails to perform separation properly. This is because of inherent limitations on the size of delay for which the parameters of the DUET algorithm can be estimated accurately as discussed in [45]. It can also be seen that both Cauchy and KL updates again offer similar performance for all metrics, and that SAR is effectively constant regardless of the delay size.

These results clearly show that all the proposed methods give good informed separation results, especially given that the only information presented is the source position. This suggests that PROJET has considerable utility for informed source separation, since the amount of side information that has to be transmitted is negligible. Further, as the informed punctual case does not require updates on the panning distribution Q, this means that this setup is suitable to run in an online manner, with source estimates obtained on a frame by frame basis and resynthesised via overlap-add, potentially permitting on-the-fly stereo to surround upmixing, particularly in the angle-only case. The performance of the oracle algorithm also suggests that a variant based on a peak-picking strategy or user assistance to determine object positions, will also perform well provided the peaks are identified accurately. Again, there is the potential to have an on-line version of the algorithm as the estimates can be obtained on a frame by frame basis.

D. Blind Source Separation

Having dealt with oracle/informed separation using PROJET, we now focus on the completely blind case, where no information about the object positions in the stereo space is provided. Results are presented for Cauchy Heuristic and KL updates. In the angle-only case, the source position set $\overline{\mathcal{L}} = (\phi_1, \tau_1), \ldots, (\phi_J, \tau_J)$, in this case consisted of $L = 30$ equally spaced panning positions spanning the range $[0, \pi/2]$, and with all $\tau_j = 0$. The projection matrix for the blind

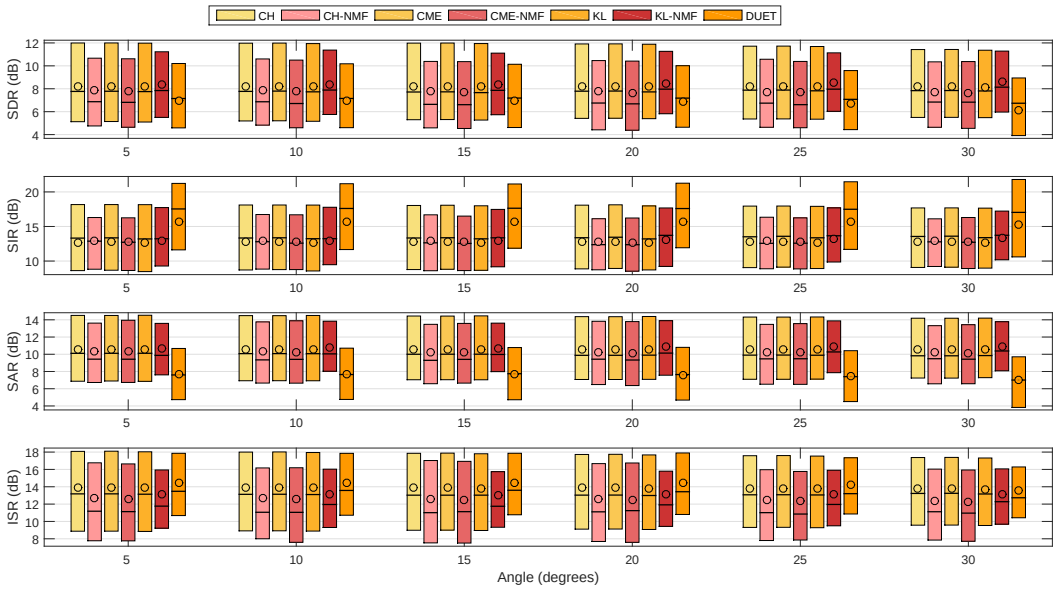

Figure 1. Separation Metrics vs Angle - Oracle/Informed Source Separation for Cauchy Heuristic updates (CH), Cauchy Heuristic updates with NMF (CH-NMF), Cauchy ME (CME), CauchyME with NMF (CME-NMF) generalised Kullback-Leibler divergence (KL),generalised Kullback-Leibler divergence (KL-NMF) and DUET .

