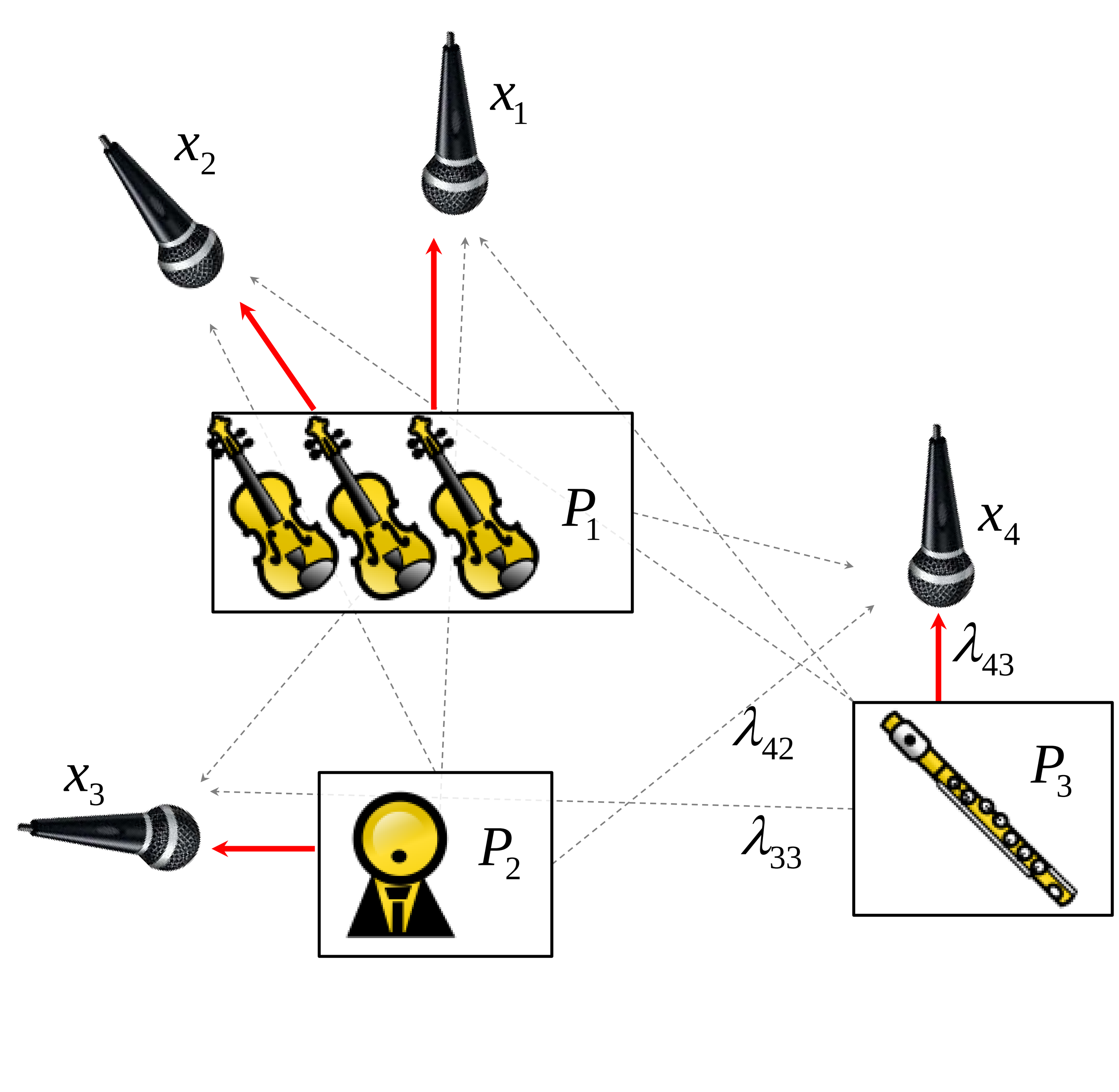
KERNEL ADDITIVE MODELING FOR INTERFERENCE REDUCTION IN MULTI-CHANNEL MUSIC RECORDINGS

