Journal of Machine Learning Research 12 (2011) 1111-1148 Submitted 11/08; Revised 8/10; Published 3/11

# Anechoic Blind Source Separation Using Wigner Marginals

**Lars Omlor** LARS.OMLOR@MEDIZIN.UNI-TUEBINGEN.DE
**Martin A. Giese**[*] MARTIN.GIESE@UNI-TUEBINGEN.DE
*Section for Computational Sensomotorics, Department of Cognitive Neurology*
*Hertie Institute for Clinical Brain Research & Center for Integrative Neuroscience*
*University Clinic Tübingen*
*Frondsbergstrasse 23*
*72070 Tübingen, Germany*

**Editor:** Daniel Lee

## Abstract

Blind source separation problems emerge in many applications, where signals can be modeled as superpositions of multiple sources. Many popular applications of blind source separation are based on linear instantaneous mixture models. If specific invariance properties are known about the sources, for example, translation or rotation invariance, the simple linear model can be extended by inclusion of the corresponding transformations. When the sources are invariant against translations (spatial displacements or time shifts) the resulting model is called an anechoic mixing model. We present a new algorithmic framework for the solution of anechoic problems in arbitrary dimensions. This framework is derived from stochastic time-frequency analysis in general, and the marginal properties of the Wigner-Ville spectrum in particular. The method reduces the general anechoic problem to a set of anechoic problems with non-negativity constraints and a phase retrieval problem. The first type of subproblem can be solved by existing algorithms, for example by an appropriate modification of non-negative matrix factorization (NMF). The second subproblem is solved by established phase retrieval methods. We discuss and compare implementations of this new algorithmic framework for several example problems with synthetic and real-world data, including music streams, natural 2D images, human motion trajectories and two-dimensional shapes.

**Keywords:** blind source separation, anechoic mixtures, time-frequency transformations, linear canonical transform, Wigner-Ville spectrum

## 1. Introduction

Blind source separation is an important approach for the modeling of data by unsupervised learning (Choi et al., 2005; Cichocki and Amari, 2002; Ogrady et al., 2005; Comon and Jutten, 2010). The most elementary class of such methods is based on linear mixture models that combine source signals or mixture components as weighted linear sum. Such linear blind source separation methods have a wide spectrum of applications. Examples include speech processing (Anthony and Sejnowski, 1995; Torkkola, 1996a; Smaragdis et al., 2007; Ogrady et al., 2005), spectral analysis (Nuzillard and Bijaoui, 2000; Chen and Wang, 2001), and the interpretation of biomedical and geophysical data (Hu and Collins, 2004; Aires et al., 2000). Popular approaches exploit typically two classes of generative models: instantaneous and convolutive mixtures (Choi et al., 2005; Ogrady

∗. Martin A. Giese is the corresponding author.

et al., 2005; Pedersen et al., 2007). The *instantaneous mixture model* is defined by the equations:

$$x_i(t) = \sum_{j=1}^{n} \alpha_{ij} s_j(t) \quad i = 1, \cdots, m.$$

The time-dependent signals $x_i(t)$ are approximated by the linear superposition of a number of hidden source signals $s_j(t)$. This superposition is computed separately, time-point by time-point. Contrasting with this model, *convolutive mixtures* assume that the source signals are filtered assuming the filter kernels $\alpha_{ij}(t)$ prior to the superposition, resulting in the mixing model:

$$x_i(t) = \sum_{j=1}^{n} \left( \int_{-\infty}^{\infty} \alpha_{ij}(\tau) s_j(t-\tau) d\tau \right) \quad i = 1, \cdots, m. \tag{1}$$

It is obvious that the instantaneous mixture model is a special case of the convolutive model, where the filter kernels are constrained to be proportional to delta functions $\alpha_{ij}(t) = \alpha_{ij}\delta(t)$. In between these two model classes are *anechoic mixture models* which linearly combine time-shifted and scaled versions of the sources, without permitting multiple occurrences of the same source in the mixture. These models are equivalent to convolutive models for which the filter kernels are constrained to the form $\alpha_{ij}(t) = \alpha_{ij}\delta(t - \tau_{ij})$, resulting in the equation:

$$x_i(t) = \sum_{j=1}^{n} \alpha_{ij} s_j(t - \tau_{ij}) \quad i = 1, \cdots, m. \tag{2}$$

Compared to full covolutive models, anechoic models constrain substantially the space of admissible filter functions. This reduces dramatically the amount of data that is necessary for the estimation of the model parameters. In addition, this restriction of the parameter space often results in model parameters that are easier to interpret, which is critical for many applications that use mixture models for data analysis.

Apart from the problem of robust parameter estimation with limited amounts of data, all blind source separation methods suffer from intrinsic ambiguities. For all discussed models the ordering of the recovered sources is arbitrary. For fully convolutive models the sources can be recovered only up to an unknown filter function (*filter ambiguity*). The distortion of the source shape by this arbitrary filter hampers interpretability of the source signals. This makes it necessary to further constrain the estimation of the sources by introduction of additional auxiliary assumptions, such as minimal distortion (see Matsuoka, 2002 for details). In contrast, for anechoic mixtures (see Equation (2)) the filter ambiguity is limited to an unknown scaling and arbitrary additive shift that can be applied to all time delays $\tau_{ij}$, while the form of the source functions is uniquely defined, except for absolute position. This implies that anechoic mixtures can be advantageous for the modeling of data that are consistent with the corresponding generative model. This advantage is particularly strong if only small amounts of data are available or if the models are employed to interpret the statistical structure of the data.

The presence of shifts (translations) is a common problem in many scientific or technical applications (e.g., spectral displacements due to doppler-shifts in astronomy, asynchronous signal transmission in electrical engineering, or spatial displacements of features in images). The anechoic model provides thus an attractive alternative for the common instantaneous model, which can model shifts only by an introduction of additional sources, typically degrading the stability of the estimates

and the interpretability of the model parameters. Also the assumption of the single occurrence of the sources in the individual components of the mixture is reasonable for many applications. Examples from biology include human motion analysis (Barliya et al., 2009; d'Avella et al., 2008), where the same control signal might influence several muscles or joints with different delays, or in functional magnetic resonance imaging, where time shift occur naturally due to hemodynamic delays (Mørup et al., 2008).

The major part of previous work on anechoic mixtures has considered under-determined (over-complete) source separation problems ($m \leq n$), where the sources outnumber the dimensions of the available data. This is typically the case for acoustic data, for example in the case of the 'cocktail party problem', where the signals of many speakers have to be recovered from a small number of microphones. Such under-determined problems require additional assumptions about the sources (for example sparseness Georgiev et al., 2005; Bofill, 2003; Yilmaz and Rickard, 2004), in order to obtain unique and stable solutions. Since most of existing algorithms for the under-determined case rely on such additional constraints for the estimation of the sources, they cannot be easily generalized for the undercomplete case. This case, where the number of sources is smaller than the dimensionality of he data, is typical for data reduction problems. The application of algorithms developed for the over-determined problem may lead to erroneous results for this case. For example strong sparseness assumptions (like $W$-disjoint orthogonality Yilmaz and Rickard, 2004) lead to an overestimation of the number of the sources (in the overdetermined case), compared to methods that only assume statistical independence of the sources.
In this paper we present a present a new algorithmic framework for the solution of arbitrary anechoic mixture problems, which is independent of the number of sources and the dimensionality of the data. Contrasting with most previous approaches addressing the model (2), our method is suitable for dimension reduction since it is applicable for the solution of over-determined problems ($m \geq n$). The key idea of the novel framework is to transform the original mixture problem into the time-frequency domain, exploiting the Wigner-Ville transformation (WVT). The resulting transformed problem is completely equivalent to the original problem, but more suitable for an efficient algorithmic solution. Exploiting the fact that for the WVT the knowledge of a limited number of marginals allows the complete and unambiguous reconstruction of the original signal, we devise an algorithm that replaces then original problem by a set lower-dimensional anechoic demixing problems with positivity constraints and a phase retrieval problem. The positive demixing problems are solved by approximative methods, such as nonnegative matrix factorization (NMF) or positive ICA. The projection onto lower-dimensional problems leads to an efficient solution even of higher-dimensional problems with multi-dimensional translations. The obtained solution in time-frequency space is then transformed back into signal space, where in a second step the full solution of the original problem is determined by solution of a phase retrieval problem.

Our method exploits specifically the advantageous mathematical properties of the Wigner-Ville transform. A particular role in this context plays the relationship between the Wigner time-frequency representation and the linear canonical transform (LCT) (also called special affine Fourier transform or ABCD transform). This class of linear integral transformations generalizes classical transformations, like the Fourier or the Gauss-Weierstrass integral transform. It is crucial for the transformation of higher-dimensional problems into a (coupled) set of lower-dimensional problems and provides a theoretical basis for the phase retrieval in the second step of the algorithm. In addition, the choice of appropriate Linear Canonincal Transformations can improve the separability of the source signals.

The paper is structured as follows: After a discussion of related approaches and an introduction of the notation, Section 2 gives a short introduction into the theory of the stochastic Wigner-Ville distribution and its expected value, the Wigner-Ville spectrum. Though most of these results are well established in mathematics they are not so well-known in the machine learning community. The second part of this section introduces the linear canonical transform (LCT) and its properties that are fundamental for our algorithm. For the solution of the non-negative anechoic mixture problem a modification of non-negative matrix factorization (NMF) (Lee and Seung, 1999) is presented in 3.2. Finally, we discuss several phase retrieval methods in 3.3. Section 4 presents in detail three concrete implementations of the novel algorithmic framework. Section 5 presents a validation of the developed method on several data sets, including music streams, natural 2D images, human motion trajectories, and two-dimensional shapes. Finally, conclusions are presented in Section 6.

A preliminary version of the algorithm and some applications have been previously published as conference papers (Omlor and Giese, 2007a,b). However, the complete theoretical framework with a comparison between different implementations has never been published before.

### 1.1 Related Approaches

Anechoic mixtures have frequently been used in acoustics to model reverberation-free environments. Such models have been treated in several papers focusing on the under-determined case, often in the context of the 'cocktail party problem'. The work in Torkkola (1996a,b) extended the information maximization approach by Anthony and Sejnowski (1995), using the adaptive delay architecture described in Platt and Faggin (1992) in order to unmix anechoic $2 \times 2$ mixtures. Another approach by Emile and Comon (1998) is to estimate the unknown parameters directly in the time domain, with the additional assumption of predefined constant mixing weights ($\alpha_{ij} = 1$). Frequency or time-frequency methods, like the DUET algorithm (Yilmaz and Rickard, 2004) or the scatter plot method by Bofill (2003), exploit sparsity properties of the sources in these domains. For the even-determined case ($n = m$) the weights and delays can be estimated by joint diagonalization of specific spectral matrices, as demonstrated in Yeredor (2003). Also a two-dimensional version of the AC-DC joint diagonalization algorithm has been successfully applied for the separation of images that appeared with unknown spatial-shifts (Be'ery and Yeredor, 2008). Other work on the under-determined case is summarized in Ogrady et al. (2005), Arberet et al. (2007) and Namgook and Jay Kuo (2009).
The over-determined case, which is most important for dimension reduction applications has been treated only very rarely so far. In Harshman et al. (2003) this problem has been addressed using an alternating least squares (ALS) algorithm (Shifted Factor Analysis). This algorithm has been revised and improved in Mørup et al. (2007), exploiting the Fourier shift theorem and information maximization in the complex domain (SICA, Shifted Independent Component Analysis).

The performance of independent component analysis and blind source separation methods is critically dependent on the non-Gaussianity of the source distributions (Cardoso, 1998), and the possibility of a sparse representation of the data, which in turn is related to the super-gaussianity of the distribution of the sources. This implies that preprocessing of the signals, for example, by application of bilinear time-frequency transformations before source separation can be essential. In this context time-frequency representations have been used quite frequently in the context of blind source separation (Karako-Eilon et al., 2003; Leung and Siu, 2007; Li et al., 2004; Seki et al., 1998), and specifically for anechoic demixing (Yilmaz and Rickard, 2004). Besides, the application

of the superposition law of the Wigner-Ville spectrum (WVS) in Belouchrani and Amin (1998), such distributions have been mainly applied for pre-processing purposes. Contrasting with this work, our approach relies on further properties of the stochastic Wigner-Ville spectrum (WVS), which to our knowledge, never have been exploited for source separation previously.

### 1.2 Notations

Throughout the paper the following well-established mathematical notations will be used:

- The notation $:=$ indicates that the left hand side is defined by the right hand side of the equation.
- $\mathbb{i}$ denotes the complex unit.
- $E$ denotes the expectation operator.
- For a scalar, or a function $x$ , $x^*$ denotes the complex conjugate.
- The operators $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier and inverse Fourier transform respectively, defined by:

$$\mathcal{F}x(f) := \int x(t)e^{-2\pi\mathbb{i}tf}dt = X(f),$$
$$\mathcal{F}^{-1}X(t) = \int X(f)e^{2\pi\mathbb{i}tf}df.$$

  In the case of discrete time variables the same symbols signify the Discrete Fourier Transform (DFT).
- The notation $T_{ij}x$ indicates the shift operator $T_{ij}x(t) = x(t-\tau_{ij})$ with $\tau_{ij} \in \mathbb{R}$.
- If not noted otherwise, $t \in \mathbb{R}^d$ and $x : \mathbb{R}^d \to \mathbb{R}, t \mapsto x(t)$, that is, in general $x$ denotes a multivariate function. In order to distinguish the variable of integration from the functional variable, in addition to the variable $t$, the variable $t'$ is used. (Thus $x(t')$ denotes the function $x$ at point $t'$.)
- The symbol $*$ marks the convolution.
- If $A, B$ are two matrices then $\frac{A}{B}$ denotes the entrywise fraction $(\frac{A}{B}) = (\frac{a_{ij}}{bij})_{ij}$.
- $x'(t)$ is short for the derivative $\frac{dx}{dt}(t)$.
- $x \leftarrow y$ implies that $x$ is replaced by $y$ in the current iteration of an algorithm.

