Separation of Spatial Audio using Projection-based Method

Narges Mohammadi

Signal, Data, and Imaging Sciences Laboratory, University of Rochester

Introduction

Audio objects are characterized by spatial and spectro-temporal modeling which spatial modeling which neither assume an object has only one underlying source point nor attempt to model complex room acoustics. The main goal of this project is to unmix multi-channel audio signals to different constitutive spatial objects. In this case, the geometry of microphone array is unknown and using anechoic mixing model, we would be able to analyze the projection of mixture signal on many spatial direction.

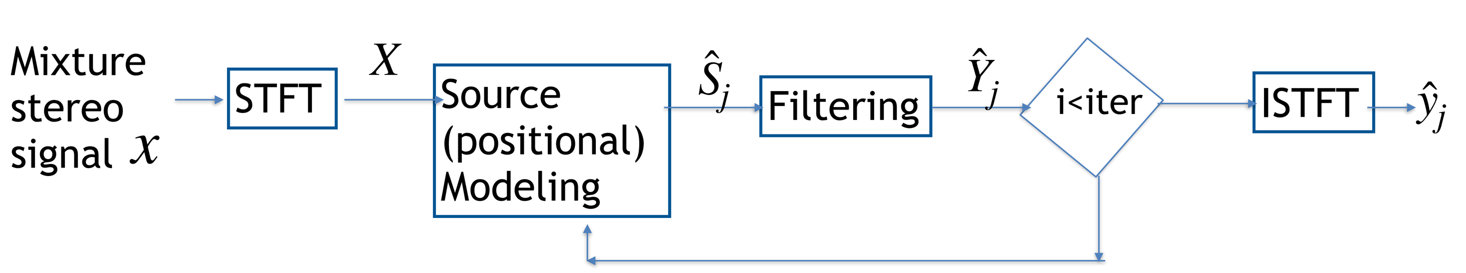
Typical Source Separation Problem

We assume that for each object image Y_j there is one single underlying monophonic signal called object source where its STFT is written as S_j . In Punctual anechoic model, each channel of object model is obtained by a simple delay and gain applied to this unique source as below:

$$\forall(f,t) \quad X(f,t) = \sum_j Y_j(f,t)$$

$$\forall(f,t) \quad y_j(f,t) = h(\phi, \tau \mid f) * s_j(f,t)$$

In common Sound Source Separation procedures, source models are obtained from the spectrogram of the audio mix which is often done in an iterative manner:



Source Position Models :

In the case of multichannel (in this case 2-channel/stereo) audio signals, the spatial position of the sources has often been exploited to perform sound source separation. It is assumed that the spatial positioning of source j has been achieved using a constant power panning law, defined by a single parameter, the panning angle φj ∈ [0,π/2].

