ACOUSTIC SOURCE SEPARATION USING GEOMETRICALLY CONSTRAINED ICA

Mirko Knaak † *Dieter Filbert* ‡

Technical University Berlin, Measurement Technology Lab
Einsteinufer 19, 10587 Berlin, Germany
†e-mail: Mirko.Knaak@tu-berlin.de
‡e-mail: Dieter.Filbert@tu-berlin.de

ABSTRACT

This contribution introduces a new algorithm using independent component analysis (ICA) with a geometrical constraint. The new algorithm solves the permutation problem, and it is significantly less sensitive to the precision of the geometrical constraint than an adaptive beamformer (ABF). The effectiveness of the algorithm for real-world mixtures is also shown in the case of three sources and three microphones.

1. INTRODUCTION

For many signal processing tasks, such as speech recognition, transmission, or classification of signals, a very good reconstruction of the target signal is essential when the target signal is disturbed by other sources. Adaptive beamformers and blind source separation (BSS) are very effective tools for multichannel signal reconstruction.

Although the utility of ABFs is well established [1], they have limited robustness against erroneous parameters. This is very troublesome since the steering vector is always roughly estimated in a reverberant environment, as is shown in this paper. The methods traditionally used to overcome this sensitivity mostly broaden the directivity pattern, resulting in a trade-off between the signal suppression performance and the parameter sensitivity.

Independent component analysis ICA is an emerging technique for finding independent components in a multi-channel signal. The main application is BSS which has been shown to be capable of recovering multiple sources from their linear mixture if the sources are independent [2].

In the field of acoustics, convolutive mixtures need to be separated, which involves estimating of many more parameters than a separation of a scalar mixture. Most approaches simplify the problem to instantaneous separation problems for the frequency components. The scaling and permutation ambiguities left in the recovered frequency components become a serious problem, particularly when the number of sources and microphones becomes larger than two. Different permutations of the frequency components lead to mixed outputs and degraded separation results. There are several approaches to overcome this problem, however, they are restricted to two sources. Hence, the number of real-world applications in the acoustics field is still very limited, and the separation performance is mostly insufficient.

Current publications indicate an equivalence between ABF and BSS, e.g., [3]. BSS is only an intelligent set of ABFs with an adaptive null directivity aimed in the direction of the unnecessary sounds. This equivalence suggests an application of a geometrical constraint on ICA to solve the permutation and scaling problem.

A geometrically constrained algorithm has been proposed by [4]. They only employ the constraint with the assumption that it is estimated correctly. This assumption is very limiting because precise information about the steering vector is very difficult to obtain. The major advantage of using ICA and geometrical information appears when only a rough estimation is possible.

2. BLIND SOURCE SEPARATION AND ADAPTIVE BEAMFORMERS

2.1. Signals and BSS algorithm

In a set $\mathbf{s}^b(t) = \left[s^b_{target}(t), s^b_{i1}(t), \ldots, s^b_{iN-1}(t)\right]^T$ of broadband sources, the first source is the target sound and the others are interfering sources. The sound is measured with an array of M microphones $\mathbf{x}^b(t) = [x^b_1(t), \ldots, x^b_M(t)]^T$. The observed signals are filtered and mixed because the room acoustics impose a different impulse response h^b_{mn} between each source s^b_n and each microphone x^b_m.

In the frequency domain, the convolutive mixture can be written as $\mathbf{x}^f = \mathbf{H}^f \cdot \mathbf{s}^f + \mathbf{n}^f$, where $\mathbf{x}^f$ is a narrowband signal component filtered from $\mathbf{x}^b$ with a band pass centered at f. For simplicity, the index f is omitted hereafter. $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_N]$ consists of the steering vectors $\mathbf{h}_N$, where $\mathbf{h}_1$ is the steering vector of the target sound. Only under anechoic conditions, they can be approximated by the phase shifts caused by the time delays τ_{mn} with $\mathbf{h}_n = [e^{j2\pi f\tau_{1n}}, \ldots, e^{j2\pi f\tau_{Mn}}]^T$. When considering echoes and reverberation, $\mathbf{h}_n$ is the sum of all echo paths.

The goal of the algorithms discussed here is to find an optimal estimation $y_1(t)$ of the target signal s_{target}. To achieve this goal, an unmixing matrix $\mathbf{W}$ or a coefficient vector $\mathbf{w}_1$ is applied to the vector of observations as follows:

$$\mathbf{y} = \mathbf{W} \cdot \mathbf{x} \qquad y_1 = \mathbf{w}_1^H \cdot \mathbf{x}, \tag{1}$$

where $\langle\rangle^H$ is the hermitian (conjugate transposed).

Blind source separation uses ICA to estimate the unmixing matrix $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_N]^H$ by making the output signals as independent as possible. Essentially, ICA has two steps ($\mathbf{W} = \mathbf{T}^H \cdot \mathbf{V}$). In the first step (sphering), the matrix $\mathbf{V}$ is determined by the principle component analysis (PCA). In the second step (rotation with $\mathbf{T}$), maximization of nongaussianity, nonlinear decorrelation, non-stationary decorrelation, or spatio-temporal decorrelation can be used to determine the rotation matrix $\mathbf{T}$ [2].

