ON THE ROBUSTNESS OF THE MULTIDIMENSIONAL STATE COHERENCE TRANSFORM FOR SOLVING THE PERMUTATION PROBLEM OF FREQUENCY-DOMAIN ICA

Benedikt Loesch[1], *Francesco Nesta*[2,3], *and Bin Yang*[1]

[1] Chair of System Theory and Signal Processing, University of Stuttgart
[2] Fondazione Bruno Kessler - Irst, Trento Italy, [3] UNITN, Trento Italy
Email: {benedikt.loesch,bin.yang}@LSS.uni-stuttgart.de, {nesta}@fbk.eu

ABSTRACT

A common problem in frequency domain independent component analysis (ICA) is the so called permutation problem which arises due to the independent demixing in each frequency bin. This paper evaluates the robustness of a an extension of a recently proposed method for permutation correction based on the time difference of arrival (TDOA) of the sources. First, we discuss the permutation problem, review the proposed method, and give an intuitive model to predict the number of permutations. Then the theoretical performance using perfect knowledge of the TDOAs of the sources as well as the practical performance using the TDOAs estimated from a multidimensional state coherence transform (SCT) are evaluated through extensive simulations. In our experiments, ICA with SCT based permutation correction outperforms Independent Vector Analysis (IVA).

***Index Terms*—** blind source separation, independent component analysis, permutation problem, state coherence transform

1. INTRODUCTION

The goal of blind source separation is to separate M convolutive mixtures $y_m(i), m = 1, \ldots, M$ into N statistically independent source signals. In frequency domain blind source separation (BSS) a time-frequency representation of the signals is generally derived by means of a short-time Fourier transform (STFT) and the signals observed at microphones are modeled as follows:

$$\mathbf{y}(k,l) = \mathbf{H}(k)\mathbf{x}(k,l). \tag{1}$$

Here k denotes the frequency bin index, l is a time index related to the analysis frame, $\mathbf{y}(k,l)$ is a vector of observed mixtures, $\mathbf{x}(k,l)$ is a vector of original signals, and $\mathbf{H}(k)$ is a mixing matrix. If $N = M$, the BSS problem is determined and the original signals $\mathbf{x}(k,l)$ can be recovered by applying a complex-valued ICA to $\mathbf{y}(k,l)$. The separation is done by estimating a set of demixing matrices $\mathbf{W}(k)$:

$$\hat{\mathbf{x}}(k,l) = \mathbf{W}(k)\mathbf{y}(k,l) \tag{2}$$

Due to the intrinsic indeterminacy of ICA, the matrix $\mathbf{W}(k)$ is an estimate of $\mathbf{H}(k)^{-1}$ up to a scaling and permutation ambiguity:

$$\mathbf{W}(k) = \mathbf{\Lambda}(k)\mathbf{\Pi}(k)\hat{\mathbf{H}}(k)^{-1} \tag{3}$$

where $\mathbf{\Lambda}(k)$ are diagonal scaling matrices, $\mathbf{\Pi}(k)$ are permutation matrices and $\hat{\mathbf{H}}(k)^{-1}$ is the estimated inverse of the true mixing matrix $\mathbf{H}(k)$. The matrix $\mathbf{\Pi}(k)$ is responsible for the so called permutation problem. Many methods have been proposed to resolve this permutation but a widely accepted solution is still not available. One way to resolve the permutation is to estimate the multidimensional propagation model [1], which can be approximated in terms of the time differences of arrival (TDOA). In [2] the state coherence transform (SCT) was proposed as an one-dimensional TDOA estimator. Recently, it has been extended to the multidimensional case by using a state vector instead of a scalar state variable [3, 4]. It was used for multidimensional source localization of multiple sources. In this paper, we propose the multidimensional SCT as an effective solution for the permutation problem. We study its robustness w.r.t to reverberation and noise and discuss its advantages and limitations.

The paper is organized as follows: In Sec. 2, the multidimensional SCT is formulated. Sec. 3 derives a model for the permutation error. Experimental results are presented in Sec. 4, and Sec. 5 gives a comparison with IVA [5].