## 2. The Wigner-Ville Distribution and Its Stochastic Generalizations

The Wigner-Ville distribution was originally defined by Wigner (1932) in the context of quantum mechanics. It was later reintroduced in signal analysis by Ville (1948), with the basic idea of defining a joint distribution of the signal energy simultaneously in time and frequency (in physics corresponding to coordinates and momentum). For a continuous scalar real or complex signal (wave function) $x(t)$ the Wigner-Ville distribution is defined as the bilinear transformation:

$$W_x(t,f) := \int x\left(t+\frac{\tau}{2}\right) x^*\left(t-\frac{\tau}{2}\right) e^{-2\pi \mathbb{i} \tau f} d\tau. \tag{3}$$

Unfortunately, this expression cannot be interpreted as a true probability density, since it can become negative. A variety of mathematical properties have been proven for the Wigner-Ville distribution (Mecklenbruker and Hlawatsch, 1997), making it a widely used tool in signal analysis, with generalizations to linear signal spaces, linear time-varying systems or frames (see, e.g., Matz and Hlawatsch, 2003 for review). While the Wigner distribution was developed in the probabilistic framework of quantum mechanics, definition Equation (3) applies to deterministic functions $x$. In the works of Janssen (1979), Martin (1982) as well as Martin and Flandrin (1985) the deterministic definition Equation (3) has been extended to the very general class of harmonizable stochastic processes. The only requirement for a zero mean random signal to be harmonizable is the existence of a Fourier representation $\Phi$ of its autocovariance function $r_x(t,t')$ that is defined by:

$$r_x(t,t') := \mathrm{E}\{x(t)x^*(t')\} = \int\int e^{2\pi \mathbb{i}(\lambda t - \mu t')} \Phi(\lambda,\mu) d\lambda d\mu.$$

The probabilistic analogue to the deterministic Wigner distribution for a stochastic process $x$ is given by the stochastic integral:

$$w_x(t,f) := \int x\left(t+\frac{\tau}{2}\right) x^*\left(t-\frac{\tau}{2}\right) e^{-2\pi \mathbb{i} \tau f} dP(\tau).$$

In this formula $dP$ signifies a probability measure, defining a stochastic integral. (See Papoulis et al., 2001 for further details.) In this case, the distribution $w_x(t,f)$ is again a stochastic process. This stochastic integral exists in quadratic mean if the absolute forth order moments $E\{|x|^4\}$ exist (Martin, 1982). Furthermore, the existence of these moments guarantees that the expected value and the integration can be exchanged:

$$\begin{aligned} E\{w_x(t,f)\} &= E\left\{\int x\left(t+\frac{\tau}{2}\right) x^*\left(t-\frac{\tau}{2}\right) e^{-2\pi \mathbb{i} \tau f} dP(\tau)\right\} \\ &= \int E\left\{x\left(t+\frac{\tau}{2}\right) x^*\left(t-\frac{\tau}{2}\right)\right\} e^{-2\pi \mathbb{i} \tau f} d\tau = \int r_x\left(t+\frac{\tau}{2}, t-\frac{\tau}{2}\right) e^{-2\pi \mathbb{i} \tau f} d\tau \\ &=: \overline{W_x(t,f)}. \end{aligned} \tag{4}$$

The last expression $\overline{W_x(t,f)}$, which can be defined for a more general class of random processes, is called the Wigner-Ville spectrum (WVS). The invertibility of the Fourier integral Equation (4) assures that the WVS is equivalent to the covariance function $r_x(t,t')$. Therefore, it contains full information about the second-order statistics of $x$. The WVS represents a time-dependent spectrum that is commonly used to study the local and global non-stationarity of random processes.

### 2.1 Basic Examples for the WVS

**White noise process**

The white noise process $x$ is defined by:

$$E\{x(t)\} = 0$$
$$r_x(t,t') = \delta(t-t').$$

Then it is easy to see that the WVS is given by:

$$\overline{W_x(t,f)} = \int \delta(\tau) e^{-2\pi \mathrm{i} \tau f} d\tau = 1.$$

**Signal plus noise**

Defining a process $x(t) = y(t) + n(t)$ by the superposition of a deterministic signal $y(t)$ and a zero-mean noise component $n(t)$, the WVS is given by:

$$\begin{aligned}
\overline{W_x(t,f)} &= \int E\left\{x\left(t+\frac{\tau}{2}\right)x^*\left(t-\frac{\tau}{2}\right)\right\} e^{-2\pi \mathrm{i} \tau f} d\tau \\
&= \int \left[E\{yy^*\} + yE\{n^*\} + y^*E\{n\} + E\{nn^*\}\right] e^{-2\pi \mathrm{i} \tau f} d\tau = \int [yy^* + E\{nn^*\}] e^{-2\pi \mathrm{i} \tau f} d\tau \\
&= \underbrace{\int yy^* e^{-2\pi \mathrm{i} \tau f} d\tau}_{\text{deterministic}} + \int E\{nn^*\} e^{-2\pi \mathrm{i} \tau f} d\tau = W_y(t,f) + \overline{W_n(t,f)}.
\end{aligned}$$

This shows that for signal plus noise the WVS is given by the sum of the deterministic Wigner distribution and the WVS of the noise.

### 2.2 Properties of the WVS and the Wigner Distribution

The stochastic as well as the deterministic definition share many properties. Consequently, in the following only the properties of the WVS will be listed, unless the properties of the deterministic WV transformation are different. (Proofs for the properties can be found, for example, in Cohen 1989).

– `Real:` The WVS is a real function of time and frequency:

$$\begin{aligned}
\overline{W_x(t,f)}^* &= \int E\left\{x^*\left(t+\frac{\tau}{2}\right)x\left(t-\frac{\tau}{2}\right)\right\} e^{2\pi \mathrm{i} \tau f} d\tau \\
&= \int E\left\{x^*\left(t-\frac{\tau}{2}\right)x\left(t+\frac{\tau}{2}\right)\right\} e^{-2\pi \mathrm{i} \tau f} d\tau = \overline{W_x(t,f)}.
\end{aligned}$$

– `Time-frequency shift covariant:` The Wigner-Ville spectrum of a time frequency shifted signal $\tilde{x}(t) = x(t-t_0)e^{2\pi \mathrm{i} f_0(t-t_0)}$ is the shifted WVS of the original signal:

$$\overline{W_{\tilde{x}}(t,f)} = \overline{W_x(t-t_0, f-f_0)}. \tag{5}$$

– `Correct marginals:` The marginals in time and frequency of the WVS reflect the second order properties of the process:

$$\int \overline{W_x(t,f)}\,df = r_x(t,t) = E\{|x(t)|^2\}, \tag{6}$$

$$\int \overline{W_x(t,f)}\,dt = r_{\mathcal{F}x}(f,f) = E\{|\mathcal{F}x(f)|^2\}. \tag{7}$$

– `Cross terms:` Due to its quadratic nature, the deterministic Wigner-distribution of a multi-component signal $x = s_1 + s_2$ always contains cross terms of the form:

$$W_{s_1,s_2}(t,f) := \int s_1\left(t+\frac{\tau}{2}\right) s_2^*\left(t-\frac{\tau}{2}\right) e^{-2\pi \mathbb{i} \tau f}\,d\tau.$$

Geometrically, these terms will always occur in the time-frequency plane midway between two (auto-)components $W_{x_{1,2}}$ of the deterministic Wigner distribution (Mecklenbruker and Hlawatsch, 1997). For the stochastic WVS existence of cross terms is determined by the time-frequency correlations of the processes. Is for example $x(t) = \sum_{j=1}^{n} \alpha_j \cdot s_j(t)$, the sum of $n$ uncorrelated zero mean random processes $s_j(t)$, then it is obvious from $r_{s_k,s_l}(t,t') = E\{s_k(t)s_l(t')\} = 0\ \forall k \neq l$ that $\overline{W_{s_k,s_l}(t,f)} = 0\ \forall k \neq l$ and thus:

$$\overline{W_x(t,f)} = \sum_j^n |\alpha_j|^2 \overline{W_{s_j}(t,f)}. \tag{8}$$

A similar superposition law also holds for the sum $x(t) = \sum_{j=1}^{n} \alpha_j \cdot s_j(t)$ of deterministic signals $s_j(t)$ if the weights $\alpha_j$ are uncorrelated zero mean random factors:

$$\overline{W_x(t,f)} = \sum_j^n E\{|\alpha_j|^2\} \overline{W_{s_j}(t,f)}.$$

– `Instantaneous frequency:` For a univariate random signal $x$ which is square mean differentiable, the instantaneous frequency can be defined as:

$$f_x(t) = \frac{\mathrm{Im}(x'(t)x(t)^*)}{2\pi \mathrm{var}(x(t))}.$$

The expectation is given by the following relation:

$$E\{f_x(t)\} = \frac{\int f \overline{W_x(t,f)}\,df}{\mathrm{var}(x(t))} = \frac{E\{\mathrm{Im}(x'(t)x(t)^*)\}}{2\pi \mathrm{var}(x(t))}. \tag{9}$$

A similar property also holds for the group delay (Cohen, 1989).

– `Symplectic covariance:` For our algorithm it will be central that signals are related that correspond to Wigner Ville Spectra resulting from each other by linear transformations in the time-frequency plane. More specifically, the WVS is covariant against area-preserving linear

transformations $M$ of the time-frequency plain. Such transformations $M \in \mathbb{R}^{2d\times 2d}$ belong to the symplectic group, that is, $M$ has the form:

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \text{ with } \det(M) = 1\ ,\ A^T C = C^T A\ ,\ D^T B = B^T D.$$

The transformed WVS is again the WVS of another signal $\tilde{x}(t)$, which is given by the relationship:

$$\begin{aligned} \overline{W_x(t,f)} &= \overline{W_{\tilde{x}}((t,f)M^T)} \text{ or equivalently} \\ \overline{W_x((t,f)(M^T)^{-1})} &= \overline{W_{\tilde{x}}(t,f)}. \end{aligned} \tag{10}$$

The processes $x$ and $\tilde{x}$ are related by the so called linear canonical transform (LCT), discussed in the following section.

### 2.3 WV Spectrum Estimation for Non-stationary Signals

In most practical applications, it is necessary to estimate the WVS from a single realization of a process. For this reason ensemble averages are often unavailable. As the WVS is the Fourier transform of the autocovariance function, any estimator for $r_x(t,t')$ is sufficient in the sense of statistics. All such estimators have a fundamental bias-variance tradeoff. Smoothing reduces the variance of the estimate but introduces a bias. One way to derive an estimator for the WVS is to assume the process $x(t)$ is semi-stationary, implying that its characteristics are changing slowly with time. This so called quasi-stationary assumption allows the approximation of the autocovariance function $r_x(t,t')$ with a local average:

$$r_x\left(t+\frac{\tau}{2}, t-\frac{\tau}{2}\right) \approx \int F(s-t,\tau) x\left(s+\frac{\tau}{2}\right) x^*\left(s-\frac{\tau}{2}\right) ds.$$

Here $F$ is an arbitrary window function that determines the local time-averaging used to estimate the autocovariance function $r_x(t,t')$. If this approximation is used to form an estimate of the WVS (replacing $r_x$ in Equation (4)) this leads to the equation:

$$\begin{aligned} \overline{W_x(t,f)} &\approx \int W_x(t',f')\vartheta(t-t', f-f')dt'df' \text{ with} \\ F(t,\tau) &= \int \vartheta(t,\nu) e^{-2\pi \mathrm{i}\nu\tau} d\nu. \end{aligned} \tag{11}$$

Therefore, all distributions belonging to Cohen's class (Cohen, 1989) form estimators for the WVS (Martin and Flandrin, 1985). An alternative way to justify this particular class of estimators can be found in Akbar and Douglas (1995). Other ways to estimate the WVS include multitaper reassignment (Xiao and Flandrin, 2007) or soft wavelet thresholding (Baraniuk, 1994).

The marginal properties (Equation 6, Equation 7) impose certain conditions on the kernel. This implies that not all members of Cohen's class share all properties of the WVS. From this point of view the choice $\vartheta \equiv \delta(t)$, which approximates the WVS by the deterministic Wigner distribution of the samples, seems a natural choice for the estimator. It should also be noted that, as long as certain properties preserved by the chosen kernels, such as the correct marginals (Equation 6, Equation 7), the derivations derived in the following sections remain valid even if the estimator does not share

all properties of the WVS. The choice of the deterministic Wigner Ville Distribution as estimator is thus less restrictive as it might originally appear.

While the algorithms derived in the following sections do not require at any point the explicit numerical computation of a WVS, numerically estimators have been proposed that allow to implement the computation of the transformation on a computer, for example Martin and Flandrin (1985):

$$\overline{W_x(t,f)} \approx 2 \sum_{\tau=-\infty}^{\infty} \sum_{\tau'=-\infty}^{\infty} F(\tau,2\tau') x(t+\tau+\tau') x^*(t-\tau+\tau') e^{-2\mathbb{i}\omega\tau}.$$

### 2.4 The Linear Canonical Transform (LCT)

At several points of the algorithms that will be derived in the following we exploit the symplectic covariance property (Equation 10) of the WVS. This properties links signals that correspond to WVS that are related by a symplectic trasformation in the time-frequency domain. The relationship between $x$ and $\tilde{x}$ as specified by the symplectic covariance property (Equation 10) can be derived explicitly in the case of a nonsingular submatrix $B$ (i.e., $\det(B) \neq 0$). It is given by the following integral transformation, which is called linear canonical transform (LCT) or ABCD-transform (Bultheel and Martínez-Sulbaran, 2006):

$$\mathcal{L}(M)[x](f) := (\det \mathbb{i}B)^{1/2} \int_{-\infty}^{\infty} x(t) \exp\left(\mathbb{i}\pi(t^T B^{-1} A t - 2t^T B^{-1} f + f^T D B^{-1} f)\right) dt. \tag{12}$$

In the case of singular matrices $B$ the general description is more involved and can be found in Alieva and Bastiaans (2007).

Integral transformations of this type play an important role in optics and information processing, as they specify affine transformations in phase space. For example Equation (12) describes the behavior of a wave function $x$ by a propagation through a system of thin lenses, free space or if focused with a satellite dish. For special cases of matrices $M$ many classical integral transformations used in signal processing can be realized. Some examples are, considering a function $x(t)$ with one-dimensional argument $t$:

– Fourier and fractional Fourier transforms are special cases given by the one-parameter subgroup $M(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$. The fractional Fourier transform with parameter $\theta$ is an operator power of the normal Fourier transform with the exponent $\frac{2\theta}{\pi}$. The Integral representation (Ozaktas et al., 2001) is given by:

$$\mathcal{F}^{\theta} x(f) := \sqrt{\frac{1-\mathbb{i}\cot(\theta)}{2\pi}} e^{\mathbb{i}\cot(\theta) f^2/2} \int_{-\infty}^{\infty} e^{-\mathbb{i}\csc(\theta) f t + \mathbb{i}\cot(\theta)\frac{t^2}{2}} x(t) dt. \tag{13}$$

The case $\theta = \frac{\pi}{2}$ corresponds to the classical Fourier transform. This form was used in the context of one implementation of the proposed algortihmic framework (see Section 4.2).

