Improved Frequency Domain Blind Source Separation for Audio Signals via Direction of Arrival Knowledge

Ricardo K. Miranda\textsuperscript{a}, Soyuj K. Sahoo\textsuperscript{b}, Ricardo Zelenovsky\textsuperscript{a} and João Paulo C. L. da Costa\textsuperscript{a}

\textsuperscript{a}Department of Electrical Engineering, University of Brasília, Brasília, Brazil
\textsuperscript{b}Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg, Erlangen, Germany

rickehrle@gmail.com, sahoo@LNT.de, \{zele, joaopaulo.dacosta\}@ene.unb.br

\textbf{ABSTRACT:} The sound pollution is a common and serious problem in modern cities. Examples of sources of sound pollution are other speakers, traffic and household devices such as vacuum cleaner, air conditioner, TV and radios. These sound interference sources degrade the performance of audio devices such as hearing aids, smart TVs and forensic recorders. One way to tackle this problem is via Blind Source Separation (BSS) algorithms using microphone arrays. In this paper, we propose a technique to improve the performance of frequency domain BSS (FDBSS) by taking into account the direction of arrival (DOA) knowledge. With our proposed scheme, a significant improvement is obtained for end fire angles.

\section{Introduction}

A typical scenario for a speech and audio Blind Source Separation (BSS) is obtained when two sound sources are placed in the same room, both simultaneously active. The sources interfere one another and degrade the performance of audio devices such as hearing aids, smart TVs and forensic recorders, etc., [1–3]. The goal of the BSS algorithms is to reduce the power of one of the sources in order to focus on only one, the desired signal. Various works proposed solutions for this problem. In [4], the solution is derived based on the maximization of the mutual information (MI), [5] suggests the minimization of the Kullback-Leibler divergence in a similar way and [6] combines three properties of the signals to form a generalized cost function and minimizes it in order to exploit the statistical independence between the source signals.

However, such algorithms face decrease of performance when exposed to small angular spacing between the sources with high angles of arrival as shown in Figure 1. These cases characterize a low spatial resolution scenario. The favorable situation is when the sources are well spaced and close to the broadside axis.

To build a commercial application with those solutions cited in this Section, it is important to increase the overall performance of the algorithm to at least to a reasonable range so that such degradation can be reduced. For measuring the performance, the signal to interference ratio (SIR) is used. A low computational complexity is also desirable for reducing power consumption and hardware size. With these objectives in mind we intend to improve the frequency domain BSS overall performance via insertion of a pre-processing step to initialize the demixing filters. In this process, the initial demixing filters captures the geometric information of the acoustic environment, more specifically, they shall capture the direction of arrival (DOA) knowledge of the acoustic sources. The DOA is assumed to be known, but it could have been estimated through DOA algorithms, for example, the Steered Response Power using the Phase Transform (SRP-PHAT) algorithm.

In this work, we first show our data model in Section II, then explain the used FDBSS algorithm in Section III. The preprocessing step for improving the performance is proposed in Section IV. In Section V, we show the overall performance of the algorithm for several spatial resolutions and identify critical conditions. In the same Section, we also show the performance gain with the proposed improvements. Finally, conclusions are drawn in Section VI.

\section{Data Model}

For an acoustic scenario, when the $s_i$ sources for $i = 1, ..., N$ are active at the same time, one is expected to observe convolutive mixtures at the sensors. Here the mixtures are expressed by filters $h_{ij}(t)$, with $i$ and $j$ denoting the source and microphone indexes respectively. Then, the
sensor signals

\[ x_j(t) = \sum_{i=1}^{N} \sum_{l=1}^{L} h_{ij}(l)s_i(t - l) \]  \hspace{1cm} (1)

should be deconvolved by estimating deconvolution filters \(w_{ji}(t)\) of finite length \(L\) to find the estimated source signals

\[ y_i(t) = \sum_{j=1}^{N} \sum_{l=1}^{L} w_{ji}(l)x_j(t - l) \]  \hspace{1cm} (2)

For the sake of simplicity, in this work we consider the same number \(N\) of sources and sensors. That explains why we write the same index \(i\) for the source and estimated signals.