2. MULTIDIMENSIONAL SCT

According to its physical meaning, a mixing matrix $\mathbf{H}(k)$ under anechoic conditions can be modeled as

$$\mathbf{H}(k) = [h_{mn}(k)]_{\substack{1 \leqq m \leqq M \\ 1 \leqq n \leqq N}}, \quad h_{mn}(k) = |h_{mn}| e^{-j2\pi f_k T_{mn}}. \tag{4}$$

f_k is the frequency corresponding to the *k-th* frequency bin, T_{mn} is the time of arrival (TOA) from *n-th* source to *m-th* microphone, $|h_{mn}|$ is the amplitude attenuation between *n-th* source and *m-th* microphone. Now we define a state for the *n-th* source, *k-th* frequency bin and microphone pair (a, b):

$$c_n^{(a,b)}(k) = \frac{h_{an}(k)}{h_{bn}(k)} = \frac{|h_{an}|}{|h_{bn}|} e^{-j2\pi f_k \tau_n^{(a,b)}} \tag{5}$$

where $\tau_n^{(a,b)} = T_{an} - T_{bn}$ is the true TDOA of the *n-th* source with respect to the microphone pair (a, b). Assuming comparable amplitude attenuations (e.g comparable source-microphone distances), (5) simplifies to

$$c_n^{(a,b)}(k) = e^{-j2\pi f_k \tau_n^{(a,b)}}. \tag{6}$$

Let $\boldsymbol{\tau} = [\tau^1, \cdots, \tau^S]^T$ be a TDOA vector containing the TDOA values of S microphone pairs related to a generic source. The ideal acoustic propagation can be represented by combining the states of these microphone pairs into a single column vector

$$\mathbf{c}(k, \boldsymbol{\tau}) = [e^{-j2\pi f_k \tau^s}]_{1 \leqq s \leqq S}. \tag{7}$$

We define the estimated normalized state for the *n-th* source, *k-th* frequency bin and sensor pair (a, b) as:

$$\overline{r}_n^{(a,b)}(k) = \frac{r_n^{(a,b)}(k)}{|r_n^{(a,b)}(k)|}, \quad r_n^{(a,b)}(k) = \frac{[\mathbf{W}(k)^{-1}]_{an}}{[\mathbf{W}(k)^{-1}]_{bn}}. \tag{8}$$

It can be shown that if $\hat{\mathbf{H}}(k)^{-1} = \mathbf{H}(k)^{-1}$ (i.e. perfect BSS) and $\mathbf{\Pi}(k) = \mathbf{I}$ (i.e. no permutation), $\overline{r}_n^{(a,b)}(k)$ is equivalent to $c_n^{(a,b)}(k)$ regardless of the diagonal scaling matrix $\mathbf{\Lambda}(k)$. Even if $\mathbf{\Pi}(k) \neq \mathbf{I}$ the equivalence holds except for a permutation of the source indices. Similarly as above, we combine S estimated states for the *n-th* source, *k-th* frequency bin and S sensor pairs into a single column vector

$$\overline{\mathbf{r}}_n(k) = [\overline{r}_n^s(k)]_{1 \leqq s \leqq S}. \tag{9}$$

Then, the multidimensional SCT is defined as

$$SCT(\boldsymbol{\tau}) = \sum_k \sum_{n=1}^{N} \left[1 - g\left(D[\mathbf{c}(k, \boldsymbol{\tau}), \overline{\mathbf{r}}_n(k)]\right)\right] \tag{10}$$

where $g(\cdot)$ is a suitable nonlinear function, $D[\cdot, \cdot]$ is a generic distance metric. It has been shown that the N maxima of $SCT(\boldsymbol{\tau})$ give the TDOA vector estimates $\hat{\boldsymbol{\tau}}_n$ of the N sources regardless of the permutation matrix [3, 4]. Therefore, the function in (10) can be considered a likelihood of the multidimensional source location in the TDOA domain.

