A Study of ICA Based DOA Estimation with Respect to Permutation Ambiguity, Scaling Ambiguity and Sensor Gain Mismatch

Benedikt Loesch and Bin Yang
Chair of System Theory and Signal Processing, University of Stuttgart
Email: {benedikt.loesch, bin.yang}@LSS.uni-stuttgart.de

***Abstract*—Recently, direction-of-arrival (DOA) and position estimation for acoustic signals have been studied intensively and many different algorithms have been proposed. Among different approaches for multiple sources, independent component analysis (ICA) based methods have drawn much attention. In this paper, we study the effects of permutation ambiguity, source scaling ambiguity and sensor gain mismatch on source localization based on an estimate of the channel matrix or its inverse. We show that the source scaling ambiguity can be removed from the estimate of the inverse channel by proper normalization, but not the sensor gain mismatch. We evaluate the influence of source scaling ambiguity and sensor gain mismatch on two frequency domain ICA based localization algorithms, the averaged directivity pattern (ADP) and the state coherence transform (SCT) using simulations. We show that the original ADP is very sensitive to source scaling ambiguity and sensor gain mismatch. In contrast, SCT is completely insensitive to these effects and shows a superior localization performance.**

***Index Terms*—source localization, independent component analysis, averaged directivity pattern, state coherence transform, scaling ambiguity, sensor gain mismatch**

I. Introduction

The task of acoustic source localization is to estimate the position of one or multiple sound sources by using an array of microphones. There are indirect localization methods over the explicit estimation of the time-difference of arrival (TDOA) and direct spatial scanning methods. Among the latter ones, some methods like ADP [1] and SCT [2] are model-based since they explicitely model and estimate the acoustic channel for localization by using ICA. Other algorithms like SRP-PHAT [3] are not model-based as they perform the localization without a channel model. [1, 2, 4, 5] discuss some advantages of ICA based localization over SRP-PHAT, such as better performance and increased spatial resolution for multiple sources, and improved robustness to noise. The contribution of this paper is a theoretical and experimental study of the effects of permutation ambiguity, source scaling ambiguity and sensor gain mismatch on ICA based localization for multiple broadband sources.

II. Signal Model

ICA aims at separating N convolutives mixtures into N statistically independent source signals. We can model the convolutive mixing in the time-domain by an impulse response between each source n and microphone m:

$$h_{mn}(t) = a_{mn} \cdot \delta(t - t_{mn}) + \tilde{h}_{mn}(t) \tag{1}$$

where a_{mn} is the amplitude of the direct-path, t_{mn} denotes the propagation delay of the direct-path and $\tilde{h}_{mn}(t)$ is the remaining part of the impulse response. To simplify the derivations in this paper, we assume $a_{mn} \approx 1$. Combining this model for all microphones and sources, we can write in the frequency domain

$$\mathbf{H}[k] = \mathbf{\Delta}[k] + \tilde{\mathbf{H}}[k], \tag{2}$$

where $k = 1, \cdots, K$ is the frequency bin index. $\mathbf{H}[k] \in \mathbb{C}^{N \times N}$ is a square matrix of the room frequency responses from all sources to all microphones. $\mathbf{\Delta}[k] = \left[e^{j\omega_k t_{mn}}\right]$, where ω_k is the center frequency of the k-th frequency bin. $\tilde{\mathbf{H}}[k]$ models the remaining part of the room frequency response.

The mixing model in the time-frequency domain is

$$\mathbf{X}[k, l] \approx \mathbf{G}[k]\mathbf{H}[k]\mathbf{S}[k, l] \tag{3}$$

where l is the time frame index. $\mathbf{S}[k, l] \in \mathbb{C}^N$ and $\mathbf{X}[k, l] \in \mathbb{C}^N$ are two column vectors containing the short-time Fourier transform of the source and microphone signals, respectively. The diagonal matrix $\mathbf{G}[k] = \mathrm{diag}(g_1[k], \cdots, g_N[k]) \in \mathbb{R}^{N \times N}$ models possible microphone gain mismatch.

Using the model (3), we can apply any frequency domain ICA algorithm to separate the signals in the time-frequency domain:

$$\mathbf{Y}[k, l] = \mathbf{W}[k]\mathbf{X}[k, l]. \tag{4}$$

$\mathbf{W}[k]$ is a square demixing matrix to be determined. Since each frequency bin is treated independently from the others, a permutation and scaling ambiguity may occur at each frequency bin. To achieve a good separation, both ambiguities need to be resolved.

