POST-PROCESSING FOR BSS ALGORITHMS TO RECOVER SPATIAL CUES

S. Wehr, M. Zourub, R. Aichner, W. Kellermann

University Erlangen-Nürnberg
Multimedia Communications and Signal Processing
Cauerstraße 7, 91058 Erlangen, Germany
{Wehr,Zourub,Aichner,WK}@LNT.de

ABSTRACT

This paper addresses the problem of recovering spatial cues after microphone array processing by blind source separation. Based on the known demixing system determined by the blind source separation, we derive two spatialization algorithms. One algorithm exploits the inverse of the demixing system, while the other algorithm exploits the adjoint of the demixing system. Both algorithms are evaluated by objective and subjective measures. We therefore consider the recovered time difference of arrival and the subjective perception of the spatialized signals.

1. INTRODUCTION

This paper is motivated by the so-called *cocktail-party problem* which arises when mixtures of multiple simultaneously active speakers are recorded by multiple microphones. In many applications (e.g. hands-free human-machine interfaces, [1]), we need to focus on one single source and try to suppress interfering sources. We address this problem here by *blind source separation* (BSS) algorithms which can deal well with unknown microphone and source positions [2]. Furthermore, BSS provides us with separated source signals which may be individually selected for further processing. Unfortunately, the spatial cues of the output signals are lost as conventional BSS algorithms provide a monaural representation of each separated output. In this paper, we propose and compare two algorithms, which are able to recover the spatial cues in the BSS output signals by post-filtering. For this purpose, we exploit the demixing filters obtained by the BSS algorithm. For example, other approaches recover spatial cues based on the input and output signals of the separation system [3], preserve the binaural cues during noise reduction [4], or model the binaural cues by head-related impulse responses [5].

Figure 1 illustrates our concept of post-filtering the BSS output signals in order to recover the spatial cues. Throughout this section, we derive the algorithms in the DFT domain and we omit the time-dependency. Furthermore, we assume a maximum number of Q simultaneously active sources and Q microphones. The $Q \times 1$ column vectors $\mathbf{s}(\omega)$, $\mathbf{x}(\omega)$ and $\mathbf{y}(\omega)$ capture the source signals $s_i(\omega)$, the sensor signals $x_i(\omega)$, and the BSS output signals $y_i(\omega)$, respectively ($i = 1, \ldots, Q$). The $Q \times Q$ matrices $\mathbf{H}(\omega)$ and $\mathbf{W}(\omega)$ represent the unknown mixing system in the form of room impulse responses $h_{ij}(\omega)$ and the demixing system determined by BSS in the form of FIR filters $w_{ij}(\omega)$ ($j = 1, \ldots, Q$). Our proposed spatialization concept is not restricted to a specific BSS algorithm. We selected the time-domain BSS algorithm described in [6].

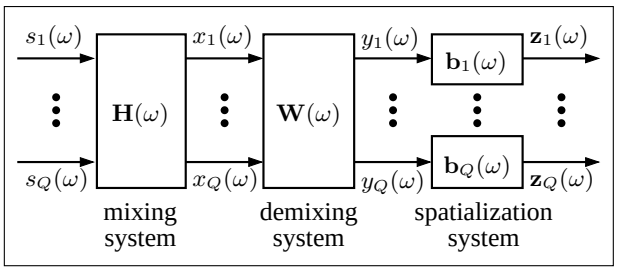

Figure 1: Q-channel mixing and demixing system with post-filtering for spatialization

In the DFT domain, the BSS output signals are

$$\begin{bmatrix} y_1 \\ \vdots \\ y_Q \end{bmatrix} = \begin{bmatrix} w_{11} \cdots w_{1Q} \\ \vdots \ \ddots \ \vdots \\ w_{Q1} \cdots w_{QQ} \end{bmatrix} \cdot \begin{bmatrix} h_{11} \cdots h_{1Q} \\ \vdots \ \ddots \ \vdots \\ h_{Q1} \cdots h_{QQ} \end{bmatrix} \cdot \begin{bmatrix} s_1 \\ \vdots \\ s_Q \end{bmatrix}. \quad (1)$$

