Permutation alignment for Frequency Domain ICA using subspace beamforming methods.

Nikolaos Mitianoudis[1] and Mike E. Davies[2]

[1] Imperial College London, Electrical and Electronic Engineering, Exhibition Road, SW7 2AZ London, UK
n.mitianoudis@imperial.ac.uk
[2] Queen Mary London, Centre for Digital Music, Mile End Road, E1 4NS London, UK
michael.davies@elec.qmul.ac.uk

Abstract. In this paper, the authors address the *permutation ambiguity* that exists in *frequency domain Independent Component Analysis* of convolutive mixtures. Many methods have been proposed to solve this ambiguity. Recently, a couple of beamforming approaches have been proposed to address this ambiguity. The authors explore the use of *subspace methods* for permutation alignment, in the case of equal number of sources and sensors.

1 Introduction

Assume an array of M sensors $\underline{x}(n) = [x_1(n)\ x_2(n)\ \dots\ x_M(n)]^T$ placed in a real room, capturing an auditory scene. Assume there are N sources in the auditory scene $\underline{s}(n) = [s_1(n)\ s_2(n)\ \dots\ s_N(n)]^T$. To model the recording environment, one could use *FIR convolutive mixtures*.

$$x_i(n) = \sum_{j=1}^{N} \underline{a}_{ij} * s_j(n) \qquad i = 1, \dots, M \tag{1}$$

where $\underline{a}_{ij}$ represents an FIR filter modelling the transfer function between the i^{th} sensor and the j^{th} source. For the rest of the analysis, we will consider only the case of equal number of sensors and sources.

The convolutive mixtures problem can be addressed in the *time domain*, by estimating unmixing FIR filters $\underline{w}_{ij}$, assuming that the sources are *statistically independent*. The filters are adaptively estimated in the time domain, using the general framework of *Independent Component Analysis* (ICA).

$$u_i(n) = \sum_{j=1}^{N} \underline{w}_{ij} * x_j(n) \qquad i = 1, \dots, N \tag{2}$$

A more robust approach is to transfer the problem in the *frequency domain*. Consequently, the convolutive mixtures problem is transformed into several instantaneous mixtures problems. Many frequency domain ICA (FD-ICA) methods were proposed in literature. In [4], a fast FD-ICA framework was proposed

with fast and robust results, compared to gradient-based methods. In frequency-domain methods, we encounter two interdeterminancies: the *scale* and the *permutation ambiguity*. The *scale ambiguity* (arbitrary source scaling) is rectified by mapping the separated sources to the observation space [3]. The *permutation ambiguity* (inherent ordering ambiguity of the instantaneous ICA model) produces an arbitrary ordering of sources along frequency. To tackle this problem, one should apply some mechanism to couple the sources along frequency. Some *source modelling* solutions exploit the coherence and the information between the frequency bands to align the permutations. There also exist some *channel modelling* solutions, assuming smooth filters, as a contraint to the unmixing algorithm.

In fact, the blind source separation systems can be considered array signal processing systems. A set of sensors arranged randomly in a room to separate the sources present is effectively a beamformer. Some methods [2, 5, 6] were proposed to solve the permutation problem using beamforming. In this paper, we investigate the idea of using subspace methods for permutation alignment in FD-ICA. Subspace methods produce more accurate alignment compared to the previously proposed methods using directivity patterns. We show that subspace methods even work in the case of equal number of sources and sensors.

2 Beamforming and Frequency-Domain ICA

A narrowband linear array of M sensors $\underline{x}(n)$, is defined as follows:

$$\underline{x}(n) = \sum_{i=1}^{N} \underline{a}(\theta_i) s_i(n) = [\underline{a}(\theta_1)\ \underline{a}(\theta_2)\ \dots\ \underline{a}(\theta_N)]\, \underline{s}(n) \tag{3}$$

where $\underline{a}(\theta_i) = [1\ \alpha e^{-j2\pi f T_i} \dots \alpha e^{-j2\pi f (M-1) T_i}]^T$, $T_i = dsin\theta_i/c$, θ_i are the DOA, d is the intra-sensor distance and $c = 340m/sec$. The array model is similar to the general Blind Source Separation model. The main objective is to estimate a filter $\underline{w}_i(f)$ to separate each source i. The *directivity pattern* (gain pattern) of the beamformer $\underline{w}_i(f) = [w_{i1} \dots w_{iN}]$, can be expressed as follows:

$$F_i(f, \theta) = \sum_{k=1}^{N} w_{ik}^{ph}(f) e^{j2\pi f (k-1) d \sin\theta / c} \tag{4}$$