Assuming a successfull blind source separation, we obtain

$$\begin{aligned}\mathbf{W}[k] &= \mathbf{\Pi}[k]\mathbf{D}[k](\mathbf{G}[k]\mathbf{H}[k])^{-1} \\ &= \mathbf{\Pi}[k]\mathbf{D}[k] \cdot \mathbf{H}^{-1}[k] \cdot \mathbf{G}^{-1}[k] \\ &= \mathbf{\Pi}[k]\mathbf{D}[k] \cdot (\mathbf{\Delta}[k] + \tilde{\mathbf{H}}[k])^{-1} \cdot \mathbf{G}^{-1}[k].\end{aligned} \tag{5}$$

$\mathbf{\Pi}[k]$ is the permutation ambiguity matrix and $\mathbf{D}[k] = \mathrm{diag}(d_1[k], \cdots, d_N[k]) \in \mathbb{C}^{N \times N}$ is the scaling ambiguity matrix. $\mathbf{\Pi}[k]$ permutes the rows of $\mathbf{H}^{-1}[k]$, $\mathbf{D}[k]$ scales the rows of $\mathbf{H}^{-1}[k]$ and $\mathbf{G}^{-1}[k]$ scales the columns of $\mathbf{H}^{-1}[k]$.

Many algorithms have been proposed to solve the permutation problem, see [6] for an overview. The source scaling ambiguity is usually solved using the so-called minimum distortion principle (MDP) [7]. To the knowledge of the authors, there is no literature covering the effect of sensor gain mismatch on ICA.

To perform ICA based localization, we want to infer the positions of the sources by using $\mathbf{W}[k]$. This information is embedded in the phase of $\mathbf{\Delta}[k]$. The next section introduces the averaged directivity patterns (ADP) and the state coherence transform (SCT) for ICA based broadband localization averaging K frequency bins.

III. ADP and SCT

A. ADP

Directivity patterns of the demixing filters have been used for a long time to correct permutations in each frequency bin. [1] proposed ADP for 2D DOA estimation. This method is based on the observation that ICA forms spatial nulls to the position of the unwanted sources in order to suppress them. To perform localization, we infer the source positions from the positions of the spatial nulls. Strictly speaking, this holds only in anechoic environments, but the idea can also be applied to reverberant rooms when the direct path is dominant. The directivity pattern for each source n at position $\mathbf{p}$ is defined by the squared amplitude response of the n-th demixing filter, the n-th row of $\mathbf{W}[k]$:

$$B_n[k, \mathbf{p}] = \left| \sum_{i=1}^{N} w_{ni}[k] e^{-j\omega_k \tau_i(\mathbf{p})} \right|^2. \tag{6}$$

$\tau_i(\mathbf{p})$ is the time delay from a source at position $\mathbf{p}$ to microphone i. Each directivity pattern has in the ideal case $N-1$ spatial nulls or minima and hence would allow to localize $N-1$ sources. If we average the directivity patterns for all frequency bins and n, except for the directivity pattern with the highest amplitude for a given frequency bin [1], we obtain

$$n^*[k,\mathbf{p}] = \arg\max_n B_n[k,\mathbf{p}],$$

$$\tilde{B}(\mathbf{p}) = \frac{1}{NK}\sum_{k=1}^{K}\sum_{n\neq n^*[k,\mathbf{p}]} B_n[k,\mathbf{p}]. \quad (7)$$

$\tilde{B}(\mathbf{p})$ should ideally have minima at the positions of the N sources. [1] proposed to transform $\tilde{B}(\mathbf{p})$ using a nonlinear function in order to sharpen the minima:

$$B(\mathbf{p}) = \tanh\left(4\cdot\frac{\tilde{B}(\mathbf{p})-\min\tilde{B}(\mathbf{p})}{\max\tilde{B}(\mathbf{p})-\min\tilde{B}(\mathbf{p})}\right). \quad (8)$$

This contrast enhancement can be applied to all methods and does not change the position of the minima. To compare ADP and SCT we will use $\mathcal{H}_{\text{ADP}}(\mathbf{p}) = 1 - B(\mathbf{p})$ in Section IV.