– The Fresnel transform is defined by the expression:

$$[\text{Fresnel}(z,l)x](\xi) = \frac{e^{\mathbb{i}\pi z/l}}{\sqrt{\mathbb{i}lz}} \int_{-\infty}^{\infty} e^{\mathbb{i}\pi/zl(\xi-t)^2} x(t) dt.$$

is obtained in the case $M = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \frac{zl}{2\pi} \\ 0 & 1 \end{pmatrix}$.

– Chirp multiplication is given by $M = \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$.

– Scaling can be achieved by $M = \begin{pmatrix} a & 0 \\ 0 & \frac{1}{a} \end{pmatrix}$.

These relationships make it possible to adopt the developed framework to different common representations of data in the frequency space, in order to optimize the performance of the algorithm.

### 2.5 Main Properties of the Linear Canonical Transform

Independent of the matrix $M$, all linear canonical transforms share the following properties:

– `Linearity:` Obviously the linear canonical transform is a linear transformation, for example:

$$\mathcal{L}(M)[\lambda x + \mu y](f) = \lambda \mathcal{L}(M)[x](f) + \mu \mathcal{L}(M)[y](f) \; \forall \, \lambda, \mu \in \mathbb{R}.$$

– `Unitarity:` The LCT is a unitary operation. Assuming $(.)^*$ denotes the adjoint operator then:

$$(\mathcal{L}(M))^{-1} = (\mathcal{L}(M))^* = \left(\mathcal{L}(M^{-1})\right).$$

– `Group structure:` The product of two LCT operators with matrices $M_1$ and $M_2$ is again a LCT with matrix $M_3 = M_1 M_2$:

$$(\mathcal{L}(M_1)\mathcal{L}(M_2)) = \mathcal{L}(M_1 M_2).$$

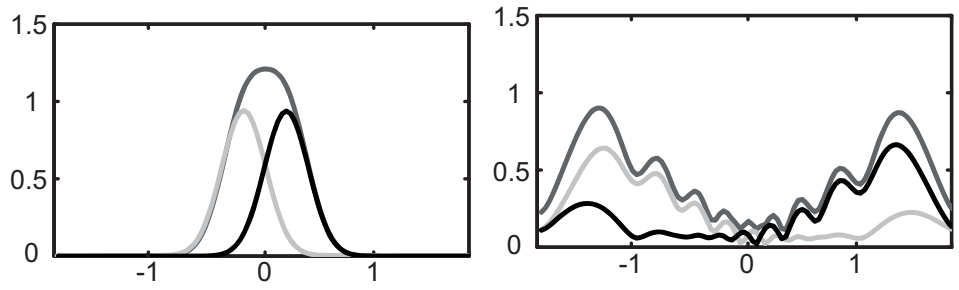
– `Shift Theorem:` Of particular interest for the anechoic mixing problem is the behavior of the linear canonical transform under shifts of the input signal:

$$\mathcal{L}(M)[x(t-\tau)](f) = \exp\left(\mathbb{i}\pi(2f - A\tau)^T C\tau\right) \; \mathcal{L}(M)[x](f - A\tau). \tag{14}$$

Since the output signal is translated by the term $A\tau$ a good choice of the matrix $A$ can optimize the separability of the the signal $x$ from it is shifted counterparts in particular LCT domains (Sharmaa and Joshi, 2006). To illustrate this point, consider the example of two gaussian signals with a small difference in means (see Figure 1(a)). Linear canonical transformation with $A = 3$ and $B = 0$ increases the mean difference, and thus leads to an improved separability, as shown in Figure 1(b).

## 3. Algorithmic Framework: Anechoic Demixing Using Wigner Marginals

In the following, we will first derive the basic algorithmic framework by application of the mathematical properties of the WVS, as discussed in the last section, to the anechoic mixing model. The resulting algorithm consists of two major steps: the solution of positive anechoic demixing problems and phase retrieval. In the following, we will discuss specific methods that implement these two main steps of our algorithmic framework. Section 3.2 discusses a special form of Non-negative Matrix Factorization, which is particularly suited for the algorithm. Finally, Section 3.3 discusses different methods for the implementation of the phase retrieval step.




(a) Two gaussian signals with a small difference in mean and their sum (dark grey).

(b) Absolute value of the LCT transformation of the two gaussians with $A = 3, B = 0$.

Figure 1: Example for increased separability of signals in appropriate LCT-domain.

### 3.1 Application of the WVS to the Mixture Model

We assume that the observed random signals (processes) $x_i(t)\,, i = 1,\ldots,m$ are delayed superpositions of uncorrelated source signals $s_j(t)$, that is:

$$x_i(t) = \sum_{j=1}^{n} \alpha_{ij} \cdot s_j(t-\tau_{ij}) \quad i = 1,\cdots,m \tag{15}$$
$$\text{with } r_{s_i,s_j}(t,t') = E\{s_i(t)s_j^*(t')\} = 0 \text{ for } i \neq j.$$

The shift covariance (5) of the WVS and the superposition law (8) imply the following relation between the spectra of the sources and observations:

$$\overline{W_{x_i}(t,f)} = \sum_{j=1}^{n} |\alpha_{ij}|^2 \overline{W_{s_j}(t-\tau_{ij},f)} = \sum_{j=1}^{n} |\alpha_{ij}|^2 \overline{W_{T_{ij}s_j}(t,f)}. \tag{16}$$

Since the relationship (16) holds pointwise for all points of the time-frequency plain $(t,f)$ it is also fulfilled after application of an area preserving symplectic transformation $M^{-1}$:

$$\begin{aligned} \overline{W_{x_i}\big((t,f)(M^T)^{-1}\big)} &= \sum_{j}^{n} |\alpha_{ij}|^2 \overline{W_{T_{ij}s_j}\big((t,f)(M^T)^{-1}\big)} \\ \overset{(10)}{\Longrightarrow} \overline{W_{\mathcal{L}[M](x_i)}(t,f)} &= \sum_{j}^{n} |\alpha_{ij}|^2 \overline{W_{\mathcal{L}[M](T_{ij}s_j)}(t,f)}. \end{aligned} \tag{17}$$

In order to make use of relation (17) an estimator for the WVS is required (cf. Section 2.3). The simplest bias-free estimator for the WVS is just the empirical Wigner transformation. Alternative estimators belong to Cohen's class of time-frequency distributions. In the following only the deterministic Wigner transformation is applied, since it has the advantage of preserving the properties of the WVS. Depending on the smoothing kernel $\vartheta$ in Equation (11), these properties are shared by large number of time-frequency-distributions in Cohen's class, consequently leading to the same expression (20) (see below). Replacing the WVS by its deterministic counterpart, Equation (17) transforms into:

$$W_{\mathcal{L}[M](x_i)}(t,f) = \sum_{j=1}^{n} |\alpha_{ij}|^2 W_{\mathcal{L}[M](T_{ij}s_j)}(t,f). \tag{18}$$

This equation could directly serve as basis for the separation of the source signals $s_j$. However, for higher-dimensional cases the computational costs make a direct solution prohibitive. Since the joint description of the involved function in time and frequency is highly redundant, an equivalent description can be derived by computing the marginals of Equation (18):

$$\begin{aligned}|\mathcal{L}(M)[x_j]|^2(t) &= \int W_{\mathcal{L}(M)[x_j]}(t,f)df = \sum_j^n |\alpha_{ij}|^2 \int W_{\mathcal{L}(M)[T_{ij}s_j]}(t,f)df \\ &= \sum_j^n |\alpha_{ij}|^2 |\mathcal{L}(M)[T_{ij}s_j]|^2(t) \overset{(14)}{=} \sum_j^n |\alpha_{ij}|^2 |\mathcal{L}(M)[s_j](t-A\tau_{ij})|^2. \end{aligned} \tag{19}$$

Note that by introduction of the general matrix $M$ into Equation (17) time and frequency marginals lead to equivalent expressions Equation (19), but with different symplectic transformations $\tilde{M}$. The two types of marginal are are related to each other by the product of $M$ with a $\frac{\pi}{2}$-rotation in the time-frequency plain, resulting in an interchange of the time and the frequency axis. The individual marginals have the form of an anechoic mixture problem with an additional nonnegativity constraint:

$$|\mathcal{L}(M)[x_i]|^2(t) = \sum_{j=1}^n |\alpha_{ij}|^2 |\mathcal{L}(M)[s_j](t-A\tau_{ij})|^2 + n(t). \tag{20}$$

The additional noise term $n(t)$ emphasizes the fact that Equation (20) is an approximation, which only holds precisely under the asymptotic condition $r_{s_i,s_j}(t,t') = 0$. In practice deviation from exact time-frequency disjointness (e.g., due to the finite sample size) result non-zero noise terms $n(t)$. However, the (approximate) solution of Equation (20) gives an estimate for the power spectra of the sources in LCT domain $|\mathcal{L}(M)[s_j]|^2$, the scaled delays $A\tau_{ij}$ and the absolute value of the weights $|\alpha_{ij}|$. Methods for the solution of the positive shifted mixture problem will be discussed in Section 3.2. The missing phase information for $\mathcal{L}(M)[s_j]$ can be recovered by computing multiple marginals depending on different matrices $M$. Phase retrieval methods for this purpose are discussed in the following Section 3.3. This operation exploits for a second time the special mathematical properties of the WVS with respect to the symplectic covariance.

Since the projections onto LCT domains (19) are also anechoic mixtures, at first glance, the positive system of Equations (20) seems not to provide any advantage compared to the direct solution of Equations (15) or (18). However, a more thorough analysis shows that the transformation into the coupled set of equations for the marginals has the following advantages:

(i) A single marginal described by Equation (19) is not sufficient to reconstruct the signal or its WVS. It is thus necessary to compute a set of different marginals, and the choice of the matrix $M$ in the LCT allows the selection of different marginals that correspond, for example, to different slanted lines in the time-frequency space in the case of fractional Fourier transform. This procedure is different from a source separation directly in the time-frequency plane (Belouchrani and Amin, 1998). In this case (separation in the TF plane), application of the LCT would not result in a facilitation of source separation because it corresponds to an area preserving deformation $M$ (Healy et al., 2008). The chosen approach to work on the marginals however, has the immediate advantage of a much higher computational efficiency since it avoids the high computational cost of the WVS and allows the reduction of multi-variate problems into an uni-variate problems.

(ii) For each single marginal (19) the linear canonical transform, that is, the choice of $M$, determines the distribution of the estimated sources, and especially their sparseness. This can be exploited to improve the interpretability and compactness of recovered source models. Since the LCT includes a lot of different transformations, such as the Fourier or the fractional Fourier transformation, it provides much more flexibility for an optimization of the representation space of the data than methods relying on a single representation (e.g., Healy et al., 2008). A nice example of this advantage is data that can be represented efficiently by a superposition of chirps, given by functions of the form $e^{2\alpha\pi\mathbb{i}t^2}$. These functions are neither sparse in time nor frequency. However, they have compact support in a correctly chosen LCT domain. A particular useful approach is to chose the matrix $M$ in a way that maximizes the sparsity of the power spectral density within the signal space that is defined by the corresponding LCT. (The WVS in different linear spaces is discussed in Hlawatsch and Kozek (1993).) Formally, one might thus determine $M$ for a given data set that spans the linear space $\mathcal{X} = \text{span}(x_1, \ldots, x_n)$ by maximizing the sparseness of the approximation, formally:

$$\max_M \text{sparsity}\left(|\mathcal{L}(M)[\mathcal{X}]|^2\right).$$

(iii) The magnitude of the effective (scaled) delays of the mixture components in Equation (19) $\tau_{ij}$ directly depends on the parameter $A$ of the LCT. An appropriate choice of $A$ results thus in improved separation of the signals (Sharmaa and Joshi, 2006). Furthermore the choice $A = 0$ simplifies the mixture by transforming it into an instantaneous mixture. As the multivariate LCT can be the tensorial product of several one-dimensional LCTs, $A$ can be adjusted for each dimension individually, making it possible to transform a multivariate anechoic mixture into several coupled one-dimensional mixtures (Omlor and Giese, 2007a). This is the core idea behind the implementation of our algorithm for the multi-variate case, which is discussed in Section 4.2.

Additional remarks about the proposed framework:

1. Since both the deterministic and the stochastic WVS share the property of correct marginals, Equation (20) holds also for the exact power-spectra $E\{|\mathcal{L}(M)s_i|^2\}$. The positivity in (20) is thus a direct consequence of the assumption of uncorrelated sources and does not result from an additional approximation. Similarly, the fact that the cross-terms vanish in Equation (18) represent a consequence of this assumption. With respect to these points, the derivation of Equation (18) exploits the same properties as joint diagonalization approaches (e.g., Belouchrani and Amin, 1998).

2. The uncorrelatedness assumption for the sources leads to vanishing cross terms in the full time-frequency plane. However, a single projection (20) requires only that the cross terms in that particular LCT domain vanish. In practice, this simplifies the choice of appropriate LCT domains in order to improve the separability of the sources.

3. The linear weights $\alpha_{ij}$ and the delays $\tau_{ij}$ can be estimated exploiting multiple marginals simultaneously. This makes the estimation more robust against the influence of noise.

4. For the special case $\mathcal{L}[M] = \mathcal{F}$ (Fourier transformation), Equation (20) can be derived without the theoretical background of time-frequency analysis. However, the obtained single equation

is not sufficient for the reconstruction of the sources, and without the mathematical theory of the WVS it is not clear how the missing phase information can be recovered.

### 3.2 Nonnegative Matrix Factorization

In the following we will introduce a modified version of convolutive non-negative matrix factorization as one method that is particularly suitable for the numerically efficient solution of the positive mixture problem (19). Opposed to standard non-negative matrix factorization (NMF), convolutive NMF assumes a non-negative convolutive model (see Equation 1).