separation algorithm $\mathbf{N}_f = \{h_1(f), \ldots, h_M(f)\}^\top$ contained $M = 10$ projections, where the projection angles equally span the range $[-\pi/2, 0]$ and all projection delays were set to 0. The number of sources to be separated was again $J = 4$, and the algorithms again ran for 200 iterations. In order to benchmark the spatial projection separation algorithms against the state of the art, we also tested the separation performance of multichannel NMF (MNMF) [32], using a publicly available implementation of the algorithm[9], on the same mixtures. Tests were also conducted using DUET, but performance was extremely poor, mainly due to the difficulty of identifying the peaks under wildly different conditions, and so these results are not included. Note that in the blind case, the parameters Q are not constrained to be punctual, and are also estimated using the updates described in Section III.

Figure 3 shows the results obtained for the blind spatial projection algorithm as a function of the angle between sources for Cauchy Heuristic, KL, and MNMF. As the oracle algorithm showed that performance was very stable with respect to changing the angle, these tests were only ran for 3 sets of angles between the sources, with the angle varying from 10 to 30 degrees in steps of 10. Further, the MNMF algorithm used in this instance was that designed for instaneous mixtures using EM updates, as this set of tests does not contain any delays between the channels. It can be seen that in the blind angle-only case, the best performing algorithm is KL-PROJET, regardless of the angle, and that this performance is robust with respect to the angle between the sources, with a drop of less than 1 dB in performance between 30 degrees and 10 degrees. It also shows that the blind version of KL-PROJET gives good overall performance, coming within less than 1.5 dB of the performance of the corresponding oracle algorithm. Further,

[9]http://www.irisa.fr/metiss/ozerov/Software/multi_nmf_toolbox.zip

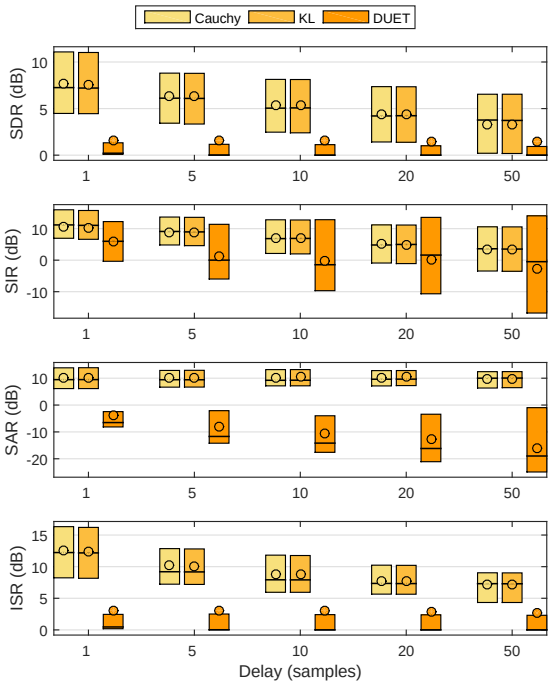

Figure 2. Separation Metrics vs. Delay - Oracle/Informed Source Separation for Cauchy Heuristic updates (Cauchy), generalised Kullback-Leibler divergence (KL), and DUET.

MNMF is the worst performing algorithm, again regardless of angle. This demonstrates the advantage of PROJET for the demixing of multichannel audio mixtures.

KL is the also the best performing of the algorithms with respect to SIR, and again shows robustness with respect to

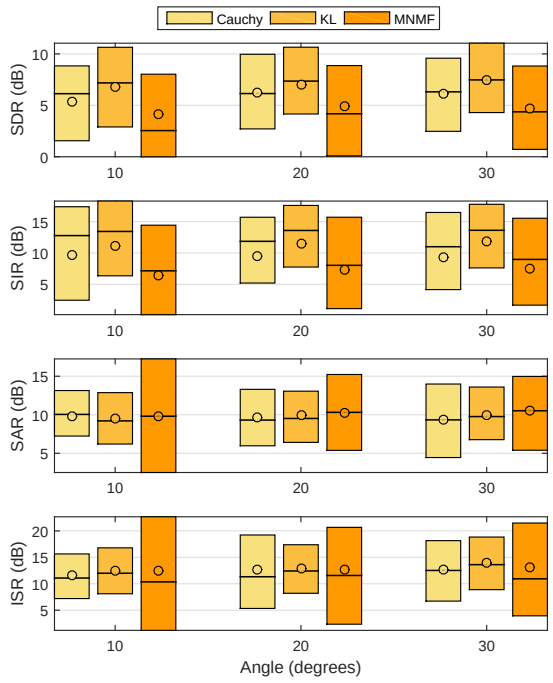

Figure 3. Separation Metrics vs Angle - Blind Source Separation for Cauchy Heuristic updates (Cauchy), generalised Kullback-Leibler divergence (KL), and Multichannel NMF (MNMF).