B. SCT

The second approach for ICA based localization is to use the inverse of the estimated demixing matrix $\mathbf{W}^{-1}[k] \sim \mathbf{H}[k]$ instead of $\mathbf{W}[k] \sim \mathbf{H}^{-1}[k]$. By calculating $\mathbf{W}^{-1}[k]$, we can compare an estimated propagation model with an assumed propagation model using the SCT approach [2]. The main idea of the state coherence transform is to compare the "state" $e^{-j\omega\tau}$ from the propagation model $\mathbf{H}[k]$ against its estimate from the result of ICA. We define the state

$$r_{ab}[k,\mathbf{p}] = e^{-j\omega_k\tau_{ab}(\mathbf{p})}. \quad (9)$$

$\tau_{ab}(\mathbf{p})$ is the TDOA of a source at $\mathbf{p}$ observed at the microphone pair (a,b). For a microphone pair (a,b) with small enough distance, the impulse responses from the source to both microphones will look similar up to a delay τ_{ab} and an amplitude scaling. In terms of the frequency responses $H_i[k](i=a,b)$ from the source to both microphones, we can write

$$\frac{H_a[k]}{H_b[k]} = \frac{|H_a[k]|}{|H_b[k]|}e^{-j\omega_k\tau_{ab}(\mathbf{p})}. \quad (10)$$

Generally, we could also use the amplitude ratio $\frac{|H_a[k]|}{|H_b[k]|}$ for localization, but this ratio is much less reliable than TDOA, especially in reverberant environments and when the sources and microphones have directivity patterns. A comparison with (9) shows

$$r_{ab}[k,\mathbf{p}] = \exp\left\{j\arg\left(\frac{H_a[k]}{H_b[k]}\right)\right\} \quad (11)$$

Now we estimate this state for source n from the result of ICA

$$\hat{r}_{ab,n}[k] = \exp\left\{j\arg\left(\frac{[\mathbf{W}^{-1}[k]]_{an}}{[\mathbf{W}^{-1}[k]]_{bn}}\right)\right\}. \quad (12)$$

The notation $[\mathbf{A}]_{ij}$ denotes the (i,j)-th element of the matrix $\mathbf{A}$. Taking more than just one sensor pair into account, we define a state column vector and a corresponding estimate for certain sensor pairs $(a,b)\in\mathcal{I}\subseteq\{(a,b)|1\le a<b\le M\}$:

$$\mathbf{r}[k,\mathbf{p}] = [r_{ab}[k,\mathbf{p}]]_{(a,b)\in\mathcal{I}}, \quad \hat{\mathbf{r}}_n[k] = [\hat{r}_{ab,n}[k]]_{(a,b)\in\mathcal{I}}. \quad (13)$$

This leads to the objective function to be maximized

$$\mathcal{H}_{\text{SCT}}(\mathbf{p}) = \sum_{n=1}^{N}\sum_{k=1}^{K}\rho(\|\hat{\mathbf{r}}_n[k]-\mathbf{r}[k,\mathbf{p}]\|) \quad (14)$$

for a single source at $\mathbf{p}$. $\rho(t)$ is a locally confined kernel function that puts more weight on $t\approx 0$ and increases resolution.

The state estimates $\hat{r}_{ab,n}[k]$ calculated from $\mathbf{W}^{-1}[k] = \mathbf{G}[k]\mathbf{H}[k]\mathbf{D}^{-1}[k]\mathbf{\Pi}^{-1}[k]$ by using (12) are up to a permutation identical to those calculated from $\mathbf{W}^{-1}[k] = \mathbf{H}[k]$ as long as $\mathbf{G}[k]\in\mathbb{R}^{N\times N}$, i.e. the sensors have only amplitude and no phase mismatch. The scaling ambiguity $\mathbf{D}[k]$ scales each column of $\mathbf{H}[k]$ individually. This is irrelevant for the state estimate in (12). The gain mismatch $\mathbf{G}[k]$ affects the row scaling of $\mathbf{H}[k]$, but does not introduce any phase distortion. In other words, the SCT is robust to both scaling ambiguity and sensor gain mismatch.

C. Role of Nonlinear Summation

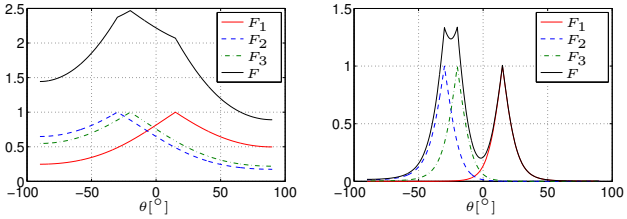
ADP uses a posterior contrast enhancement, i.e. applying a nonlinear function after the two summations in (7). This is quite different from the nonlinear transform in the SCT method [4] where it is done before the summation over n and k. In contrast to ADP (8), the transform in the SCT method can change the position of the maxima. It can be seen as a robust distance metric and serves to suppress outliers and to increase the resolution.