Thomas Prätzlich, Rachel M. Bittner, Antoine Liutkus, and Meinard Müller

Abstract

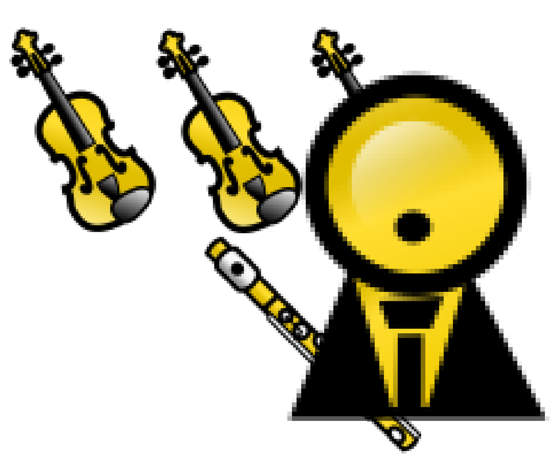
It is difficult to acoustically shield microphones when recording a live music performance. In practice, each microphone also contains interferences from the other voices. In this paper, we aim to reduce these interferences in multi-channel recordings to recover only the isolated voices. Following the recently proposed Kernel Additive Modeling framework, we present a method that iteratively estimates both the power spectral density of each voice and the corresponding strength in each microphone signal. With this information, we build an optimal Wiener filter, strongly reducing interferences. The trade-off between distortion and separation can be controlled by the user through the number of iterations of the algorithm. Furthermore, we present a computationally effective approximation of the iterative procedure. Listening tests demonstrate the effectiveness of the method.

Input Data: Multi-Channel Recording

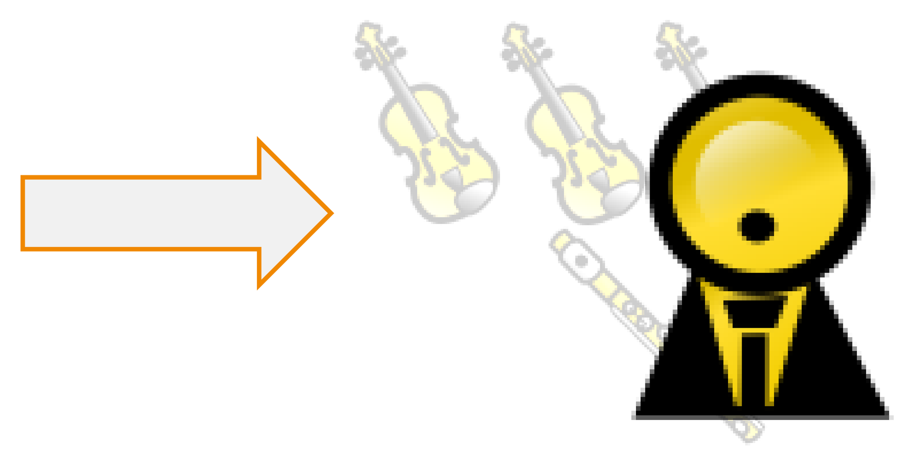


Goal 1: Interference Reduction

Example task: Remove interferences from strings and flute in the singer's microphone channel