Interestingly, the estimated TDOAs can be used to solve the permutation problem of frequency domain ICA without knowledge of the microphone array geometry. Let $\Pi_k(\cdot)$ be a permutation function for frequency bin k which defines the mapping between the indices of the true sources and indices of the demixed sources. $\Pi_k(\cdot)$ is another but equivalent notation of the permutation matrix $\mathbf{\Pi}(k)$. If e.g. $\mathbf{\Pi}(k) = \mathbf{I}$, then $\Pi_k(n) = n,\ \forall n$. Given the estimated state vectors $\bar{\mathbf{r}}_n(k)$ and the TDOA estimates $\hat{\boldsymbol{\tau}}_n$, we determine the permutation $\Pi_k(\cdot)$ at frequency bin k by the optimization

$$\hat{\Pi}_k = \underset{\Pi_k}{\operatorname{argmin}} \sum_{n=1}^{N} D[\mathbf{c}(k, \hat{\boldsymbol{\tau}}_n), \bar{\mathbf{r}}_{\Pi_k(n)}(k)]. \qquad (11)$$

It is a combinatoric optimization problem.

3. MODEL OF PERMUTATION ERRORS

The optimization in (11) aims to find the best match between the estimated state vectors $\bar{\mathbf{r}}_n(k)$ and the ideal state vectors $\mathbf{c}(k, \hat{\boldsymbol{\tau}}_n)$ by finding the permutation which minimizes the sum of the distance metrics $D[\mathbf{c}(k, \hat{\boldsymbol{\tau}}_n), \bar{\mathbf{r}}_{\Pi_k(n)}(k)]$. The number of permutation errors P_k after permutation correction in each frequency bin k can be calculated by counting the number of zero elements on the diagonal of $\mathbf{\Pi}(k)\hat{\mathbf{\Pi}}(k)^{-1}$, where $\mathbf{\Pi}(k)$ is the true permutation matrix at frequency bin k. In the simulations, $\mathbf{\Pi}(k)$ is determined by correlating the ideally demixed sources (using $\mathbf{H}^{-1}(k)$) with the separated source signals $\hat{\mathbf{x}}(k, l)$ and finding the permutation which gives the largest correlation. The percentage of permutation errors is then

$$P = \frac{1}{K \cdot N} \sum_{k=1}^{K} P_k. \qquad (12)$$

After permutation correction using (11), there might remain permutation errors if $D[\mathbf{c}(k, \hat{\boldsymbol{\tau}}_n), \bar{\mathbf{r}}_{\Pi_k(n)}(k)] < D[\mathbf{c}(k, \hat{\boldsymbol{\tau}}_n), \bar{\mathbf{r}}_{\Pi_k(m)}(k)]$ for any $n \neq m$. The permutation error P is expected to assume a low value if the estimated states $\bar{\mathbf{r}}_n(k)$ for different sources are sufficiently separated, which means that there is enough spatial diversity between the propagation of each source. Therefore, the effectiveness of (11) is strictly dependent on the acoustic conditions and on the locations of sources and microphones which intrinsically modify such a spatial diversity. If the acoustic conditions are bad (high reverberation, high noise, closely spaced sources) the state vector estimates $\bar{\mathbf{r}}_n(k)$ will have a large variance and hence the remaining permutation error will be high.

To derive a model of the permutation errors, we assume the reverberation as a diffuse noise field which is applicable if certain conditions are met [6]. We model the estimated TDOA vectors $\hat{\boldsymbol{\tau}}_n$ as Gaussian random vectors and hence use the following model for the estimated states

$$\bar{\mathbf{r}}_{\Pi_k(n)}(k) = \mathbf{c}(k, \hat{\boldsymbol{\tau}}_n) \qquad (13)$$

where $\hat{\boldsymbol{\tau}}_n \sim \mathcal{N}(\boldsymbol{\tau}_n, \mathbf{C})$. $\boldsymbol{\tau}_n$ is the vector of true TDOAs of source n and $\mathbf{C}$ is the covariance matrix of the noise. We can then estimate the expected value of P by Monte-Carlo simulation as follows:

- For each $\boldsymbol{\tau}_n$, generate Q realizations $\hat{\boldsymbol{\tau}}_n \sim \mathcal{N}(\boldsymbol{\tau}_n, \mathbf{C})$.
- Calculate Q corresponding state vectors according to (13).
- Average the percentage of permutation errors P for all Q realizations using $\mathbf{c}(k, \boldsymbol{\tau}_n)$ instead of $\mathbf{c}(k, \hat{\boldsymbol{\tau}}_n)$ in (11).