As an illustrating example, we consider a uniform linear array (ULA) with $M=3$ microphones and a microphone spacing of $d = 0.1\,\text{m}$ for 1D DOA estimation. $K=39$ frequency bins are equally spaced between 100 Hz and 4 kHz. Assuming an anechoic far-field mixing model, we can write $\mathbf{H}[k]$ as

$$\mathbf{H}[k] = \begin{bmatrix} 1 & 1 & 1 \\ e^{-j\omega_k\tau_1(\mathbf{p}_1)} & e^{-j\omega_k\tau_1(\mathbf{p}_2)} & e^{-j\omega_k\tau_1(\mathbf{p}_3)} \\ e^{-j\omega_k\tau_2(\mathbf{p}_1)} & e^{-j\omega_k\tau_2(\mathbf{p}_2)} & e^{-j\omega_k\tau_2(\mathbf{p}_3)} \end{bmatrix} \quad (15)$$

where $\tau_m(\mathbf{p}_n) = m\cdot dc^{-1}\sin(\theta_n)$, $\mathbf{p}_n=\theta_n$ is the DOA of source n and c is the sound propagation speed. We assume $\mathbf{W}[k]=\mathbf{H}^{-1}[k]$ for this section. The true DOAs are $\theta_{1,2,3} = -30°, -20°, 15°$.

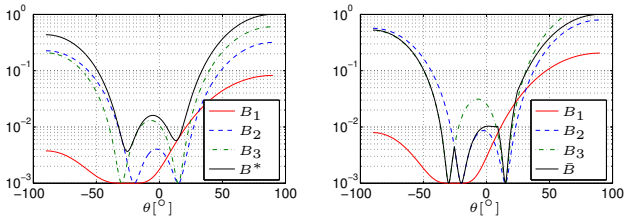
SCT: Fig. 1 shows the individual terms $F_n(\mathbf{p}) = \rho(\|\hat{\mathbf{r}}_n[k]-\mathbf{r}[k,\mathbf{p}]\|)$ for a single frequency bin k, as well as $F=\sum_n F_n$ for the case of $\rho(t)=1-t/2$ and $\rho(t)=1-\tanh(5\cdot t/2)$. Clearly, the nonlinear transform increases the localization resolution.



(a) $\rho(t)=1-t/2$ (b) $\rho(t)=1-\tanh(5\cdot t/2)$

Fig. 1: SCT for a single frequency bin

ADP: The ADP method implicitly contains a nonlinearity in the summation process: The exclusion of the directivity pattern with maximum amplitude from the summation in (7) is a nonlinear operation. This exclusion is necessary to obtain a function $\bar{B}[k,\mathbf{p}]$ with minima at the correct positions as shown in Fig.2 which plots $B_n[k,\mathbf{p}]$, $B^*=\sum_n B_n[k,\mathbf{p}]$ and $\bar{B}=\sum_{n\neq n^*[k,\mathbf{p}]}B_n[k,\mathbf{p}]$.



(a) without exclusion (b) with exclusion

Fig. 2: ADP for a single frequency bin

IV. Effect of the Different Ambiguities

A. Analysis of Permutation Ambiguity

If there was no permutation, we could estimate the location of source n by using the n-th row of $\mathbf{W}[k]$ along all frequency bins k. However, the permutation ambiguity changes the order of the rows of $\mathbf{W}[k]$ differently in each frequency bin k. If we do the estimation

step in each frequency bin individually and then combine the results, permutation ambiguity becomes relevant. If we, however, first sum over all sources and frequency bins and then do the estimation step, permutation ambiguity is completely irrelevant since the permutation only changes the order of the terms in the sum:

$$\sum_k \sum_n f\left([\mathbf{\Pi}[k]\mathbf{W}[k]]_{n:}\right) = \sum_k \sum_n f\left([\mathbf{W}[k]]_{n:}\right) \quad (16)$$

$$\sum_k \sum_n f\left([(\mathbf{\Pi}[k]\mathbf{W}[k])^{-1}]_{:n}\right) = \sum_k \sum_n f\left([\mathbf{W}^{-1}[k]]_{:n}\right) \quad (17)$$

$[\cdot]_{n:}$ selects the n-th row of a matrix, while $[\cdot]_{:n}$ selects the n-th column. The ADP cost function is of the type in (16), while the SCT cost function is of the type in (17). This means, the permutation ambiguity is irrelevant for both ADP and SCT.