In the context of FD-ICA, at a given frequency bin, the unmixing matrix can be interpreted as a *null-steering beamformer* that uses a *blind algorithm* (ICA) to place nulls on the interfering sources. The source separation framework does not use any information concerning the geometry of the auditory scene, but only the sources statistical profile. Inclusion of this additional information can help in aligning the permutations. Although, we are dealing with real room recordings, we assume that there is a consistent DOA along frequency for each source, belonging to the direct path signal signal. This is equivalent of approximating

the room's transfer function with a single delay. The permutations of the unmixing matrices are flipped so that the directivity pattern of each beamformer is approximately "aligned". More specifically, having estimated the unmixing matrix $W(f)$ using FD-ICA, we permute the rows of $W(f)$, in order to align the permutations along the frequency axis. We form the directivity pattern (2), where $w_{ik}^{ph}(f) = W_{ik}(f)/|W_{ik}(f)|$ is the phase of the unmixing filter coefficient between the k^{th} sensor and the i^{th} source at frequency f. This approach can be considered a *channel modelling* technique.

However, in audio source separation, the sensors capture more than a single delay. The room's reflections tend to shift the "actual" DOA by a small arbitrary amount at each frequency. However, the average shift of DOA along frequency is not so significant and usually we can spot a main DOA. This implies that we can align the permutations in FD-ICA, using the DOA.

The reason why we are using beamforming for permutation alignment and not for separation is the poor estimate for DOA along frequency. The ICA algorithm can give very accurate separation. Instead, the slightly "shifted" DOA can help us in identifying the correct permutation of separated sources.

Next, we will address some ambiguities in DOA estimation and permutation alignment using directivity patterns, plus a novel mechanism to apply subspace techniques for permutation alignment.

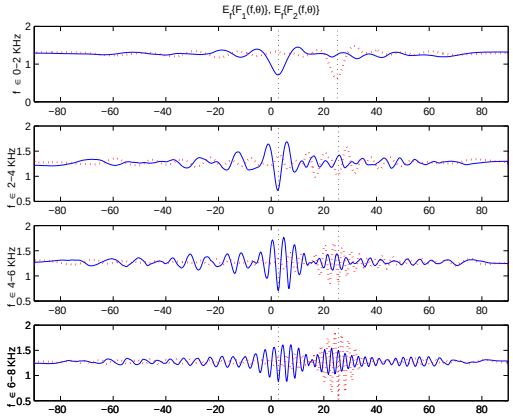

Fig. 1. Average Beampatterns along certain frequency bands for both sources.

2.1 DOA estimation ambiguity

Saruwatari et al [6] estimated the DOA by taking the statistics with respect to the direction of the nulls in all frequency bins and then tried to align the

permutations by grouping the nulls that exist in the same DOA neighbourhood. On the other hand, Ikram and Morgan [2] proposed to estimate the sources DOA in the lower frequencies, as it is less noisy than in higher frequencies. Parra and Alvino [5] used more sensors than sources along with *known* source locations and added this information as a geometric constraint to their unmixing algorithm.

In figure 1, we plot the average beampatterns along a certain frequency range $\mathcal{F}$, assuming a two sensor setup in a real room, where $d = 1m$. More specifically, we plot the average beampatterns between $0 - 2kHz$, $2 - 4kHz$, $4 - 6kHz$ and $6 - 8kHz$. We can see that in the lower frequencies, we get clear peaks denoting the directions of arrival. However, in higher frequencies, we get peaks at the same angle, but also multiple peaks around the main DOA. Observing the higher frequencies, we can not really define which of the peaks is the actual DOA. As a result, we may want to use only the lower subband $(0 - 2kHz)$ for DOA estimation.