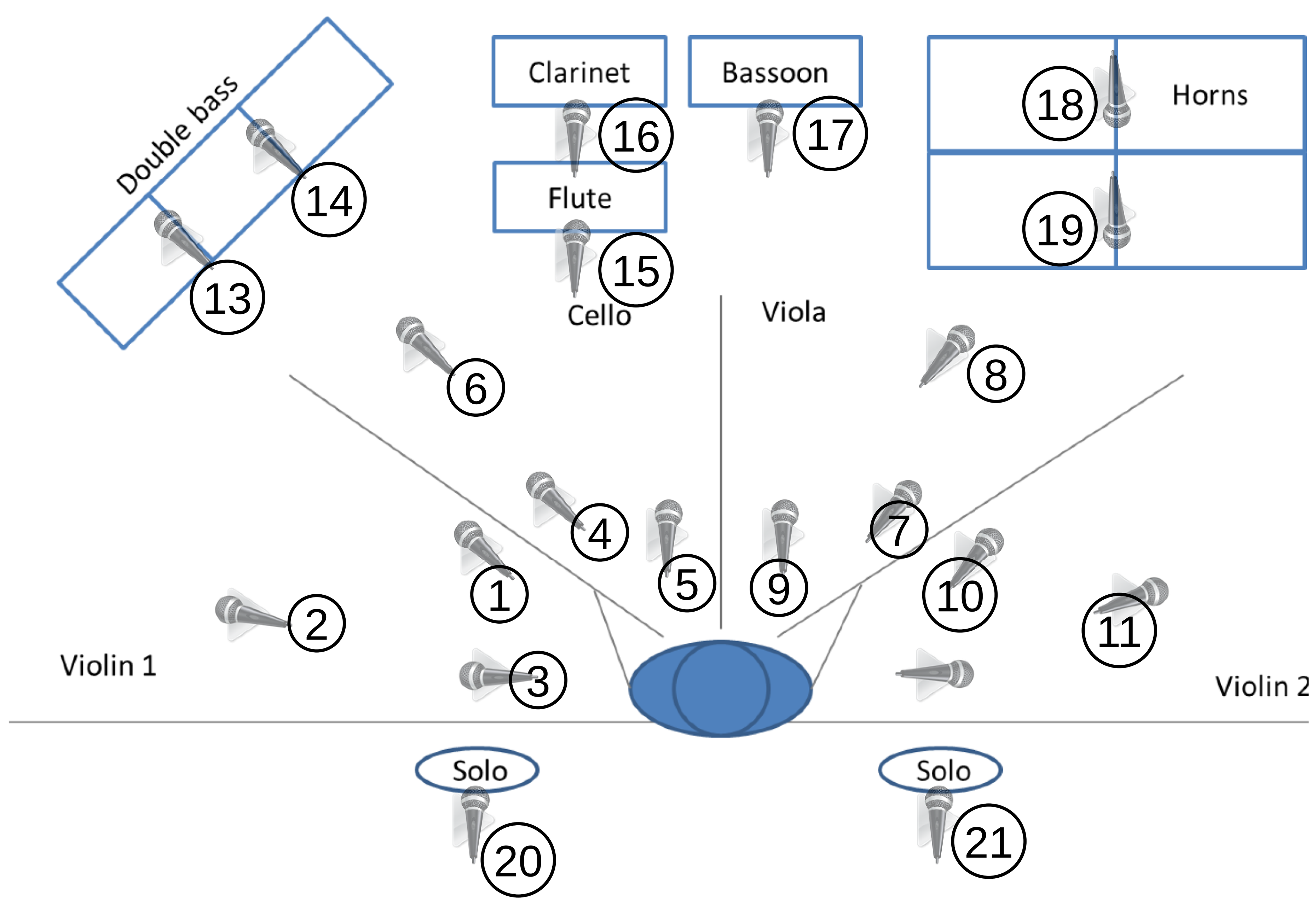


X_3

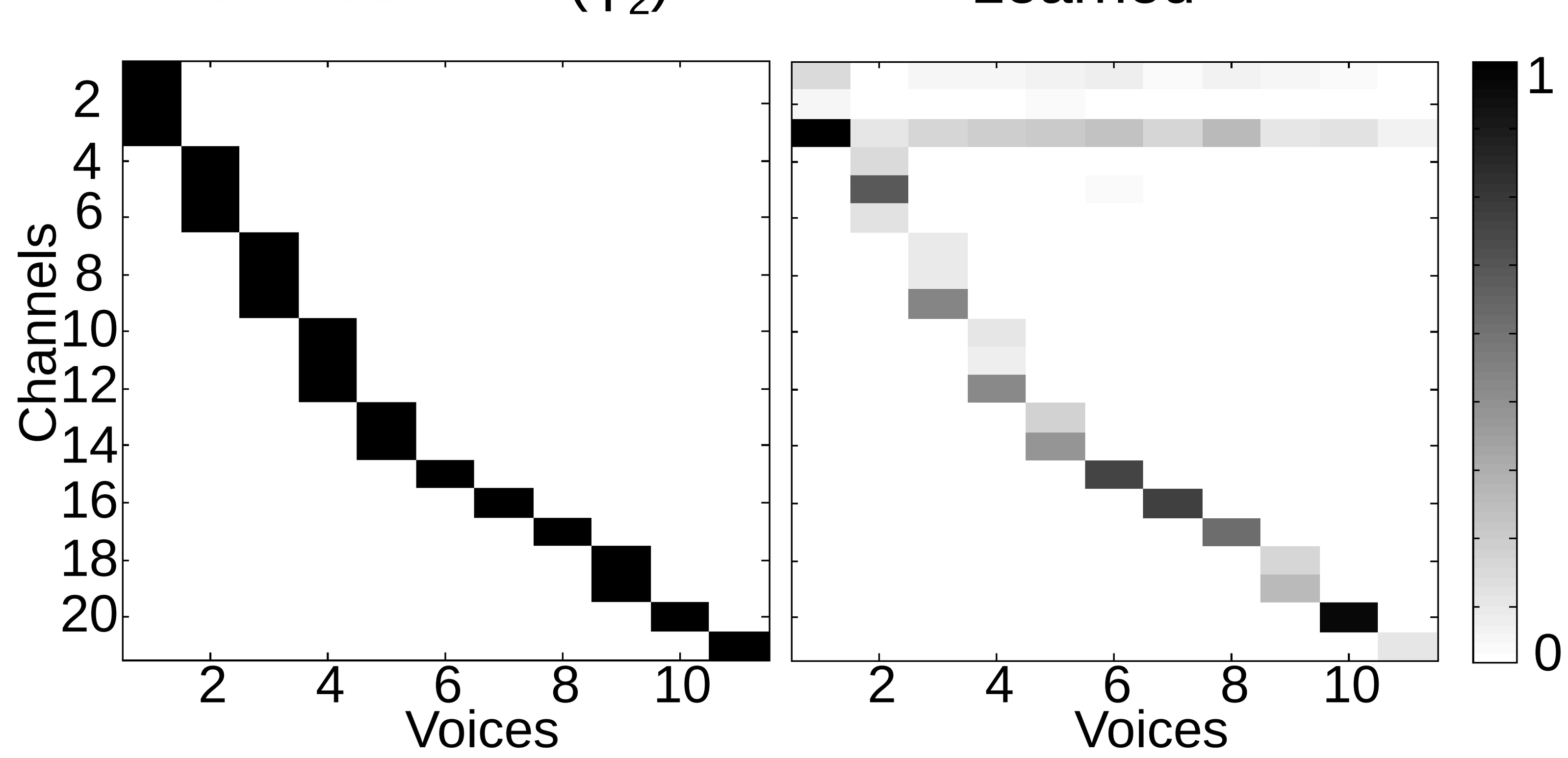


$\hat{Y}_{32}$

Goal 2: Learning of Interference Matrix



Interference Matrix



Learn contribution of each voice into each channel with Non-Negative Matrix Factorization (NMF)

INTERNATIONAL AUDIO LABORATORIES ERLANGEN
A joint institution of Fraunhofer IIS and Universität Erlangen-Nürnberg

AUDIO LABS

ICASSP 2015 Brisbane Australia

KERNEL ADDITIVE MODELING FOR INTERFERENCE REDUCTION IN MULTI-CHANNEL MUSIC RECORDINGS

Thomas Prätzlich, Rachel M. Bittner, Antoine Liutkus, and Meinard Müller

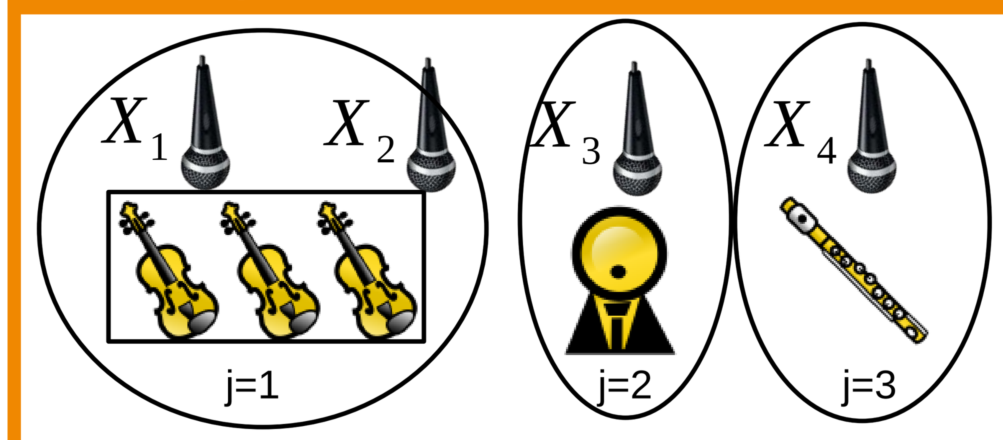
Algorithm

1. **Input:**
   - STFT for each channel i
   - Assignment of each voice j to a set of channels according to φ
   - Minimal interference ρ
   - Optional: A kernel for each voice
2. **Initialization**
   - For each frequency, initialize interference matrix with voice channel assignment and minimal interference ρ
   - Set initial estimate of voice's images to STFT of associated microphone signal
3. **Parameter Fitting**
   a) For each voice: Update the Power Spectral Densities (PSD) according to frequency dependent interference matrix
   b) *Optional: Apply a voice specific kernel median filter on each PSD that captures characteristics of each voice*
   c) Update interference matrix with NMF
   d) Rescale each PSD and normalize
4. **Separation Step / Output:** Update the voice's images