B. Theoretical Analysis of Scaling Ambiguity and Sensor Gain

In this section, we want to theoretically study the effects of scaling ambiguity and sensor gain mismatch on ADP. In the following, we assume a direct-path only and neglect permutation. Furthermore, we consider a single frequency bin k:

$$\mathbf{W}[k] = \mathbf{D}[k](\mathbf{G}[k]\mathbf{\Delta}[k])^{-1} = \mathbf{D}[k]\mathbf{\Delta}^{-1}[k]\mathbf{G}^{-1}[k] \quad (18)$$

If $\mathbf{W}[k] = \mathbf{\Delta}^{-1}[k]$, then $B_n[k,\mathbf{p}]$ would have $M-1$ zeros at the correct locations of the sources. We found out:

- The scaling ambiguity $\mathbf{D}[k]$ changes the scaling of the individual directivity patterns $B_n[k,\mathbf{p}]$, but not the positions of the zeros of both the $B_n[k,\mathbf{p}]$ in (6) and the total ADP $B(\mathbf{p})$ in (8).
- The sensor gain mismatch $\mathbf{G}[k]$, however, changes the position of the zeros as we proof in the appendix. As a consequence, the ADP $B[\mathbf{p}]$ does not have its zeros at the correct locations if $\mathbf{G}[k] \neq c \cdot \mathbf{I}$.

We also perform a sensitivity analysis of the zero positions of $B_n[k,\mathbf{p}]$ with respect to sensor gain mismatch (see appendix). From the sensitivity analysis we conclude:

- The closer the sources (in relation to the microphone spacing), the more sensitive the positions of the zeros of $B_n[k,\mathbf{p}]$ with respect to sensor gain.
- The larger the sensor gain mismatch, the larger the change in zero positions of $B_n[k,\mathbf{p}]$.
- The sign of the change in zero positions depends on $g_q \gtrless 1$.

$\mathbf{G}[k] \neq c \cdot \mathbf{I}$ can destroy the ADP as shown in the next section.

C. Practical Analysis of Scaling Ambiguity and Sensor Gain

As shown in Sec. III-B, scaling ambiguity and sensor gain mismatch do not influence the SCT. Hence, we study only the effects on ADP and give a comparison of ADP and SCT later on. Below we use the exact model matrix $\mathbf{H}[k]$ from (15) for the calculation of $\mathbf{W}[k] = \mathbf{D}[k]\mathbf{H}^{-1}[k]\mathbf{G}^{-1}[k]$. Fig. 3 plots the averaged $\mathcal{H}_{\text{ADP}}(\mathbf{p})$ for $K = 39$ for different cases of $\mathbf{D}[k]$ and $\mathbf{G}[k]$. The true DOAs $\theta_{1,2,3} = -30°, -20°, 15°$ are marked with dashed vertical lines.

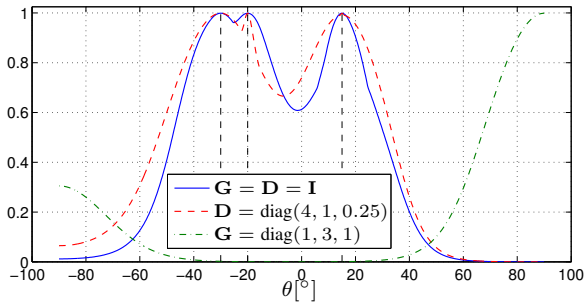


Fig. 3: ADP for different scenarios

We observe different effects on $\mathcal{H}_{\text{ADP}}(\mathbf{p})$:

- The scaling ambiguity matrix $\mathbf{D}[k]$ changes the norm of the rows of $\mathbf{H}^{-1}[k]$. This does not change the shape of the individual directivity patterns $B_n[k,\mathbf{p}]$ over $\mathbf{p}$, but their relative amplitude to each other and hence the determination of the row $n^*[k,\mathbf{p}]$ which is excluded from the overall ADP $\tilde{B}(\mathbf{p})$ in (7). The result is a modified $\mathcal{H}_{\text{ADP}}(\mathbf{p})$ as shown in Fig. 3.
- The sensor gain matrix $\mathbf{G}[k]$ changes the scaling of the columns of $\mathbf{H}^{-1}[k]$. Since the elements $w_{ni}[k]$ in (6) are subject to different column scalings, each individual directivity pattern $B_n[k,\mathbf{p}]$ changes significantly. The result is a serious change of $\mathcal{H}_{\text{ADP}}(\mathbf{p})$ as shown in Fig. 3 where a gain mismatch of factor 3 in the second microphone completely destroys $\mathcal{H}_{\text{ADP}}(\mathbf{p})$.

D. Different Normalization Methods for ADP

To improve the robustness of ADP with respect to scaling ambiguity and sensor gain mismatch, we can use different normalization strategies. These normalizations are proposals by the authors except for the first one (MDP).