It is simple to show that averaging beampatterns over a lower frequency band $\mathcal{F}$ will emphasize the position of the two DOAs. Hence, the following mechanism can be used for DOA estimation, without sorting the permutations along frequency.

1. Unmix the sources using an FD-ICA algorithm
2. For each frequency bin f and source i estimate the beamforming pattern $F_i(f, \theta)$.
3. Form the following expression for $\mathcal{F} = [0 - 2kHz]$

$$P(\theta) = \sum_{f \in \mathcal{F}} \sum_{i=1}^{N} |F_i(f, \theta)|^2 \tag{5}$$

The minima of $P(\theta)$ will give an accurate estimate of the Directions of Arrival. The exact low-frequency range $\mathcal{F}$ we can use for DOA estimation is mainly dependent on the microphone spacing d. If we choose a small microphone spacing ($\sim$ cm), the ripples will start to appear at higher frequencies, as $f_{ripple} \sim c/2d$. However, as the microphones will be closer, the signals that will be captured will be more similar. Thus, the source separation SNR will decrease considerably, as our setup will degenerate to the less sensors than sources case. Therefore, the choice of sensor spacing is a tradeoff between *separation quality* and *beamforming pattern clarity*.

2.2 Permutation alignment ambiguity

Once we have estimated the DOA, we want to align the permutations along the frequency axis to solve the permutation problem in frequency domain ICA. There is a slight problem with that. Basically, all nulls, as explained in an earlier section, are slightly drifted due to reverberation. As a result, the classification of the permutations cannot be accurate.

One solution can be to look for nulls in a "neighbourhood" of the DOA. Then, we can do some classification, however, the definition of the neighbourhood is

arbitrary. Hu and Kobatake [1] observed that for a room impulse response around $300ms$, the drift from the real DOA maybe $1-3$ degrees on average (this may be generally different at various frequencies). As a result, we can define the neighbourhood as 3 degrees around the DOA. However, in mid-higher frequencies there might be more than one null, making the classification even more difficult.

3 Permutation alignment using the MuSIC algorithm

Another idea is to introduce *subspace methods*, as they tend to produce more "spiky" directivity patterns. The multiple nulls ambiguity still exists, however, the DOAs are more distinct and the permutation alignment should be more efficient. Although, in theory, we need to have more sensors than sources, it is possible to apply subspace methods in the case of equal number of sources and sensors. In our case, we will look at the MuSIC algorithm [7]. According to the MuSIC algorithm, one gets very localised estimates for the DOA by plotting the following function $M(\theta)$:

$$M(\theta) = \frac{1}{|P^{\perp}\underline{a}(\theta)|^2} \qquad \forall\ \theta \in [-\pi/2, \pi/2] \tag{6}$$

where $P^{\perp} = (I - E_s E_s^H) = E_n E_n^H$, where $E_s = [\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_N]$ contains the eigenvectors of $C_x = \mathcal{E}\{\underline{x}\underline{x}^H\}$ that correspond to the desired source and $E_n = [\underline{e}_{N+1}, \ldots, \underline{e}_M]$ contains the eigenvectors of C_x that correspond to noise. The N peaks of the function $M(\theta)$ will denote the DOA of the N sources.

In [4], we proposed to rectify the *scale ambiguity* by mapping the separated sources back to the microphones' domain. Therefore, we have an observation of each source at each sensor, i.e. a more sensors than sources scenario. If we do not take any steps for the permutation problem, the ICA algorithm will unmix the sources at each frequency bin, however, the permutations will not be aligned along frequency. It is simple to demonstrate that mapping back to the observation space is not influenced by the permutation ambiguity [3]. Hence, after mapping we will have observations of each source at each microphone, however, the order of sources will not be the same along frequency. Using the observations of all microphones for each source, we can use MuSIC to find a more accurate estimation for the DOAs, using (6).

