

RESEARCH

Open Access

Using information theoretic distance measures for solving the permutation problem of blind source separation of speech signals

Eugen Hoffmann^{*}, Dorothea Kolossa, Bert-Uwe Köhler and Reinhold Orglmeister

Abstract

The problem of blind source separation (BSS) of convolved acoustic signals is of great interest for many classes of applications. Due to the convolutive mixing process, the source separation is performed in the frequency domain, using independent component analysis (ICA). However, frequency domain BSS involves several major problems that must be solved. One of these is the permutation problem. The permutation ambiguity of ICA needs to be resolved so that each separated signal contains the frequency components of only one source signal. This article presents a class of methods for solving the permutation problem based on information theoretic distance measures. The proposed algorithms have been tested on different real-room speech mixtures with different reverberation times in conjunction with different ICA algorithms.

Keywords: blind source separation, independent component analysis, permutation problem

1 Introduction

Blind source separation (BSS) is a technique of recovering the source signals using only observed mixtures when both the mixing process and the sources are unknown. Due to a large number of applications for example in medical and speech signal processing, BSS has gained great attention. This article considers the case of BSS for acoustic signals observed in a real environment, i.e., convolutive mixtures, focusing on speech signals in particular. In recent years, the problem has been widely studied and a number of different approaches have been proposed [1,2]. Many state-of-the-art unmixing methods of acoustic signals are based on independent component analysis (ICA) in the frequency domain, where the convolutions of the source signals with the room impulse response are reduced to multiplications with the corresponding transfer functions. So for each frequency bin, an individual instantaneous ICA problem arises [2].

Due to the nature of ICA algorithms, obtaining a consistent ordering of the recovered signals is highly unlikely. In case of frequency domain source separation, this means that the ordering of outputs may change for each

frequency bin. In order to correctly estimate source signals in the time domain, all separated frequency bins need to be put in a consistent order. This problem is also known as the permutation problem.

There exist several classes of algorithms giving a solution for the permutation problem. Approaches presented in [3-6] try to find permutations by considering the cross statistics (such as cross correlation or cross cumulants etc.) of the spectral envelopes of adjacent frequency bins. In [7] algorithms were proposed, that make use of the spectral distance between neighboring bins and try to make the impulse response of the mixing filters short, which corresponds to smooth transfer functions of the mixing system in the frequency domain. The algorithm proposed by Kamata et al. [8] solves the problem using the continuity in power between adjacent frequency components of the same source. A similar method was presented by Pham et al. [9]. Baumann et al. [10] proposed a solution by comparing the directivity patterns resulting from the estimated demixing matrix in each frequency bin. Similar algorithms were presented in [11-13]. In [14] it was suggested to use the direction of arrival (DOA) of source signals, determined from the estimated mixing matrices, for the problem solution. The approach in [15] is to exploit the continuity of the frequency response of

^{*} Correspondence: eugen.hoffmann@tu-berlin.de
Berlin Institute of Technology, Chair of Electronics and Medical Signal Processing, Einsteinufer 17, 10587 Berlin, Germany

the mixing filter. A similar approach was presented in [16] using the minimum of the L_1 -norm of the resulting mixing filter and in [17] using the minimum distance between the adjacent filter coefficients. In [18] the authors suggest to use the cosine between the demixing coefficients of different frequencies as a cost function for the problem solution. Sawada et al. [19] proposed an approach based on basis vector clustering of the normalized estimated mixing matrices. In [20] a hybrid approach combines spectral continuity, temporal envelope and beamforming alignment with a psychoacoustic post-filter, and in [21] the permutation problem was solved using a maximum-likelihood-ratio between the adjacent frequency bins.

However with growing number of the independent components, the complexity of the solution grows. This is true not only because of the factorial increase of permutations to be considered, but also because of the degradation of the ICA performance. So not all of the approaches mentioned above perform equally well for an increasing number of sources.

The goal of this article is to investigate the usefulness of information theoretic distance measures for the solution of the permutation ambiguity problem. For this purpose it is assumed that the amplitudes of the estimated independent signals possess a Rayleigh distribution [22] and the logarithms of the amplitudes possess a generalized Gaussian distribution (GGD). It should be noted that the approach in [23] is based on a similar assumption, namely that the extracted signals are generalized Gaussian distributed. The authors handle the problem by comparing the parameters of the GGD of each frequency bin. However the resulting algorithm solves the permutation problem only partially and requires a combination with another approach, for instance [24].^a In contrast, the algorithms proposed in this article deal with the problem in a self-contained way and require no completion by other approaches.

The resulting approaches will be tested on different speech mixtures recorded in real environments with different reverberation times in combination with different ICA algorithms, such as JADE [25], INFOMAX [4,26], and FastICA [27,28].

2 Problem formulation

This section provides an introduction into the problem of blind separation of acoustic signals.

At first a general situation will be considered. In a reverberant (real) room, N acoustic signals $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]$ are simultaneously active (t represents the time index). The vector of the source signals $\mathbf{s}(t)$ is recorded with M microphones placed in the room, so that an observation vector $\mathbf{x}(t) = [x_1(t), \dots, x_M(t)]$ results. Due to

the time delay and to the signal reflections, the resulting mixture $\mathbf{x}(t)$ is a result of a convolution of the source signal $\mathbf{s}(t)$ with an unknown filter tensor $\underline{\mathbf{a}} = (\underline{\mathbf{a}}_1 \dots \underline{\mathbf{a}}_K)$ where $\underline{\mathbf{a}}_k$ is the k -th ($k \in [1 \dots K]$) $M \times N$ matrix with filter coefficients and K is the filter length. This problem can be summarized by

$$\mathbf{x}(t) = \sum_{k=0}^{K-1} \underline{\mathbf{a}}_{k+1} \mathbf{s}(t-k) + \mathbf{n}(t). \quad (1)$$

The term $\mathbf{n}(t)$ denotes the additive sensor noise. Now the problem is to find a filter matrix $\underline{\mathbf{w}} = (\underline{\mathbf{w}}_1 \dots \underline{\mathbf{w}}_{K'})$ so that by applying it to the observation vector $\mathbf{x}(t)$, the source signals can be estimated via

$$\mathbf{y}(t) = \sum_{k'=0}^{K'-1} \underline{\mathbf{w}}_{k'+1} \mathbf{x}(t-k'). \quad (2)$$

In other words, for the estimated vector $\mathbf{y}(t)$ and the source vector $\mathbf{s}(t)$, $\mathbf{y}(t) \approx \mathbf{s}(t)$ should hold.

This problem is also known as cocktail-party-problem. A common way to deal with the problem is to reduce it to a set of instantaneous separation problems, for which efficient approaches exist.

For this purpose, the time-domain observation vectors $\mathbf{x}(t)$ are transformed into a frequency domain time series by means of the short time Fourier transform (STFT)

$$\mathbf{X}(\Omega, \tau) = \sum_{t=-\infty}^{\infty} \mathbf{x}(t) w(t - \tau R) e^{-j\Omega t}, \quad (3)$$

where Ω is the angular frequency, τ represents the frame index, and $w(t)$ is a window function (e.g., Hanning window) of length N_{FFT} , τ represents the frame index and corresponds to the time shift of the window and R is the shift size, in samples, between successive windows [29]. Transforming Equation (1) into the frequency domain reduces the convolutions to multiplications with the corresponding transfer functions, so that for each frequency bin an individual instantaneous ICA problem

$$\mathbf{X}(\Omega, \tau) \approx \mathbf{A}(\Omega) \mathbf{S}(\Omega, \tau) + \mathbf{N}(\Omega, \tau) \quad (4)$$

arises. $\mathbf{A}(\Omega)$ is the mixing matrix in the frequency domain, $\mathbf{S}(\Omega, \tau) = [S_1(\Omega, \tau), \dots, S_N(\Omega, \tau)]$ represents the source signals, $\mathbf{X}(\Omega, \tau) = [X_1(\Omega, \tau), \dots, X_M(\Omega, \tau)]$, denotes the observed signals, and $\mathbf{N}(\Omega, \tau)$ is the frequency domain representation of the additive sensor noise. In order to reconstruct the source signals unmixing matrix $\mathbf{W}(\Omega) \approx \mathbf{P}^{-1}(\Omega) \mathbf{A}^{-1}(\Omega)$ is derived using complex-valued ICA, so that

$$\hat{\mathbf{Y}}(\Omega, \tau) = \mathbf{W}(\Omega) \mathbf{X}(\Omega, \tau) \quad (5)$$

holds. Here $\hat{\mathbf{Y}}(\Omega, \tau) = [\hat{Y}_1(\Omega, \tau), \dots, \hat{Y}_N(\Omega, \tau)]$ is the time frequency representation of the permuted ICA outputs. In order to solve the permutation problem a correction matrix $\mathbf{P}(\Omega)$ for each frequency bin has to be found, which is the main topic of this article. The data flow of the whole application is shown in Figure 1.

3 Permutation correction

This section gives an overview over the applied permutation correction methods. To resolve the permutations, the probability density functions (pdfs) of the magnitudes or of the logarithms of the magnitudes of the resulting frequency bins are compared. At this point, the assumption is made that adjacent frequency bins of the same source signal possess similar distributions.