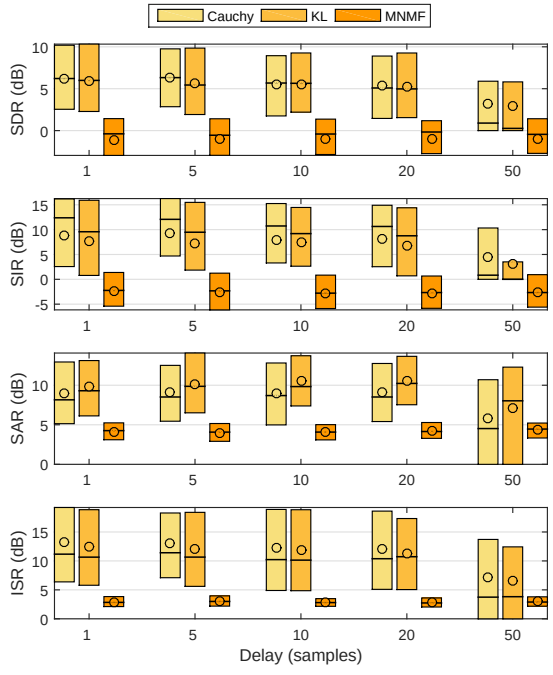

Figure 4. Separation Metrics vs Delay - Blind Source Separation for Cauchy Heuristic updates (Cauchy), generalised Kullback-Leibler divergence (KL), and Multichannel NMF (MNMF).

the angle between objects. Further, it again comes within less than 1.5 dB of the oracle KL spatial projection algorithm, further demonstrating the overall robustness of the KL-based algorithm. MNMF is again the worst performing of all the algorithms tested showing a high amount of interference due to the presence of other sources. With respect to SAR, it is MNMF that performs best, followed by KL. However, KL performs here within 0.5 dB of the MNMF algorithm, at a performance within less than 1.2 dB of the oracle KL performance, and with a drop in performance of less than 0.5 dB between 30 degrees and 10 degrees. Furthermore, the KL-based algorithm shows much less spread in the SAR scores obtained, suggesting an increased robustness over MNMF. For ISR, KL is the best performing, again achieving within 1.5 dB of the oracle KL performance, followed by MNMF.

It can be observed that in the angle-only case, the blind results in general exhibit greater variance than the oracle results regardless of the update equations used, which was expected since more parameters are to be estimated. It is interesting to note that the KL update shows less deviation with regards to ISR and SAR in both the oracle and blind separation cases. This is particularly noteworthy in the blind case, where it is near the top in terms of performance for these two metrics, suggesting that it is more consistent than the other methods in terms of artefact reduction and spatial distortions in the angle only case. It can be observed that MNMF has the greatest variance for ISR, SAR and SDR, showing that MNMF works quite well in some cases and not in others, suggesting that initialisation may be a problem for MNMF.

Figure 4 then shows the results obtained when using an angle of 20 degrees between sources, and varying the baseline delay as described previously. In this case, 20 equally spaced pans were used for the source positions, and the delays were taken from $-2\tau_B$ to $2\tau_B$ by steps of $\tau_B/4$, yielding a set of 17 delays. For the projection set, 10 equally spaced pan projection positions were used, and the projection delays were taken with the same range but with steps of $\tau_B/2$, yielding a set of 9 projection delays. MNMF was again used as a baseline to compare the performance of PROJET, this time using the convolutive EM version of MNMF, which was designed to deal with delay between channels. Again, the proposed PROJET algorithms give improved performance over the baseline MNMF, which cannot successfully deal with the large delay sizes that PROJET can. In this case, the Cauchy update outperforms KL for all metrics except for SAR, suggesting that it is better able to deal with delays between sources than KL, which performs better when no delay is present. As with the oracle case, performance degrades with increased delay sizes. A possible reason for this is that PROJET as currently implemented makes use of the circularity assumption when projecting delays in the frequency domain. This is increasingly violated at larger delays and affects the cancellation of individual sources during projection. Nonetheless, the technique clearly works over a wide range of delays.

