Realtime Capable Beamforming-Based Convolutive Source Separation[1]

Wolf Baumann, Dorothea Kolossa, Bert-Uwe Köhler und Reinhold Orglmeister
Technische Universität Berlin

Fachgebiet Elektronik und medizinische Signalverarbeitung

Email: {w.baumann,d.kolossa,b.koehler,orglmeister}@ee.tu-berlin.de

Abstract

A realtime algorithm for the separation of two acoustically mixed speech signals in a car environment is proposed. The algorithm is based on two parallel frequency domain beamformers, each of which cancels the signal from one interfering source by frequency dependent null-beamforming.

The zero-directions of the beamformers are determined for each frequency band separately using independent component analysis (ICA). They are found by optimizing a higher order statistics cost function such that the output signals, stemming from the driver and the co-driver, are as independent as possible.

Interpreting the elements of the separation matrix as the coefficients of a beamformer and hence determining the directions of arrival (DOA) allows the assignment of the output signals to the different sources. At the same time, the algorithm is retaining the major advantage of blind methods that do not require an external estimate of the DOA.

Introduction

Independent component analysis (ICA) has successfully been applied to many problems including the separation of e.g. biomedical data, sonar or seismographic data.

Depending on the mixing process, e.g. linear/nonlinear or instantaneous/convolved, the task of source separation can be very difficult. In case of convolutive mixtures, separation algorithms are usually computationally expensive and often restricted to certain room conditions.

The convolutive mixing process can be expressed as

$$\mathbf{x}(t) = \mathbf{A} * \mathbf{s}(t) \qquad \text{eq. 1}$$

with **x(t)**, **A** and **s(t)** representing the recorded signals, the mixing matrix and the source signals, respectively. Here, the mixed signals **x(t)** are superpositions of the filtered source signals **s(t)**, and the mixing matrix **A** contains the room impulse responses between each source and microphone. The convolution is denoted by the asterisk. For a known mixing matrix **A**, separation can be achieved with an unmixing matrix **W**, ideally a stable approximation of the inverse of **A**

$$\mathbf{y}(t) = \mathbf{W} * \mathbf{x}(t)\,. \qquad \text{eq. 2}$$

Subsequently, we restrict the model to a 2×2 mixing system, i.e. two simultaneously talking speakers are recorded by two microphones. One commonly practised technique for convolutive source separation is to transform equation (1) to the frequency domain and to solve the resulting instantaneous source separation problem, i.e. $\mathbf{Y}(j\omega)=\mathbf{W}(j\omega)\mathbf{X}(j\omega)$, for each frequency band ω.

Due to the permutation indeterminacy inherent in this technique, the elements of the reconstructed signal vector $\mathbf{Y}(j\omega)$ are randomly arranged, which constitutes a severe problem. To overcome such permutation problems, recent approaches apply geometrical constraints in ICA algorithms e.g. [1], while other approaches iterate between beamforming and ICA stages, like [2].

Algorithm

We suggest a model (see Figure 1), consisting of two parallel frequency dependent null-beamformers, where the null-directions are adjusted to make output signals as independent as possible.

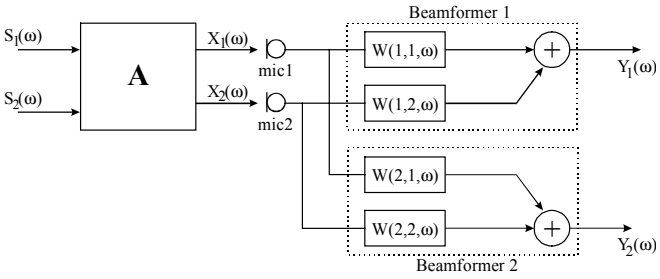

Figure 1: Model in the frequency domain

The two beamformers, implemented by the separation matrix $\mathbf{W}(j\omega)$, have to be optimized jointly, because the unattenuated direction of the first one is the zero direction of the second one and vice versa. This ensures a constant transfer function in one look direction while forcing that of the interfering source to be zero.

The separation matrix $\mathbf{W}(j\omega)$ for frequency dependent null-beamforming is given by

$$\mathbf{W}(j\omega) = \frac{1}{e_1 - e_2} \cdot \begin{bmatrix} -e_2 & 1 \\ e_1 & -1 \end{bmatrix} \qquad \text{eq. 3}$$

where

$$e_1 = \exp\left[-i2\pi f \frac{d}{c}\sin(\varphi_1(\omega))\right] \qquad \text{eq. 4}$$

$$e_2 = \exp\left[-i2\pi f \frac{d}{c}\sin(\varphi_2(\omega))\right], \qquad \text{eq. 5}$$

with c denoting the propagation velocity of sound waves, d the distance between microphones and φ the direction of the impinging wave relative to broadside (perpendicular to the array).

As it is common to all beamforming algorithms, the model in eq. 3 does not consider attenuations caused by the room impulse

[1] patent pending

responses but assumes the mixing process to be adequately described as a superposition of delayed source signals. Though this is not an exact reproduction of the real world conditions, it has proven to be sufficient for close microphone arrangements, and furthermore decreases the number of parameters to estimate by reducing the search space to the unit circle.

However, the major difference to a standard delay and sum beamforming model is that the angles φ_1 and φ_2 are not restricted to be the same for all frequencies. This is absolutely necessary for the compensation of phase distortions, caused by the reverberations inherent in room impulse responses and by the neglection of the magnitudes within the null-beamforming model of eq. (3).