4. SIMULATION RESULTS

To evaluate the robustness of the proposed approach, we have conducted extensive simulations with an L-shaped array with different microphone spacings d and three sources shown in Fig. 1. We fix one source at a direction-of-arrival (DOA) of $\theta_3 = 0^\circ$ and vary the two other DOAs in the range of $[0^\circ, 360^\circ]$ in 10° steps. For each scenario, we simulate the room impulse reponses assuming omnidirectional sources and sensors using the ISM RoomSim toolbox [7] with a sampling frequency of $f_s = 16\,\mathrm{kHz}$ and a room size of $6.0\,\mathrm{m} \times 6.0\,\mathrm{m} \times 2.5\,\mathrm{m}$. We then generate mixtures by convolving three speech signals of length $5\,\mathrm{s}$ with the simulated impulse responses. The first three experiments consider the noiseless case, while the fourth experiment uses SNRs of $(0, 10, 20)\,\mathrm{dB}$. In the following evaluations, we use the states of the microphone pairs $(1, 2)$ and $(1, 3)$. We use the recursive Scaled Infomax algorithm from [8] with FFT size 1024. Please note that, although we use the DOA for the source positions, the permutation correction is performed without knowledge of the array geometry by evaluating the SCT as a function of the TDOAs. We consider two evaluation criteria:

- The percentage P of remaining permutation errors
- The distribution of the permutation errors P_k over frequency, but averaged for all considered source positions

We plot P as a function of the two DOAs θ_1, θ_2. This plot is called a (remaining) permutation map.

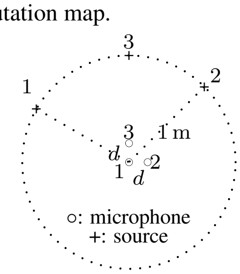
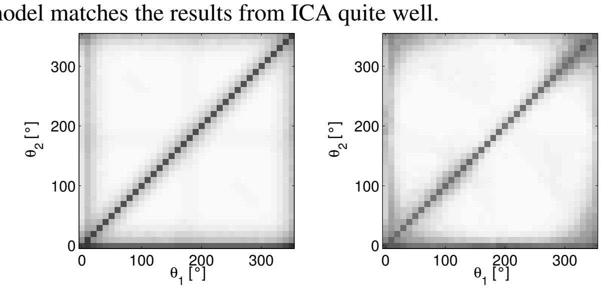


Fig. 1: Experimental setup

4.1. Comparison of Model with ICA Results

In this experiment, we want to verify the validity of the model proposed in the previous section. We assume the measurements from different microphone pairs to be independent. Furthermore, we assume the source-to-microphone distance to be comparable for all microphone pairs and hence we model the vector of TDOAs $\hat{\boldsymbol{\tau}}_n$ as a Gaussian random vector with covariance matrix $\mathbf{C} = \sigma^2\mathbf{I}$. In Fig. 2, we compare the permutation map obtained using the model for the estimated states (13) with the permutation map obtained using ICA to estimate the states (8). For both cases, we use the ideal model states $\mathbf{c}(k, \boldsymbol{\tau}_n)$ for the permutation correction step (11). Clearly, the model matches the results from ICA quite well.