Taken together, these results show that PROJET clearly outperform the baseline algorithms and can handle delays which these cannot. It is worth noting that in the proposed algorithms, separation is achieved mainly on a spatial model,

possibly with no constraints on the object spectrograms.

We also informally tested PROJET on a number of professionally produced recordings taken from commercial CDs. The results are available for listening on the webpage accompanying this paper. It can be noted that, in some cases, multiple objects are separated together because they were panned to the same direction in the original mixture and so were impossible to separate using the projection model presented in this paper.

V. Conclusion

We have presented a novel spatial projection-based method called PROJET for the separation of multichannel audio. Here, we used a mixing model that assumes that the spatial image of each object is a weighted sum of independent contributions originating from all pan-delay directions. We showed how to estimate the parameters efficiently and to proceed to separation by first projecting the multichannel mixture signal onto a range of spatial-delay directions, yielding an augmented set of observations in which some objects are enhanced or attenuated.

We have derived a number of inference methods for these parameters of the spatial projection method, based on the assumptions that the individual TF bins of the Short Term Fourier Transforms are independent and distributed with respect to a complex isotropic α-stable distribution, which generalises the classical Gaussian model to the case of impulsive signals, that are common in audio. We then derived updates for the special cases of $\alpha = 1$ (multivariate Cauchy distribution) and $\alpha = 2$ (complex isotropic Gaussian distribution). In the general case of other choices of α, where no analytical expressions for the α-stable distribution are available, we discussed heuristics using fractional lower order moment fitting.

In an evaluation section, we then demonstrated the effectiveness of PROJET for the demixing of music under 2 cases, firstly an oracle/informed sound separation scenario, where the spatial positions and associated delays of the objects are known a-priori, and secondly, a blind separation scenario where no knowledge is available regarding the direction of the objects. In all these tests, the PROJET method was observed to outperform the other competing methods, namely DUET and multichannel-NMF. This shows that PROJET has potential for use in informed source separation, as the amount of side information to be transmitted with this technique is minimal, or as a user assisted separation algorithm, where the user picks the panning directions of the objects, or where they are provided via a peak picking technique.

In the blind separation scenario, we showed that the proposed PROJET method permits remarkable separation performance, mostly using spatial and delay diversity of the objects, with possibly no constraints at all on their spectro-temporal characteristics. We also informally demonstrated the effectiveness of PROJET in demixing professionally produced music recordings taken from commercial CDs.

There are a number of possible directions for future work based on the projection approach introduced in this paper. The first such direction is the inclusion of further constraints placed on the objects spectro-temporal characteristics. For example, this could include imposing sparsity or smoothness-promoting priors, as well as combining PROJET with popular deep-learning based spectrogram models. Another direction is to extend the mixing model to overcome the circularity assumption currently limiting the performance of the algorithm. Our preliminary tests on this topic proved promising. We also intend investigating the extension of the proposed technique to deal with reverberant environments. It is also intended to investigate optimising the projections. Finally, an online version of the informed setup could be investigated to allow for real-time supervised demixing of music.

References

[1] J. Carabias-Orti, M. Cobos, P. Vera-Candeas, and F. Rodríguez-Serrano. Nonnegative signal factorization with learnt instrument models for sound source separation in close-microphone recordings. *EURASIP Journal on Advances in Signal Processing*, 2013(1):1–16, 2013.

[2] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Wiley Publishing, September 2009.

[3] O. Dikmen and A.T. Cemgil. Unsupervised single-channel source separation using Bayesian NMF. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 93–96, NY, USA, October 2009.