A scheme showing the mixing and demixing filters is seen in Figure 2. Extending the model to the frequency domain and considering each frequency (or bin in the discrete case) as a linear mixture, one can write:

\[ X(f, m) = A(f)S(f, m) \]  \hspace{1cm} (3)

where \(A\) is a linear mixing matrix at frequency \(f\) and \(m\) is the frame index. The search for a demixing matrix characterizes an Independent Component Analysis (ICA) problem [4]. Its goal is to find a demixing matrix \(W\) for each frequency such that

\[ Y(f, m) = W(f)X(f, m) \]  \hspace{1cm} (4)

is the estimated demixed signal. In the next Section we cover how the demixing matrices \(W(f)\) can be computed in an adaptive manner.

III. A FREQUENCY DOMAIN BSS ALGORITHM

The algorithm used in this work is based on [11]. The basic idea of this algorithm is to divide the signal into time frames and then transform them to the frequency domain using a short-time Fourier Transformation. The update rule is derived by [5] through the information maximization combined with the natural gradient algorithm, which results in the following:

\[ \Delta W = \mu[I - \langle g(Y)Y^H \rangle]W, \]  \hspace{1cm} (5)

where \(\mu\) is the step size, \(\langle \cdot \rangle\) is the time average operator and \(g(\cdot)\) is a nonlinear function dependent on the Probability Density Function (PDF) of the desired signals. For the complex signal \(Y\), the nonlinear function \(g(Y)\), proposed in [7], may have the following form:

\[ g(Y) = -\frac{1}{Y} \log p(Y) e^{j \arg(Y)} \]  \hspace{1cm} (6)

assuming that its density \(p(Y)\) is independent from its argument.

Unfortunately, after using this algorithm, we must solve the ICA scaling and permutation problems. The permutation problem is very important because all the frequency bins belonging to a filter must stay aligned within the same filter. For complex filters, a detailed approach for solving the permutation problem is shown in [8]. The solution for the used scaling problem can be solved by applying the Minimum Distortion Principle (MDP) [9].

IV. IMPROVED FDBSS VIA DOA KNOWLEDGE

If the information of the environment is available, one can use it to initialize the algorithm described in the previous Section in order to improve its performance. With this purpose we add a filter initialization step as seen in the block diagram of Figure 3. First a DOA estimation is done. After that, using the estimated angles, initial filters for the next step are generated. Then, the ICA natural gradient adaptive algorithm, as seen in the previous Section, is executed. After all, the bins are scaled and checked whether they were permuted or not. Similar work was done in [10]. There, this scheme was used for a time domain BSS algorithm. Here we use the scheme for a frequency domain BSS algorithm. Moreover, a broader spatial investigation is made and the spatial power is taken into consideration for a better overall performance. The starting point is trying to

\[ x_1 \xrightarrow{DOA \text{ Estimation}} y_1 \]

\[ x_2 \xrightarrow{Scaling \text{ and } \text{permutation fix}} y_2 \]

\[ x_N \xrightarrow{\text{Filter Initialization}} w_{ji}^0 \xrightarrow{\text{ICA Algorithm}} w_{ji} \]

\[ y_N \xrightarrow{\text{Estimate } w_{ji}} \]

\[ \text{DOA Estimation} \]

\[ \text{Scaling and permutation fix} \]

\[ \text{Filter Initialization} \]

\[ \text{ICA Algorithm} \]

\[ \text{Estimate } w_{ji} \]

Fig. 3. Proposed method diagram.
estimation algorithms. With the DOA knowledge, the filters \( w_{ji} \) can be initialized. Consequently, an improved performance is expected.

Note that for an environment with impulse response filters \( h'_{ij} \), that are simplified version of the filter \( h_{ij}(t) \) with just one delay, i.e. no reflection and no signal spread over time are taken into account. So we can imagine the impulse response in the format of a shifted dirac impulse \( \delta(t - \tau) \), where \( \tau \) is the delay for the source signals to reach the microphones of the array taking the microphone at the origin \( d_0 \) as reference. The delay \( \tau \) can be calculated by geometrical inspection of Figure 4. A plane sound wave emitted by the source \( s_i \) and coming from a direction \( \theta_i \) reaches the microphone \( x_i \) with a delay

\[
\tau_{ij} = c^{-1}d_{j-1} \sin \theta_i, \quad (7)
\]

where \( c \) is the speed of sound. Taking the first microphone as reference and using the delays calculated from Equation (7), we can form the simplified room impulse responses

\[
h_{ij}'(t) = \begin{cases} \delta(t) & \text{if } j = 1, \\ \delta(t + \tau_{ij}) & \text{if } j \neq 1. \end{cases} \quad (8)
\]