3.1 Speech density modeling

3.1.1 Distribution of the speech magnitudes

As shown in [22], for speech signals the distribution of the magnitudes of spectral components can be described by the Rayleigh distribution. The pdf of the Rayleigh distribution of a random variable x is given by

$$f(x|\sigma) = \frac{x}{\sigma^2} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}, \quad (6)$$

where σ is a shape parameter that can be estimated e.g., by using the maximum likelihood estimator [30].

For the vector of random variables $\mathbf{x} = (x_1, x_2, \dots, x_N)$, the multivariate Rayleigh distribution can be written as follows

$$f(\mathbf{x}|\Sigma) = \det \Sigma^{-1/2} \prod_{i=1}^N \frac{\tilde{x}_i}{\tilde{\sigma}_i^2} \exp\left\{-\frac{\tilde{x}_i^2}{2\tilde{\sigma}_i^2}\right\} \left[\sum_{j=1}^N \exp\left\{-\frac{\tilde{x}_j^2}{2\tilde{\sigma}_j^2}\right\} - N + 1\right]^{-(N+1)} \quad (7)$$

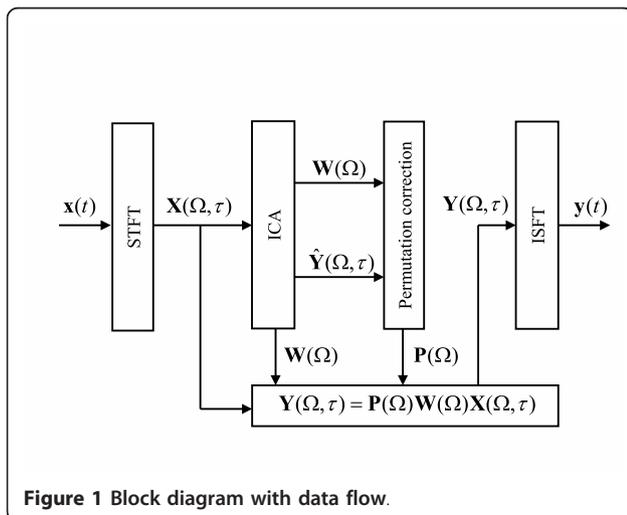


Figure 1 Block diagram with data flow.

where

$$\Sigma = (\mathbf{x} - \mu)^T (\mathbf{x} - \mu) \quad (8)$$

is the symmetric positive definite covariance matrix of \mathbf{x} ,

$$\tilde{\mathbf{x}} = \Sigma^{-1/2} (\mathbf{x} - \mu) \quad (9)$$

is a vector of the decorrelated random variables and $\tilde{\sigma}_i$ is the shape parameter for the signal \tilde{x}_i [31][32].^b

3.1.2 Distribution of the logarithms of the speech magnitudes

For the approximation of the logarithms of the speech magnitudes the GGD is applied. The PDF of the GGD of a random variable x is given by

$$f(x|\mu, \sigma, \beta) = \frac{\beta}{2a\Gamma(1/\beta)} \exp\left\{-\left|\frac{x - \mu}{a}\right|^\beta\right\}, \quad (10)$$

where μ is the mathematical expectation of x . The scale parameter a is obtained by

$$a = \sigma \sqrt{\frac{\Gamma(1/\beta)}{\Gamma(3/\beta)}}, \quad (11)$$

and the Gamma function is given by

$$\Gamma(z) = \int_0^\infty u^{z-1} e^{-u} du. \quad (12)$$

The β -parameter describes the distribution shape and σ is the standard deviation of x . However, the β -parameter is unknown and needs to be estimated e.g., by using the maximum likelihood estimator [33] or the moment estimator [34,35].

For the vector of random variables $\mathbf{x} = (x_1, x_2, \dots, x_N)$, the multivariate generalized Gaussian PDF can be written as follows

$$f(\mathbf{x}|\mu, \Sigma, \beta_1, \dots, \beta_N) = \det \Sigma^{-1/2} \prod_{i=1}^N \left[\frac{\beta_i}{2a_i \Gamma(1/\beta_i)} \exp\left\{-\left|\frac{\tilde{x}_i}{a_i}\right|^{\beta_i}\right\}\right], \quad (13)$$

where Σ is the covariance matrix of \mathbf{x} and $\tilde{\mathbf{x}}$ is a vector of the decorrelated random variables (Equation (9)) [33].

3.2 Distance measures

Suppose the pdfs of magnitudes in two adjacent frequency bins

$$f\left(\left|\hat{\mathbf{Y}}(\Omega_k, \tau)\right|\right) = \left[f\left(\left|\hat{Y}_1(\Omega_k, \tau)\right|\right), \dots, f\left(\left|\hat{Y}_N(\Omega_k, \tau)\right|\right)\right] \quad (14)$$

and

$$f\left(\left|\hat{\mathbf{Y}}(\Omega_{k+1}, \tau)\right|\right) = \left[f\left(\left|\hat{Y}_1(\Omega_{k+1}, \tau)\right|\right), \dots, f\left(\left|\hat{Y}_N(\Omega_{k+1}, \tau)\right|\right)\right] \quad (15)$$

of separated speech signals $\hat{Y}(\Omega, \tau)$ are known. To solve the permutation ambiguity problem, it is necessary to define a pairwise similarity measure $d(\cdot, \cdot)$ between two PDFs, so the overall dependence (distance) results in

$$D(f(\hat{Y}^P(\Omega_k, \tau)), f(\hat{Y}_n(\Omega_{k+1}, \tau))) = \sum_{n=1}^N d(f(\hat{Y}_n^P(\Omega_k, \tau)), f(\hat{Y}_n(\Omega_{k+1}, \tau))), \quad (16)$$

where $k \in [1, N_{\text{FFT}} - 1]$ is the frequency index,

$$\hat{Y}^P(\Omega_k, \tau) = \pi(\hat{Y}(\Omega_k, \tau)) \quad (17)$$

is a permutation of $\hat{Y}(\Omega_k, \tau)$, $\pi(x)$ defines a permutation of the components of the vector x and N is the number of separated signals. The total distance D between a permuted vector of frequency bins, $\hat{Y}^P(\Omega_k, \tau)$, and a reference vector in bin $k + 1$, is a sum of distances between each pair $\hat{Y}_n^P(\Omega_k, \tau)$ and $\hat{Y}_n(\Omega_{k+1}, \tau)$.

Below, several information theoretic similarity measures will be considered, which seem to be suitable for the solution of the permutation ambiguity problem. But first a definition of entropy or “self-information” is necessary.

The generalized formulation of entropy was given by Rényi and is known as the Rényi entropy in information theory [36,37]. The Rényi differential entropy of order α , where $\alpha \geq 0$, for a random variable with a pdf $f(x)$ whose support is a set \mathbb{X} , is defined as

$$H_\alpha(f(x)) = \frac{1}{1-\alpha} \log \left(\int_{\mathbb{X}} f^\alpha(x) dx \right). \quad (18)$$

It can be shown, that in limit for $\alpha \rightarrow 1$, $H_\alpha(f(x))$ converges to the Shannon entropy [37,38],

$$H_1(f(x)) = - \int_{\mathbb{X}} f(x) \log f(x) dx. \quad (19)$$

Similarly to the marginal entropy above, the joint entropy of a vector of random variables $\mathbf{x} = (x_1, x_2, \dots, x_N)$ is defined as

$$H_\alpha(f(\mathbf{x})) = \frac{1}{1-\alpha} \log \left(\int_{\mathbb{X}} f^\alpha(\mathbf{x}) d\mathbf{x} \right), \quad (20)$$

where $f(\mathbf{x})$ is the multivariate pdf.

At this point it is possible to introduce the necessary dependence measures that will be used as the pairwise similarity measure $d(\cdot, \cdot)$ in Equation (16):

- **Rényi generalized divergence** between two distributions $f(x)$ and $g(x)$ of order α , where $\alpha \geq 0$, is defined [36] as

$$d_\alpha(f(x)||g(x)) = \frac{1}{\alpha-1} \log \left(\int f^\alpha(x) g^{1-\alpha}(x) dx \right). \quad (21)$$

Special cases of Equation (21) [39] are the
 - Bhattacharyya coefficient

$$d_{1/2}(f(x)||g(x)) = -2 \log \left(\int \sqrt{f(x)g(x)} dx \right), \quad (22)$$

- Kullback-Leibler divergence

$$d_1(f(x)||g(x)) = \int f(x) \log \frac{f(x)}{g(x)} dx, \quad (23)$$

- Log distance

$$d_2(f(x)||g(x)) = \log E \left[\frac{f(x)}{g(x)} \right], \quad (24)$$

where $E[\cdot]$ denotes the statistical expectation according to $f(x)$,

- and log of the maximum ratio

$$d_\infty(f(x)||g(x)) = \log \sup_x \frac{f(x)}{g(x)}. \quad (25)$$

Rényi’s divergence describes the likeness between two distributions. The smaller the Rényi divergence, the more similar the distributions are. The main advantage of the Rényi divergence is the small computational burden. The problem in using the Rényi divergence is the fact that this measure is not symmetric, so typically $d_\alpha(f(x)||g(x)) \neq d_\alpha(g(x)||f(x))$, and not bounded, so infinite values can arise.