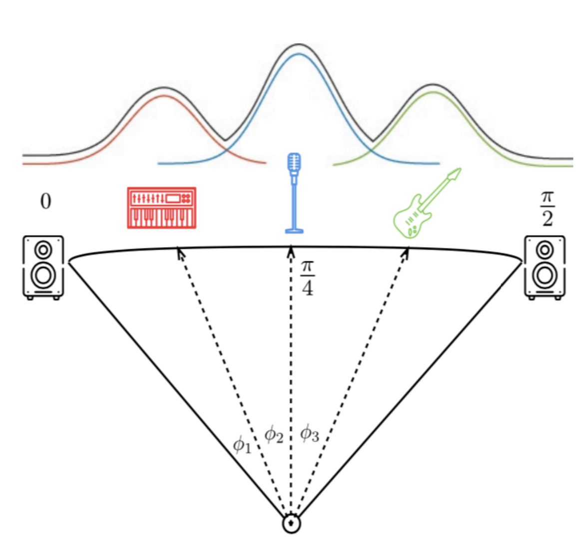
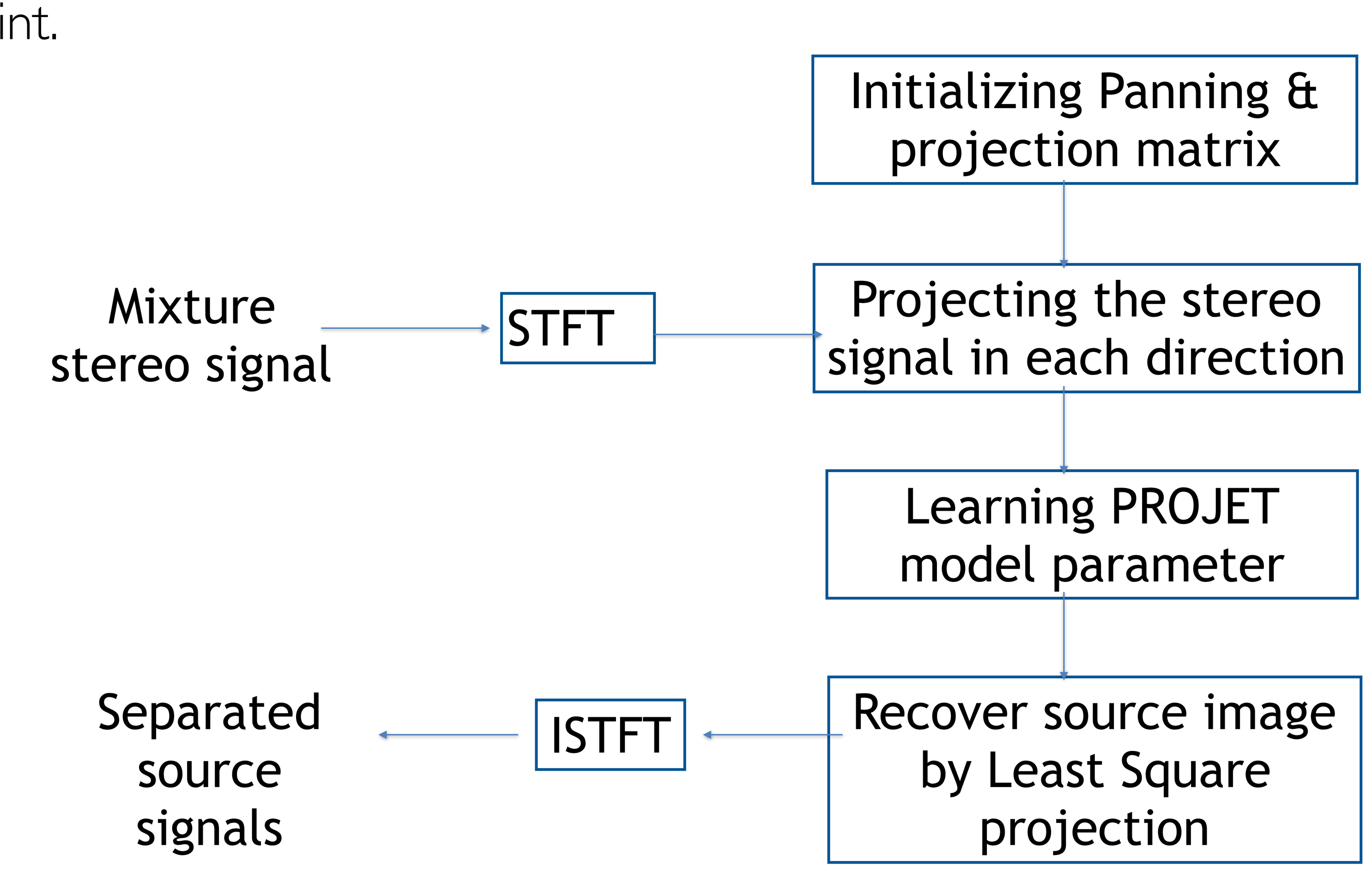


Illustration of standard Pan Law

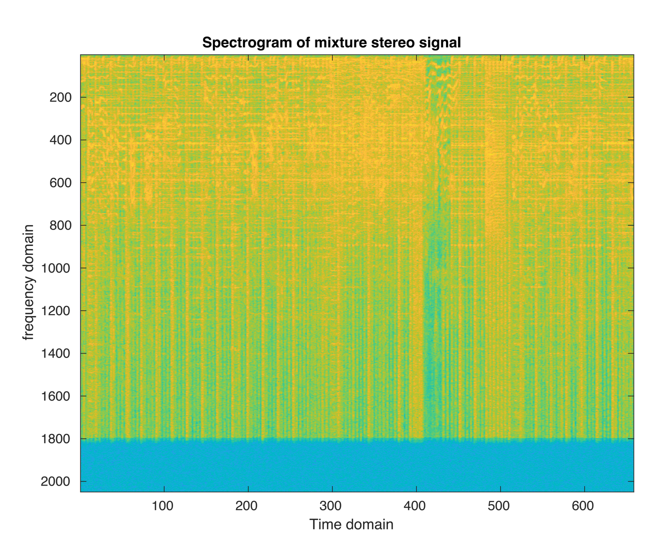
Approach: Source Position Modeling

In this method, it is assumed that the time-frequency representations of the sources have very little overlap. In this case, a single source j will contribute most of the energy at a single point in the time-frequency representations, and so R(k, n) ≈ cos φj / sin φj .