The joint adjustment of φ_1 and φ_2 is simplified when the source locations are restricted to different quadrants. In this case, no permutations between frequency bands are possible, because φ_1 and φ_2 can only take positive or negative values, respectively. While restricting the source directions, one additionally avoids non-invertible constellations, e.g. if the sources impinge from the same direction, which, to our experience, are not separable in general.

Cost Function

In the proposed algorithm, the cost function *J*, which is optimized to obtain the null-directions of the two beamformers, consists of a fourth order cross-cumulant [3]:

$$J(Y_1', Y_2') = |\, Cum(Y_1', Y_2')\,| \cdot \qquad \text{eq. 6}$$

$Cum(Y_1', Y_2')$ refers to the cross-cumulant of Y_1' and Y_2' defined by

$$Cum(Y_1', Y_2') = E\left[|Y_1'|^2 \cdot |Y_2'|^2\right] - E\left[|Y_1'|^2\right] \cdot E\left[|Y_2'|^2\right] - \left|E\left[Y_1' \cdot Y_2^{*\prime}\right]\right|^2 - \left|E\left[Y_1' \cdot Y_2'\right]\right|^2 \qquad \text{eq. 7}$$

where Y_1' and Y_2' are the centered and normalized output variables:

$$Y' = \frac{Y - E[Y]}{\sqrt{E\left[(Y - E[Y])^2\right]}} . \qquad \text{eq. 8}$$

The optimization with respect to the elements of the separation matrix e_1 and e_2 can be carried out regardless of the complex factor $1/(e_1 - e_2)$ in equation (3). Taking the complex gradient [4] of the four terms of the cost function in eq. (7) leads to the partial derivatives $\partial J / \partial e_1^*$ and $\partial J / \partial e_2^*$ as follows:

$$\frac{\partial J}{\partial e_1^*} = \mathrm{sgn}(J) \cdot \left(E\left[Y_1 Y_1^* X_1^* Y_2\right] - E\left[X_1^* Y_2\right] - E\left[Y_1^* Y_2\right] E\left[X_1^* Y_1\right] - E\left[Y_1 Y_2\right] E\left[X_1^* Y_1^*\right]\right) \qquad \text{eq. 9}$$

$$\frac{\partial J}{\partial e_2^*} = \mathrm{sgn}(J) \cdot \left(E\left[Y_2 Y_2^* X_1^* Y_1\right] - E\left[X_1^* Y_1\right] - E\left[Y_2^* Y_1\right] E\left[X_1^* Y_2\right] - E\left[Y_1 Y_2\right] E\left[X_1^* Y_2^*\right]\right) \qquad \text{eq. 10}$$

The adaptation of φ_1 and φ_2 is performed by a gradient descent algorithm wherby the complex gradient is formed by the projection of the the partial derivatives in the equations (9) and (10) onto the unit circle.

Results

The beampattern of one of the resulting null-beamformers is depicted in Figure 2. The direction of the null-beam varies over frequency around a principal direction. The ratio of signal duration to processing time, using Matlab and a sampling rate of 22.4 kHz, is about 1:1. The separation result is, depending on room characteristics, comparable to or better than [1].

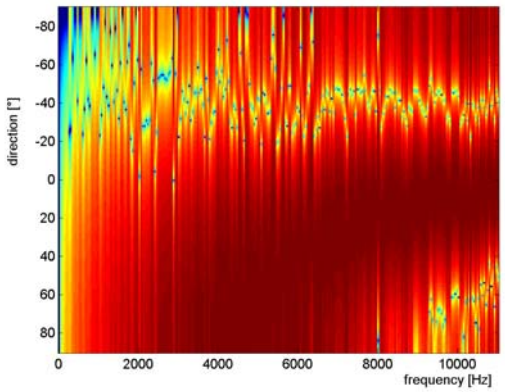

Figure 2: Beampattern of one of the resulting beamformers

Discussion

Considering convolutive ICA as a beamformer brings new insight into the capabilities of convolutive ICA algorithms. In comparison to standard convolutive ICA algorithms, the simple beamforming model achieves a reduction of the search space to the unit circle. This reduction turns out to be extremely helpful in finding the principal direction of the beamformer and the appropriate deviations within each frequency band. Avoiding the permutation problem by assigning the sources to different quadrants is a reasonable approach at least in car environments and has been shown to be successful in practice.

Conclusion

The concept of frequency dependent null-beamforming leads to real-time capable algorithms that can achieve successful signal separation under real room conditions.

References

[1] Lucas Parra and Christopher Alvino, „Geometric source separation: Merging convolutive source separation with geometric beamforming“, in IEEE Transaction on Speech and Audio Processing, September 2002, vol. 10, pp. 352-362.

[2] Hiroshi Saruwatari, Toshiya Kawamura and Kiyohiro Shikano, „Blind Source Separation Based on Fast-Convergence Algorithm Using ICA and Array Signal Processing“, ICA 2001, pp. 412-417.

[3] J.-F. Cardoso, „High order contrasts for independent component analysis“, Neural Computation, vol. 11, pp. 157-192, 1999.

[4] D. Brandwood, „A complex gradient operator and its application in adaptive array theory“, IEE Proc., vol. 130, no. 1, pp. 11-16, Feb. 1983.