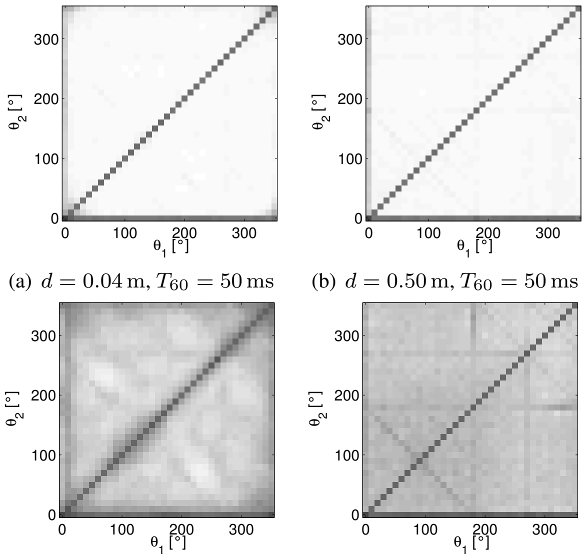
(a) Model (b) ICA

Fig. 2: Permutation map for $T_{60} = 150\,\mathrm{ms}$, $d = 0.02\,\mathrm{m}$

4.2. Perfect Knowledge of TDOA

The second experiment evaluates the proposed approach with perfect knowledge of the TDOAs, i.e. the permutation correction step (11) uses the ideal model states $\mathbf{c}(k, \boldsymbol{\tau}_n)$ instead of $\mathbf{c}(k, \hat{\boldsymbol{\tau}}_n)$, using different microphone distances $d = (0.02, 0.04, 0.1, 0.2, 0.5)\,\mathrm{m}$ in three different reverberation conditions $T_{60} = 50\,\mathrm{ms}, 150\,\mathrm{ms}$ and $300\,\mathrm{ms}$. Fig. 3 shows the permutation map for $T_{60} = 50\,\mathrm{ms}$ and $T_{60} = 300\,\mathrm{ms}$ for two different spacings: $d = 0.04\,\mathrm{m}, 0.50\,\mathrm{m}$.

For $T_{60} = 50\,\mathrm{ms}$, the proposed approach has very little permutation errors except for the degenerate cases where two of the three sources arrive from exactly the same direction. For more reverberant



(a) $d = 0.04\,\mathrm{m}, T_{60} = 50\,\mathrm{ms}$ (b) $d = 0.50\,\mathrm{m}, T_{60} = 50\,\mathrm{ms}$

(c) $d = 0.04\,\mathrm{m}, T_{60} = 300\,\mathrm{ms}$ (d) $d = 0.50\,\mathrm{m}, T_{60} = 300\,\mathrm{ms}$

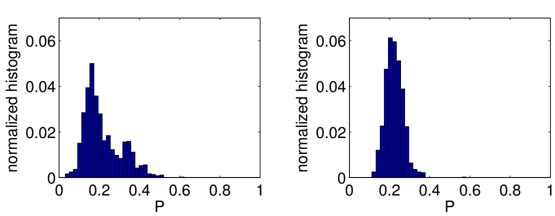
Fig. 3: Permutation map with known TDOAs

environments, a larger microphone spacing provides less permutation error for closely spaced sources. This is due to the fact that with increasing reverberation time, the estimated states have a larger variance and a larger microphone spacing d provides a better distinction between the states of different sources.

To quantitatively evaluate the permutation errors, we calculate the mean μ_P and the standard deviation σ_P of P over all considered source positions excluding the degenerate cases. The results are summarized in Table 1, histograms of P for $T_{60} = 300\,\mathrm{ms}$ are shown in Fig. 4. Clearly, a larger microphone spacing shows a lower fluctuation in the permutation errors. This means that the performance of the permutation correction is more constant over all considered source positions. We also see that with $T_{60} = 300\,\mathrm{ms}$ and the assumption of omni-directional sources and sensors, the reverberation leads to a large variance in the states and hence even after permutation correction with perfect knowledge of the TDOAs, we cannot completely solve the permutation problem.

mic spacing	$T_{60} = 50\,\mathrm{ms}$	$T_{60} = 150\,\mathrm{ms}$	$T_{60} = 300\,\mathrm{ms}$
$d = 0.02\,\mathrm{m}$	1.62(4.79)	8.44(10.58)	20.86(13.54)
$d = 0.04\,\mathrm{m}$	1.31(2.78)	7.43(7.04)	21.59(9.28)
$d = 0.10\,\mathrm{m}$	1.16(1.83)	6.59(3.56)	20.20(4.79)
$d = 0.20\,\mathrm{m}$	1.32(1.68)	7.01(2.83)	21.37(3.80)
$d = 0.50\,\mathrm{m}$	1.43(1.69)	7.20(2.71)	22.38(4.44)