#### 3.2.1 CONVOLUTIVE NON-NEGATIVE MATRIX FACTORIZATION

The simple standard non-negative matrix factorization model

$$\mathbf{X} \approx \mathbf{AS} \text{ with } \mathbf{A} \geq 0 \ , \ \mathbf{S} \geq 0 \tag{21}$$

assumes that the data $\mathbf{X}$ arises from a linear superposition of positive vectors. For many applications, such as spectral analysis or blind image deblurring, the data can be described more exactly as a (discrete) positive convolution of the form:

$$x_i(t) = \sum_{j=1}^{n} (a_{ij} * s_j)[t] = \sum_{j=1}^{n} \sum_{u} a_{ij}[u] s_j[t-u] \tag{22}$$
$$a_{ij} \geq 0 \ , \ s_j \geq 0 \ \forall i \in 1, \cdots, m.$$

Depending on the definition of the convolution operator $*$ in Equation (22), the weights $a_{ij}$ are either periodic functions or are characterized by a compact support, referred to as circular or linear convolution. It is always possible to interpret a linear convolution with compact support as a circular convolution with additional zero padding (Rabiner and Gold, 1975). We thus present here the convolutive NMF algorithm in the following for the circular case.

If both, the filters $a_{ij}$ and the sources $s_j$ are unknown, Equation (22) implies the estimation problem:

$$\min_{a_{ij}, b_j} \| x_i(t) - \sum_{j=1}^{n} (a_{ij} * s_j)(t) \| \tag{23}$$
$$\text{subject to } s_j \geq 0 \ , \ a_{ij} \geq 0 \ \forall i, j.$$

Since (22) is both finite-dimensional and linear, it can be expressed in matrix form. In the simple case of one-dimensional (vectorial) data the convolution $a_{ij} * s_j$ is equal to the matrix products:

$$
\begin{aligned}
(a_{ij} * s_j) &= \begin{bmatrix} a_{ij}(0) & a_{ij}(N) & a_{ij}(N-1) & \cdots \\ a_{ij}(1) & a_{ij}(0) & a_{ij}(N) & \dots \\ a_{ij}(2) & a_{ij}(1) & \ddots & \ddots \\ \vdots & \vdots & \vdots & \vdots \\ a_{ij}(N) & a_{ij}(N-1) & \dots & \dots \end{bmatrix} \cdot \begin{bmatrix} s_j(0) \\ s_j(1) \\ \vdots \\ \\ s_j(N) \end{bmatrix} =: \mathbf{A}^{ij}\mathbf{S}^j \\
&= \begin{bmatrix} s_j(0) & s_j(N) & s_j(N-1) & \cdots \\ s_j(1) & s_j(0) & s_j(N) & \dots \\ s_j(2) & s_j(1) & \ddots & \ddots \\ \vdots & \vdots & \vdots & \vdots \\ s_j(N) & s_j(N-1) & \dots & \dots \end{bmatrix} \cdot \begin{bmatrix} a_{ij}(0) \\ a_{ij}(1) \\ \vdots \\ \\ a_{ij}(N) \end{bmatrix} =: \mathbf{S}_j\mathbf{A}_{ij}. \qquad (24)
\end{aligned}
$$

For higher dimensional data the vectors $\mathbf{S}_j, \mathbf{S}^j$ and the Toeplitz matrices $\mathbf{A}_{ij}, \mathbf{A}^{ij}$ are replaced with column vectors and block circular matrices respectively. Adopting this matrix notation, one can reformulate the model (22):

$$
\begin{aligned}
\mathbf{X} := \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = (x_i)_i = \left( \sum_{j=1}^{n} a_{ij} * s_j \right)_i \quad &= \left( \sum_{j=1}^{n} \mathbf{A}^{ij}\mathbf{S}^j \right)_i = \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} & \dots & \mathbf{A}^{1n} \\ \mathbf{A}^{21} & \mathbf{A}^{22} & \dots & \mathbf{A}^{2n} \\ \vdots & \vdots & \ddots & \end{bmatrix} \cdot \begin{bmatrix} \mathbf{S}^1 \\ \vdots \\ \mathbf{S}^n \end{bmatrix} \\
=: \mathbf{A} \cdot \mathbf{S}. \qquad (25)
\end{aligned}
$$

and the optimization problem (23):

$$
\begin{aligned}
&\min_{\mathbf{A},\mathbf{S}} \|\mathbf{X} - \mathbf{A}\mathbf{S}\| \\
&\text{subject to } \mathbf{A}, \mathbf{S} \geq 0 \; \forall i, j.
\end{aligned}
$$

Writing the deconvolution problem in this form immediately shows the connection to nonnegative matrix factorization. Indeed Equation (25) is just a special case of Equation (21) with a specific structure of the matrix $\mathbf{A}$. Thus according to the multiplicative updates derived by Lee and Seung (1999) Equation (23) can be optimized by repeated iteration of the two steps:

$$
\mathbf{A}_{ij} \leftarrow \mathbf{A}_{ij} \frac{(\mathbf{X}\mathbf{S}^T)_{ij}}{(\mathbf{A}\mathbf{S}\mathbf{S}^T)_{ij}}, \qquad (26)
$$

$$
\mathbf{S}_j \leftarrow \mathbf{S}_j \frac{(\mathbf{A}^T\mathbf{X})_j}{(\mathbf{A}^T\mathbf{A}\mathbf{S})_j}. \qquad (27)
$$

In Equations (26,27) it is important to note that none of the operators $\mathbf{A}, \mathbf{A}^T, \mathbf{S}$ and $\mathbf{S}^T$ appears alone, but always in combination with its influence on another operator. For example, in order to compute

$$
\mathbf{A}\mathbf{S} = \left( \sum_{j=1}^{n} a_{ij} * s_j \right)_i
$$

it is not necessary to represent either **A** or **S** as high-dimensional matrix. Additionally the convolution can be implemented very efficiently exploiting Fast Fourier Transform (FFT), instead of an expensive matrix multiplication. From relations (24) and (25) it is easy to derive that the effect of the transposed operators $\mathbf{A}^T$, $\mathbf{S}^T$ can be expressed as:

$$\mathbf{A}^T\mathbf{X} = \left(\sum_{i=1}^{n} a_{ij}^T * x_i\right)_j = \left(\sum_{i=1}^{n} \mathcal{F}^{-1}\big((\mathcal{F}a_{ij})^* \mathcal{F}x_i\big)\right)_j,$$
$$\mathbf{X}\mathbf{S}^T = \big(x_i * s_j^T\big)_{ij} = \big(\mathcal{F}^{-1}\big((\mathcal{F}s_j)^* \mathcal{F}x_i\big)\big)_{ij}.$$

Here we exploited the fact that the impulse response of the adjoint filter is the complex conjugate of the original signal. Replacing the matrices in Equation (26,27) by their corresponding representations in the discrete Fourier domain, results in the following fast *update rules for convolutive NMF*:

$$a_{ij} \leftarrow a_{ij} \frac{\big(\mathcal{F}^{-1}\left((\mathcal{F}s_j)^* \cdot \mathcal{F}x_i\right)\big)}{\left(\mathcal{F}^{-1}\left(\sum_k (\mathcal{F}s_j)^* \cdot \mathcal{F}a_{ik} \cdot \mathcal{F}s_k\right)\right)} \forall i, j,$$
$$s_j \leftarrow s_j \frac{\mathcal{F}^{-1}\left(\sum_k (\mathcal{F}a_{kj})^* \cdot \mathcal{F}x_k\right)}{\mathcal{F}^{-1}\big(\sum_{p,l} (\mathcal{F}a_{pj})^* \cdot \mathcal{F}a_{pl} \cdot \mathcal{F}s_l\big)} \forall j.$$

Though the non-negative deconvolution problem (22) is formally equivalent to the regular NMF the matrix product **AB** does not define a low rank decomposition. As a consequence Equation (22) has, in contrast to Equation (21), no prospect of uniqueness. For example, every member of the family $a_{ij} = x_i(t-\tau_j)$ and $s_j = \delta_{-\tau_j}$ is a (trivial) possible solution. In order to find unique solutions for the optimization problem (22) it is thus necessary to introduce some form of regularization, such as sparseness assumptions about the filters (e.g., Chen and Cichocki, 2005).

### 3.2.2 CONVOLUTIVE NMF FOR ANECHOIC DEMIXING (ANECHOIC NMF (ANMF))

The positive anechoic demixing problem

$$x_i(t) = \sum_{j=1}^{n} \alpha_{ij} s_j(t-\tau_{ij}) \text{ with } x_i, \alpha_{ij}, s_j \geq 0 \; \forall \; i, j$$

can be written as a special case of a convolution, by defining the filters as $\nu_{ij} = \alpha_{ij}\delta(t-\tau_{ij})$:

$$x_i = \sum_{j=1}^{n} \nu_{ij} * s_j \text{ with } x_i, \nu_{ij}, s_j \geq 0 \; \forall \; i, j. \tag{28}$$

The anechoic problem (28) can thus be solved using the convolutive NMF algorithm for the solution of the problem (22), if sparseness of the filters is enforced. Replacing the update rules for the squared euclidian distance with rules which minimize Amari's alpha-divergence (Cichocki et al., 2008) results in the update rules:

$$\nu_{ij} \leftarrow \left[\nu_{ij}\left(\mathcal{F}^{-1}\left((\mathcal{F}s_j)^* \mathcal{F}\left[\frac{y_i}{\mathcal{F}^{-1}\left(\sum_k \mathcal{F}\nu_{ik}\mathcal{F}s_k\right)}\right]^{\beta}\right)\right)^{\frac{\mu}{\beta}}\right]^{1+\lambda_1}, \tag{29}$$

$$s_j \leftarrow \left[s_j\big(\mathcal{F}^{-1}\sum_k (\mathcal{F}\nu_{jk})^* \mathcal{F}\left[\frac{y_k}{\mathcal{F}^{-1}\big(\sum_l \mathcal{F}\nu_{kl}\mathcal{F}s_l\big)}\right]^{\beta}\big)^{\frac{\mu}{\beta}}\right]^{1+\lambda_2} \tag{30}$$

which can be adjusted for sparseness. The specific choice $\mu = 1.9, \beta = 2, \lambda_1 = 0.02, \lambda_2 = 0$ results in sparse features (Cichocki et al., 2008). Alternative approaches for the introduction of sparseness include, for example, bayesian regularization (Lin and Lee, 2005).

One disadvantage of delay estimation by deconvolution, is the difficulty to locate the exact peak of the sparse filter, especially for fractional delays (non-integer delays). This can be avoided if the multiplicative updates for the sparse filter given by Equation (29), are replaced with a constrained standard estimator for the delays (see Appendix).

The positive anechoic mixture problem that is given by Equation (20)

$$|\mathcal{L}(M)[x_i]|^2(t) = \sum_{j=1}^{n} |\alpha_{ij}|^2 |\mathcal{L}(M)[s_j](t - A\tau_{ij})|^2 + n(t).$$

can be solved with this anechoic non-negative matrix factorization approach (ANMF). The derivation of the last equation assumed vanishing cross-terms in the Wigner-Ville spectrum (equivalent to the assumption of uncorrelated sources). Similar as in Vollgraf et al. (2000), this together with the marginal properties of the Wigner-Ville cross spectrum implies vanishing LCT cross-spectra, that is:

$$\begin{aligned} W_{s_i,s_j}(t,f) \overset{i\neq j}{=} 0 \;&\Rightarrow\; \mathcal{L}(M)(s_i)\mathcal{L}(M)(s_j)^* \overset{i\neq j}{=} 0 \\ &\Rightarrow\; |\mathcal{L}(M)(s_i)|^2 |\mathcal{L}(M)(s_j)|^2 \overset{i\neq j}{=} 0. \end{aligned} \tag{31}$$

The fact that the cross-spectra are not present in Equation 31 permits a reformulation as a constraint for the anechoic NMF problem (19) of the form:

$$\mathbf{S}\mathbf{S}^T = I. \tag{32}$$

This orthogonality constraint guarantees a unique solution NMF-factorization problem (Ding et al., 2006). In addition, it is closely linked to positive independent component analysis (Yang and Yi, 2007; Plumbley and Oja, 2004). Equation (32) represents a special property of the estimator, which it is not necessarily fulfilled for the expectations $E\{\mathcal{L}(M)(s_i)\}$ because they are not zero mean. This implies that the condition (32) guaranties source separation, but only if one imposes an additional asymptotic constraint on the sources. This orthogonality constraint (32) guaranties the uniqueness of the solution of the anechoic demixing problem, making the new algorithm a true blind source separation method. In practice, however, it is often useful to relax this restriction for the sources and to replace the orthogonal NMF update rules with more classical updates, for example, based on least squares. The resulting model can capture also sources that are not perfectly uncorrelated (independent), often resulting in better approximation of the data.

### 3.3 Phase Retrieval

The solution of the positive anechoic mixture problem:

$$|\mathcal{L}(M)[x_i]|^2(t) = \sum_{j=1}^{n} |\alpha_{ij}|^2 |\mathcal{L}(M)[s_j](t - A\tau_{ij})|^2$$

that is given by Equation (20) yields the LCT power spectra $|\mathcal{L}(M)[s_j]|^2$. In order to reconstruct the sources $s_j$, or equivalently their LCT transforms $\mathcal{L}(M)[s_j]$, the missing phase information needs to be recovered. Depending on the choice of $M$ in (20) several phase recovery techniques can be applied for this purpose.

### 3.3.1 DECONVOLUTION

In many applications, such as computer tomography or image processing, the linear weights $\alpha_{ij}$ can be assumed to be nonnegative. This implies that the absolute value operation on the weights can be dropped in (20). For $A \neq 0$ in this case the anechoic mixture (2) reduces to a simple deconvolution problem with known filters $\nu_{ij} = a_{ij}\delta(t-\tau_{ij})$. If the weights are real and have arbitrary signs the filter can determined only up to an arbitrary sign, and if the weights are complex a constant complex phase factor of the filter remains undetermined. In all cases (with appropriate reguarization) the filters can be estimated applying applying a standard deconvolution algorithms, such as the Wiener filter (Wiener, 1964), and the sources $s_j$ can be retrieved by least squares estimation.