[4] N.Q.K Duong, E. Vincent, and R. Gribonval. Under-determined convolutive blind source separation using spatial covariance models. In *Acoustics, Speech and Signal Processing, IEEE Conference on (ICASSP'10)*, pages 9–12, Dallas, United States, March 2010.

[5] N.Q.K. Duong, E. Vincent, and R. Gribonval. Under-determined reverberant audio source separation using a full-rank spatial covariance model. *Audio, Speech, and Language Processing, IEEE Transactions on*, 18(7):1830 –1840, Sept. 2010.

[6] J.L. Durrieu and J.P. Thiran. Musical audio source separation based on user-selected F0 track. In *International Conference on Latent Variable Analysis and Signal Separation*, Tel-Aviv, Israel, March 12-15 2012.

[7] S. Ewert and M. Müller. Using score-informed constraints for NMF-based source separation. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, March 2012.

[8] C. Févotte, N. Bertin, and J.-L. Durrieu. Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis. *Neural Computation*, 21(3):793–830, March 2009.

[9] C. Févotte and J. Idier. Algorithms for nonnegative matrix factorization with the beta-divergence. *Neural Computation*, 23(9):2421–2456, Sep. 2011.

[10] D. FitzGerald. User assisted source separation using non-negative matrix factorisation. In *IET Irish Signals and Systems Conference*, Dublin, 2011.

[11] D. FitzGerald, M. Cranich, and E. Coyle. Extended non-negative tensor factorisation models for musical sound source separation. *Computational Intelligence and Neuroscience, vol. 2008, Article ID 872425*, 2008.

[12] D. FitzGerald, M. Cranitch, and E. Coyle. Non-negative tensor factorisation for sound source separation. In *Proceedings of the Irish Signals and Systems Conference*, 2005.

[13] D. FitzGerald, M. Cranitch, and E. Coyle. Shifted 2d non-negative tensor factorisation. In *IET Irish Signals and Systems Conference*, 2006.

[14] D. FitzGerald, M. Cranitch, and E. Coyle. On the use of the beta divergence for musical source separation. In *Irish Signals and Systems Conference (ISSC)*, Galway, Ireland, June 2008.

[15] B. Fuentes, R. Badeau, and G. Richard. Blind Harmonic Adaptive Decomposition Applied to Supervised Source Separation. In *Proceedings European Signal Processing Conference (EUSIPCO 2012)*, Bucharest, Romania, 2012.

[16] P. Georgiou, P. Tsakalides, and C. Kyriakakis. Alpha-stable modeling of noise and robust time-delay estimation in the presence of impulsive noise. *IEEE Transactions on Multimedia*, 1(3):291–301, September 1999.

[17] R. Hennequin, B. David, and R. Badeau. Score informed audio source separation using a parametric model of non-negative spectrogram. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Prague, Czech Republic, May 2011.

[18] R. Jaiswal, D. FitzGerald, D. Barry, Coyle E., and S. Rickard. Clustering NMF basis functions using shifted NMF for monaural sound source separation. In *International Conference on Acoustics, Speech, and Signal Processing*, Prague, 2011.

[19] B. King, C. Févotte, and P. Smaragdis. Optimal cost function and magnitude power for NMF-based speech separation and music interpolation. In *Machine Learning for Signal Processing (MLSP), 2012 IEEE International Workshop on*, pages 1–6. IEEE, 2012.

[20] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems (NIPS)*, volume 13, pages 556–562. The MIT Press, April 2001.

[21] S.N. Lee, S.H. Park, and K. Sung. Beamspace-domain multichannel nonnegative matrix factorization for audio source separation. *Signal Processing Letters, IEEE*, 19(1):43–46, Jan 2012.

[22] A. Liutkus and R. Badeau. Generalized Wiener filtering with fractional power spectrograms. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.

[23] A. Liutkus, R. Badeau, and G. Richard. Gaussian processes for underdetermined source separation. *IEEE Transactions on Signal Processing*, 59(7):3155 –3167, July 2011.

[24] A. Liutkus, R. Badeau, and G. Richard. Low bitrate informed source separation of realistic mixtures. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013.

[25] A. Liutkus, J-L. Durrieu, L. Daudet, and G. Richard. An overview of informed audio source separation. In *Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, pages 1–4, Paris, France, July 2013.