Table 1: $\mu_P(\sigma_P)$ in % with known TDOAs



(a) $d = 0.04\,\mathrm{m}$ (b) $d = 0.50\,\mathrm{m}$

Fig. 4: Histograms of P, known TDOAs, $T_{60} = 300\,\mathrm{ms}$

Furthermore a large spacing d distributes the permutation errors more equally across the frequency bins and hence has lower permutation error in the low frequency region than a small spacing d. This is due to the fact that the estimated states $\bar{\mathbf{r}}_n(k)$ can be better separated at low frequencies using a large spacing d than with small spacing d. Fig. 5 shows the permutation errors across the frequency range $(0, 4)\,\mathrm{kHz}$ for $d = 0.02\,\mathrm{m}, 0.1\,\mathrm{m}, 0.5\,\mathrm{m}$ smoothed with a moving average filter of order 32. This corresponds to a window size of 500 Hz.

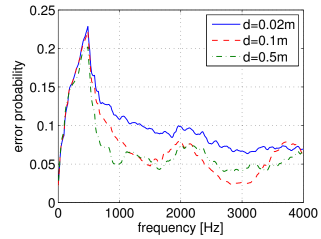


Fig. 5: Permutation error across frequency, $T_{60} = 150\,\mathrm{ms}$

4.3. Estimation of TDOAs using SCT

In this section, we compare the performance of permutation correction based on the TDOAs estimated by the SCT, i.e. permutation correction using the estimated model states $\mathbf{c}(k, \hat{\boldsymbol{\tau}}_n)$, and based on the perfectly known TDOAs, i.e. permutation correction using the ideal model states $\mathbf{c}(k, \boldsymbol{\tau}_n)$. In both cases, we use ICA to obtain the estimated states $\bar{\mathbf{r}}_n(k)$. For easy comparison with the true source locations, Fig. 6 shows the SCT as a function of the DOA by mapping the TDOAs to the corresponding DOA. The true DOAs are $\theta_1 = 40°, \theta_2 = 240°, \theta_3 = 0°$ and are denoted with dashed lines.

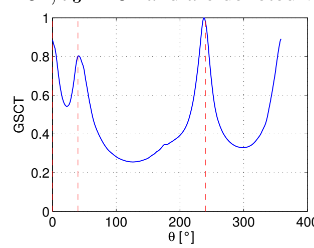


Fig. 6: SCT for $T_{60} = 300ms$ and $d = 0.02\,\mathrm{m}$

Clearly, the SCT is able to estimate the TDOAs very precisely. Hence, permutation correction using the estimated DOA works almost as well as with the perfectly known TDOAs. This is reflected in Figs. 7 and 8 which show the permutation error map and the histograms of permutation error P using SCT. However, the SCT sometimes fails to estimate the TDOAs correctly for closely spaced sources and a very small spacing of $d = 0.02\,\mathrm{m}$. Hence, in the following we will not consider the case of $d = 0.02\,\mathrm{m}$. Comparing μ_P and σ_P for the SCT (Table 2) with the results for perfect knowledge of the TDOAs (Table 1) we note that the results match quite well.

mic spacing	$T_{60} = 50\,\mathrm{ms}$	$T_{60} = 150\,\mathrm{ms}$	$T_{60} = 300\,\mathrm{ms}$
$d = 0.04\,\mathrm{m}$	1.50(4.50)	9.25(11.04)	23.39(11.70)
$d = 0.10\,\mathrm{m}$	1.17(1.82)	6.59(3.60)	20.48(5.40)
$d = 0.20\,\mathrm{m}$	1.33(1.72)	6.99(2.83)	21.47(3.83)
$d = 0.50\,\mathrm{m}$	1.41(1.63)	7.17(2.73)	22.39(4.46)

