INTERFERENCE REDUCTION ON FULL-LENGTH LIVE RECORDINGS

Diego Di Carlo[1], Antoine Liutkus[2], Ken Déguernel[3]

[1]Inria, *Multispeech team*, Villers-lès-Nancy, France; [2]Inria and LIRMM, Montpellier, France;
[3]IRCAM STMS Lab (CNRS, UPMC, Sorbonne Universitès), Paris, France.

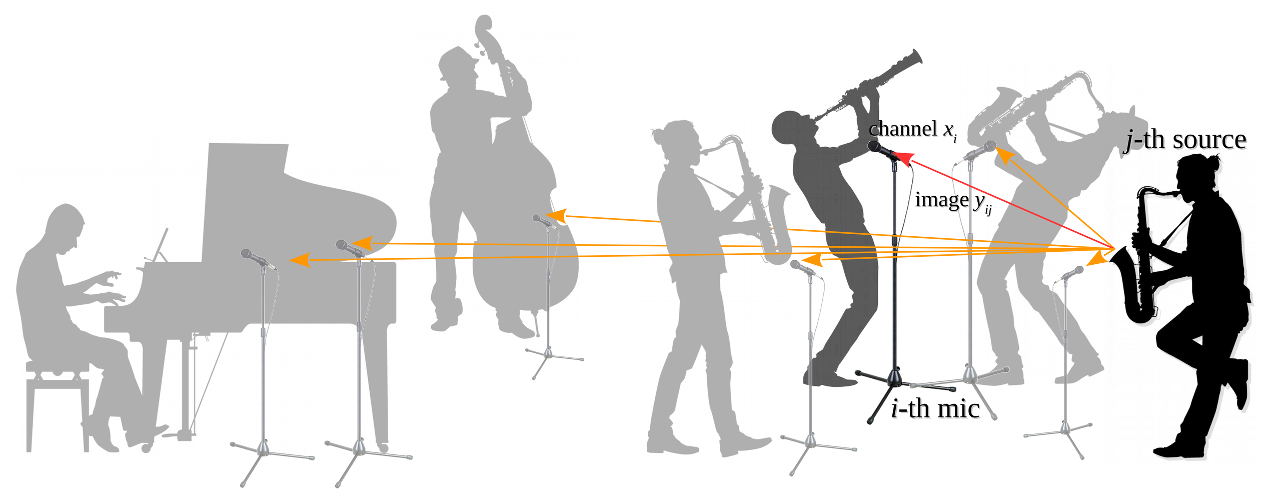
Summary

Live concert recordings ofter features significant **interference** between channels. Recently, we proposed a NFM-based algorithm reduce this effect. However this method is too computationally demanding for full-length concerts.
We show how **Random Projections** of the data can be leveraged for effective estimation of the NMF parameters in acceptable time.

Notation and Model

Time-Frequency domain representation:

$$X_i(f,t) = \sum_{j=1}^{J} Y_{ij}(f,t), \quad V_i(f,t) \triangleq |X_i(f,t)|^2,$$



Interference Reduction goal:
$\forall\ i$ and j, estimate $\hat{Y}_{ij}$ of the images Y_{ij} from the observation of $X_i(f,t)$.

Hypotheses

- *Signals* from different sources *are independent*
- *Neglecting phase dependencies* between channels [1]
  Channels are related only through their energies

Local Gaussian Model and source separation[2]:
The STFT of the image y_{ij}:

$$Y_{ij}(f,t) \sim \mathcal{N}_c(0,\ \lambda_{ij}(f)\ P_j(f,t))$$

Posterior distribution given the mixture through *Wiener filtering*:

$$\hat{Y}_{ij}(f,t) \triangleq \mathbb{E}[Y_{ij}(f,t) \mid X_i(f,t), \Theta] = \frac{\lambda_{ij} P_j(f,t)}{\sum_{j=1}^{J} \lambda_{ij} P_j(f,t)} X_i(f,t).$$

Close-mics assumption:

- Close-mis signals features already good separation quality
- Estimation only for the images of interest