1) Minimum Distortion Principle (MDP): Generally, frequency domain ICA methods remove $\mathbf{D}[k]$ using MDP [7]:

$$\begin{aligned}\mathbf{W}[k] &\leftarrow \text{diag}(\mathbf{W}^{-1}[k])\mathbf{W}[k] \\ &= \mathbf{G}[k]\text{diag}(\mathbf{H}[k])\mathbf{D}^{-1}[k]\mathbf{D}[k]\mathbf{H}^{-1}[k]\mathbf{G}^{-1}[k] \\ &= \mathbf{G}[k]\text{diag}(\mathbf{H}[k])\mathbf{H}^{-1}[k]\mathbf{G}^{-1}[k]. \end{aligned} \quad (19)$$

However, this operation introduces the sensor gain mismatch $\mathbf{G}[k]$ as a row scaling matrix. The net effect is to replace the scaling ambiguity matrix $\mathbf{D}[k]$ by the sensor gain mismatch $\mathbf{G}[k]$. Hence this normalization method does not work in the presence of sensor gain mismatch.

2) Row normalization of $\mathbf{W}$:

$$\mathbf{w}_n[k] \leftarrow \frac{\mathbf{w}_n[k]}{\|\mathbf{w}_n[k]\|} e^{-j \arg w_{nn}[k]} \quad (20)$$

$\mathbf{w}_n[k]$ is the n-th row of $\mathbf{W}[k]$. This normalization compensates $\mathbf{D}[k]$ but not $\mathbf{G}[k]$. It avoids the calculation of $\mathbf{W}^{-1}[k]$. In principle, we can use any vector norm $\|\mathbf{w}_n[k]\|$. In our experiments, we used the l_1-norm.

3) Row and column normalization of $\mathbf{W}$: Similar to the row normalization, we can use a joint row and column normalization to simultaneously normalize the rows and columns of $\mathbf{W}[k]$ to unit norm. By this procedure, we aim to reduce the influence of the scaling ambiguity $\mathbf{D}[k]$ and the differing row norms of $\mathbf{H}^{-1}[k]$ as well as the sensor gain mismatch $\mathbf{G}[k]$. To achieve a joint row and column normalization

$$\mathbf{W}[k] \leftarrow \text{diag}(\beta_1, \cdots, \beta_N)\mathbf{W}[k]\text{diag}(\gamma_1, \cdots, \gamma_N), \quad (21)$$

we need to solve a set of nonlinear equations such that the l_1-row-norm and the l_1-column-norm of $\mathbf{W}[k]$ become unit after the normalization, i.e.

$$\beta_m \cdot \sum_{n=1}^{N} |w_{mn}|\gamma_n = 1, \quad \gamma_n \cdot \sum_{m=1}^{N} |w_{mn}|\beta_m = 1, \quad (22)$$

for all $n = 1, \cdots, N, m = 1, \cdots, N$. To find the normalization weights β_m and γ_n we use a numerical equation solver.

Afterwards, we make the diagonal elements of $\mathbf{W}[k]$ real by

$$\mathbf{W}[k] \leftarrow \text{diag}(e^{-j \arg w_{11}[k]}, \cdots, e^{-j \arg w_{NN}[k]})\mathbf{W}[k] \quad (23)$$

4) Phase normalization of $\mathbf{W}^{-1}$: From (5), we see that the TDOA of the sources is embedded in the phase of $\mathbf{\Delta}[k]$. Hence, we should use only the phase from $\mathbf{W}^{-1}$

$$\hat{\mathbf{H}}[k] = \mathbf{W}^{-1}[k], \ \hat{\mathbf{\Delta}}[k] = \left[e^{j \arg\left(\frac{\hat{h}_{mn}}{\hat{h}_{Jn}}\right)} \right]_{\substack{1 \le m \le N \\ 1 \le n \le N}},$$

$$\tilde{\mathbf{W}}[k] = \hat{\mathbf{\Delta}}^{-1}[k], \quad (24)$$

where J denotes a reference microphone. After the calculation of $\tilde{\mathbf{W}}[k]$, we perform row normalization as in (20) and use this result to compute ADP as previously.

E. Effect of Normalization Methods on ADP

The MDP normalization and the row normalization do not correct the sensor gain mismatch. Hence, Fig. 4 compares only the original ADP (orig ADP), ADP with row and column normalization (r&c ADP), ADP with phase normalization (phase ADP) and SCT for $\mathbf{W}$ estimated by ICA [8]. We have considered the scenario with $\mathbf{G} = \mathrm{diag}(1, 3, 1)$ from before. We used speech signals with a sampling rate of 8 kHz and 128 equally spaced frequency bins for averaging in ADP and SCT.