### 3.3.2 GERCHBERG-SAXTON (GS) ALGORITHM

In typical phase retrieval applications the parameters known about the signal are the phase and some other properties like the support. A simple procedure to recover the signal from these constraints is based on the projection onto convex sets (POCS). This method iteratively projects random starting values onto constrained sets that are specified by the given data. This approach can been applied for the reconstruction of signals from multiple power-spectra (Zalevsky et al., 1996). For the phase retrieval problem in our algorithm, an estimate of the signal is transformed back and forth between different LCT domains, always replacing the estimated amplitude spectrum by the known amplitude spectrum from the solution of the positive mixture problem and re-eastimating the phase. This fundamental algorithm is called Gerchberg-Saxton (GS) procedure (Gerchberg and Saxton, 1972). Given the spectra for the LCT domains $M_0,\ldots,M_k$, it can be characterized by the following pseudo-code:

**Algorithm 1**: Simple GS algorithm for the linear canonical transform

**Input**: Spectra $|\mathcal{L}(M_0)[x]|,\ldots,|\mathcal{L}(M_k)[x]|$
$y = |\mathcal{L}(M_0)[x]|$;
$y_{\text{old}} = -\infty$; *% Initial vector with entries* $-\infty$
**while** $\|y - y_{old}\| > \varepsilon$ **do**
  **for** $i = 0 : k$ **do**

$$y_{\text{old}} = y$$

$$y = |\mathcal{L}(M_{\text{mod}(i+1,k+1)})[x]| \frac{\mathcal{L}(M_{\text{mod}(i+1,k+1)} M_i^{-1})[y]}{|\mathcal{L}(M_{\text{mod}(i+1,k+1)} M_i^{-1})[y]|}$$

  **end**
**end**

According to the properties of the LCT $\mathcal{L}(M_{i+1}M_i^{-1})$ transforms a signal from the domain $M_i$ into the domain belonging to $M_{i+1}$. The modulo operation $\text{mod}(i+1,k+1)$ just ensures a cyclic run through all parameters.

### 3.3.3 HIGHER-ORDER WIGNER MOMENTS

Given a one-dimensional signal $x(t) = |x(t)|\exp(\mathbb{i}\varphi(t))$ the instantenous frequency property of the WVS (9) is given by:

$$\begin{aligned}\int f W_x(t,f)df &= \mathrm{Im}(x'(t)x^*(t)) = \mathrm{Im}\Big(\big(|x(t)|'e^{\mathbb{i}\varphi(t)} + |x(t)|e^{\mathbb{i}\varphi(t)}\mathbb{i}\varphi'(t)\big)|x(t)|e^{-\mathbb{i}\varphi(t)}\Big)\\ &= |x(t)|^2\varphi'(t).\end{aligned}$$

This link between the derivative of the phase of a signal and the first moment, with respect to frequency, of it is Wigner distribution can be applied to Equation (17). It is thus possible to compute the missing phase information from the relation:

$$\begin{aligned}|\mathcal{L}(M)[x_i(t)]|^2\varphi'_{L_T^C(M)[x_i]}(t) &= \int f W_{\mathcal{L}(M)[x_i]}(t,f)df = \sum_{j=1}^{n}|\alpha_{ij}|^2\int f W_{\mathcal{L}[M](T_{ij}s_j)}(t,f)df\\ = \sum_{j=1}^{n}\underbrace{|\alpha_{ij}|^2|\mathcal{L}(M)[s_j(t-A\tau_{ij})]|^2}_{\text{known from (20)}}&\big(\varphi'_{\mathcal{L}(M)[s_j]}(t-A\tau_{ij}) + 2\pi C\tau_{ij}\big).\end{aligned} \tag{33}$$

Using the estimators from Equation (20) in relation (33), allows the computation of the phase derivatives $\varphi'_{\mathcal{L}(M)[s_j]}$. Thus by integration, the missing phase information can be theoretically recovered up to the natural ambiguity of an arbitrary timeshift of the sources. In practice, this approach has the disadvantage that only the product of the power spectrum and the phase derivative can be computed. Zeros in the power spectrum thus prevent a direct integration of the phase information. However, this approach is feasible if it is combined with the GS algorithm (presented in 3.3.2). For this purpose, not only the correct power-spectra are enforced during each step of the GS algorithm, but in addition also also the constraint given by (33).

### 3.3.4 PHASE RETRIEVAL FROM TWO SIMILAR LCT SPECTRA

In theory, the Gerchberg-Saxton algorithm is suitable for computing the phase from arbitrary two LCT power spectra. In practice however (Cong et al., 1998), the speed of convergence highly depends on the distance between the power spectra specified by different values of the matrix $M$. For spectra defined by very similar matrices the GS procedure might fail. In this case, the following procedure permits to recover the phase from two very similar power spectra.

For simplicity consider again the one-dimensional signal case. This makes it possible to reformulate the instantaneous frequency property of the WVS:

$$\begin{aligned}\mathcal{F}|\mathcal{L}(M)[x(t)]|^2(\omega) &= \int|\mathcal{L}(M)[x(t)]|^2e^{-2\pi\mathbb{i}\omega t}dt \overset{(10)}{=} \iint W_x\big((t,f)(M^T)^{-1}\big)df\,e^{-2\pi\mathbb{i}\omega t}dt\\ &= \iint W_x(u,v)e^{-2\pi\mathbb{i}\omega(Au+Bv)}dudv =: \mathcal{A}_x(A\omega,B\omega).\end{aligned} \tag{34}$$

The two-dimensional Fourier transform of the Wigner distribution $\mathcal{A}_x(A\omega,B\omega)$ is called the Ambiguity-function (Mecklenbruker and Hlawatsch, 1997). Due to the properties of the Fourier transform we obtain:

$$t^k f^l W_x(t,f) = \left(\frac{1}{2\pi\mathbb{i}}\right)^{k+l}\mathcal{F}\left[\frac{\partial^{k+l}\mathcal{A}_x(\alpha,\beta)}{\partial^k\alpha\,\partial^l\beta}\right](t,f).$$

Now the instantaneous frequency property can be formulated as follows:

$$\begin{aligned}
|x(r)|^2\varphi_x'(r) &= \int f W_x(r,f)\,df = \int \frac{1}{2\pi\mathbb{i}}\mathcal{F}\left[\frac{\partial \mathcal{A}_x(\alpha,\beta)}{\partial\beta}\right](r,f)e^{-2\pi\mathbb{i}0f}\,df \\
&= \frac{1}{2\pi\mathbb{i}}\int\left(\frac{\partial \mathcal{A}_x(\tau,u)}{\partial u}\right)_{u=0} e^{-2\pi\mathbb{i}r\tau}\,d\tau \\
&\overset{\tau=A\omega}{=} \frac{1}{2\pi A\mathbb{i}}\int\left(\frac{\partial \mathcal{A}_x(A\omega,u)}{\partial u}\right)_{u=0} e^{-2\pi\mathbb{i}rA\omega}\,d\omega \\
&= \frac{1}{2\pi A\mathbb{i}}\int\left(\frac{\partial \mathcal{A}_x(A\omega,B\omega)}{\partial B}\frac{1}{\omega}\right)_{B=0} e^{-2\pi\mathbb{i}rA\omega}\,d\omega \\
&\overset{(34)}{=} \frac{1}{2\pi A\mathbb{i}}\iint\left[\frac{\partial|\mathcal{L}(M)[x(t)]|^2}{\partial B}\right]_{B=0}\frac{1}{\omega}e^{-2\pi\mathbb{i}\omega t}\,dt\,e^{-2\pi\mathbb{i}rA\omega}\,d\omega \\
&= \int\left[\frac{\partial|\mathcal{L}(M)[x(t)]|^2}{\partial B}\right]_{B=0}\left(\frac{1}{2\pi A\mathbb{i}}\int\frac{1}{\omega}e^{-2\pi\mathbb{i}(\omega t+rA\omega)}\,d\omega\right)dt \\
&= \frac{1}{2}\int\left[\frac{\partial|\mathcal{L}(M)[x(t)]|^2}{\partial B}\right]_{B=0}\operatorname{sgn}\left(\frac{t}{A}+r\right)dt. \qquad (35)
\end{aligned}$$

Equation (35) links the derivative of LCT intensities $|\mathcal{L}(M)[x]|^2$ with the instantaneous frequency of the signal. Given two close intensities the derivative $\left(\partial|\mathcal{L}(M)[x(t)]|^2/\partial B\right)_{B=0}$ can be approximated by a difference quotient, so that the local frequency can be recovered (Bastiaans and Wolf, 2003).

## 4. Example Algorithms

Summarizing the theoretical considerations from Section 3.1, we conclude that with the assumption of uncorrelated sources general the anechoic mixture problem for mixtures of the form

$$x_i(t) = \sum_{j=1}^{n}\alpha_{ij}s_j(t-\tau_{ij}) \quad i = 1,\cdots,m$$

can be solved by the following two step procedure:

**Algorithm 2**: Generic algorithm for anechoic demixing based on Wigner-marginals

**Input**: Observed data $x_i$ , $i = 1,\ldots,m$, LCT domains $M_k$ , $k = 1,\ldots,K$

**Step 1:** Solve the positive anechoic demixing problems given by:

$$|\mathcal{L}(M_k)[x_j]|^2(t) = \sum_{j}^{n}|\alpha_{ij}|^2|\mathcal{L}(M_k)[s_j](t-A\tau_{ij})|^2$$

for example, using the ANMF algorithm (1) or the bayesian Probabilistic Latent Component Analysis (PLCA) algorithm (Smaragdis et al., 2007).

**Step 2:** Recover the phase information for $s_j$ by one of the methods discussed in Section 3.3.

The general algorithm 2 is quite flexible, and it can be implemented by many choices for the LCT and the phase retrieval method. To illustrate this generality we discuss in the following three example implementations that cover a wide spectrum of applications.

### 4.1 Example 1: Time Series Demixing using Higher-order Wigner Moments

The choice $M = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ reduces the LCT to the standard Fourier transform. With $A = 0$ this simplifies the first step in algorithm 2, reducing the positive anechoic mixture to an instantenous mixture problem:

$$|\mathcal{F}x_i|^2(f) = \sum_j^n |\alpha|_{ij}^2 |\mathcal{F}s_j|^2(f).$$

In this case the phase retrieval (step 2) can be implemented by an alternating least squares approach that is given by the following iteration:

---

**for** *iter=1:maxiter* **do**

Compute the instantaneous phase $\varphi'_{(\mathcal{F}s_j)}$ by solving numerically the following equation:

$$|\mathcal{F}x_i(f)|^2 \varphi'_{(\mathcal{F}x_i)}(f) = \sum_j^n |\alpha|_{ij}^2 |\mathcal{F}s_j(\xi)|^2 \left(\varphi'_{(\mathcal{F}s_j)}(f) - 2\pi\tau_{ij}\right)$$

Update the delays $\tau_{ij}$ using an arbitrary delay estimator (e.g., Swindelhurst, 1998.).

**end**

---

### 4.2 Example 2: Multivariate Demixing using Fractional Fourier Transform

In the case of multivariate data (e.g., images) the simplest choice for a linear canonical transform is just the tensorial product of one dimensional LCTs, corresponding to the choice of diagonal matrices for $A, B, C, D$. This has the advantage that by selection of $\mathrm{diag}(A) = (a_{ii})_i$ the multivariate positive demixing problem simplifies to a set of one-dimensional problems.

As an illustrative example we discuss the demixing of two-dimensional image data $x_i(t_1, t_2)$, where $t_1$ and $t_2$ are the pixel coordinates. A simple example for a tensorial linear canonical transform are the fractional Fourier transforms $\mathcal{F}^{(\theta_1,\theta_2)} = \mathcal{F}_1^{\theta_1}\mathcal{F}_2^{\theta_2}$ . In this notation $\mathcal{F}_i^{\theta}$ represents the fractional Fourier transform $\mathcal{F}^{\theta}$ (as defined by (13)) in the $i$th variable (see (13) or Ozaktas et al., 2001). The system of equations (20) then reduces to:

$$|\mathcal{F}^{(\pi/2,\theta)}x_i(\omega_1,\omega_2)|^2 = \sum_{j=1}^n |\alpha|_{ij}^2 |\mathcal{F}^{(\pi/2,\theta)}s_j\big((\omega_1,\omega_2) - (0, \tau_{ij_2}\cos\theta)\big)|^2,$$

$$|\mathcal{F}^{(\theta,\pi/2)}x_i(\omega_1,\omega_2)|^2 = \sum_{j=1}^n |\alpha|_{ij}^2 |\mathcal{F}^{(\theta,\pi/2)}s_j\big((\omega_1,\omega_2) - (\tau_{ij_1}\cos\theta, 0)\big)|^2.$$

Note that for multivariate data the shifts $\tau_{ij}$ are vectors, denoted here with $\tau_{ij} = (\tau_{ij_1}, \tau_{ij_2})$. Since in this context the variables $\alpha_{ij}$ model intensities of basis images $s_j$, one can assume $\alpha_{ij} \geq 0$. Hence phase retrieval can be accomplished by a simple deconvolution procedure, as described in 3.3.1.

### 4.3 Example 3: LCT Based Positive Demixing of Time Series

For the univariate case a particularly attractive parameter choice for the linear canonical transform is $A = 1$ and $B = 1$. For $A = 1$ the values of the delays remain unchanged, and $B = 1$ has the

considerable numerical advantage that the integral transform (12) does not involve scaling. Thus this transformation just converts the original anechoic mixture problem into a nonnegative mixture problem that can be solved using the update rules given by (29) and (30).

## 5. Applications

In its generic form algorithm 2 has a very broad spectrum of applications. The inclusion of shifts into the mixing model makes it suitable for the shift-invariant learning of features or sources. This problem is relevant in a variety of areas. Furthermore, the theoretical results in Section 3.1 are valid independent of the dimensionality of the data and of the number of extracted sources. However, the ANMF algorithm is better suited for the over-determined case. We constrained the testing of our framework to dimension-reduction for 1D and 2D data, since these are the most frequent problems outside of acoustics. A version of the code that implements the basic version of the algorithm can be downloaded from the web page: `http://www.compsens.uni-tuebingen.de/pub/download/software/anmf`.

### 5.1 Sound Mixtures

Since the formulation of the "cocktail party problem" by Cherry (1953) acoustics has been a major field for the application of blind source separation methods. The cocktail party problem refers to the separation of individual human speakers in a noisy environment from a limited number of signals, recorded by microphones. In this context, many methods for the unmixing of sound signals have been developed over the years (e.g., Anthony and Sejnowski, 1995; Torkkola, 1996a). The most realistic linear models for BSS in acoustics are convolutive. However, anechoic mixtures are relevant for the case of reverberation-free environments (Bofill, 2003; Yilmaz and Rickard, 2004).