Table 2: $\mu_P(\sigma_P)$ in % with SCT

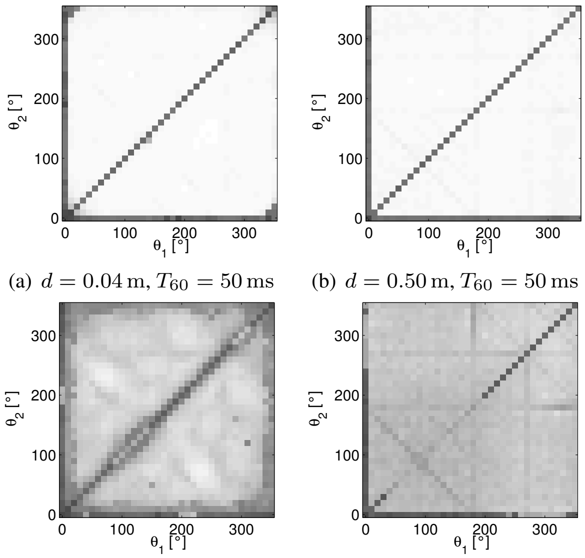
The low permutation error is also reflected in the SIR gain values shown in Table 3. Considering the SIR results as well as the permuation errors, a microphone spacing of $d = 0.1\,\mathrm{m}$ is a good tradeoff for practical applications.

mic spacing	$T_{60} = 50\,\mathrm{ms}$	$T_{60} = 150\,\mathrm{ms}$	$T_{60} = 300\,\mathrm{ms}$
$d = 0.04\,\mathrm{m}$	20.50(5.59)	11.81(5.45)	7.13(4.02)
$d = 0.10\,\mathrm{m}$	20.45(5.39)	11.84(5.12)	7.17(4.13)
$d = 0.20\,\mathrm{m}$	20.27(5.22)	12.00(4.36)	7.41(3.59)
$d = 0.50\,\mathrm{m}$	19.09(5.50)	11.79(3.89)	6.93(3.29)

Table 3: Mean and standard dev. of SIR gain for SCT

4.4. Robustness to Noise

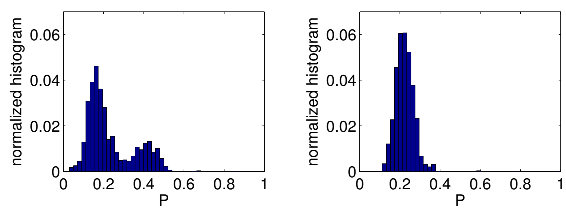
In this section we study the robustness of the permutation correction with respect to noise. We consider the case of $T_{60} = 150\,\mathrm{ms}$ and $d = 0.1\,\mathrm{m}$. Table 4 summarizes the mean μ_P and standard deviation



(a) $d = 0.04\,\mathrm{m}$, $T_{60} = 50\,\mathrm{ms}$ (b) $d = 0.50\,\mathrm{m}$, $T_{60} = 50\,\mathrm{ms}$

(c) $d = 0.04\,\mathrm{m}$, $T_{60} = 300\,\mathrm{ms}$ (d) $d = 0.50\,\mathrm{m}$, $T_{60} = 300\,\mathrm{ms}$

Fig. 7: Permutation map for SCT



(a) $d = 0.04\,\mathrm{m}$ (b) $d = 0.50\,\mathrm{m}$

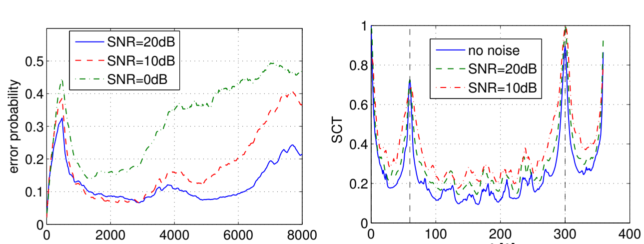
Fig. 8: Histogram of P with SCT, $T_{60} = 300\,\mathrm{ms}$

SNR	known DOA	SCT
$20\,\mathrm{dB}$	$12.19(4.52)$	$12.27(4.98)$
$10\,\mathrm{dB}$	$24.79(5.01)$	$27.84(8.12)$
$0\,\mathrm{dB}$	$39.19(4.00)$	$51.65(6.60)$

Table 4: $\mu_P(\sigma_P)$ in $\%$ for $T_{60} = 150\,\mathrm{ms}$ and $d = 0.1\,\mathrm{m}$

σ_P of the permutation errors P for perfect knowledge of the TDOAs and for the SCT for different noise conditions.