Model parameter to estimate:

$$\Theta = \{\Lambda(f), \{P_j(f,t)\}\}$$

Music Interference Removal Algorithm - `MIRA`

`MIRA` estimates from the parameters likelihood given the observation $X_i(f,t)$[3]:

$$\hat{\Theta} \leftarrow \arg\min_{\Theta} \sum_{f,t,i,j} d_{\text{IS}}\left(V_i(f,t) \| \sum_j \lambda_{ij}(f)\, P_j(f,t)\right)$$

Nonnegative updates:

$$P_j(f,t) \leftarrow P_j(f,t) \cdot \frac{\sum_{i=1}^{I} P_i(f,t)^{-2} V_i(f,t)\, \lambda_{ij}(f)}{\sum_{i=1}^{I} P_i(f,t)^{-1} \lambda_{ij}(f)}$$

$$\lambda_{ij}(f) \leftarrow \lambda_{ij}(f) \cdot \frac{\sum_{t=1}^{T} P_i(f,t)^{-2} V_i(f,t) P_j(f,t)}{\sum_{t=1}^{T} P_i(f,t)^{-1} P_j(f,t)}$$

Initialization:
Close-mics information is used to initialize both $\Lambda(f)$ and $P_j(f,t)$

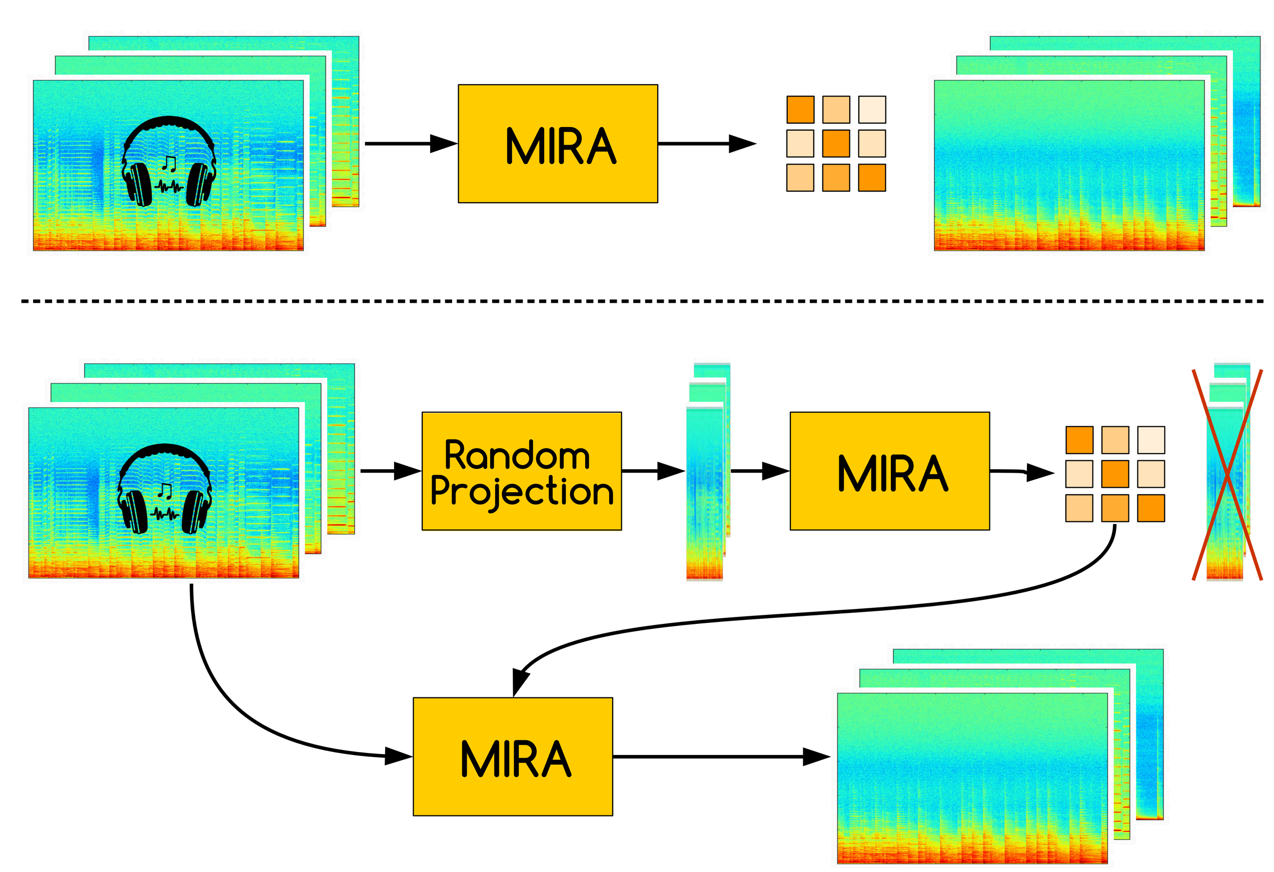
Computational load

- $\Lambda(f)$ requires parsing the whole data
- $P_j(f,t)$ can be estimated online