2.2. ABF with an imprecise estimation of the steering vector

An ABF minimizes the power of the output signal with a constraint: the energy of a signal coming from the direction of the target is passed without changes $\mathbf{w}_1^H \hat{\mathbf{h}}_1 = c_1$. This means that the source position has to be known in advance. It can be estimated by sound localization methods (e.g., MUSIC [1]), determined by image processing, or simply known by geometry.

A major drawback of ABFs is that they rely on the correct estimation of the steering vector. Since the impulse response of a room is normally not available, the steering vector $\hat{\mathbf{h}}_1$ is estimated by time delays of the direct sound only. An estimation error of $\hat{\mathbf{h}}_1$ is caused by two reasons: a wrong position or direction of arrival, or the existence of strong reverberations. The latter reason is due the multiple directions of arrival while only the direct sound is used for the estimation. The estimation becomes rough even when the source position might be well known.

3. GEOMETRICALLY CONSTRAINED ICA

3.1. Derivation of new algorithm

The algorithm is based on negentropy maximization proposed by [2]. In this approach, the negentropy is approximated by the nonlinear function $G()$ with the derivation $g()$. As usual in ICA approaches, a PCA is applied first. Then, the following iteration [5] is being used

$$\mathbf{t}^{k+1} = \mathbf{t}_k - \frac{\mathrm{E}\{\mathbf{z}g(\mathbf{t}_k^H\mathbf{z})\} + \beta\mathbf{t}_k}{\mathrm{E}\{\mathbf{z}g(\mathbf{t}_k^H\mathbf{z})\} + \beta} \quad (2)$$

$$\mathbf{t}^{k+1_{new}} = \frac{\mathbf{t}_{k+1}}{|\mathbf{t}_{k+1}^H \mathbf{V}\hat{\mathbf{h}}_1|} \quad (3)$$

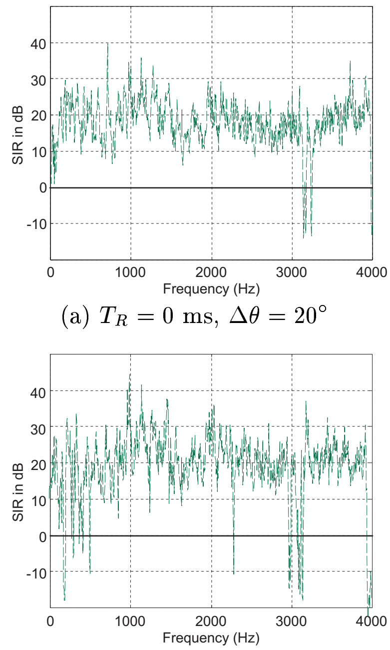
$$\beta = \frac{\mathrm{E}\{\mathbf{t}_k^H\mathbf{z} \cdot g(\mathbf{t}_k^H\mathbf{z})\}}{|\mathbf{t}_k|^2}. \quad (4)$$

The new algorithm starts with $\mathbf{t}^0 = \hat{\mathbf{t}}_1 = \mathbf{V}\hat{\mathbf{h}}_1$. If the estimation of $\hat{\mathbf{h}}_1$ is correct, the estimation itself is already the correct solution. Then, the algorithm converges according to (2)-(4) to a saddle point of Lagrangian.

3.2. Real-world assessment

Several tests with realistic mixtures were executed. Japanese language sounds were mixed with room impulse responses measured in real rooms. The reverberation times were 0, 150 and 300 ms. The impulse responses where $N = M = 3$ were from the RWCP database, and a complete description is available at http://tosa.mri.co.jp/ sounddb/index.htm.

Incorrect steering vectors caused by an incorrect look direction (a) and by reverberation (b) were used in all plots of Fig. 1. A high improvements of the frequency SIR (signal-to-interference-ratio) over 10dB were achieved, even in the 3×3 ($N = M = 3$) case. This demonstrates the effectiveness of the new algorithm in its major domain when a rough estimation of the steering vector is available. Although the computational cost has not been analyzed yet, the convergence is very fast due to the Newton method of the underlying FastICA algorithm.



(a) $T_R = 0$ ms, $\Delta\theta = 20°$

(b) $T_R = 150$ ms, $\Delta\theta = 0°$

Figure 1: Real-world simulation results in $N = M = 3$

4. CONCLUSION

We introduced a new ICA algorithm with a geometrical constraint and showed its effectiveness by using impulse responses from a real room. The new algorithm solves the permutation problem of BSS of acoustic mixtures, particularly when the number of sources and microphones becomes larger than two.

5. ACKNOWLEDGMENT

We thank Shoji Makino and Shoko Araki for valuable discussions, and the NTT Communication Science Lab for financial support.

REFERENCES

[1] D. Johnson and D. Dudgeon, *Array Signal Processing: Concepts and Techniques*, Prentice Hall, Englewood Cliffs, NJ, 1993.

[2] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, John Wiley, New York, 2001.

[3] S. Araki, S. Makino, R. Mukai, and H. Saruwatari, “Equivalence between frequency domain blind source separation and frequency domain adaptive null beamformers,” in *Eurospeech*, September 2001, pp. 2595–2598.

[4] L. Parra and C. Alvino, “Geometric source separation: Merging convolutive source separation with geometric beamforming,” in *NNSP*, 2001, pp. 273–282.

[5] M. Knaak, S. Araki, and S. Makino, “Geometrically Constrained ica for convolutive mixtures of sound,” in *ICASSP*, 2003.