For brevity, we have omitted (ω). By filtering the BSS-output $y_i(\omega)$ with the spatialization filters $b_{ji}(\omega)$, we (approximately) recover the spatial cues given by the unknown mixing system. The $Q \times 1$ column vector $\mathbf{z}_i(\omega)$ captures the Q spatialized channels $z_{ji}(\omega)$ derived from BSS-output $y_i(\omega)$. The necessary spatialization filters $b_{ji}(\omega)$ are captured in the $Q \times 1$ column vector $\mathbf{b}_i(\omega)$. Then, the spatialized output signal $z_{ji}(\omega)$ is given by

$$z_{ji}(\omega) = b_{ji}(\omega) \cdot y_i(\omega). \quad (2)$$

The problem of determining the spatialization filters $b_{ji}(\omega)$ is addressed in this paper, which is structured as follows: In Section 2, we first illustrate how the unknown mixing system – and thus the spatial cues – can be approx-

imated by the known demixing system. This approximation then allows the determination of the spatialization filters by exploiting the *inverse* and *adjoint* demixing matrix (Sec. 2.2, 2.3). After presenting the results of objective and subjective evaluations in Section 3, we finally draw conclusions in Section 4.

2. RECOVERING SPATIAL CUES

In this section, we first approximate the unknown mixing system $\mathbf{H}(\omega)$ based on the available demixing system $\mathbf{W}(\omega)$. We then introduce a spatialization approach which determines the spatialization filters based on the *inverse demixing matrix*. The second considered algorithm derives the spatialization filters from the *adjoint demixing matrix*.

2.1. Mixing System Approximation

Here, we introduce an approximation of the unknown mixing system $\mathbf{H}(\omega)$ based on the available demixing system $\mathbf{W}(\omega)$.

Referring to (1), we substitute the propagation from the sources to the BSS outputs by the $Q \times Q$ matrix $\mathbf{C}(\omega)$:

$$\mathbf{C}(\omega) = \mathbf{W}(\omega) \cdot \mathbf{H}(\omega) = \begin{bmatrix} c_{11}(\omega) \cdots c_{1Q}(\omega) \\ \vdots \quad \ddots \quad \vdots \\ c_{Q1}(\omega) \cdots c_{QQ}(\omega) \end{bmatrix}. \quad (3)$$

Neglecting permutations of the BSS output signals and assuming perfect source separation, matrix $\mathbf{C}(\omega)$ simplifies to

$$\text{offdiag}\{\mathbf{C}(\omega)\} = 0 \quad (4)$$

and the output signals are then

$$y_i(\omega) = c_{ii}(\omega) \cdot s_i(\omega), \quad (5)$$

where the operator offdiag$\{\mathbf{C}(\omega)\}$ returns the off-diagonal elements of matrix $\mathbf{C}(\omega)$.

Assuming invertibility of matrix $\mathbf{W}(\omega)$, we may multiply (3) with the inverse of matrix $\mathbf{W}(\omega)$ and we thus obtain

$$\mathbf{H}(\omega) = \mathbf{W}^{-1}(\omega) \cdot \mathbf{C}(\omega) = \frac{\text{adj}\{\mathbf{W}(\omega)\}}{\det\{\mathbf{W}(\omega)\}} \cdot \mathbf{C}(\omega). \quad (6)$$

The required invertibility of matrix $\mathbf{W}(\omega)$ is addressed in Sections 2.2 and 2.3. For perfect separation, i.e., when (4) is fulfilled, we may reformulate (6) to:

$$\frac{1}{\det\{\mathbf{W}(\omega)\}} \left[\text{adj}\{\mathbf{W}(\omega)\}\right]_{ij} = \frac{1}{c_{jj}(\omega)} h_{ij}(\omega). \quad (7)$$

The operator $[.]_{ij}$ returns element ij of a matrix. With (7), we may exploit the known demixing system $\mathbf{W}(\omega)$ in order to approximately recover the spatial cues given by the unknown mixing matrix $\mathbf{H}(\omega)$ (Sections 2.2, 2.3). Note that the matrix $\mathbf{C}(\omega)$ is unknown but inherently given by the BSS output signals.

2.2. Inverse Demixing Matrix

We now present an algorithm which defines the spatialization filters based on the *inverse* of the demixing matrix.