- **Mutual information** for a vector of random variables $\mathbf{X} = (X_1, X_2, \dots, X_K)$ is defined as the Kullback-Leibler divergence between the product of the distribution functions $\prod_{i=1}^K f_{X_i}(x_i)$ and the multivariate distribution $f_{\mathbf{x}}(\mathbf{x})$

$$I(\mathbf{X}) = d_1 \left(f_{\mathbf{x}}(\mathbf{x}) \left\| \prod_{i=1}^K f_{X_i}(x_i) \right. \right) \quad (26)$$

$$= \int f_{\mathbf{x}}(\mathbf{x}) \log \frac{f_{\mathbf{x}}(\mathbf{x})}{\prod_{i=1}^K f_{X_i}(x_i)} d\mathbf{X} \quad (27)$$

$$= \int f_{\mathbf{x}}(\mathbf{x}) \log f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} - \int f_{\mathbf{x}}(\mathbf{x}) \log \prod_{i=1}^K f_{X_i}(x_i) d\mathbf{x} \quad (28)$$

$$= \sum_{i=1}^K H_1(f_{X_i}(x_i)) - H_1(f_{\mathbf{x}}(\mathbf{x})) \quad (29)$$

where $H_1(f_{X_i}(x_i))$ is the marginal entropy and $H_1(f_{\mathbf{x}}(\mathbf{x}))$ is the joint entropy of \mathbf{X} .

Mutual information gives the amount of information contained in the random variables of \mathbf{X} . Since for the computation of the term $f_{\mathbf{X}}(\mathbf{x})$ is taken into account i.e., the dependencies are considered, the mutual information is a stronger cost function than Rényi divergence and using it for resolving the permutations, better results are to be expected.

- **The Jensen-Rényi divergence** of the vector of random variables $\mathbf{X} = (X_1, X_2, \dots, X_K)$ of order α , where $\alpha \geq 0$, is defined [40] as

$$d_{JR\alpha}(\mathbf{X}) = H_\alpha \left(\frac{1}{K} \sum_{i=1}^K f_{X_i}(x_i) \right) - \frac{1}{K} \sum_{i=1}^K H_\alpha(f_{X_i}(x)). \quad (30)$$

The Jensen-Rényi divergence is based on the Kullback-Leibler divergence and can be seen as an extension of it with the difference that it is symmetric^c and always of finite value. On the other hand, due to the fact that the distributions of the random variables are compared indirectly using the average $\frac{1}{K} \sum_{i=1}^K f_{X_i}(x)$, the Jensen-Rényi divergence can be seen as an alternative to the mutual information [41]. In fact, as shown in [42], both measures show similar characteristics.

- **The modified Jensen-Rényi divergence.** The Jensen-Rényi-divergence from the Equation (30) measures the distance between two distributions $f_X(x)$ and $f_Y(x)$ in respect to a third point in the distribution space. In this case, the third point is chosen as the average of the two distributions. This approach is justified because of the concavity of the entropy in distribution space

$$H_\alpha \left(\frac{f_X(x) + f_Y(x)}{2} \right) \geq \frac{H_\alpha(f_X(x)) + H_\alpha(f_Y(x))}{2}. \quad (31)$$

In principle, it is possible to define the distance in respect to any other point, if the assumption of the concavity for this point holds. Such a point can be chosen as an average over the random variables, the distributions of which are currently analyzed.

For the entropy of a random variable X

$$H_\alpha(f_X(x)) \propto \|f_X(x)\|_\alpha \quad (32)$$

holds, and for the entropy of the sum of two random variables X and Y [38,43]

$$H_\alpha(f_{X+Y}(x)) \propto \|f_X(x) * f_Y(x)\|_\alpha. \quad (33)$$

$\|\cdot\|_\alpha$ denotes the α norm operator and \boxtimes stands for convolution. Using the entropy power inequality [38] for the case of $\alpha = 1$, and extending Young's inequality [44] for the case of $\alpha \neq 1$, it can be shown [45], that

$$H_\alpha(f_{X+Y}(x)) \geq \max(H_\alpha(f_X(x)), H_\alpha(f_Y(x))). \quad (34)$$

Since

$$\max(a, b) \geq \frac{a + b}{2} \quad (35)$$

holds, the inequality in Equation (34) can be rewritten as

$$H_\alpha(f_{X+Y}(x)) \geq \frac{H_\alpha(f_X(x)) + H_\alpha(f_Y(x))}{2}. \quad (36)$$

So, at this point a modification of the Jensen-Rényi divergence is proposed. This distance measure of the vector of random variables $\mathbf{X} = (X_1, X_2, \dots, X_K)$ of order α , where $\alpha \geq 0$, is defined as

$$d_{mJR\alpha}(\mathbf{X}) = H_\alpha(f_{\bar{X}}(x)) - \frac{1}{K} \sum_{i=1}^K H_\alpha(f_{X_i}(x)) \quad (37)$$

where $\bar{X} = \frac{1}{K} \sum_{i=1}^K X_i$ In the way the modified Jensen-Rényi divergence is used here, this distance measure describes the amount of new information coming to a spectrogram if an adjacent frequency bin $Y(\Omega_{k+1}, \tau)$ is included. The lesser the new information provided, the closer the frequency bins are. This modification has less computational burden than the classical Jensen-Rényi divergence, since for $H_\alpha(f_{\bar{X}}(x))$, only one pdf has to be calculated instead of K in the Jensen-Rényi divergence. Furthermore, for the entropy $H_\alpha(f_{\bar{X}}(x))$ there exists an analytical solution, which improves the accuracy of the results.

3.3 The Permutation correction algorithm

In this section the actual permutation correction algorithm will be discussed. As mentioned before, it will be assumed that subsequent frequency bins of the same source signal possess similar distributions. The similarity between the frequency bins is measured by applying the measures given in Equations (21),(29), (30), and (37) in the optimization of Equation (16).

However, as mentioned in [14] the use of only one frequency bin as a reference bin for the correction causes a risk of a misalignment of the algorithm. To avoid this problem, the approach presented in [5] uses an average value of the already corrected frequency bins. So, the Equation (16) will be redefined as

$$D \left[f(|\hat{Y}^p(\Omega_b, \tau)|), f \left(\frac{1}{L} \sum_{i=b+1}^{b+L} |\hat{Y}(\Omega_i, \tau)| \right) \right] = \sum_{n=1}^N d \left[f \left(|\hat{Y}_n^p(\Omega_b, \tau)| \right), f \left(\frac{1}{L} \sum_{i=b+1}^{b+L} |\hat{Y}_n(\Omega_i, \tau)| \right) \right] \quad (38)$$

where L is the number of the already corrected frequency bins to be used for the averaging. Then the correction algorithm can be implemented as described in Algorithm 1.

Algorithm 1

1. Initialization: Start with the frequency^d Set $k = N_{\text{FFT}}/2$.

2. Estimate the parameters of the Rayleigh distribution of $|\hat{Y}(\Omega_k, \tau)|$ and of the average of L already corrected bins $\frac{1}{L} \sum_{l=k+1}^{\hat{L}} |\hat{Y}_n(\Omega_l, \tau)|$ with $\hat{L} = \min(k + 1 + L, N_{\text{FFT}}/2 + 1) - (k + 1)$ using Equations (6)-(9).

3. Calculate $D \left[f \left(|\hat{Y}^P(\Omega_k, \tau)| \right), f \left(\frac{1}{L} \sum_{l=k+1}^{\hat{L}} |\hat{Y}(\Omega_l, \tau)| \right) \right]$ as defined in Equation (38) for all possible permutations of $|\hat{Y}(\Omega_k, \tau)|$.

4. Choose the permutation $\pi_+(|\hat{Y}(\Omega_k, \tau)|)$ with the most dependent value of D .

5. Correct the current frequency bin in order with the best permutation $\pi_+(|\hat{Y}(\Omega_k, \tau)|)$.

6. Decrement k and if $k \neq 0$ go to Step 2.

The same scheme can be applied on the logarithms of the spectral magnitudes of the signals $\log |\hat{Y}(\Omega_k, \tau)|$ instead of $|\hat{Y}(\Omega_k, \tau)|$ and using generalized Gaussian instead of Rayleigh distributions. In that case Algorithm 2 results.

Algorithm 2

1. Initialization: Start with the frequency $k = N_{\text{FFT}}/2$.

2. Estimate the GGD parameters of $\log |\hat{Y}(\Omega_k, \tau)|$ and of the average of L already corrected bins $\log \left(\frac{1}{L} \sum_{l=k+1}^{\hat{L}} |\hat{Y}_n(\Omega_l, \tau)| \right)$ with $\hat{L} = \min(k + 1 + L, N_{\text{FFT}}/2 + 1) - (k + 1)$ using Equations (10)-(13).^e

3. Calculate $D \left[f \left(\log |\hat{Y}^P(\Omega_k, \tau)| \right), f \left(\log \left(\frac{1}{L} \sum_{l=k+1}^{\hat{L}} |\hat{Y}(\Omega_l, \tau)| \right) \right) \right]$ as defined in Equation (38) for all possible permutations of $|\hat{Y}(\Omega_k, \tau)|$.

4. Choose the permutation $\pi_+(\log |\hat{Y}(\Omega_k, \tau)|)$ with the most dependent value of D .

5. Correct the current frequency bin in order with the best permutation $\pi_+(\log |\hat{Y}(\Omega_k, \tau)|)$.

6. Decrement k and if $k \neq 0$ go to Step 2.

The Algorithms 1 and 2 will be used in the following sections for the experimental comparison of the distance measures given in Equations (21),(29), (30), and (37).