To demonstrate that algorithm 2 is applicable to sound mixtures we present three different separation problems for synthetically generated delayed mixtures (model (2)) using speech and sound segments from the ICA benchmark described in Cichocki and Amari (2002). In total the data set consisted of 14 signals with a length of 8000 time points. In order to obtain statistically representative results, data sets were recomputed 20 times with random selection of the source signals, and/or of the mixing and delay matrices. Three types of mixtures were generated:

(I) Mixtures of 2 source segments with random mixing and delay matrices ($2 \times 2$), each resulting in two simulated signals $x_{1,2}$.

(II) Mixtures of 2 randomly selected segments from the speech data basis using the constant mixing matrix $A$ and the constant delay matrix $T$:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 10 & 5 \\ 1 & 2 \\ 1 & 1 \end{pmatrix}, \; T = (\tau_{ij})_{ij} = \begin{pmatrix} 0 & 4000 \\ 2500 & 5000 \\ 100 & 200 \\ 1 & 1 \\ 500 & 333 \end{pmatrix}.$$

Data set (II) with fixed mixing and delay matrices was included since completely random generation sometimes produced degenerated anechoic mixtures (instantaneous mixtures or ill-conditioned mixing matrices).

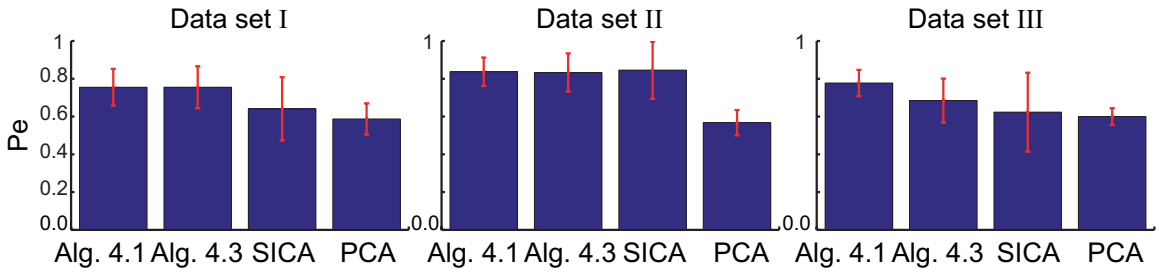



Figure 2: Comparison of different blind source separation algorithms for synthetic mixtures of sound signals with delays (data sets I-III, see text).

(III) A third data set was generated by mixing two randomly selected source segments with random mixture matrices and random delay matrices.

We compared the implementations of our algorithm described in section 4.1 and 4.2 with principal component analysis (as baseline) and shifted independent component analysis (SICA) (Mørup et al., 2007). The performance *Pe* of each algorithm was measured by the maximum of the cross-correlations between extracted sources $s_{\text{extract},j}$ and original sources $s_{\text{orig},j}$ (after appropriate matching of individual sources, since the recovered sources are not ordered in a specific manner):

$$Pe = (1/n)\sum_{j=1}^{n} \max_{\tau} \left| E\{s_{\text{approx.},j}(t)s_{\text{orig},j}(t+\tau)\} \right|.$$

#### 5.1.1 RESULTS FOR THE SOUND DEMIXING

The results for the sound demixing are summarized in Figure 2. The bar plots show the mean and the standard deviation of the performance measure *Pe* determined for the twenty simulations, comparing the methods for the data sets I-III. As expected, the performance *Pe* of the instantaneous mixture model assumed by principal component analysis (PCA) is worse than the performance achieved by the three compared anechoic methods. Overall the classical Fourier domain method 4.1 with $Pe \approx 80\%$ is superior to both the LCT-method 4.3 ($Pe \approx 75\%$) as well as SICA ($Pe \approx 70\%$). This reflects the well-known fact that the Fourier frequency domain is well suited for the separation of sound signals.

### 5.2 Human Motion Data

Characterizing manifolds that parameterize human motion is an important task for both, the analysis and the synthesis of movement trajectories. In analysis (Flash and Hochner, 2005) a popular idea is that such manifold representations reflects some lower dimensional space that results from the fact that the central nervous system reduces the exploited degree of freedoms by application of appropriate control strategies. A related important interpretation is that the set of source signals, that are appropriate for the reconstruction of human movement trajectories, or electromyography (EMG) signals, might reflect control units, synergies or movement primitives that are exploited by the central nervous system to simplify the underlying control problem by avoiding the curse of

dimensionality (Bellman, 1957). An overview of BSS methods in this field can be found in Tresch et al. (2006) or Chau (2001).

In synthesis the main purpose of an approximation of such a manifold is to represent a set of natural movements, in order to guarantee realistic looking animations. For this purpose many different methods have been suggested, ranging from linear or multilinear methods like PCA (Safonova et al., 2004) to nonlinear methods like isomap (Jenkins and Mataric, 2002), locally linear embedding (LLE) (Roweis and Saul, 2000), or methods based on space-time correspondence (Ilg et al., 2004).

Our second data set consists of human motion data. We show that anechoic mixtures have the potential for a strong dimensionality reduction. A first data set consisted of joint angle trajectories (Euler angles) that were computed from motion capture data (mocap), recorded from twenty-five lay-actors who executed walking with five basic emotional styles (neutral, happy, angry, sad and fear). A second data set consisted of the shoulder and elbow trajectories of five actors executing various arm movements (right-handed throwing, golf swing and tennis swing). All mocap data was recorded using a VICON 612 motion capture system with seven, respectively nine cameras. The system has a sampling frequency of 120 Hz and determines the three-dimensional positions of reflective markers (diameter: 1.25 cm) with spatial error below 1.5 mm. The markers were attached to skin or tight clothing with double-sided adhesive tape, according to the positions of VICON's Plug-In-Gait marker set. Commercial VICON software was used to reconstruct and label the markers, and to interpolate short missing parts of the trajectories.

### 5.2.1 RESULTS FOR THE MOTION DATA

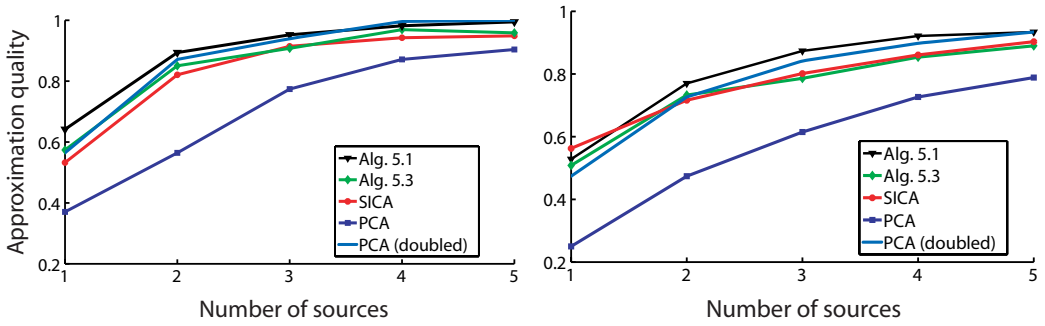
To quantify the performance of different methods for dimension reduction, we measured the quality of approximation as a function of the number of sources. The quality of an approximation $F$ to the data matrix $X$ can be quantified by the quotient:

$$Q := 1 - \frac{\|X - F\|}{\|X\|}$$

where $\|.\|$ is the Frobenius norm. This measure is related to explained variance, which is defined by $1 - \left(\|X-F\|/\|X\|\right)^2$, but it has the advantage of linear scaling with the residual-norm. For small residuals explained variance is hard to interpret since, due to the fact that it is quadratic in the residual norm it results in values close to one even for mediocre approximations. Figure 3 shows this measure for approximation quality as a function of the number of sources comparing principal component analysis (as baseline), shifted independent component analysis (SICA) (Mørup et al., 2007) and the example algorithms described in Sections 4.1 and 4.3.

The results of the comparison are shown in Figure 3. Overall, the best approximation is achieved by the Fourier-domain algorithm 4.1, exceeding the quality of PCA with the double number of sources (light blue line in Figure 3(a)). Slightly worse performance is obtained with algorithm 4.3, likely explained by the additional positivity constraint $\alpha_{ij} \geq 0$. Qualitatively the same results are obtained for the non periodic arm movements (figure 3(b)). The total approximation quality is lower in the second case, reflecting the higher variability of this data set.

In general, for both classes of movements a very accurate approximation of the trajectory sets can be achieved with very few ($\leq 5$) sources. This makes the method very interesting for the classification of movements (Omlor and Giese, 2007b), but also for the synthesis of realistic looking human movement data in computer graphics (Park et al., 2008; Mukovskiy et al., 2008).




(a) Approximation quality as a function of the number of extracted sources for periodic gait data.

(b) Approximation quality as a function of the number of extracted sources for non periodic arm movements.

Figure 3: Comparison of different blind source separation algorithms. PCA doubled refers to principle component analysis using twice as many sources.

### 5.3 Image Processing

Blind source separation is interesting for a wide range of applications in image processing. This includes watermarking (Bounkong et al., 2004), denoising (Hoyer and Oja, 2000), deblurring (Bar et al., 2005), or the extraction of image features (Lee and Seung, 1999; Draper et al., 2003). In most of these applications only instantenous models and algorithms have been applied.

With anechoic mixtures spatial displacements can be explicitly modeled (Omlor and Giese, 2007a; Be'ery and Yeredor, 2008). To demonstrate this, a first set of test images (Figure 4) was generated by pasting two objects at random positions in images with a resolution of $150\times150$ pixels. Ideally, algorithm 4.2 should be able to extract the original pictures of the two objects as features. The result of the feature extraction are shown in Figure 4 for both, the Gerchberg-Saxton phase retrieval method 3.3.2 and deconvolution 3.3.1.

Clearly, deconvolution is superior to the phase retrieval. This is partially due to the very slow convergence of the Gerchberg-Saxton algorithm, especially for small fractional powers of the fractional Fourier transform and for inaccurate estimates of the power spectra. In addition, the deconvolution method exploits, for a second time, the specific structure of the mixture model. A quantitative comparison shows that images predicted from the extracted components predict 95% of the variance of the original images for the deconvolution method, but only 72% for the phase retrieval method.

More interesting is the application of this feature extracting method to real images. Our second image data set consisted of four gray-scale images taken with a digital camera and resampled with a resolution of 200 $\times$ 200 pixels (cf. Figure 5). The photographs show two objects (scissors and a cup) that were placed at different positions on a wooden surface. Before the application of the algorithms the images were whitened (Gluckman, 2005) in order to compensate for the correlation statistics of natural images. This procedure removes strong correlations between features on small spatial scales. In this case only the deconvolution method was implemented. The reconstruction explains 85% of the of the pre-whitened training images and recovers the original objects with reasonable accuracy (see Figure 5).

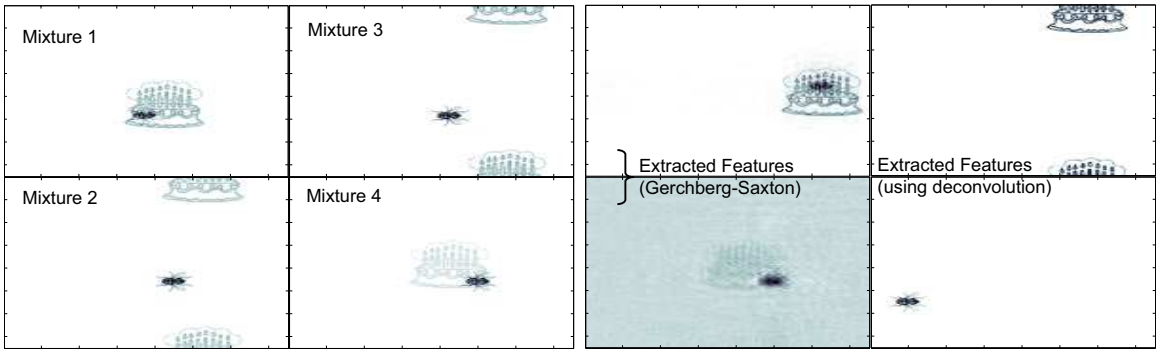



Figure 4: Left: synthetic example images defining an (over-determined) anechoic mixture in two dimensions. Right: extracted features from the image set on the left using GR phase retrieval or deconvolution.

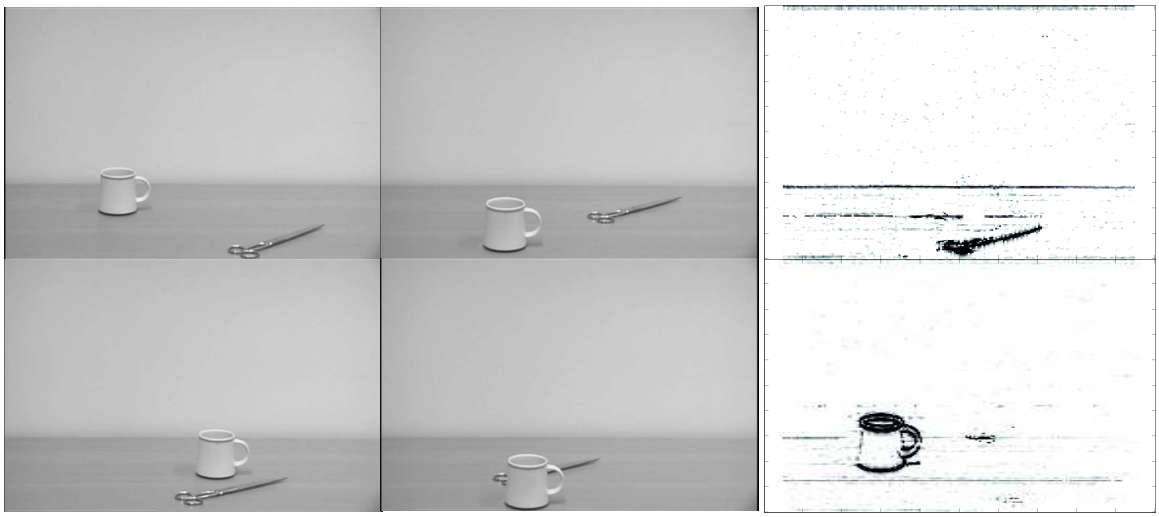


Figure 5: Left: Real Images. Right: Extracted Features from pre-whitened images

### 5.4 Scale and Rotation Invariant Shape Separation

Automatic identification of image content is an important challenge for many applications such as medical imaging, surveillance or the search in image databases. Often the shape of objects in images has to be recognized independently of object size or orientation. Many supervised and unsupervised learning approaches have been proposed for the classification of images using object shapes (see, e.g., Veltkamp and Hagedoorn, 1999; Mohanty et al., 2005; Ling and Jacobs, 2007; Ahmad and Ibrahim, 2006). Essential for the performance of such algorithms is the underlying shape representation (Zhang and Lu, 2004).