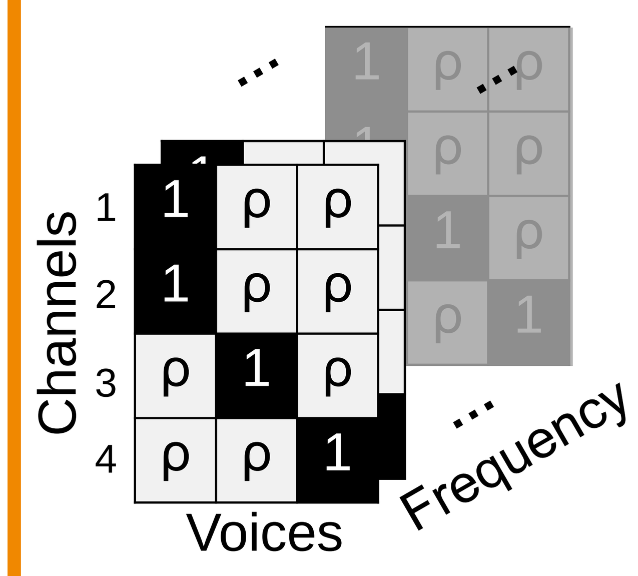
$$\hat{Y}_{ij}(\omega,t) = \frac{\lambda_{ij}(\omega)\,P_j(\omega,t)}{\sum_{j'=1}^{J}\lambda_{ij'}(\omega)\,P_{j'}(\omega,t)}X_i(\omega,t) \triangleq W_{ij}(\omega,t)\,X_i(\omega,t)\,,$$

5. **Iteration:** For another iteration, return to step (3)

Example

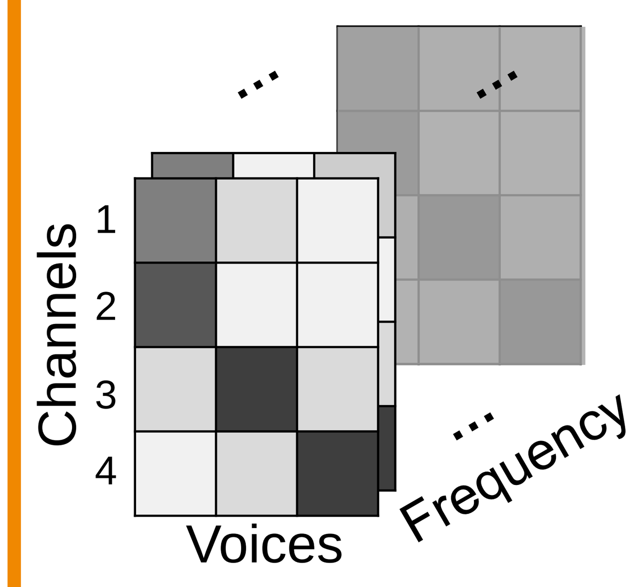


$\varphi(1) = \{1,2\}$
$\varphi(2) = \{3\}$
$\varphi(3) = \{4\}$



$$\forall (i,j,\omega),\ \lambda_{ij}(\omega) = \begin{cases} 1 & \text{if } i \in \varphi(j) \\ \rho & \text{otherwise,} \end{cases}$$

$\hat{Y}_{11} = X_1 \quad \hat{Y}_{21} = X_2$	
$\hat{Y}_{32} = X_3$	
$\hat{Y}_{43} = X_4$	