Random Projection - `fastMIRA`

Contribution

- Learn Λ, on a **small random projection** of the data
- Use $\tilde{\Lambda}(f)$ to estimate $\{P_j(f,t)\}_j$ online



Derivation:

$$M_i(f,r) = \sum_{t=1}^{T} X_i(f,t)\, Q_i(r,t) \text{ with } Q_i(r,t) \sim \mathcal{N}(0,1) \text{ and } R \ll T$$

Thanks to the Gaussian Model of the mixtures, it holds:

$$M_i(f,r) \sim \mathcal{N}(0, \textstyle\sum_j \lambda_{ij}(f)\, S_j(f,r)) \text{ with } S_j(f,r) = \sum_t P_j(f,t)\, Q_i(r,t)^2$$

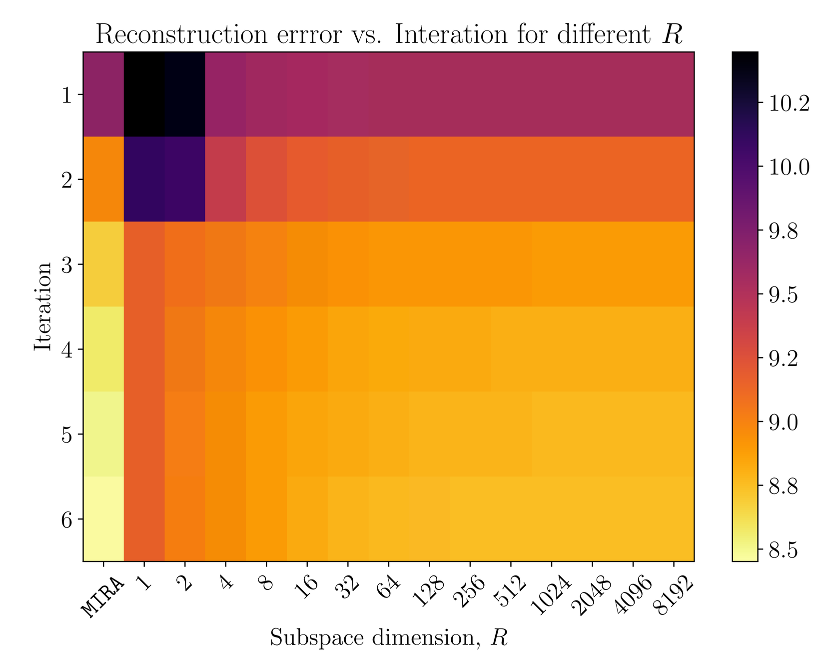
$\Rightarrow$ **MIRA can be used to estimate Λ on M instead of X**