4 Experiments and results

4.1 Conditions

For the evaluation of the proposed approaches, two different sets of recordings were used. In the first data set, different audio files from the TIDigits database [46] were used and mixtures with up to four speakers were recorded under real room conditions. The distance between speakers and the center of a linear microphone array was varied between 0.9 and 2 m. The second

dataset was recorded by Sawada [47]. Here also mixtures with up to four speakers are presented. All of the mixtures were made with the same number of microphones as the number of speakers in the mixture ($M = N$), i.e., in each mixture a determined problem is considered so the classical ICA algorithms for source separation can be applied. The experimental setups are presented schematically in Figure 2 and the experimental conditions are summarized in Tables 1, 2, 3, and 4.

4.2 Parameter settings

The algorithms were tested on all recordings, which were first transformed to the frequency domain at a resolution of $N_{\text{FFT}} = 1,024$. For calculating the spectrogram, the signals were divided into overlapping frames with a Hanning window and an overlap of $3/4 \cdot N_{\text{FFT}}$.

4.3 ICA performance measurement

For calculation of the effectiveness of the proposed algorithm, the improvement ΔSIR of the signal to interference ratio

$$\Delta\text{SIR}_i = 10 \log_{10} \frac{\sum_n y_{i,s_i}^2(n)}{\sum_{j \neq i} \sum_n y_{i,s_j}^2(n)} - 10 \log_{10} \frac{\sum_n x_{i,s_i}^2(n)}{\sum_{j \neq i} \sum_n x_{i,s_j}^2(n)} \quad (39)$$

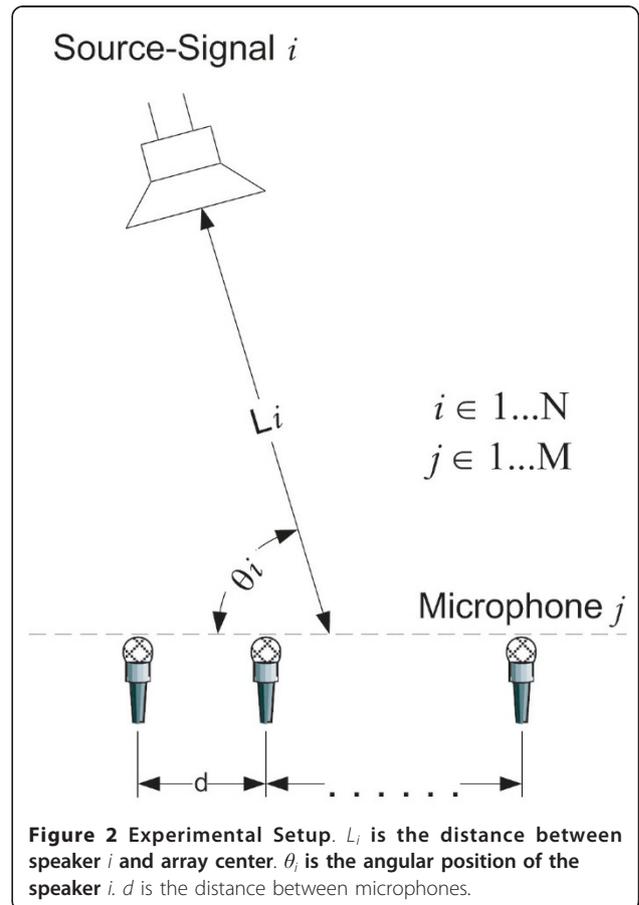


Figure 2 Experimental Setup. L_i is the distance between speaker i and array center. θ_i is the angular position of the speaker i . d is the distance between microphones.

Table 1 Mixture characteristics

Mixture	Mix. 1 TU Berlin	Mix. 2 TU Berlin	Mix. 3 TU Berlin
Reverberation time T_R	159 ms	159 ms	159 ms
Distance between two sensors d	3 cm	3 cm	3 cm
Sampling rate f_s	11 kHz	11 kHz	11 kHz
Number of speakers N	2	3	4
Number of microphones M	2	3	4
Distance between speaker i and array center	$L_1 = L_2 = 0.9$ m	$L_1 = L_2 = L_3 = 0.9$ m	$L_1 = L_2 = L_3 = L_4 = 0.9$ m
Angular position of the speaker i	$\theta_1 = 50^\circ$ $\theta_2 = 115^\circ$	$\theta_1 = 30^\circ$ $\theta_2 = 80^\circ$ $\theta_3 = 135^\circ$	$\theta_1 = 25^\circ$ $\theta_2 = 80^\circ$ $\theta_3 = 130^\circ$ $\theta_4 = 155^\circ$
Mean input SIR in [dB]	-0.1 dB	-3 dB	-5 dB

Table 2 Mixture characteristics

Mixture	Mix. 4 TU Berlin	Mix. 5 TU Berlin	Mix. 6 TU Berlin
Reverberation time T_R	189 ms	189 ms	189 ms
Distance between two sensors d	3 cm	3 cm	3 cm
Sampling rate f_s	11 kHz	11 kHz	11 kHz
Number of speakers N	2	3	4
Number of microphones M	2	3	4
Distance between speaker i and array center	$L_1 = L_2 = 2.0$ m	$L_1 = L_2 = L_3 = 2.0$ m	$L_1 = L_2 = L_3 = L_4 = 2.0$ m
Angular position of the speaker i	$\theta_1 = 75^\circ$ $\theta_2 = 165^\circ$	$\theta_1 = 35^\circ$ $\theta_2 = 80^\circ$ $\theta_3 = 165^\circ$	$\theta_1 = 30^\circ$ $\theta_2 = 75^\circ$ $\theta_3 = 125^\circ$ $\theta_4 = 165^\circ$
Mean input SIR in [dB]	-0.04 dB	-3.4 dB	-6.9 dB

Table 3 Mixture characteristics

Mixture	Mix. 7 NTT	Mix. 8 NTT	Mix. 9 NTT
Reverberation time T_R	130 ms	130 ms	130 ms
Distance between two sensors d	4 cm	4 cm	4 cm
Sampling rate f_s	8 kHz	8 kHz	8 kHz
Number of speakers N	2	3	4
Number of microphones M	2	3	4
Distance between speaker i and array center	$L_1 = L_2 = 1.2$ m	$L_1 = L_2 = L_3 = 1.2$ m	$L_1 = L_2 = L_3 = L_4 = 1.2$ m
Angular position of the speaker i	$\theta_1 = 75^\circ$ $\theta_2 = 165^\circ$	$\theta_1 = 35^\circ$ $\theta_2 = 80^\circ$ $\theta_3 = 165^\circ$	$\theta_1 = 30^\circ$ $\theta_2 = 75^\circ$ $\theta_3 = 125^\circ$ $\theta_4 = 165^\circ$
Mean input SIR in [dB]	0.02 dB	-2.9 dB	-4.7 dB

was used as a measure of the separation performance and the signal to distortion ratio (SDR)

$$\text{SDR}_i = 10 \log_{10} \frac{\sum_n x_{k_i}^2(n)}{\sum_n (x_{k_i}(n) - \alpha y_{i_s_i}(n - \delta))^2} \quad (40)$$

as a measure of the signal quality. Here y_{i,s_j} is the i -th separated signal with only the source s_j active, and x_{k,s_j} is the observation obtained by microphone k when only s_j is active. α and δ are parameters for phase and

Table 4 Mixture characteristics

Mixture	Mix. 10 TU Berlin	Mix. 11 TU Berlin	Mix. 12 TU Berlin
Reverberation time T_R	159 ms	159 ms	159 ms
Distance between two sensors d	12 cm	12 cm	12 cm
Sampling rate f_s	11 kHz	11 kHz	11 kHz
Number of speakers N	2	3	4
Number of microphones M	2	3	4
Distance between speaker i and array center	$L_1 = L_2 = 0.9$ m	$L_1 = L_2 = L_3 = 0.9$ m	$L_1 = L_2 = L_3 = L_4 = 0.9$ m
Angular position of the speaker i	$\theta_1 = 30^\circ$ $\theta_2 = 70^\circ$	$\theta_1 = 30^\circ$ $\theta_2 = 70^\circ$ $\theta_3 = 150^\circ$	$\theta_1 = 30^\circ$ $\theta_2 = 70^\circ$ $\theta_3 = 115^\circ$ $\theta_4 = 170^\circ$
Mean input SIR in [dB]	0.02 dB	-2.5 dB	-4.2 dB

amplitude chosen to optimally compensate the difference between y_{i,s_j} and x_{k,s_j} [19].

For measuring the performance of the proposed algorithms on all speakers present in a mixture recording, an average Δ SIR

$$\Delta \text{SIR} = \frac{1}{N} \sum_{i=1}^N \Delta \text{SIR}_i \quad (41)$$

and SDR

$$\text{SDR} = \frac{1}{N} \sum_{i=1}^N \text{SDR}_i \quad (42)$$

were used, where N is the number of speakers in the considered mixture.

4.4 Experimental results

In this section the experimental results of the signal separation will be compared. All the mixtures from Tables 1, 2, 3, and 4 were separated by JADE, INFO-MAX, and the FastICA algorithm and the permutation problem was solved using either Algorithm 1 or 2 from Section 3.3 and distance measures from Equations (21), (29), (30), and (37). For each result the performance is calculated using Equations (39) and (40).