Firstly, we assume perfect source separation. With (2), (5) and (7), filtering the output $y_i(\omega)$ with the spatialization filter

$$b_{ji}(\omega) = \left[\mathbf{W}^{-1}(\omega)\right]_{ji} \quad (8)$$

yields the spatialized signal $z_{ji}(\omega)$:

$$z_{ji}(\omega) = \left[\mathbf{W}^{-1}(\omega)\right]_{ji} \cdot y_i(\omega) = h_{ji}(\omega) \cdot s_i(\omega). \quad (9)$$

Here, the spatialization filter equalizes the unknown filtering by $\mathbf{C}(\omega)$ and it exactly recovers the spatial information given by the unknown mixing matrix $\mathbf{H}(\omega)$. Using the inverse of the demixing system has been already applied for monaural BSS in [7].

Secondly, we consider non-perfect source separation and we therefore investigate the effect of the spatialization filters defined by (8) on the BSS output signals. Multiplying (3) with the inverse mixing system and inverting again, we can reformulate (3) leading to

$$\mathbf{W}^{-1}(\omega) = \mathbf{H}(\omega)\mathbf{C}^{-1}(\omega). \quad (10)$$

According to (9), the spatialized signal $z_{ji}(\omega)$ is

$$z_{ji}(\omega) = [\mathbf{H}(\omega)\mathbf{C}^{-1}(\omega)]_{ji} \cdot \sum_{q=1}^{Q} c_{iq}(\omega) s_q(\omega). \quad (11)$$

We illustrate (11) by calculating the spatialized output signal $z_{11}(\omega)$ for $Q = 2$:

$$\begin{aligned} z_{11} &= \frac{c_{11}c_{22}h_{11} - c_{11}c_{21}h_{12}}{c_{11}c_{22} - c_{12}c_{21}} \, s_1 \\ &+ \frac{c_{12}c_{22}h_{11} - c_{12}c_{21}h_{12}}{c_{11}c_{22} - c_{12}c_{21}} \, s_2 \end{aligned} \quad (12)$$

For brevity, we have omitted (ω). In the case of sufficient source separation, i.e. offdiag$\{\mathbf{C}(\omega)\} \approx 0$ (see Eq. (4)), we may assume

$$||c_{11}c_{22}|| >> ||c_{12}c_{21}||, \quad (13)$$

i.e. the suppressed sources do not significantly contribute to the BSS output signals. Hence, (12) simplifies to

$$z_{11} = h_{11}s_1 + \frac{c_{12}}{c_{11}} h_{11}s_2 - \frac{c_{21}}{c_{22}} h_{12}s_1 \quad (14)$$

As desired, the spatialization filter recovers the spatial cues given by the mixing system for source signal $s_1(\omega)$ (term $h_{11}s_1$). Unfortunately, these spatial cues are also imposed on source signal $s_2(\omega)$, which is additionally

attenuated (term $\frac{c_{12}}{c_{11}} h_{11} s_2$). The most disturbing phenomenon is caused by the term $\frac{c_{21}}{c_{22}} h_{12} s_1$: Source 1 is perceived attenuated at the location of source 2.

Note that requiring invertibility of the demixing matrix restricts this approach defined by (8). If the determinant of the demixing matrix $\det\{\mathbf{W}(\omega)\}$ equals zero, we are not able to calculate the spatialization filters for this frequency component. We therefore compute the matrix inverse by regularizing the determinant:

$$\widetilde{\mathbf{W}}^{-1}(\omega) = \frac{\text{adj}\,\{\mathbf{W}(\omega)\}}{\det\{\mathbf{W}(\omega)\} + \varepsilon}. \tag{15}$$

ε denotes a small constant which avoids divisions by zero and stability problems due to very small values of the determinant.

2.3. Adjoint Demixing Matrix

Motivated by the optimum demixing filters $\mathbf{W}_{\text{opt}}(\omega) = \text{adj}\,\{\mathbf{H}(\omega)\}$, which may be derived from [8], we now define the spatialization filters based on the *adjoint* of the demixing matrix.