We can form "MuSIC directivity patterns" using $M(\theta)$ (6), instead of the original directivity patterns. To find more accurate DOA estimates, we can form $P(\theta)$ as expressed in (5), using $M(\theta)$ instead of the original directivity pattern. Finally, we can use the DOAs to align the "sharper" "MuSIC directivity patterns". The proposed algorithm can be summarised as follows:

1. Unmix the sources using the FD-ICA framework.
2. Map the sources back to the observation space, i.e. observe each source at each microphone.
3. Having observations of each source at each microphone, we apply the MuSIC algorithm to have more accurate DOA estimates along frequency.
4. Align permutations now, according to the DOAs estimated by MuSIC.

4 Experiments

In this section, we perform two experiments to verify the ideas analysed so far in this paper. The Fast FD-ICA algorithm [4] is used to unmix the data, without the Likelihood Ratio solution.

4.1 Experiment 1 - Single Delay

In the first experiment, two speech signals are mixed artificially using single delays between $5-6$ msecs at $16kHz$. We test the performance of the proposed solutions for the permutation problem, in terms of beamforming. In figure 2 (left), we can see a plot of $P(\theta)$ (5) for this case of a single delay. We averaged the directivity patterns over the lower frequency band $(0-2kHz)$ and as a result we can see two Directions of Arrival. The estimated DOAs will be used to align the permutations. Since we are modeling a single delay, we will not allow any deviations from the estimated DOAs. In figure 3 (left), we can see the general performance of this scheme for one of the sources. We can spot some mistakes in the mid-higher frequencies, verifying that it might be difficult to align the permutations there. In figure 2 (right), we can see a plot of $P(\theta)$ (5) using the MuSIC algorithm. We averaged the MuSIC directivity patterns over the lower frequency band $(0-2kHz)$. Now the peaks indicating the Directions of Arrival are now a lot more distinct and "spiky". In figure 3 (right), we can see that the permutations are correctly aligned using the more accurate MuSIC directivity plots.

4.2 Experiment 2 - Real room recording

Next, we perform a real world experiment. We used a university lecture room $\sim 7.5 \times 6\text{m}^2$ to record a 2 sources - 2 sensors experiment. We investigate the nature of real room directivity patterns as well as explore the performance of the proposed schemes for permutation alignment. In figure 4 (left), we can see a plot of $P(\theta)$ (5) for this case of real room recording. Averaging over the lower $2kHz$, we seem to get a very clear image of the main DOAs, giving us an accurate measure for this estimation task. We try to align to the permutations around the estimated DOAs allowing $\pm 3^\circ$ deviation. In figure 5 (left), we see the results for one of the sources. We can spot that generally this scheme can perform robust permutation alignment in the lower frequencies, but considerable confusion exists in higher frequencies, as expected from our theoretical analysis. In figure 4 (right), we can see a plot of $P(\theta)$ (5), averaging the MuSIC directivity patterns over the lower frequency band $(0-2kHz)$. The two Directions of Arrival are more clearly identified from this graph. In figure 5 (right), we can see that most of the permutations are correctly aligned using the more accurate MuSIC directivity plots.

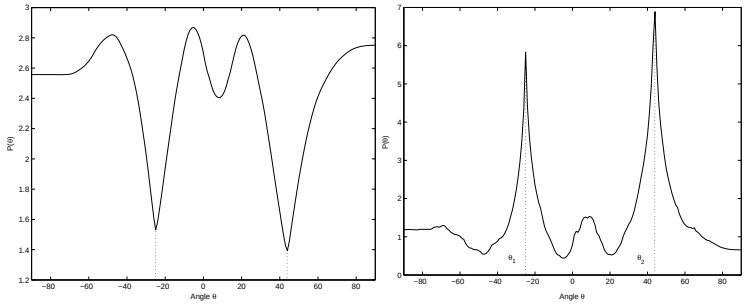

Fig. 2. Plotting P(θ) (eq. 5) using directivity patterns (left) and MuSIC directivity patterns (right) for the first 2kHz for the single delay case. Two distinct DOAs are visible.