If we want to remove one signal from one of the outputs, we should try to form a null system response for the undesired signal, i.e. \( h_{ij}'(t) + w_{ji}^0(t) = 0 \). In that way, the filters \( w_{ji}^0 \) should look similar to \( h_{ij}' \) so we just create diracs with delays in a similar manner as in Equation (8). The difference is that now we take for each output its corresponding microphone as the reference, e.g. microphone 1 is the reference for output 1, microphone 2 for output 2 and so forth. With that in mind we can calculate the delays

\[
\tau_{ij}^0 = c^{-1}(d_{(j-1)} - d_{(i-1)}) \sin \theta_i \quad (9)
\]

using the two positions \( i \) and \( j \) to interpret that reference and form the filters

\[
w_{ji}^0(t) = \begin{cases} \delta(t) & \text{if } i = j, \\ -\delta(t + \tau_{ij}^0) & \text{if } i \neq j. \end{cases} \quad (10)
\]

Figure 5 illustrates how the filters \( h_{ij}' \) and \( w_{ji}^0 \) should look like for two sources at 20 and 70°. The first row shows the \( h_{ij}' \) filters and the second row the \( w_{ji}^0 \) filters. The goal is to produce the null system response by initializing the filters \( w_{21}^0 \) and \( w_{12}^0 \). First we look back at Figure 2 and take the straight paths from input 1 to output and input 2 to output 2 as references. The delays are calculated so in output 1 we cancel filter \( h_{22}'(t) \) and in the output 2 we cancel filter \( h_{11}'(t) \). By calculating the filters according to Equation (10) and then convolving and summing the with the filters \( w_{ji}^0 \), one should observe the null response for the interfering source each output.

Fig. 4. Angle of incidence.

Fig. 5. Filter initialization. Delays \( \tau_{21} \) and \( \tau_{12} \) are calculated according to Equation (10).

**V. EXPERIMENTS AND RESULTS**

For the experiments we set an artificial environment where the sources are one meter away from the array. The algorithm parameters are set according to Table I. To create the impulse responses the image method [13] was used. They were created as if the sources were positioned over a circle of radius 1m for a span of 170°. This corresponds to angles of arrival \( \theta \) varying from -85 to 85°. For testing purposes we chose positions within that range in steps of 5°. The angular spacing \( \Delta \theta \) between the sources varies from 20 to 60° in steps of 10°. The whole scenario is depicted in Figure 6. In practical applications it is often better to “point” the initial filters to the greater DOAs then the DOA from the actual position where the sources stand on. The DOA from a source at -40° is seen in the frequency domain.
in Figure 7. In the picture the DOA was estimated for each frequency bin. We focus on the low frequencies, i.e. where speech is concentrated and no spatial aliasing is present. In the picture it represents the frequencies from -2000 to 2000 Hz. There, we see that many peaks extrapolate the -40° angle. Therefore, we obtained a better performance when “pointing” toward angles smaller than -40°.

To measure the

both sources on the same side of the plane. This significant drawback of the FDBSS algorithm is what the proposed scheme aims in order to increase SIR\textsuperscript{gain}. The dashed lines show that, at least when exposed to good separation angles (above 30°), the SIR is significantly increased. Nevertheless, some improvement can still be seen for a small separation.

VI. Conclusion

In this paper, it was seen that the FDBSS suffers from a decrease of performance when sources are close to the end fire axis. In order to improve the separation in such conditions a scheme for taking into account the direction of arrival knowledge was proposed. This knowledge is inserted into the algorithm through the initialization of the separation filters. With the proposed scheme, a significant improvement in the blind source separation is obtained for endfire angles.

ACKNOWLEDGEMENTS

The authors would like to thank Professor Dr.-Ing. Walter Kellerman from the University of Erlangen-Nuremberg for his support. They also thank the Centre for Research on Architecture of Information (CPAI) of the University of Brasilia (UnB) for their partial funding.

REFERENCES