[26] A. Liutkus, S. Gorlow, N. Sturmel, S. Zhang, L. Girin, R. Badeau, L. Daudet, S. Marchand, and G. Richard. Informed source separation : a comparative study. In *Proceedings European Signal Processing Conference (EUSIPCO 2012)*, August 2012.

[27] Y. Mitsufuji and A. Roebel. On the use of a spatial cue as prior information for stereo sound source separation based on spatially weighted non-negative tensor factorization. *EURASIP Journal on Advances in Signal Processing*, 2014(1), 2014.

[28] M. Nakano, J. Le Roux, H. Kameoka, T. Nakamura, N. Ono, and S. Sagayama. Bayesian nonparametric spectrogram modeling based on infinite factorial infinite hidden markov model. In *Proceedings of the WASPAA 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 2011.

[29] C. Nikias and M. Shao. *Signal processing with alpha-stable distributions and applications*. Wiley-Interscience, 1995.

[30] J. Nikunen and T. Virtanen. Direction of arrival based spatial covariance model for blind sound source separation. *IEEE/ACM Transactions on Audio, Speech & Language Processing*, 22(3):727–739, 2014.

[31] J. Nolan, A. Panorska, and J. McCulloch. Estimation of stable spectral measures. Technical report, American University, 1997.

[32] A. Ozerov and C. Févotte. Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation. *IEEE Transactions on Audio, Speech and Language Processing*, 18(3):550–563, March 2010.

[33] A. Ozerov, E. Vincent, and F. Bimbot. A general flexible framework for the handling of prior information in audio source separation. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(4):1118–1133, May 2012.

[34] Z. Rafii, A. Liutkus, and B. Pardo. A simple user interface for recovering patterns repeating in time and frequency in mixtures of sounds. In *Proc. IEEE Intl. Conf. Acoust. Speech Signal Processing (ICASSP)*, Brisbane, Australia, April 2015.

[35] G. Samoradnitsky and M. Taqqu. *Stable non-Gaussian random processes: stochastic models with infinite variance*, volume 1. CRC Press, 1994.

[36] H. Sawada, H. Kameoka, S. Araki, and N. Ueda. Multichannel extensions of non-negative matrix factorization with complex-valued data. *Audio, Speech, and Language Processing, IEEE Transactions on*, 21(5):971–982, May 2013.

[37] U. Simsekli and A.T. Cemgil. Score guided musical source separation using generalized coupled tensor factorization. In *Proceedings European Signal Processing Conference (EUSIPCO 2012)*, August 2012.

[38] P. Smaragdis. Discovering auditory objects through non-negativity constraints. In *Proc. of ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing*, 2004.

[39] P. Smaragdis, C. Févotte, G.J. Mysore, N. Mohammadiha, and M. Hoffman. Static and dynamic source separation using nonnegative factorizations: A unified view. *IEEE Signal Processing Magazine*, 31(3):66–75, May 2014.

[40] M. Spiertz and V. Gnann. Beta divergence for clustering in monaural blind source separation. In *128th AES Convention*, London, UK, May 2010.

[41] N. Stein. Nonnegative tensor factorization for directional blind audio source separation. *arXiv preprint arXiv:1411.5010*, 2014.

[42] G.A. Tsihrintzis, P. Tsakalides, and C.L. Nikias. Spectral methods for stationary harmonizable alpha-stable processes. In *European signal processing conference (EUSIPCO)*, pages 1833–1836, Rhodes, Greece, September 1998.

[43] E. Vincent, R. Gribonval, and C. Févotte. Performance measurement in blind audio source separation. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(4):1462 –1469, July 2006.

[44] T. Virtanen. Sound source separation using sparse coding with temporal continuity objective. In *Proc. of International Computer Music Conference*, Singapore, Oct 2003.

[45] O. Yilmaz and S. Rickard. Blind separation of speech mixtures via time-frequency masking. *IEEE Transactions on Signal Processing*, 52(7):1830–1847, July 2004.

[46] S. Zhang, L. Girin, and A. Liutkus. Informed source separation from compressed mixtures using spatial Wiener filter and quantization noise estimation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013.