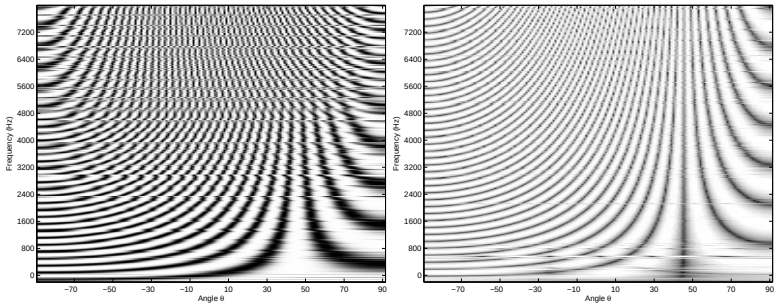

Fig. 3. Permutations aligned using the directivity patterns (left) and the MuSIC directivity patterns (right) in the single delay case.

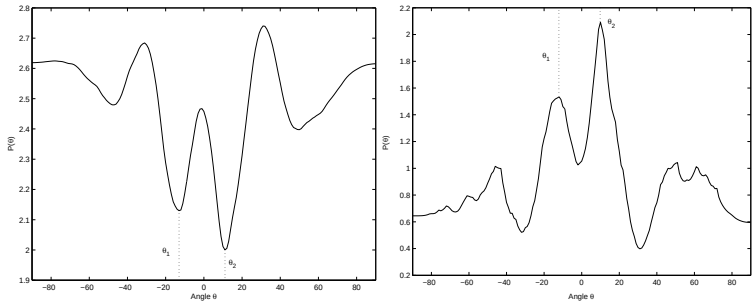

Fig. 4. Plotting P(θ) (eq. 5) using directivity patterns (left) and MuSIC directivity patterns (right) for the first 2kHz in the real room case. MuSIC enhances the positions of the DOAs.

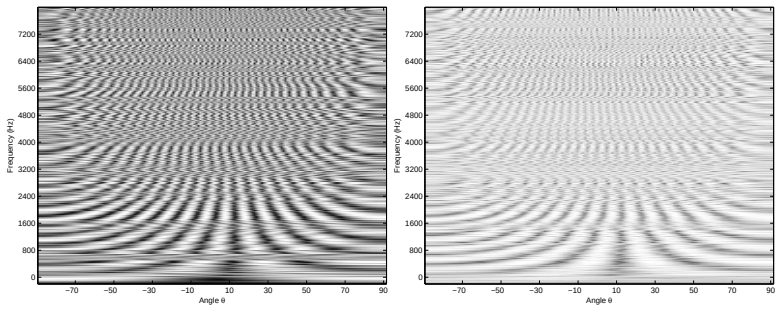

Fig. 5. Permutations aligned using the Directivity Patterns (left) and the MuSIC directivity patterns (right) in the real room case.

5 Conclusion

In this paper, we interpreted the Frequency-Domain audio source separation framework, as a Frequency-Domain beamformer. We reviewed some of the proposed methods for permutation alignment. In addition, a novel mechanism to employ *subspace methods* for permutation alignment in the frequency domain source separation framework in the case of equal number of sources and sensors was proposed. Such a scheme seems to be less computationally expensive in the general $N \times N$ case, compared to the Likelihood Ratio, as we do not have to work in pairs or even calculate the likelihood of all permutations of the N sources.

References

1. X. Hu and H. Kobatake. Blind source separation using ica and beamforming. In *Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pages 597–602, Nara, Japan, 2003.
2. M.Z. Ikram and D.R. Morgan. A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation. In *ICASSP*, 2002.
3. N. Mitianoudis. *Audio Source Separation using Independent Component Analysis*. PhD thesis, Queen Mary, University of London, 2004.
4. N. Mitianoudis and M. Davies. Audio source separation of convolutive mixtures. *Trans. Audio and Speech Processing*, 11(5):489 –497, 2003.
5. L. Parra and C. Alvino. Geometric source separation: Merging convolutive source separation with geometric beamforming. *IEEE Transactions on Speech and Audio Processing*, 10(6):352–362, 2002.
6. H. Saruwatari, T. Kawamura, and K. Shikano. Fast-convergence algorithm for ica-based blind source separation using array signal processing. In *Proc. Int. IEEE WASPAA*, pages 91–94, New Paltz, New York, 2001.
7. R.O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Trans. on Antennas and propagation*, AP-34:276–280, 1986.