Figures 3 and 4 show the behavior of three different approaches in terms of Δ SIR and SDR (Equations (41) and (42)) over the mixtures for the Infomax approach.

In Tables 5 and 6, the separation results are averaged for each distance measure for the mixtures of 2, 3, and 4 signals separately. M_2 in Tables 5 and 6 contains the

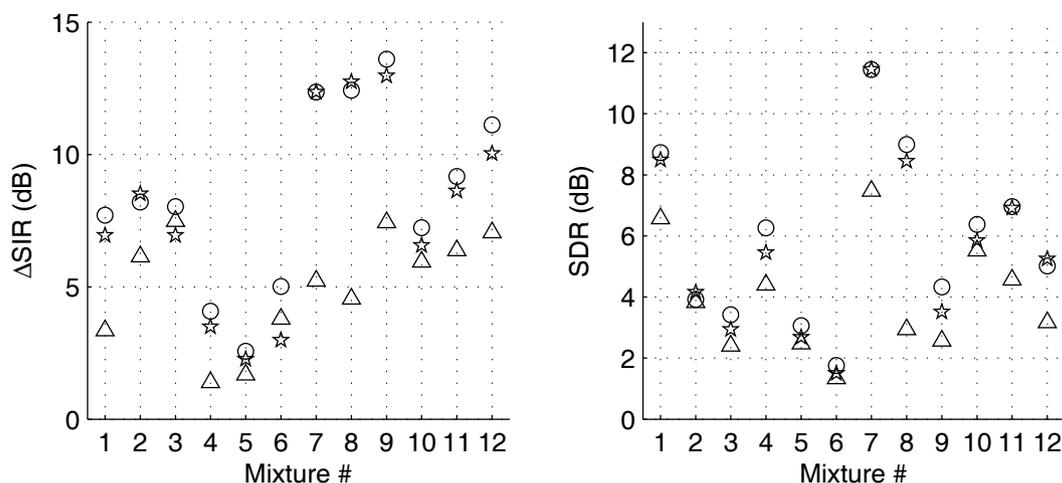


Figure 3 Results obtained by Infomax and Algorithm 1 using 'x' mutual information with $\alpha = 1$, 'Δ' Jensen-Rényi divergence with $\alpha = 1$ and 'o' modified Jensen-Rényi divergence with $\alpha = 1$ over the mixtures.

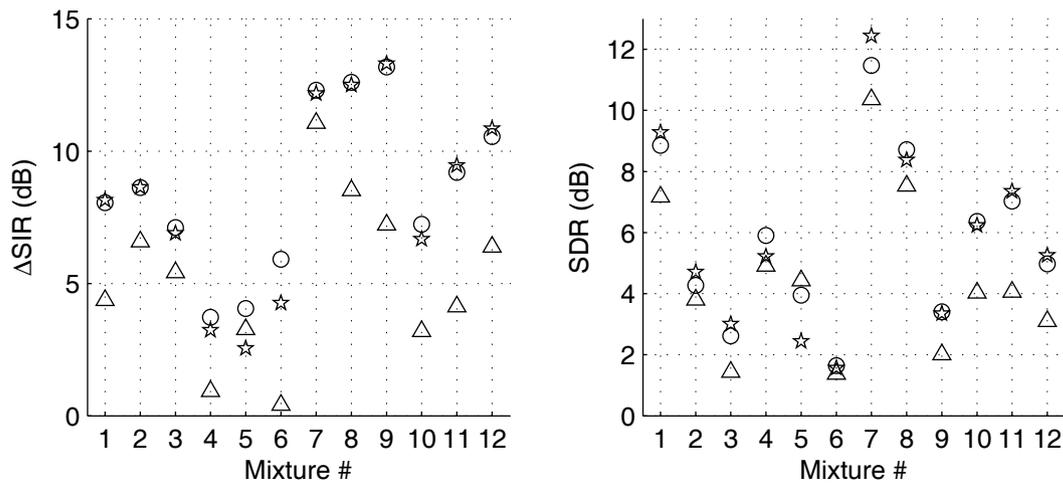


Figure 4 Results obtained by Infomax and Algorithm 2 using ‘*’ mutual information with $\alpha = 1$, ‘Δ’ Jensen-Rényi divergence with $\alpha = 1$ and ‘o’ modified Jensen-Rényi divergence with $\alpha = 1$ over the mixtures.

average Δ SIR/SDR of the separation results of Mix. 1, Mix. 4, Mix. 7 and Mix. 10, cf. Tables 1, 2, 3, and 4. Similarly, M_3 contains the separation results of Mix. 2, Mix. 5, Mix. 8 and Mix. 11, and M_4 those of Mix. 3, Mix. 6, Mix. 9 and Mix. 12.

4.5 Discussion

The calculated results show the usefulness of the proposed method for permutation correction, though not all of the applied distance measures perform equally. As already mentioned above, the best results were achieved using mutual information and the modified Jensen-

Rényi divergence, while results obtained using generalized Rényi divergence are rather poor. This is especially the case, if $\alpha = 2$ is used. Of all the applied distance measures based on the generalized Rényi divergence, the best performance was achieved in the case of the Bhattacharyya coefficient, i.e., $\alpha = 0.5$. A similar tendency can be seen with “classical” Jensen-Rényi divergence. Here the best results were achieved using $\alpha = 1$. In contrast, correction based on mutual information and the modified Jensen-Rényi divergence provides stable good results.

The poor performance in the case of generalized Rényi divergence can be explained by fact that the assumed

Table 5 Average values of the obtained results of Algorithm 1 in terms of Δ SIR and SDR for each distance measure

Distance measure	Δ SIR			SDR		
	M_2	M_3	M_4	M_2	M_3	M_4
Bhattacharyya coefficient	0.97	1.32	2.5	3.21	2.77	1.78
Kullback-Leibler divergence	0.86	2.12	1.93	5.03	3.01	1.02
Log of the maximum ratio	0.62	2.14	1.22	4.63	2.76	0.82
Jensen-Rényi divergence, $\alpha = 0.5$	3.00	3.65	5.44	5.33	3.17	1.43
Jensen-Rényi divergence, $\alpha = 1$	4.00	4.50	6.44	6.08	3.49	2.15
Jensen-Rényi divergence, $\alpha = 2$	4.01	3.94	6.09	5.75	3.29	1.45
Mod. Jensen-Rényi divergence, $\alpha = 0.5$	7.89	7.27	8.99	8.12	4.86	2.78
Mod. Jensen-Rényi divergence, $\alpha = 1$	7.89	7.27	8.97	8.12	4.87	2.74
Mod. Jensen-Rényi divergence, $\alpha = 2$	7.89	7.28	8.98	8.12	4.87	2.78
Mutual information	7.35	7.66	8.15	7.79	5.23	2.59

M_i stands for the average Δ SIR/SDR value calculated over all mixtures of $N = i$ signals (cf. Tables 1, 2, 3, and 4). The best performance for each case M_i is marked in bold.

Table 6 Average values of the obtained results of Algorithm 2 in terms of Δ SIR and SDR for each distance measure

Distance measure	Δ SIR			SDR		
	M_2	M_3	M_4	M_2	M_3	M_4
Bhattacharyya coefficient	2.21	3.23	3.54	5.53	3.65	1.10
Kullback-Leibler divergence	3.78	5.23	5.47	5.97	4.2	1.46
Log of the maximum ratio	3.52	4.99	4.14	6.32	4.12	1.14
Jensen-Rényi divergence, $\alpha = 0.5$	3.83	4.93	5.77	6.19	3.92	1.64
Jensen-Rényi divergence, $\alpha = 1$	4.00	5.04	5.45	6.39	4.12	1.44
Jensen-Rényi divergence, $\alpha = 2$	2.84	4.42	5.34	6.04	4.14	1.41
Mod. Jensen-Rényi divergence, $\alpha = 0.5$	7.31	8.14	8.53	8.01	5.63	2.44
Mod. Jensen-Rényi divergence, $\alpha = 1$	7.35	8.15	8.61	8.07	5.67	2.47
Mod. Jensen-Rényi divergence, $\alpha = 2$	7.40	8.27	8.43	8.12	5.76	2.50
Mutual information	7.31	8.50	8.37	8.18	6.00	2.60