Again, we firstly assume perfect source separation. Referring to (2), (5) and (7) and filtering the output $y_i(\omega)$ with the spatialization filter

$$b_{ji}(\omega) = [\text{adj}\,\{\mathbf{W}(\omega)\}]_{ji} \tag{16}$$

yields the spatialized signal

$$\begin{aligned} z_{ji}(\omega) &= [\text{adj}\,\{\mathbf{W}(\omega)\}]_{ji} \cdot y_i(\omega) \\ &= \det\{\mathbf{W}(\omega)\}\, h_{ji}(\omega) \cdot s_i(\omega). \end{aligned} \tag{17}$$

Comparing (17) with (9), we notice that the spatialization filter perfectly recovers the spatial cues given by the filter $h_{ji}(\omega)$, except for the scaling factor $\det\{\mathbf{W}(\omega)\}$.

We now consider the spatialized output signals in the case of non-perfect source separation. Filtering the output signal $y_i(\omega)$ with the spatialization filter defined by (16) yields

$$z_{ji}(\omega) = [\text{adj}\,\{\mathbf{W}(\omega)\}]_{ji} \cdot \sum_{q=1}^{Q} c_{iq}(\omega) s_q(\omega). \tag{18}$$

Calculating again the spatialized output signal $z_{11}(\omega)$ for $Q = 2$ illustrates (18).

$$z_{11}(\omega) = \det\{\mathbf{W}(\omega)\}\, h_{11}(\omega) \left(s_1(\omega) + \frac{c_{12}(\omega)}{c_{11}(\omega)} s_2(\omega) \right) \tag{19}$$

Except for the scaling with $\det\{\mathbf{W}(\omega)\}$, the spatialization filter perfectly recovers the spatial cues given by the mixing system for source signal $s_1(\omega)$ (term $h_{11}(\omega)s_1$). Source 2 is attenuated by the factor $\det\{\mathbf{W}(\omega)\}\,\frac{c_{12}(\omega)}{c_{11}(\omega)}$ and it is perceived at the same location as source 1. In contrast to the approach described in Section 2.2, we do not perceive source 1 at two locations. This significantly improves the perception of the spatialized signals.

Equation (19) illustrates that this algorithm is robust to not invertible demixing matrices. If the demixing matrix is not invertible in a specific frequency bin, the determinant becomes zero. Therefore, the corresponding frequency component in the spatialized signal also equals zero. Regularization as described in Section 2.2 is not necessary.

3. SIMULATIONS

We now evaluate the two algorithms which were described in Sections 2.2 and 2.3. Firstly, we describe the simulation setup. Secondly, we consider two quality criteria: On the one hand, we determine the *time difference of arrival* (TDOA) based on the spatialized signals. This objective measure illustrates the performance of both algorithms in terms of recovering spatial cues. On the other hand, we describe the subjective perception of the spatialized signals.

3.1. Setup

In a low-echoic chamber with reverberation time $T_{60} \approx 50$ms, we set up two microphones and two loudspeakers in front of the microphones. Two clean speech signals were played back by the loudspeakers and recorded by the two microphones. Both recorded microphone signals were processed by the BSS algorithm and by the two investigated spatialization algorithms, which were all implemented in Matlab. With this setup, we obtain as true TDOAs corresponding to the two sources: $\text{TDOA}_{1,2} = \pm 8.15$ samples.

3.2. Recovered TDOA

We now determine the TDOAs of both spatialized signals. Therefore, we compute the cross-correlation between the two channels of each spatialized signal and for each spatialization algorithm. Additionally, we compute the cross-correlation between the two microphone signals. Figure 2 shows the obtained cross-correlations.

In the top plot, we see the cross-correlation of the two microphone signals. Although considering only a low-echoic acoustic environment, we are able to approximately localize only one source. There is no significant peak in the cross-correlation corresponding to the second source.

The center plot and the bottom plot show the cross-correlation between the two channels of the spatialized signals. Both spatialization algorithms yield significant