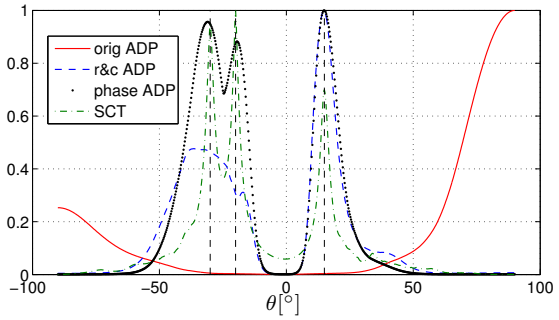


Fig. 4: Comparison of ADP and SCT for $\mathbf{W}$ estimated by ICA

As shown in Fig. 4, in the presence of sensor gain mismatch, SCT clearly outperforms ADP without proper normalization: SCT shows its peaks at the correct DOA values, while the original ADP and ADP with row and column normalization fail to detect all sources and show peaks at wrong DOA values. The reasons of the superiority of SCT over ADP are that, by working on $\mathbf{W}^{-1}$ instead of $\mathbf{W}$, SCT focuses on phase information of $\mathbf{H}$ only which allows to compensate both scaling ambiguity $\mathbf{D}[k]$ and sensor gain mismatch $\mathbf{G}[k]$. If we use $\mathbf{W} \sim \mathbf{H}^{-1}$, we cannot compensate the sensor gain mismatch $\mathbf{G}[k]$ since $\mathbf{W}$ contains the TDOA information coupled in the phase and amplitude. ADP with phase normalization (24) performs comparable to SCT but requires two inversions of $\mathbf{W}$ while SCT needs only one. Furthermore, the nonlinear transform in SCT yields an increased spatial resolution.

As discussed in Section IV-B, the effects of sensor gain mismatch depend on the spacing of the sources in relation to the microphone spacing and on the amount of gain mismatch.

V. Conclusion

In this paper, we have studied the influence of permutation ambiguity, scaling ambiguity and sensor gain mismatch on ICA based source localization. We have shown that the permutation ambiguity is irrelevant for source localization if we consider all frequency bins and sources simultaneously. If we work with the demixing matrix $\mathbf{W}$ without proper normalization, localization performance is poor. The scaling ambiguity can be removed from $\mathbf{W}$ but not the sensor gain mismatch. Working with $\mathbf{W}^{-1}$ allows a source localization that is independent of scaling ambiguity and sensor gain mismatch. We have verified these findings by comparing two different localization approaches: ADP (working on $\mathbf{W}$) and SCT (working on $\mathbf{W}^{-1}$). SCT is robust to both scaling ambiguity and sensor gain mismatch and shows a superior performance in comparison to ADP without proper normalization. If we derive a normalized $\mathbf{W}$ by just taking into account the phase of $\mathbf{W}^{-1}$, ADP performs comparably to SCT.

Appendix

A. Positions of Zeros of ADP

For a given frequency bin k, let $\mathbf{W} = \mathbf{\Delta}^{-1}\mathbf{G}^{-1}$ be the perfect estimate of the inverse mixing matrix $\mathbf{\Delta}^{-1}$ except for the sensor gain matrix $\mathbf{G} = \mathrm{diag}(g_1, \cdots, g_N)$. We now evaluate the values of the N individual directivity patterns $B_n[k, \mathbf{p}], 1 \le n \le N$ from (6) at the N true source positions $\mathbf{p}_m$. Clearly, $B_n[k, \mathbf{p}_m] = |B_{nm}|^2$ where B_{nm} is the (n, m)-th element of $\mathbf{B} = \mathbf{W}\mathbf{\Delta} = \mathbf{\Delta}^{-1}\mathbf{G}^{-1}\mathbf{\Delta}$. If there is no sensor gain mismatch, i.e. $\mathbf{G} = c \cdot \mathbf{I}$, all off-diagonal elements of $\mathbf{B}$ are zero. This means $B_n[k, \mathbf{p}_m] = 0\ \forall m \neq n$ and each directivity pattern shows $N - 1$ zeros at the position of the interfering sources. Hence, $\bar{B}(\mathbf{p}) = \sum_{n \neq n^*[k,\mathbf{p}]} B_n[\mathbf{p}]$ is zero at all N source positions.

If, however, $\mathbf{G} \neq c \cdot \mathbf{I}$ due to sensor gain mismatch, there is in each column of $\mathbf{B}$ at least one non-zero off-diagonal element. This can be easily shown by a counter proof. If $\mathbf{B} = \mathrm{diag}(b_1, \cdots, b_N)$ was diagonal, then

$$b_m \left[\delta_{1m}, \cdots, \delta_{Nm}\right]^T = \left[\delta_{1m}/g_1, \cdots \delta_{Nm}/g_N\right]^T \tag{25}$$

would follow for the m-th column of both sides of $\mathbf{\Delta}\mathbf{B} = \mathbf{G}^{-1}\mathbf{\Delta}$. This implies $g_1 = \cdots = g_N$ and hence $\mathbf{G} = c \cdot \mathbf{I}$, which violates the assumption $\mathbf{G} \neq c \cdot \mathbf{I}$.