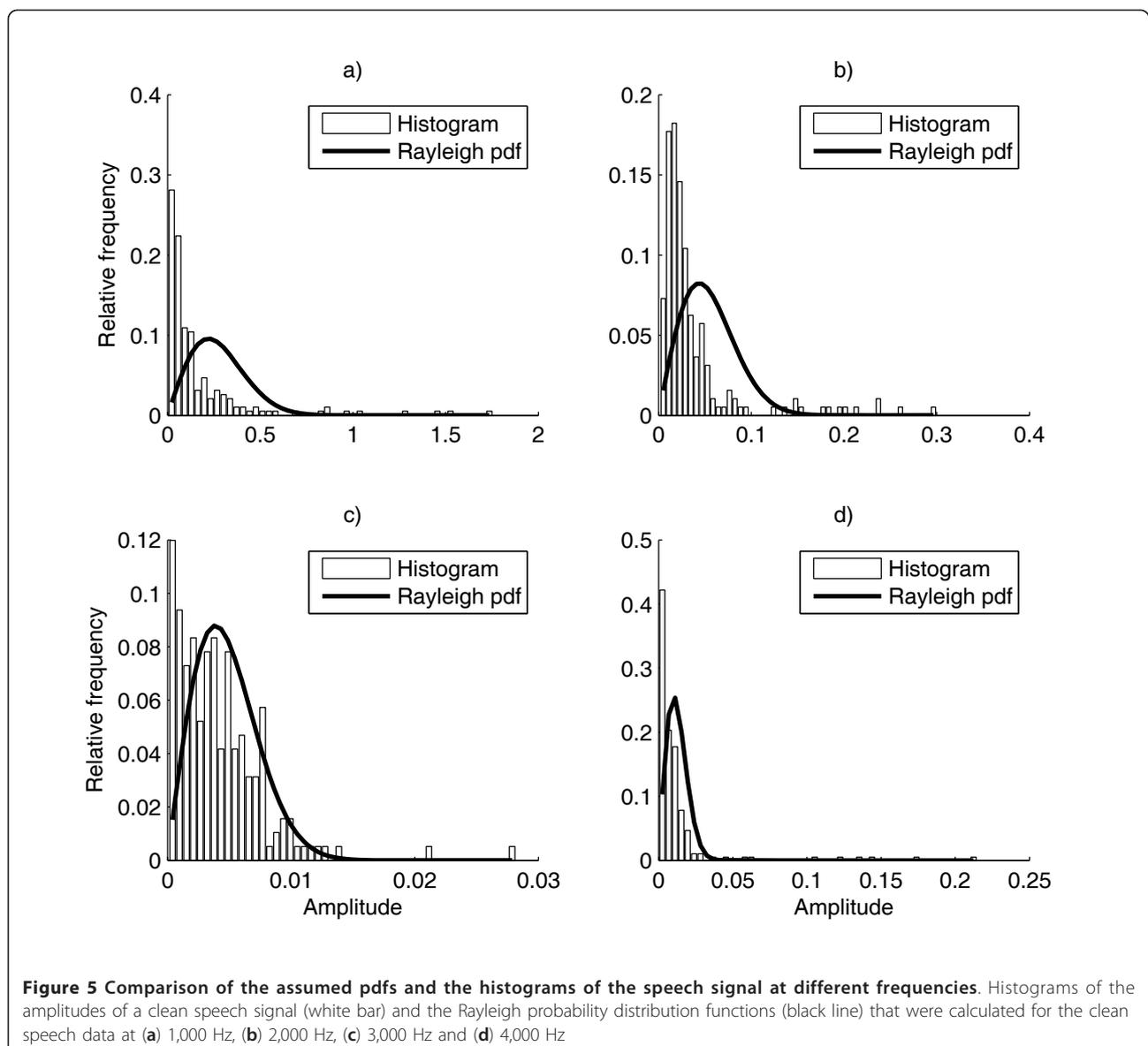
M_i stands for the average Δ SIR/SDR value calculated over all mixtures of $N = i$ signals. The best performance for each case M_i is marked in bold.

models of the probability distribution of the amplitudes and log amplitudes are not exact enough to be used with generalized Rényi divergence for a successful solution of the permutation problem. As can be seen in Figure 5, the distribution of the lower frequency bins of a speech signal can be only roughly modeled with Rayleigh pdf. In contrast the higher frequencies can be seen as Rayleigh distributed variables. This causes a high error rate in the permutation correction at lower frequencies, when the generalized Rényi divergence is used. Since, as shown in [48], the lower frequencies play a more significant role in BSS of the speech signals than the higher frequencies, the low values of the SIR and SDR are explainable. The same holds for the distribution of the log amplitudes. This problem can be solved

using a more exact model for the pdf of the speech signals such as Rayleigh mixture models (RMM) for the amplitudes or Gaussian mixture models (GMM) for the log amplitudes and is the subject of the future study.

Furthermore, the effects of the various reverberation times on the performance of different distance measures are going to be studied in the future. While, as it can be seen in Figures 3 and 4, the performance of the separation system decreases with a growing reverberation time of the environment, the effect of reverberation time on different methods of permutation correction should be analyzed in a more exact way.

On the other hand, the assumption of the Rayleigh distribution and GGD is good enough for permutation correction with mutual information and the modified



Jensen-Rényi divergence, since these distance measures are not as sensitive to the inter frequency-bin pdf perturbations as the generalized Rényi divergence. Furthermore, in this case there exists an analytical solution for modified Jensen-Rényi divergence, which reduces the computational burden of the algorithm and improves the accuracy of the solution.

As it can be seen in Tables 5 and 6, the separation performance of Algorithm 1 is slightly better than the performance of Algorithm 2. A possible explanation for this issue is the fact that for the Rayleigh pdf just one parameter has to be estimated instead of 2 parameters in case of GGD. Furthermore the estimation of the GGD parameters is more complicated than the estimation of the σ in case of Rayleigh distribution. These might cause the uncertainties and errors in the permutation correction.

In the next step, the best four of the proposed methods were compared with some other approaches from Section 1. In Table 7, the results of the comparison are shown. As can be seen, the proposed method (Algorithm 1 with the modified Jensen-Rényi divergence, $\alpha = 2$) performs better than other algorithms in terms of Δ SIR in most cases.

5 Conclusions

In this article, a method for the permutation correction in convolutive source separation has been presented. The approach is based on the assumption that magnitudes of

speech signals adhere to a Rayleigh distribution and the logarithms of magnitudes can be modeled by a GGD. The assumption of Rayleigh or GG distributed signals allows to use information theoretic similarity measures. The information theoretic distance measures are used to detect similarities in subsequent frequency bins after bin-wise source separation is completed, in order to group the frequency bins coming from the same source. Beside the existing information theoretic distance measures, a modification of the Jensen-Rényi divergence is proposed. This modified distance measure shows very good results for the considered problem.

The proposed method has been tested on different reverberant speech mixtures in connection with different ICA algorithms. The experimental results and the comparison with today's state-of-the-art approaches for permutation correction show the usefulness of the proposed method. Further, the experimental results have shown that the method performs best using either the mutual information or the modified Jensen-Rényi divergence criterion (Tables 5 and 6). This fact may be explained at least partially by the ability of the Jensen-Rényi divergence and the mutual information to utilize temporal dependence structure, which puts these two criteria ahead of the Rényi generalized divergence and its special cases of the Kullback-Leibler divergence and the log maximum ratio, which we considered as alternatives.

Table 7 Average values of the obtained results in terms of Δ SIR and SDR for each distance measure

Algorithms	Δ SIR			SDR		
	M2	M3	M4	M2	M3	M4
Proposed Algorithm 1 with Jensen-Rényi div., $\alpha = 2$	7.90	7.28	8.98	8.12	4.87	2.78
Proposed Algorithm 2 with Jensen-Rényi div., $\alpha = 0.5$	7.31	8.14	8.53	8.01	5.63	2.44
Proposed Algorithm 1 with mutual information	7.35	7.66	8.15	7.79	5.23	2.59
Proposed Algorithm 2 with mutual information	7.31	8.50	8.37	8.18	6.00	2.60
Permutation correction based on phase difference [50]	7.38	6.77	7.99	8.11	4.87	2.69
Cross-correlation [4]	7.44	7.61	7.76	8.16	5.35	2.63
Power ratio [51]	7.90	7.86	8.53	8.37	5.42	2.68
Continuity of the impulse response of the calculated mixing system [15]	3.93	1.83	2.14	5.67	2.98	0.89
Amplitude modulation decorrelation [3]	6.89	7.55	8.02	7.83	5.24	2.64
Cross-cumulants [4]	3.10	2.16	2.42	5.53	2.80	0.61
Continuity of the mixing system [17]	-0.33	0.65	0.93	4.37	2.52	0.77
Minimum of the L_1 -norm of the mixing system [16]	0.04	0.65	1.41	3.94	3.32	1.02
p -Norm distance ($p = 1$) [8]	6.05	7.60	7.68	7.37	5.51	2.38
Clustering of the amplitudes [9]	6.93	5.17	5.22	8.00	4.56	1.77
Likelihood ratio criterion between the frequency bins [21]	7.47	8.33	8.74	8.07	6.17	2.67
Basis vector clustering [19]	7.08	5.40	6.49	7.63	4.32	1.92
Minima of the beampattern [10]	7.40	3.74	2.81	7.74	3.24	0.82
Cosine distance [18]	5.33	4.78	4.56	6.76	4.26	1.74
GGD parameter comparison and cross-correlation [23]	0.45	4.64	6.39	4.13	3.71	1.77

M_i stands for the average Δ SIR/SDR value calculated over all mixtures of $N = i$ signals, (Tables 1, 2, 3, and 4).

The best performance for each case M_i is marked in bold.

Appendix 1

To calculate the distance measures from Equation (21), (29), (30), and (37), in most cases an integral has to be solved. The Rényi differential entropy (Equation (18)) in case of the Rayleigh distribution is calculated as

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left(\int_{\mathbb{X}} f_X^\alpha(x) dx \right) \quad (43)$$

$$= \log \left(\frac{\sigma}{\sqrt{2}} \right) + \frac{1}{1-\alpha} \log A + \frac{1}{1-\alpha} \log \left(\Gamma \left(\alpha - \frac{\alpha}{2} + \frac{1}{2} \right) \right), \quad (44)$$

where σ is the shape parameter of the Rayleigh distribution and

$$A = \alpha^{(-\alpha + \frac{\alpha}{2} - \frac{1}{2})}. \quad (45)$$

Setting $\alpha \rightarrow 1$ the Equation (44) becomes

$$H_1(x) = 1 + \log \frac{\alpha}{\sqrt{2}} + \frac{\gamma}{2}, \quad (46)$$

where γ is the Euler-Mascheroni constant $\gamma \approx 0.57722$.