Clearly, for increasing noise the number of permutation errors increases. However, as we can see from Fig. 9 (a), the error probability increases mainly at higher frequencies. This is due to the fact that speech signals have less power in the high frequency region and hence the local SNR in that region is much worse than for the lower frequencies. Furthermore, Fig. 9 (b) shows that the SCT itself is quite robust to noise and can estimate the source locations correctly even with considerable amount of noise. This is also the reason, why the SCT achieves approximately the same number of permutation errors except for $0\,\mathrm{dB}$ SNR in Table 4 as we get with known TDOAs.



(a) Perm. error across frequency (b) SCT

Fig. 9: Robustness to noise: $T_{60} = 150\,\mathrm{ms}$ and $d = 0.1\,\mathrm{m}$

5. COMPARISON WITH IVA

First, we repeat the simulation of Sec. 4.4 for IVA [5], an approach which is supposed to be permutation-free. We get an average permutation error of $34.7\%, 48.5\%, 58.4\%$ for SNRs of $20, 10, 0\,\mathrm{dB}$. The mean permutation error for IVA is much higher than ICA+SCT (see Table 4). This is due to the fact that IVA is sensitive to the convergence to local minima which may also correspond to wrong permutations, while ICA+SCT is more robust.

Second, similar to Fig. 1, we individually record 6 sources separated by 60° in a real office room with $T_{60} = 520\,\mathrm{ms}$. Data is available at: http://www.lss.uni-stuttgart.de/mitarbeiter/loesch/bss_signals.html. We consider all possible mixtures of 3 sources. The input SIR is $-3.0\,\mathrm{dB}$. Table 5 compares the average permutation error P and the mean SIR for different learning window sizes with the results of IVA. Due to noise and the small dimensions of the room the data is difficult to separate and hence SIR values are low. Again, ICA+SCT clearly outperforms IVA in terms of permutation error and SIR, especially for short data segments.

	$1\,\mathrm{s}$		$2\,\mathrm{s}$		$3\,\mathrm{s}$	
IVA	42.3%	$0.8\,\mathrm{dB}$	41.4%	$3.3\,\mathrm{dB}$	44.2%	$2.0\,\mathrm{dB}$
ICA+SCT	30.8%	$5.2\,\mathrm{dB}$	27.2%	$7.5\,\mathrm{dB}$	26.9%	$7.7\,\mathrm{dB}$

Table 5: Comparison of P (%), SIR (dB) for real room, $d = 0.04\,\mathrm{m}$

6. CONCLUSIONS

In this paper, we evaluated the performance of TDOA based permutation correction for frequency domain ICA. First, we evaluated the performance using perfect knowledge of the TDOA. We have shown that, for moderately reverberant environments, TDOA based permutation correction is sufficient to almost completely solve the permutation problem. The multidimensional SCT is robust to noise, source positions and short data segments and achieves for moderately noisy environments the performance bound of TDOA based permutation correction using perfect knowledge of TDOA.

7. REFERENCES

[1] H. Sawada, S. Araki, R. Mukai, and S. Makino, "Grouping separated frequency components by estimating propagation model parameters in frequency-domain blind source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 5, pp. 1592–1604, July 2007.

[2] F. Nesta, M. Omologo, and P. Svaizer, "A novel robust solution to the permutation problem based on a joint multiple TDOA estimation," *Proc. International Workshop for Acoustic Echo and Noise Control (IWAENC)*, 2008.

[3] F. Nesta and M. Omologo, "Generalized state coherence transform for multidimensional localization of multiple sources," *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2009.

[4] B. Loesch, S. Uhlich, and B. Yang, "Multidimensional localization of multiple sound sources using frequency domain ICA and an extended state coherence transform," *Proc. IEEE Workshop on Statistical Signal Processing (SSP)*, 2009.

[5] T. Kim, H. Attias, S.-Y. Lee, and T.W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 1, 2007.

[6] T. Gustaffson, B. D. Rao, and M. Trivedi, "Source localization in reverberant environments: Modeling and statistical analysis," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 791–803, Nov. 2003.

[7] E. Lehmann, "Image-source method for room impulse response simulation (room acoustics)," `http://www.watri.org.au/~ericl/ism_code.html`, 2008.

[8] F. Nesta, P. Svaizer, and M. Omologo, "Separating short signals in highly reverberant environment by a recursive frequency domain BSS," *Proc. Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, May 2008.