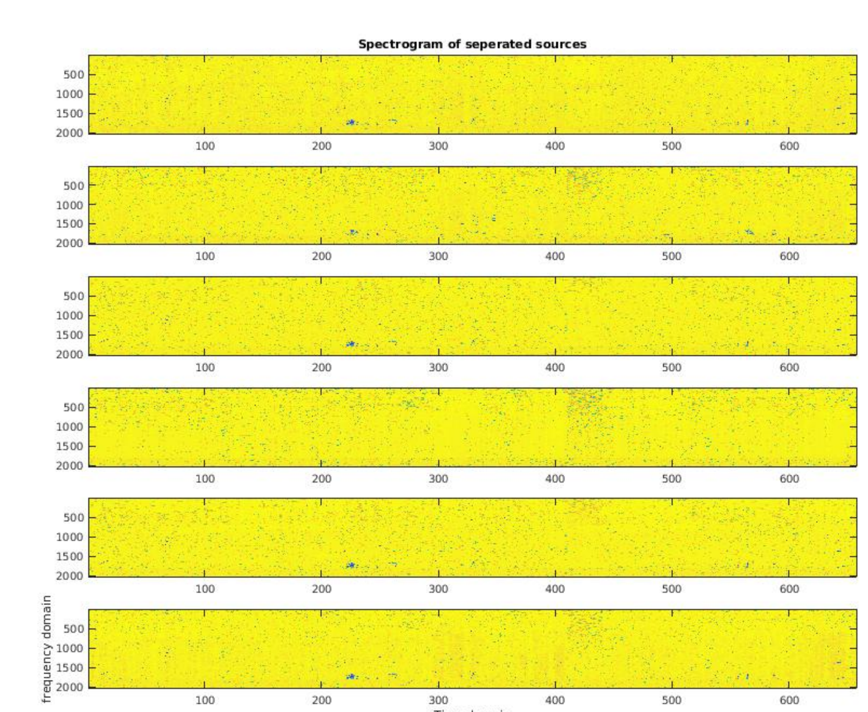
Given that R(k, n) only depends on φj under this assumption, it can therefore be used to estimate a panning angle for each time-frequency point.



Results:



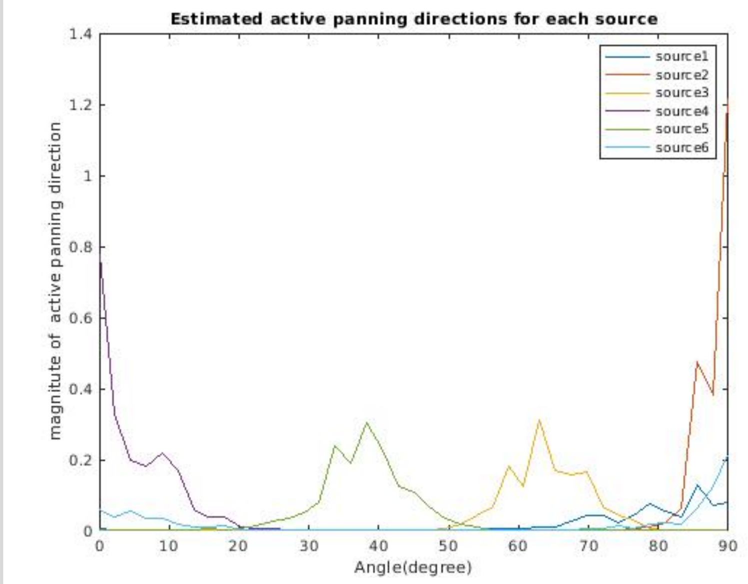
Spectrogram of mixture signal



Spectrogram of separated sources

This method has been applied on a mixed stereo signal and tried to separate 6 sources by estimating their corresponding panning angle for each time-frequency point. (The No. of panning direction is set to 41 and the No. of panning projection is set to 15)

Results:



Estimated active panning direction of each source

Angle (degree)	10	20	30
SDR	5.1	6.1	6.2
SIR	10.2	9.4	9.1
SAR	10.1	9.5	9.6
ISR	11.9	12.2	13.1

For separation evaluation performance, Signal to Distortion Ratio (SDR), Signal to Interference Ratio (SIR) and Image to Spatial Distortion Ratio (ISR) has be used.

This method has been also applied on MSD100 dataset with mixtures of 4 objects with an equal angle between them and no delay between channels.

Discussion

In this work, a spatial projection-based method was presented for application of multi-channel audio separation using a mixing model which benefit from the fact that the spatial image of each object is a weighted sum of independent contributions originating from all pan-delay directions. It is shown how to create a set of observations in which some objects are attenuated or enhanced by projecting the multi-channel mixture signal onto a range of spatial-delay direction with no more constraint on spectro-temporal characteristics of audio signals

References

1. D. FitzGerald, A. Liutkus and R. Badeau, "Projection-Based Demixing of Spatial Audio," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 9, pp. 1560-1572, Sept. 2016.
2. S. Leglaive, U. Şimşekli, A. Liutkus, R. Badeau and G. Richard, "Alpha-stable multichannel audio source separation," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, 2017, pp. 576-580.
3. E. Cano, D. FitzGerald, A. Liutkus, M. D. Plumbley and F. Stöter, "Musical Source Separation: An Introduction," in IEEE Signal Processing Magazine, vol. 36, no. 1, pp. 31-40, Jan. 2019.

nmohamm4@ur.rochester.edu