For the GG distribution, the entropies can be computed as

$$H_\alpha^R(X) = \frac{1}{1-\alpha} \log \left(\int_{\mathbb{X}} f_X^\alpha(x) dx \right) \quad (47)$$

$$= \frac{1}{1-\alpha} \log \alpha^{-1/\beta_X} - \log \left(\frac{\beta_X}{2a\Gamma(1/\beta_X)} \right) \quad (48)$$

and

$$H_1(X) = \frac{1}{\beta_X} - \log \left(\frac{\beta_X}{2a\Gamma(1/\beta_X)} \right). \quad (49)$$

The solution of the Equation (46) is given in [38]. The solutions of the Equations (44) and (48) were derived using MATHEMATICA. For the distance measures without an analytical solution the trapezoidal rule for numerical integration was applied [49].

Appendix 2

Since information theoretic similarity measures make use only of the pdfs of the signals, a question may arise, as to whether temporal dependence structures of the signals are utilized at all in the suggested framework. The temporal structure is taken into account indirectly in the applied similarity measures, since each of the measures contains a term where either the joint probability, the pdf of the mean value of the random variables (Equation (37)), the mean of the pdf or a quotient of the pdfs is considered. These are the terms where the values of the distribution functions produced at the same time domain window are “compared”.

To demonstrate this issue, the following example was constructed: We compare signal $U_1(\tau)$, which contains amplitudes of a speech signal at the frequency $f = 3219.2$ Hz, $U_2(\tau)$ which is the same signal as $U_1(\tau)$ but time delayed, and $U_3(\tau)$, which contains amplitudes of the same speech signal as $U_1(\tau)$ at the frequency $f = 3230$ Hz, the next frequency bin, with additional Gaussian noise, see Figure 6.

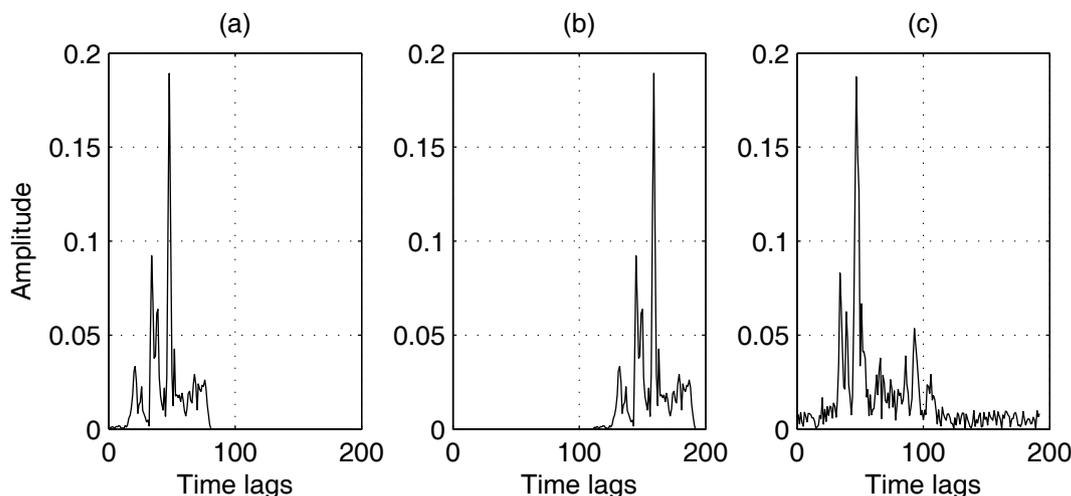


Figure 6 Constructed signals for demonstration. (a) Signal $U_1(\tau)$ is a speech signal at the frequency $f = 3, 219.2$ Hz, (b) $U_2(\tau)$ is the same signal as $U_1(\tau)$ but time delayed and (c) signal $U_3(\tau)$ contains amplitudes of the same speech signal as $U_1(\tau)$ at the frequency $f = 3, 230$ Hz (the next frequency bin) with an additional Gaussian noise.

Table 8 Comparison of signal pairs $\langle U_1(\tau), U_2(\tau) \rangle$ and $\langle U_1(\tau), U_3(\tau) \rangle$ with each distance measure

Distance measure	$\langle U_1(\tau), U_2(\tau) \rangle$	$\langle U_1(\tau), U_3(\tau) \rangle$
Bhattacharyya coefficient	0,09	0,59
Kullback-Leibler divergence	0,00	0,16
Log of the maximum ratio	0,00	0,16
Jensen-Rényi divergence with $\alpha = 0.5$	0,43	0,23
Jensen-Rényi divergence with $\alpha = 1$	1,92	0,22
Jensen-Rényi divergence with $\alpha = 2$	68,35	45,4
Modified Jensen-Rényi divergence with $\alpha = 0.5$	0,12	0,02
Modified Jensen-Rényi divergence with $\alpha = 1$	0,35	0,04
Modified Jensen-Rényi divergence with $\alpha = 2$	31,06	2,50
Mutual information	15,53	16,46

The most dependent value of each distance measure is marked in bold.

Table 9 Comparison of signal pairs $\langle \log(U_1(\tau)), \log(U_2(\tau)) \rangle$ and $\langle \log(U_1(\tau)), \log(U_3(\tau)) \rangle$ with each distance measure

Distance measure	$\langle \log(U_1(\tau)), \log(U_2(\tau)) \rangle$	$\langle \log(U_1(\tau)), \log(U_3(\tau)) \rangle$
Bhattacharyya coefficient	0,03	0,71
Kullback-Leibler divergence	0,0	0,03
Log of the maximum ratio	0,0	0,30
Jensen-Rényi divergence, $\alpha = 0.5$	17,83	7,48
Jensen-Rényi divergence, $\alpha = 1$	7,54	1,58
Jensen-Rényi divergence, $\alpha = 2$	0,99	0,01
Mod. Jensen-Rényi divergence, $\alpha = 0.5$	4,79	1,26
Mod. Jensen-Rényi divergence, $\alpha = 1$	1,23	0,26
Mod. Jensen-Rényi divergence, $\alpha = 2$	0,11	0,01
Mutual information	4,59	6,43

The most dependent value of each distance measure is marked in bold.

For each signal pair $\langle U_1(\tau), U_2(\tau) \rangle$, $\langle U_1(\tau), U_3(\tau) \rangle$, $\langle \log(U_1(\tau)), \log(U_2(\tau)) \rangle$ and $\langle \log(U_1(\tau)), \log(U_3(\tau)) \rangle$, the similarity measures from Equations (21), (29), (30), and (37) were applied. The results of the signal comparison can be found in Tables 8 and 9.

As can be seen, for this example each similarity measure that was considered in this article rates $U_1(\tau)$ more similar to $U_3(\tau)$ than to $U_2(\tau)$,^f which implies that the temporal dependencies and correlations were not ignored during the computation of the probability distribution functions.

In contrast to the other measures, in the case of the Rényi generalized divergence defined in Equation (21), and in its special cases of the Kullback-Leibler divergence and the log maximum ratio, the time dependency can not be taken into account in this manner. Still, these similarity measures can also be used for permutation correction, since the situation we considered in the example above is rather artificial and cannot be expected for realistic situations with two speech signals as the desired sources.

Endnotes

^aIn the cases where no permutation correction by the means of the comparison of the GGD parameters is

possible, the problem is handled by applying the correlation based permutation correction approach. ^bEquation (7) is a special case of the multivariate Weibull distribution with $\alpha = 1$ and $c_i = 2$ [32, Equation (14)]. ^cE.g. $d_{JR_\alpha}(X_1, X_2) = d_{JR_\alpha}(X_2, X_1)$. ^dThe proposed algorithm solves the permutations problem starting with the higher frequency bins. The first frequency bin in this case is the bin with $k = N_{FFT}/2 + 1$. Since there is no other definition of the correct order of the signals, the signal order in frequency bin $k = N_{FFT}/2 + 1$ will be assumed as correct. ^eFor the experiments from the Section 4 the parameter β was calculated using the approximation for the inverse function as proposed in [34]. ^fThe more dependent the signals are, the higher the value of the mutual information Equation (29) becomes, while simultaneously, the values of the similarity measures from (21), (30), and (37) decrease.

Competing interests

The authors declare that they have no competing interests.