Since an object should be identifiable independent of it is relative size and orientation, scaling and rotation are desired invariances for two-dimensional contour recognition. In order to apply the anechoic mixture model (2) to this problem, a planar contour can be transformed into a log-polar image. Assuming that the center of the coordinate system is given by $[t_1, t_2] = [0, 0]$, the nonlinear transformation is given by:

$$\rho = \log\left(\sqrt{t_1^2 + t_2^2}\right),$$
$$\vartheta = \arctan\left(\frac{t_2}{t_1}\right).$$

Translations in the new coordinates $(\rho, \vartheta)$ permit to model scalings and rotations in the original coordinates. A shift of the angle $\vartheta$ corresponds to a rotation of the object in the image plane. Shifts of the logarithmic radius $\rho$ can be used to model objects with different sizes. This is just one example for a transformation. In the general case, the independent variable $t \in \mathbb{R}^n$ can be replaced by the nonlinearly transformed coordinates $\phi(t)$. The mixture model (2) then transforms into:

$$\tilde{x}_i(t) = x_i(\phi(t)) = \sum_{j=1}^{n} \alpha_{ij} s_j(\phi(t - \tau_{ij})) = \sum_{j=1}^{n} \alpha_{ij} \tilde{s}_j(t - \tau_{ij}) \quad i = 1, \cdots, m.$$

Thus a shift in the new independent variable $t$ corresponds to a transformation of the source functions $\tilde{s}_j(.) = s_j(\phi(.))$, that depends on the coordinate change $\phi$.

#### 5.4.1 APPLICATION TO TWO-DIMENSIONAL SHAPES

To illustrate this application of anechoic mixtures to model rotation and scale invariance, we tested the scale and rotation invariant learning on a small sample of shapes taken from the MPEG-7 test database (Sebastian et al., 2001) (depicted in Figure 6a). After the transformation into log-polar coordinates (with coordinate centers chosen as the xy-mean of the shape), algorithm 4.2 was applied for feature extraction.

Both the extracted features (figure 6b) as well as the corresponding weights $\alpha_{ij}$ (figure 6c) show that in principle, the anechoic mixing model is capable of correctly classifying different objects, independent of their size and rotation. Due to the fact that the one-dimensional contours are treated as 2D images, this approach has a high computational cost, limiting its applicability to large databases. Besides complex objects would need more than one source for an adequate description. This is obvious for the example of the elephant in Figure 6b. An increase in the necessary number of sources would result in decrease of classification performance. There are better approaches for the parametrization of line shapes, such as level sets (Osher and Fedkiw, 2002), which might improve the results of our algorithm in this case.

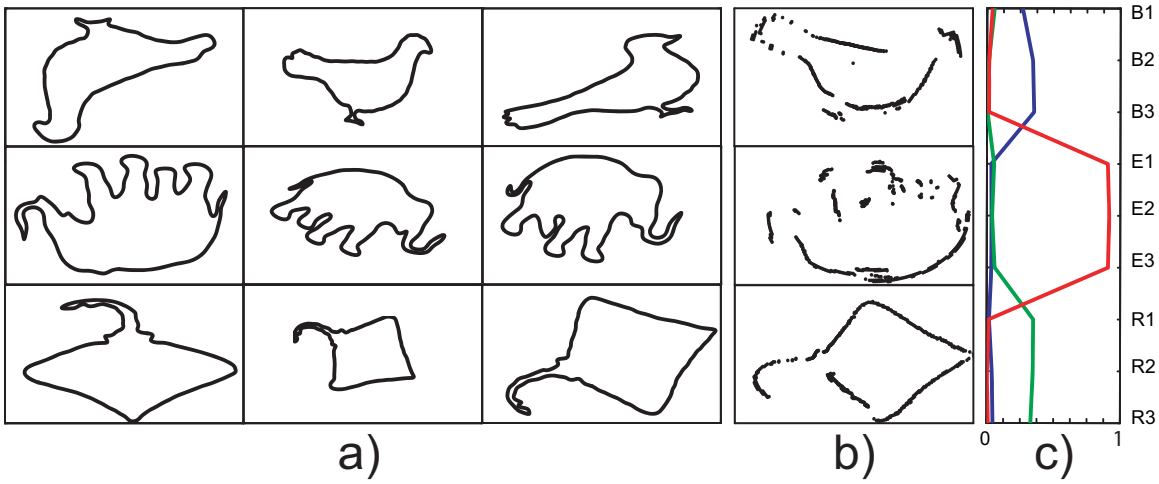



Figure 6: a) Sample object contours taken from the MPEG-7 database. b) Extracted features using algorithm 4.2. c) Amplitude normalized weights $\alpha_{ij}$ corresponding to the three shapes Bird (B1-3), Elephant (E1-3) and Ray (R1-3).

## 6. Conclusions

We have presented a novel class of algorithms for the solution of over- and under-determined anechoic mixture problems, which extend to an arbitrary number of dimensions of the time argument. The developed method exploits the marginal properties of the stochastic Wigner-Ville distribution. Proper application of this bilinear time-frequency distribution to the delayed mixture model transforms the anechoic problem into a simpler delayed mixture problem in the domain of a linear canonical transform and a phase recovery problem. Appropriate choice of this transformation enhances the separability of the signals and allows the projection of high-dimensional problems onto a system of one-dimensional problems. This results in implementations with a computational complexity that grows linearly in the number of dimensions.

The efficiency of this approach was demonstrated by a series of applications including both synthetic and real world data, such as music streams, natural 2D images, human motion trajectories as well as two dimensional shapes. These examples represent only a small subset of the many possible applications for the learning of invariant features. Due to its modularity, the proposed framework is very flexible and can be easily adjusted and optimized for special applications.

In order to optimize the computational performance of the proposed approach, future work will investigate the use of marginals of time-frequency distributions transcending the set of distributions from the convenient Cohen class. Another important direction is the inclusion of additional constraints in the optimization problem, for example, for the weights of the time delays. Such constraints will allow to model certain topological structures which are given by the data. One particular example for this are the neighborhood relations of certain muscles (d'Avella and Bizzi, 2005), for example, in the context of face movements. In addition, some aspects of the discussed time-frequency framework may be applicable to full convolutive mixtures or mixture models with other invariance properties. Finally, some steps of the proposed algorithms seems to be well-suited

for a fast implementation on Graphics Processing Units (GPUs) (David B. Kirk and Hwu, 2010). An example are the discussed variants of NMF algorithms.

## Acknowledgments

Supported by EC FP6 project COBOL, the EC FP7 projects SEARISE, AMARSI and TANGO, the Volkswagenstiftung, DFG Forschergruppe Perceptual Graphics, and the Human Frontiers Science Program (HFSP). Additional support was provided by the Hermann and Lilly-Schilling-Stiftung. We thank C.L. Roether for help with the trajectory acquisition and W. Ilg for support with the motion capturing.

## Appendix A. Delay Estimation

Time delay estimation (TDE) problems are common in many technical applications like telecommunications (Takeuchi et al., 1990), radar (Raja Rajeswari and Rani, 1998), sonar (Carter, 1981) or seismology (Du et al., 2004). The corruption of the signal with noise, degradation of the signal shape, moving signal sources, and the multiple overlaps of reflected copies of the signals (multipath), make TDE a very challenging problem. Consequently, a variety of algorithmic approaches have been developed (see Chen et al., 2006 and Bjorklund and Ljung, 2003 for review). If the observed signal $x(t)$ is the superposition of multiple delayed signals $s_j$:

$$x(t) = \sum_{j=1}^{n} \alpha_j s_j(t - \tau_{ij}) + n(t).$$

then (in case of gaussian noise $n$) the maximum likelihood (ML) delay estimator takes the form of the least squares problem:

$$\arg \min_{\alpha_j, (D_j)_j} E \left\{ \|x(t) - \sum_{j=1}^{n} \alpha_j s_j(t - D_j))\|^2 \right\}. \tag{36}$$

Though the ML estimator is statistically efficient (i.e., achieves the Cramér Rao bound (CRB)) it requires in general a $n$-dimensional search and is thus computationally inefficient (Swindelhurst, 1998.). If the signals are uncorrelated $r_{s_i,s_j}(t,t') = 0$ then the ML estimation simplifies to $n$ 1-dimensional estimation problems ($r_{x,s_j}(t)$ specifying the cross-correlation function of the signals $x$ and $s_j$):

$$\arg \min_{(D_j)_j} E \left\{ \|x(t) - \sum_{j=1}^{n} \alpha_j s_j(t - D_j))\|^2 \right\} = \left( \arg \max_{D_j} r_{x,s_j}(D_j) \right)_j. \tag{37}$$

In this case the corresponding optimal weights $\alpha_j$ can be obtained by linear regression. Though computational very efficient, this approximation is only accurate if cross-talk between the signals is minimal. It is straight-forward, to change both estimators to include the constraint $\alpha_j \geq 0$.

For the solution of the delay estimation problem in the anechoic NMF algorithm (instead of Equation (29)), the ML-cost function (36) was minimized by a nonlinear optimization based on the nonlinear Gauss-Seidel algorithm.

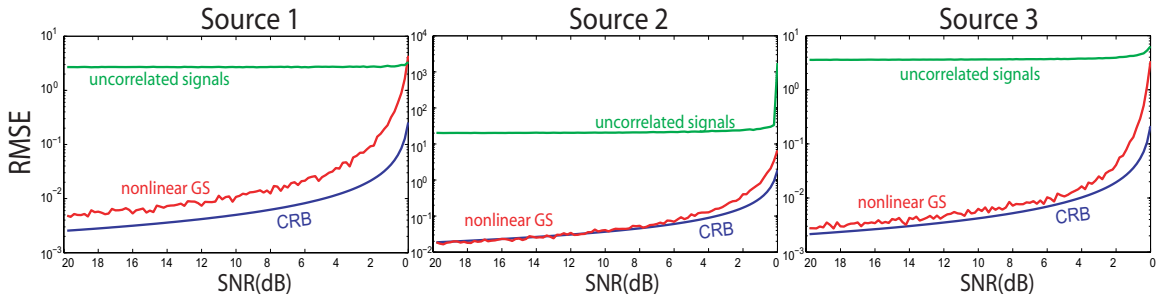



Figure 7: Comparison of different multi-source time-delay estimation methods. The red line shows the RMSE of the nonlinear Gauss-Seidel algorithm, the blue line indicates the Cramér Rao bound (CRB) and the green line shows the performance of the simple estimator (37).

The performance of this estimator was tested in several simulations, and compared to the Cramér Rao bound. The test sources were part of the music data-base used for the sound mixtures described in Section 5.1. The delays were estimated from mixtures with three randomly selected sources and constant weight $A$ and delay matrices $T = (\tau_{ij})_{ij}$:

$$A = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \; T = (\tau_{ij})_{ij} = \begin{pmatrix} 20.2 \\ 7.9 \\ 120.2 \end{pmatrix}.$$

Different amounts of noise were added in order to vary the signal to noise ratio (SNR). The root mean square error $\sqrt{E\{(\hat{\tau}_{ij} - \tau_{ij})^2\}}$, which is compared with the Cramér Rao bound, is based on 1000 simulated trials.

The results for the different estimators can be found in Figure 7. Clearly, the nonlinear Gauss-Seidel iteration outperforms the simple estimator (37). In addition, the Gauss-Seidel iteration results in estimation errors close to the Cramér Rao bound. This implies that the described method for phase retrieval performs close to the theoretically possible optimum.

## References

I. Ahmad and M. T. Ibrahim. Image classification and retrieval using correlation. In *CRV '06: Proceedings of the The 3rd Canadian Conference on Computer and Robot Vision*, page 60, Washington, DC, USA, 2006. IEEE Computer Society.

F. Aires, A. Chedin, and J.-P. Nadal. Independent component analysis of multivariate time series: Application to the tropical sst variability. *Journal of Geophysical Research*, 105:1743717455, 2000.

M.S. Akbar and L.J. Douglas. Optimal kernels for nonstationary spectral estimation. *IEEE Transactions on Signal Processing*, 43(2):478–491, 1995.

T. Alieva and M. J. Bastiaans. Properties of the linear canonical integral transformation. *Journal of the Optical Society America A*, 24(11):3658–3665, 2007.

J. B. Anthony and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

S. Arberet, R. Gribonval, and F. Bimbot. A robust method to count and locate audio sources in a stereophonic linear anechoic mixture. In *IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP 2007.*, volume 3, pages III–745 –III–748, 15-20 2007.

L. Bar, N. Sochen, and N. Kiryati. Image deblurring in the presence of salt-and-pepper noise. *Scale Space and PDE Methods in Computer Vision*, pages 107–118, 2005.

R.G. Baraniuk. Wigner-ville spectrum estimation via wavelet soft-thresholding. In *Time-Frequency and Time-Scale Analysis, 1994., Proceedings of the IEEE-SP International Symposium on*, pages 452 –455, October 1994. doi: 10.1109/TFSA.1994.467316.

A. Barliya, L. Omlor, M.A. Giese, and T. Flash. An analytical formulation of the law of intersegmental coordination during human locomotion. *Experimental Brain Research*, 193(3):371–385, 2009.

M. J. Bastiaans and K. B. Wolf. Phase reconstruction from intensity measurements in linear systems. *Journal of the Optical Society America A*, 20(6):1046–1049, 2003.

E. Be'ery and A. Yeredor. Blind separation of superimposed shifted images using parameterized joint diagonalization. *IEEE Transactions on Image Processing*, 17(3):340–353, 2008.

R. E. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ., 1957.

A. Belouchrani and M.G. Amin. Blind source separation based on time-frequency signal representations. *IEEE Transactions on Signal Processing*, 46(11):2888–2897, 1998.

S. Bjorklund and L. Ljung. A review of time-delay estimation techniques. In *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, volume 3, pages 2502 – 2507, 2003. doi: 10.1109/CDC.2003.1272997.

P. Bofill. Underdetermined blind separation of delayed sound sources in the frequency domain. *Neurocomputing*, 55:627–641, 2003.

S. Bounkong, B. Toch, D. Saad, and D. Lowe. Ica for watermarking digital images. *Journal of Machine Learning Research*, 4(7-8):1471–1498, 2004.