The non-zero off-diagonal elements in each column of $\mathbf{B}$ mean that for each source position $\mathbf{p}_m$, at least one $B_n[k, \mathbf{p}_m]$ with $n \neq m$ is non-zero in addition to $B_m[k, \mathbf{p}_m] \neq 0$. As a result, $\bar{B}(\mathbf{p}_m) = \sum_{n \neq n^*[k,\mathbf{p}_m]} B_n[k, \mathbf{p}_m] \neq 0\ \forall m$. In other words, the ADP function in (7) loses its zero-value property at all true source positions.

B. Sensitivity Analysis of Zeros of $B_n[k, \mathbf{p}]$

Assuming an ULA and taking the first microphone as reference, we can write $B_n[k, \mathbf{p}] = |B(z)|^2 = |\sum_{i=1}^{N} w_{ni} z^{-(i-1)}|^2$ with $z = e^{j\omega_k \tau(\mathbf{p})}$ for a single frequency bin k. $\tau(\mathbf{p})$ is the TDOA between neighbouring microphones in the ULA. We now find the sensitivity of $B(z)$ to sensor gain mismatch by calculating $\frac{\partial B(z)}{\partial w_{nq}}$:

$$B(z) = \sum_{i=1}^{N} w_{ni} z^{-(i-1)} = w_{n1} \prod_{i=1}^{N-1} (1 - z_{ni} \cdot z^{-1})$$

$$\left.\frac{\partial B(z)}{\partial w_{ni}}\right|_{z=z_{nq}} = z_{nq}^{-(i-1)} = \left.\frac{\partial B(z)}{\partial z_{nq}}\right|_{z=z_{nq}} \cdot \frac{\partial z_{nq}}{\partial w_{ni}} \tag{26}$$

Solving for the sensitivity $\frac{\partial z_{nq}}{\partial w_{ni}}$ of the zeros z_{nq} of $B(z)$, we get

$$\frac{\partial z_{nq}}{\partial w_{ni}} = -\frac{z_{nq}^{N-i}}{w_{n1} \cdot \prod_{\nu \neq q}(z_{nq} - z_{n\nu})}. \tag{27}$$

The change in zero positions Δz_{nq} for a small change $\Delta w_{ni} = w_{ni} \cdot (1/g_i - 1)$ of w_{ni} due to sensor gain mismatch depends on how close the zeros are ($z_{nq} \approx z_{n\nu}$) and on the amount of sensor gain mismatch $(1/g_i - 1)$. Furthermore the sign of the change depends on $g_i \gtrless 1$.

The zero positions z_{nq} of $B(z)$ are related to the zero positions of $B_n[k, \mathbf{p}]$ by $z = e^{j\omega_k \tau(\mathbf{p})}$. Closely spaced sources $\mathbf{p}_q \approx \mathbf{p}_\nu$ result in zeros $z_{nq} \approx z_{n\nu}$ and hence a large Δz_{nq} if $\mathbf{G} \neq c \cdot \mathbf{I}$.

References

[1] A. Lombard, T. Rosenkranz, H. Buchner, and W. Kellermann, "Multidimensional localization of multiple sound sources using averaged directivity patterns of blind source separation systems," *Proc. ICASSP*, 2009.

[2] B. Loesch, S. Uhlich, and B. Yang, "Multidimensional localization of multiple sound sources using frequency domain ICA and an extended state coherence transform," *Proc. SSP*, 2009.

[3] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. Brandstein, Ed. Springer Verlag, 2001.

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

[5] F. Nesta, P. Svaizer, and M. Omologo, "Cumulative state coherence transform for a robust two-channel multiple source localization," *Proc. ICA*, 2009.

[6] M. Pedersen, J. Larsen, U. Kjems, and L. Parra, "Convolutive blind source separation methods," in *Springer Handbook on Speech Processing and Speech Communication*, J. Benesty, Y. Huang, and M. Sondhi, Eds., Mar. 2008, ch. 1, pp. 1–34.

[7] K. Matsuoka, "Minimal distortion principle for blind source separation," *Proc. of the 41st SICE Annual Conference*, 2002.

[8] S. Douglas and M. Gupta, "Scaled natural gradient algorithms for instantaneous and convolutive blind source separation," *Proc. ICASSP*, 2007.