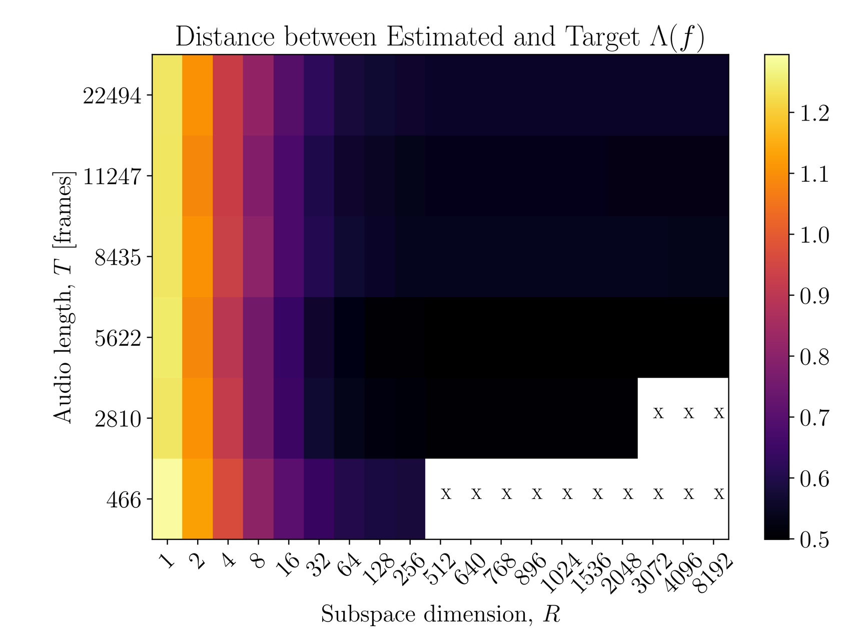
Experimental evaluation

Data: *Power of Love* by Heuy Lewys and the News

- Montreux Jazz Festival 2000
- length of 5'10"; size of 1.2 Gb.
- 48 kHz, 16 bit/sample
- 40 mics, 30 voices

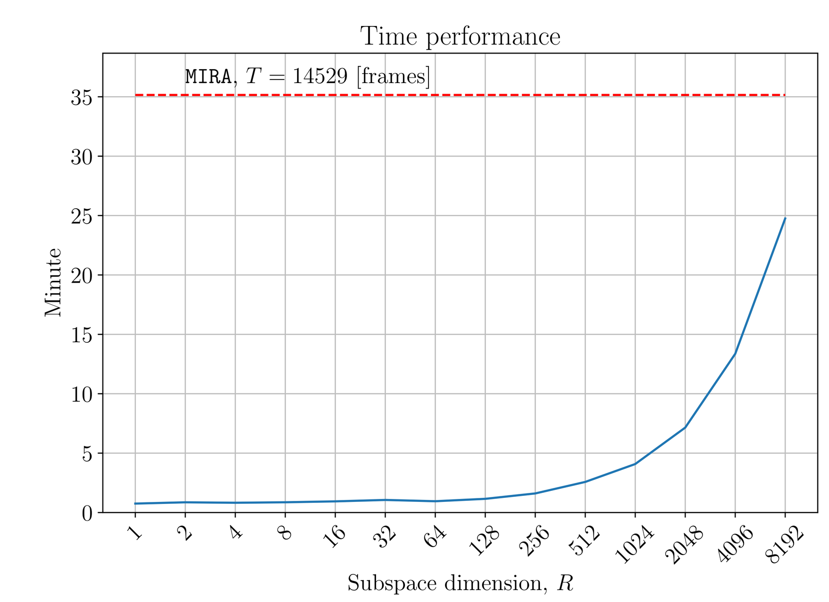
1. Parameter estimation evaluation:





- Already after few iteration `fastMira` provide similar reconstruction error
- After $R = 64$, a good estimation of $\Lambda(f)$ is achieved

2. Computational load evaluation:



Recordings	Size [GB]	Duration [min]	Time elapsed [h]	nMSE [dB]
Huey Lewis...	10	55	6 h 58 min	0.767
Sigur Rós	39.2	123	16 h 41 min	0.909

Table 1: Difference in dB between 5-minute of estimated voice images using `fastMIRA` on the full-length recordings and the same portion processed individually through `MIRA`. $R = 512$ was used.

- A good approximation is yielded in only few minutes
- This method is not particularly affected by the length of the recordings
- `MIRA` could not even run on such recordings on a studio-like workstation

Conclusion

Random Projection of the data can be leveraged for effective estimation of the parameters. Thus, interference reduction can be achieved on full-length live multi-track recordings in acceptable time and used by sound engineers.

References

[1] T. Prätzlich, R. M. Bittner, A. Liutkus, and M. Müller, "Kernel additive modeling for interference reduction in multi-channel music recordings," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pp. 584–588, IEEE, 2015.

[2] A. Liutkus, R. Badeau, and G. Richard, "Gaussian processes for underdetermined source separation," *IEEE Transactions on Signal Processing*, vol. 59, pp. 3155 –3167, July 2011.

[3] D. Di Carlo, K. Déguernel, and A. Liutkus, "Gaussian framework for interference reduction in live recordings," in *AES International Conference on Semantic Audio*, (Erlangen, Germany), p. 8, 2017.

diego.di-carlo@inria.fr