Received: 31 October 2011 Accepted: 3 April 2012

Published: 3 April 2012

References

1. A Mansour, M Kawamoto, ICA papers classified according to their applications and performances. *IEICA Trans. Fundam.* **E86-A(3)**, 620–633 (2003)
2. MS Pedersen, J Larsen, U Kjems, LC Parra, Convolutional blind source separation methods, *Springer Handbook of Speech Processing and Speech Communication*, Springer Verlag, Berlin/Heidelberg, pp. 1065–1094 (2008)
3. J Anemüller, B Kollmeier, Amplitude modulation decorrelation for convolutional blind source separation, in *Proc. ICA 2000*, Helsinki, pp. 215–220 (2000)
4. C Mejuto, A Dapena, L Castedo, Frequency-domain infomax for blind separation of convolutional mixtures, in *Proc. ICA 2000*, Helsinki, pp. 315–320 (2000)
5. N Murata, S Ikeda, A Ziehe, An approach to blind source separation based on temporal structure of speech signals. *Neurocomputing.* **41(1-4)**, 1–24 (2001)
6. VG Reju, SN Koh, IY Soon, Partial separation method for solving permutation problem in frequency domain blind source separation of speech signals. *Neurocomputing.* **71**, 2098–2112 (2008)
7. L Parra, C Spence, B De Vries, Convolutional blind source separation based on multiple decorrelation, in *Proc. IEEE NNNSP Workshop*, (Cambridge, UK, 1998), pp. 23–32
8. K Kamata, X Hu, H Kobatake, A new approach to the permutation problem in frequency domain blind source separation, in *Proc. ICA 2004*, Granada, Spain, pp. 849–856 (Sept 2004)
9. D-T Pham, C Servièrè, H Boumaraf, Blind separation of speech mixtures based on nonstationarity, in *IEEE Signal Processing and Its Applications, Proceedings of the Seventh International Symposium*, pp. 73–76 (2003)
10. W Baumann, D Kolossa, R Orglmeister, Maximum likelihood permutation correction for convolutional source separation, in *ICA 2003*, pp. 373–378 (2003)
11. S Kurita, H Saruwatari, S Kajita, K Takeda, F Itakura, Evaluation of frequency-domain blind signal separation using directivity pattern under reverberant conditions, in *ICASSP2000*, pp. 3140–3143 (2000)
12. M Ikram, D Morgan, A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation, in *ICASSP02*, pp. 881–884 (2002)
13. N Mitianoudis, M Davies, Permutation alignment for frequency domain ICA using subspace beamforming methods, in *Proc. ICA 2004, LNCS 3195*, pp. 669–676 (2004)
14. H Sawada, R Mukai, S Araki, S Makino, A robust approach to the permutation problem of frequency-domain blind source separation, in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003)*, **V**, 381–384 (2003)
15. D-T Pham, C Servièrè, H Boumaraf, Blind separation of convolutional audio mixtures using nonstationarity, in *Proc. ICA2003*, pp. 981–986 (2003)
16. P Sudhakar, R Gribonval, A sparsity-based method to solve permutation indeterminacy in frequency-domain convolutional blind source separation, in *Independent Component Analysis and Signal Separation: 8th International Conference, ICA 2009, Proceedings*, (Paraty, Brazil, 2009)
17. W Baumann, B-U Köhler, D Kolossa, R Orglmeister, Real time separation of convolutional mixtures, in *Independent Component Analysis and Blind Signal Separation: 4th International Symposium, ICA 2001, Proceedings*, (San Diego, USA, 2001)
18. F Asano, S Ikeda, M Ogawa, H Asoh, N Kitawaki, Combined approach of array processing and independent component analysis for blind separation of acoustic signals, in *IEEE Trans. Speech Audio Proc.* **11(3)**, 204–215 (2003)
19. H Sawada, S Araki, R Mukai, S Makino, Blind extraction of a dominant source from mixtures of many sources using ICA and time-frequency masking, in *Proc. ISCAS 2005*, pp. 5882–5885 (2005)
20. W Wang, JA Chambers, S Sanei, A novel hybrid approach to the permutation problem of frequency domain blind source separation, in *Proc. 5th International Conference on Independent Component Analysis and Blind Signal Separation, ICA 2004*, Granada, Spain, pp. 530–537 (2004)
21. N Mitianoudis, ME Davies, Audio source separation of convolutional mixtures. *IEEE Trans. Audio Speech Process.* **11(5)**, 489–497 (2003)
22. Y Ephraim, D Malah, Speech enhancement using a minimum mean square error log-spectral amplitude estimator. *IEEE Trans. Acoust. Speech Signal Process.* **33**, 443–445 (1985)
23. R Mazur, A Mertins, Solving the permutation problem in convolutional blind source separation, in *Proc. ICA 2007, LNCS 4666*, pp. 512–519 (2007)
24. S Ikeda, N Murata, A method of blind separation based on temporal structure of signals, in *Proc. Int. Conf. on Neural Information Processing*, pp. 737–742 (1998)
25. J-F Cardoso, High order contrasts for independent component analysis. *Neural Comput.* **11**, 157–192 (1999)
26. A Bell, T Sejnowski, An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* **7**, 1129–1159 (1995)
27. A Hyvärinen, E Oja, A fast fixed-point algorithm for independent component analysis. *Neural Comput.* **9**, 1483–1492 (1997)
28. N Mitianoudis, M Davies, New fixed-point solutions for convolved mixtures, in *Proc. ICA2001*, (San Diego, CA, 2001), pp. 633–638
29. JB Allen, LR Rabiner, A unified approach to short-time Fourier analysis and synthesis. *Proc. IEEE.* **65**, 1558–1564 (1977)
30. KR Lee, CH Kapadia, DB Brock, On estimating the scale parameter of the Rayleigh distribution from doubly censored samples. *Statistische Hefte.* **21(1)**, 14–29 (1980)
31. WC Hoffman, The joint distribution of n successive outputs of a linear detector. *J. Appl. Phys.* **25**, 1006–1007 (1954)
32. GA Darbellay, I Vajda, Entropy expressions for multivariate continuous distributions. *IEEE Trans. Inf. Theory.* **46(2)**, 709–712 (2000)
33. L Boubchir, JM Fadili, Multivariate statistical modeling of images with the curvelet transform, in *IEEE Signal Processing and Its Applications, 2005. Proc. of the Eighth International Symposium.* **2**, 747–750 (28-31 Aug 2005)
34. JA Dominguez-Molina, G Gonzalez-Farias, RM Rodriguez-Dagnino, A practical procedure to estimate the shape parameter in the generalized Gaussian distribution. *CIMAT Tech. Rep. I-01-18_eng.pdf* http://www.cimat.mx/reportes/enlinea/I-01-18_eng.pdf. [Online]
35. R Prasad, Fixed-point ICA based speech signal separation and enhancement with generalized Gaussian model. PhD Thesis <http://citeseer.ist.psu.edu/prasad05fixedpoint.html> (2005)
36. A Rényi, On measures of entropy and information, in *Selected Papers of Alfred Rényi*, vol. 2. (Akademia Kiado, Budapest, 1976), pp. 565–580
37. JC Principe, D Xu, JW Fisher III, Information-theoretic learning, in *Unsupervised Adaptive Filtering*, ed. by Haykin S (Wiley, New York, 2000), pp. 265–319
38. TM Cover, JA Thomas, *Elements of Information Theory* (Wiley, New York, 1991)
39. AO Hero, B Ma, O Michel, JD Gorman, Alpha divergence for classification, indexing and retrieval, in *Technical Report 328, Comm. and Sig. Proc. Lab., Dept. EECS, Univ. Michigan* (2001)
40. AB Hamza, H Krim, Jensen-Rényi divergence measure: theoretical and computational perspectives, in *Proc. ISIT 2003*, (Yokohama, Japan, 2003)
41. AFT Martins, MAT Figueiredo, PMQ Aguiar, NA Smith, EP Xing, Nonextensive entropic kernels, in *ICML 08: Proc. of the 25th International Conference on Machine Learning, ACM.* **307**, 640–647 (2008)
42. Y He, AB Hamza, H Krim, A generalized divergence measure for robust image registration. *IEEE Trans. Signal Process.* **51(5)**, 1211–1220 (2003)
43. C Arndt, Information measures: information and its description in science and engineering, in *Signals and Communication Technology*, 2nd edn. (Springer, Berlin, 2004)
44. F Barthe, Optimal Youngs inequality and its converse: a simple proof. *Geom. Funct. Anal.* **8(2)**, 234–242 (1998)
45. JF Bercher, C Vignat, A Rényi entropy convolution inequality with application, in *Proc. EUSIPCO*, (Toulouse, France, 2002)
46. RG Leonard, A Database for speaker-independent digit recognition, in *Proc. ICASSP 84.* **3**, 42.11 (1984)
47. H Sawada, <http://www.kecl.ntt.co.jp/icl/signal/sawada/demo/bss2to4/index.html>
48. MG Jafari, MD Plumbley, The role of high frequencies in convolutional blind source separation of speech signals, in *Proc. 7th Int. Conf. on Independent Component Analysis and Signal Separation, ICA 2007*, (London, UK, 2007)
49. HR Schwarz, *Numerische Mathematik* (B.G. Teubner, Stuttgart, 1997)
50. E Hoffmann, D Kolossa, R Orglmeister, A batch algorithm for blind source separation of acoustic signals using ICA and time-frequency masking, in *Proc. 7th Int. Conf. on Independent Component Analysis and Signal Separation, ICA 2007*, (London, UK, 2007)
51. H Sawada, S Araki, S Makino, Measuring dependence of bin-wise separated signals for permutation alignment in frequency-domain BSS, in *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on (2007)*, pp. 3247–3250 (2007)

doi:10.1186/1687-4722-2012-14

Cite this article as: Hoffmann et al.: Using information theoretic distance measures for solving the permutation problem of blind source separation of speech signals. *EURASIP Journal on Audio, Speech, and Music Processing* 2012 **2012**:14.