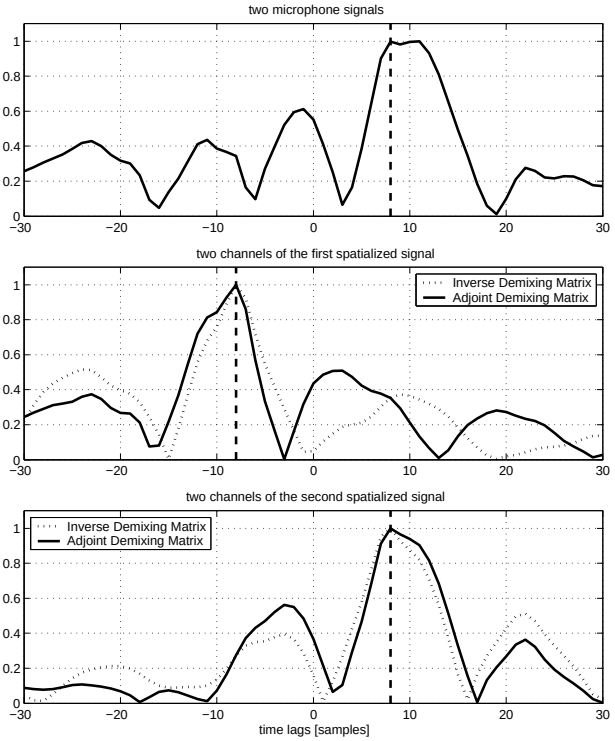

Figure 2: Normalized cross-correlation; Correlated signals are given in the titles; Dashed vertical lines mark localized sources

cross-correlation peaks. Therefore, we are able to localize both sources with the correct TDOA.

These results illustrate that – in terms of source localization by exploiting the cross-correlation – the spatial cues of both sources are correctly recovered by both spatialization algorithms.

3.3. Subjective Perception

By informal listening tests, we investigated the subjective perception of the spatialized signals. Both spatialization algorithms result in perceiving the suppressed source at the same location as the emphasized source. As long as the source separation performs well, this phenomenon hardly degrades the subjective perception.

As suggested by the theoretical considerations in Section 2.2, incorporating the inverse demixing matrix yields insufficient spatialization results for imperfect separation. Perceiving the emphasized source both in the correct direction and in the direction of the suppressed source suggests a highly reverberated acoustic environment, even in the low-echoic chamber. The subjective perception is thus significantly degraded.

The spatialization algorithm which incorporates the adjoint demixing (Section 2.3) results in subjectively well-spatialized signals. The emphasized source can be acoustically localized by the listener, with the perceived DOA matching the true DOA given by the simulation setup. Furthermore, we noticed that neither scaling the suppressed signal (Eq. (19): $\frac{c_{12}}{c_{11}}$) nor multiplying both signals with the potentially small determinant (Eq. (19)) leads to any audible degradation the subjective perception.

4. CONCLUSIONS

In this paper, we have presented two algorithms which recover the spatial cues in the output signals of blind source separation. We firstly introduced an approximation of the unknown mixing system by exploiting the demixing system of BSS. Based on this approximation, we derived an adjoint-based algorithm and an inverse-based algorithm, which both recover the spatial cues in the BSS outputs by appropriate post-filtering. This allows source localization based on the cross-correlations of the spatialized signals. In terms of subjective perception, the adjoint-based algorithm outperforms the inverse-based algorithm.

5. REFERENCES

[1] M.S. Brandstein and D.B. Ward (eds.), *Microphone Arrays: Signal Processing Techniques and Application*, Springer-Verlag, Berlin, May 2001.

[2] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, 2001.

[3] T. Takatani, S. Ukai, T. Nishikawa, H. Saruwatari, and K. Shikano, "Evaluation of SIMO separation methods for blind decomposition of binaural mixed signals," *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, pp. 233–236, Sep. 2005.

[4] S. Doclo, R. Dong, T. Klasen, J. Wouters, S. Haykin, and M. Moonen, "Extension of the multi-channel Wiener filter with localisation cues for noise reduction in binaural hearing aids," *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, pp. 221–224, Sep. 2005.

[5] N. Adams and G. Wakefield, "The binaural display of clouds of point sources," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2005.

[6] R. Aichner, H. Buchner, and W. Kellermann, "A novel normalization and regularization scheme for broadband convolutive blind source separation," *International Symposium on Independent Component Analysis and Blind Signal Separation (ICA)*, pp. 527–535, Mar. 2006.

[7] S. Ikeda and N. Murata, "A method of ICA in time–frequency domain," *International Symposium on Independent Component Analysis and Blind Signal Separation (ICA)*, pp. 365–371, Jan. 1999.

[8] H. Buchner, R. Aichner, and W. Kellermann, "Relation between blind system identification and convolutive blind source separation," *Joint Workshop for Hands-Free Speech Communication and Microphone Arrays (HSCMA)*, Mar. 2005.