A. Bultheel and H. Martínez-Sulbaran. Recent developments in the theory of the fractional Fourier transforms and linear canonical transforms. *Bulletin of the Belgian Mathematical Society- Simon Stevin*, 13:971–1005, 2006.

J.F. Cardoso. Blind singal separation: statistical principles. *Proceedings of the IEEE*, 9(10):2009–2025, 1998.

G. Carter. Time delay estimation for passive sonar signal processing. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 29(3):463–470, 1981.

T. Chau. A review of analytical techniques for gait data. part 1: fuzzy, statistical and fractal methods. *Gait & Posture*, 13(1):49–66, 2001.

J. Chen and X.Z. Wang. A new approach to near-infrared spectral data analysis using independent component analysis. *Journal of Chemical Information and Computer Sciences*, 41(4):992–1001, 2001.

J. Chen, J. Benesty, and Y. Huang. Time delay estimation in room acoustic environments: an overview. *EURASIP Journal of Applied Signal Processing*, pages 170–170, 2006.

Z. Chen and A. Cichocki. Nonnegative matrix factorization with temporal smoothness and/or spatial decorrelation constraints. In *Laboratory for Advanced Brain Signal Processing, RIKEN, Tech. Rep*, 2005.

E. C. Cherry. Some experiments on the recognition of speech, with one and with two ears. *Journal of Acoustical Society of America*, 25(5):975–979, 1953.

S. Choi, A. Cichocki, H.H. Park, and S.Y. Lee. Blind signal separation and independent component analysis: A review. *Neural Information Processing – Letters and Reviews*, 5:1–57, 2005.

A. Cichocki and S. Amari. Adaptive blind signal and image processing. *John Wiley, Chichester*, 2002.

A. Cichocki, H. Lee, Y.D. Kim, and S. Choi. Non-negative matrix factorization with alpha-divergence. *Pattern Recognition Letters*, 29(9):1433–1440, 2008.

L. Cohen. Time-frequency distributions-a review. *Proceedings of the IEEE*, 77(7):941–981, 1989.

P. Comon and C. Jutten. *Handbook of Blind Source Separation, Independent Component Analysis and Applications*. Academic Press, 2010.

W. X. Cong, N. X. Chen, and B. Y. Gu. Recursive algorithm for phase retrieval in the fractional fourier transform domain. *Applied Optics*, 37(29):6906–6910, 1998.

A. d'Avella and E. Bizzi. Shared and specific muscle synergies in natural motor behaviors. *Proc Natl Acad Sci U S A*, 102(8):3076–81, 2005.

A. d'Avella, L. Fernandez, A. Portone, and F. Lacquaniti. Modulation of Phasic and Tonic Muscle Synergies With Reaching Direction and Speed. *J Neurophysiol*, 100(3):1433–1454, 2008.

D.B. David B. Kirk and W. W. Hwu. *Programming Massively Parallel Processors: A Hands-on Approach*. Morgan Kaufman Publishers, San Fransisco, CA, 2010.

C. Ding, T. Li, W. Peng, and H. Park. Orthogonal nonnegative matrix tri-factorizations for clustering. In *Proceedings of the Twelfth ACM SIGKDD international conference on knowledge discovery and data mining*, pages 126–135, 2006.

B.A. Draper, K. Baek, M.S. Bartlett, and J.R. Beveridge. Recognizing faces with PCA and ICA. *Computer Vision and Image Understanding*, 91(1-2):115–137, 2003.

W.X. Du, C.H. Thurber, and D. Eberhart-Phillips. Earthquake relocation using cross-correlation time delay estimates verified with the bispectrum method. *Bulletin of the Seismological Society of America*, 94(3):856–866, 2004.

B. Emile and P. Comon. Estimation of time delays between unknown colored signals. *Signal Processing*, 69(1):93–100, 1998.

T. Flash and B. Hochner. Motor primitives in vertebrates and invertebrates. *Current Opinion in Neurobiology*, 15 (6):660–666, 2005.

P. Georgiev, F. Theis, and A. Cichocki. Sparse component analysis and blind source separation of underdetermined mixtures. *IEEE Transactions on Neural Networks*, 16(4):992 –996, 2005.

R. W. Gerchberg and W. O. Saxton. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237–246, 1972.

J. Gluckman. Higher order whitening of natural images. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2:354–360, 2005.

R. A. Harshman, S. Hong, and M. E. Lundy. Shifted factor analysis–part i: Models and properties. *Journal of Chemometrics*, 17:363–378, 2003.

J.J. Healy, P. O'Grady, and J.T. Sheridan. Simulating paraxial optical systems using the linear canonical transform: properties, issues and applications. *Optics and Photonics for Information Processing II*, 7072(1):70720E, 2008.

F. Hlawatsch and W. Kozek. The Wigner distribution of a linear signal space. *IEEE Transactions on Signal Processing*, 41(3):1248–1258, 1993.

P. Hoyer and E. Oja. Image denoising by sparse code shrinkage. In *Intelligent Signal Processing*. IEEE Press, 2000.

W. Hu and L.M. Collins. Classification of closely spaced subsurface objects using electromagnetic induction data and blind source separation algorithms. *Radio Science*, 39:1–13, 2004.

W. Ilg, G. H. Bakir, J. Mezger, and M.A. Giese. On the representation, learning and transfer of spatio-temporal movement characteristics. *International Journal of Humanoid Robotics*, 1:613–636, 2004.

A. J. E. M. Janssen. *Application of the Wigner Distribution to Harmonic Analysis of Generalized Stochastic Processes*. Mathematical Centre Tracts 114, Amsterdam, 1979.

O. C. Jenkins and M. J. Mataric. Deriving action and behavior primitives from human motion data. *Intelligent Robots and System, 2002. IEEE/RSJ International Conference on*, 3:2551–2556, 2002.

S. Karako-Eilon, A. Yeredor, and D. Mendlovic. Blind source separation based on the fractional fourier transform. *Proc. 4th Int. Symp. on Independent Component Analysis*, 46(11):615–620, 2003.

D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

C.-T. Leung and W.-C. Siu. A general contrast function based blind source separation method for convolutively mixed independent sources. *Signal Processing*, 87(1):107–123, 2007.

Y. Li, A. Cichocki, and S. Amari. Analysis of sparse representation and blind source separation. *Neural Computation*, 16(6):1193–1234, 2004.

Y. Lin and D.D. Lee. Bayesian regularization and nonnegative deconvolution for time delay estimation. In *in Advances in Neural Information Processing Systems 17*, pages 809–816. MIT Press, 2005.

H. Ling and D.W. Jacobs. Shape classification using the inner-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):286–299, 2007.

W. Martin. Time-frequency analysis of random signals. In *Proc. IEEE ICASSP*, pages 1325–1328, Paris, 1982.

W. Martin and P. Flandrin. Wigner-Ville spectral analysis of nonstationary processes. *IEEE Transactions on Acoustics, Speech, and. Signal Processing*, 33(6):1461–1470, 1985.

K. Matsuoka. Minimal distortion principle for blind source separation. In *SICE 2002. Proceedings of the 41st SICE Annual Conference*, volume 4, pages 2138 – 2143, 2002. doi: 10.1109/SICE.2002.1195729.

G. Matz and F. Hlawatsch. Wigner distributions (nearly) everywhere: time-frequency analysis of signals, systems, random processes, signal spaces, and frames. *Signal Processing*, 83(7):1355–1378, 2003.

W. Mecklenbruker and F. Hlawatsch. *The Wigner Distribution*. Elsevier Science, 1997.

N. Mohanty, T. Rath, A. Lee, and R. Manmatha. Learning shapes for image classification and retrieval. In W.-K. Leow, M. Lew, T. S. Chua, W.-Y. Ma, L. Chaisorn, and E. Bakker, editors, *Image and Video Retrieval*, volume 3568 of *Lecture Notes in Computer Science*, pages 589–598. Springer Berlin / Heidelberg, 2005.

M. Mørup, K. H. Madsen, and L. K. Hansen. Shifted independent component analysis. In M. E. Davies, C. J. James, S. A. Abdallah, and M. D. Plumbley, editors, *ICA2007*, pages 89–96. Springer, Berlin, 2007.

M. Mørup, L.K. Hansen, S.M. Arnfred, L. Lim, and K.H. Madsen. Shift invariant multilinear decomposition of neuroimaging data. *NeuroImage*, 42(4):1439–50, 2008.

A. Mukovskiy, A. Park, L. Omlor, and M.A. Giese. Self-organization of character behavior by mixing of learned movement primitives. In *Proceedings of the 13th Fall Workshop on Vision, Modeling, and Visualization (VMV)*, Konstanz, Germany, 2008.

C. Namgook and C.-C. Jay Kuo. Underdetermined audio source separation from anechoic mixtures with long time delay. In *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, ICASSP '09, pages 1557–1560, Washington, DC, USA, 2009. IEEE Computer Society.

D. Nuzillard and A. Bijaoui. Blind source separation and analysis of multispectral astronomical images. *Astronomy and Astrophysics Suppl. Series*, 147:129–138, 2000.

L. Omlor and M. A. Giese. Learning of translation-invariant independent components: Multivariate anechoic mixtures. In M. E. Davies, C. J. James, S. A. Abdallah, and M. D. Plumbley, editors, *ICA 2007*, pages 762–769. Springer, Berlin, 2007a.

L. Omlor and M. A. Giese. Blind source separation for over-determined delayed mixtures. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 1049–1056. MIT Press, Cambridge, MA, 2007b.

S. J. Osher and R. P. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Springer, 2002.

H. M. Ozaktas, Z. Zalevsky, and M. A. Kutay. *The Fractional Fourier Transform with Applications in Optics and Signal Processing*. John Wiley & Sons, 2001.

P. O. Ogrady, B. A. Pearlmutter, and S. T. Rickard. Survey of sparse and non-sparse methods in source separation. *International Journal of Imaging Systems and Technology*, 15:2005, 2005.

A. Papoulis, P.S. Unnikrishna, and S. Unnikrishna. *Probability, Random Variables, and Stochastic Processes*. McGraw Hill Higher Education, 2001.

A. Park, A. Mukovskiy, L. Omlor, and M.A. Giese. Self organized character animation based on learned synergies from full-body motion capture data. In *International Conference on Cognitive Systems (CogSys)*, Karlsruhe, Germany, 2008.

M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra. A survey of convolutive blind source separation methods. In *Springer Handbook of Speech Processing*, pages 1065–1094. Springer Verlag, Berlin, 2007.

J.C. Platt and F. Faggin. Networks for the separation of sources that are superimposed and delayed. *Advances in Neural Information Processing Systems*, 4:730–737, 1992.

M.D. Plumbley and E. Oja. A ”nonnegative pca” algorithm for independent component analysis. *IEEE Transactions on Neural Networks*, 15(1):66–76, 2004.

L. R. Rabiner and B. Gold. *Theory and Application of Digital Signal P'rocessing*. Englewood Cliffs, NJ : Prentice-Hall, 1975.

K. Raja Rajeswari and D.E. Rani. Time-delay estimation using mle approach for wide-band radar systems. *Fourth International Conference on Signal Processing,Proceedings,ICSP '98.*, 2:1493–1496, 1998.

S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 5500(290):2268–2269, 2000.

A. Safonova, J. K. Hodgins, and N. S. Pollard. Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. *ACM Transactions on Graphics*, 23(3):514–521, 2004.

T. B. Sebastian, P. N. Klein, and B. B. Kimia. Recognition of shapes by editing shock graphs. In *Proceedings of ICCV*, pages 550 – 571, 2001.

K. Seki, M. Narusawa, and P. Smaragdis. Blind separation of convolved mixtures in the frequency domain - signal processing. *Neurocomputing*, 22(14):21–34, 1998.

K. K. Sharmaa and S. D. Joshi. Signal separation using linear canonical and fractional fourier transforms. *Optics Communications*, 265:454–460, 2006.

P. Smaragdis, B. Raj, and M. V. S. Shashanka. Supervised and semi-supervised separation of sounds from single-channel mixtures. In M. E Davies, C. J. James, S. A. Abdallah, and M. D. Plumbley, editors, *ICA*, volume 4666 of *Lecture Notes in Computer Science*, pages 414–421. Springer, 2007.

A. Swindelhurst. Time delay and spatial signature estimation using known asynchronous signals. *IEEE Transactions on Signal Processing*, ASSP-33,no. 6:1461–1470, 1998.

T. Takeuchi, M. Sako, and S. Yoshida. Multipath delay estimation for indoor wireless communication. *Vehicular Technology Conference, 1990 IEEE 40th*, pages 401–406, 1990.

K. Torkkola. Blind separation of convolved sources based on information maximization. In *In IEEE Workshop on Neural Networks for Signal Processing*, pages 423–432, 1996a.

K. Torkkola. Blind separation of delayed sources based on information maximization. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 6:3509–3512, 1996b.

M.C. Tresch, V.C. Cheung, and A. d'Avella. Matrix factorization algorithms for the identification of muscle synergies: evaluation on simulated and experimental data sets. *Journal of Neurophysiology*, 95:2199–2212, 2006.

R. Veltkamp and M. Hagedoorn. State-of-the-art in shape matching. Technical Report UU-CS-1999-27, Utrecht University, the Netherlands, 1999.

J. Ville. Théorie et applications de la notion de signal analytique. *Câbles et Trans.*, 2:61–74, 1948.

R. Vollgraf, M. Stetter, and K. Obermayer. Convolutive decorrelation procedures for blind source separation. In *In Proceedings of the 2. International Workshop on*, pages 515–520, 2000.

N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. MIT Press, Cambridge, MA, 1964.

E. Wigner. On the quantum correction for thermodynamic equilibrium. *Physical Review*, 40(5): 749–759, 1932.

J. Xiao and P. Flandrin. Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement. *IEEE Transactions on Signal Processing*, 55(6):2851–2860, 2007.

S. Yang and Z. Yi. Nonnegative matrix factorization for independent component analysis. In *International Conference on Communications, Circuits and Systems.*, pages 769–771, 2007.

A. Yeredor. Time-delay estimation in mixtures. *Acoustics, Speech, and Signal Processing*, 5:237–40, 2003.

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

Z. Zalevsky, D. Mendlovic, and R. G. Dorsch. Gerchberg-Saxton algorithm applied in the fractional Fourier or the Fresnel domain. *Optics Letters*, 21:842–844, 1996.

D. S. Zhang and G. Lu. Review of shape. representation and description techniques. *Pattern Recognition*, 37(1):1–19, 2004.