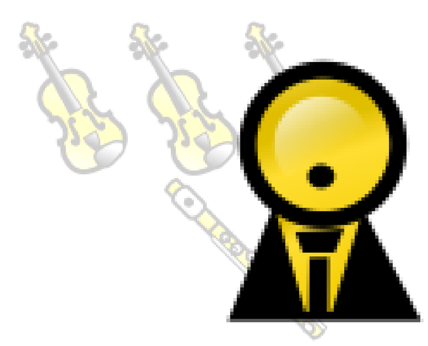


Assumption $P_j(\omega,t) \approx \frac{1}{\lambda_{ij}(\omega)}\left|\hat{Y}_{ij}(\omega,t)\right|^2$

$$P_j(\omega,t) \leftarrow \frac{1}{|\varphi(j)|}\sum_{i\in\varphi(j)}\frac{1}{\lambda_{ij}(\omega)}\left|\hat{Y}_{ij}(\omega,t)\right|^2$$

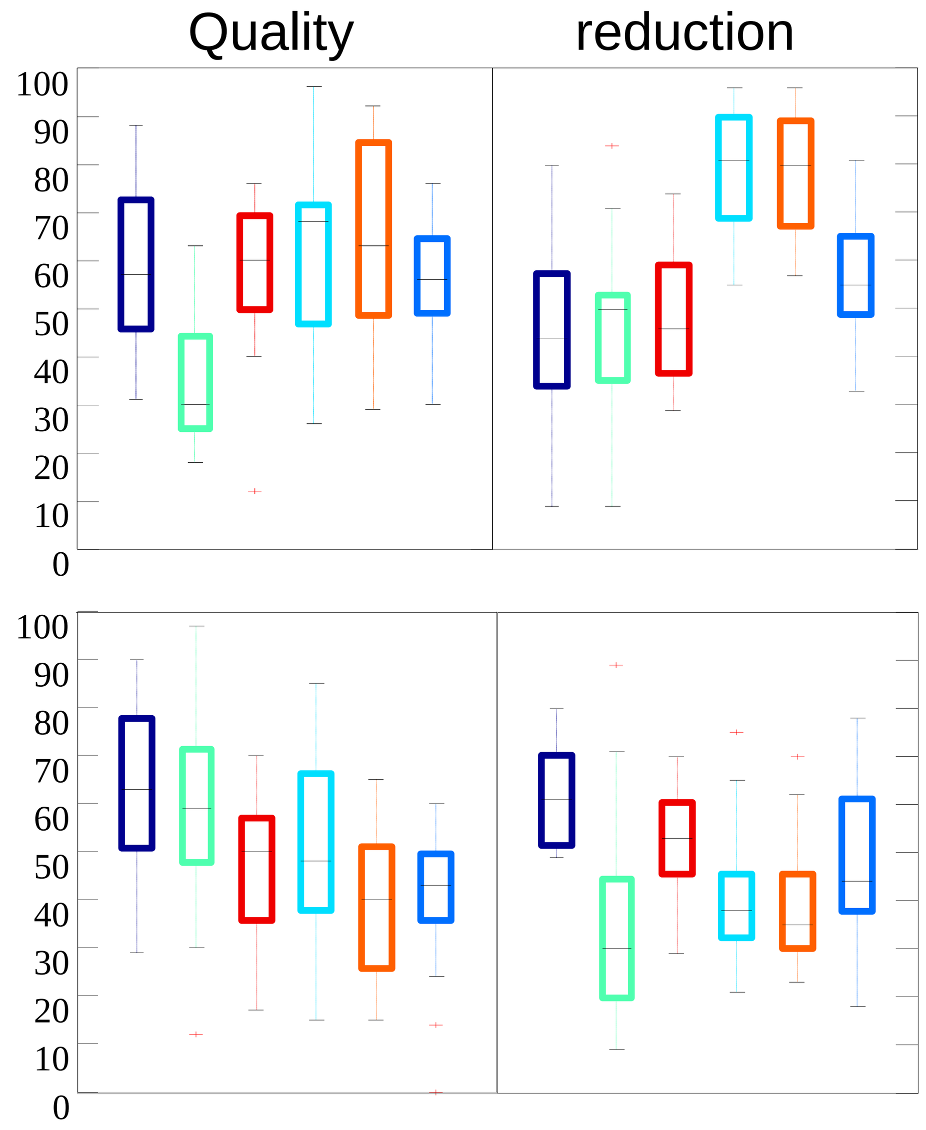
$$\lambda_{ij}(\omega) \leftarrow \lambda_{ij}(\omega)\cdot\frac{\sum_{t=1}^{N_t}\hat{V}_i(\omega,t)^{\beta-2}\,V_i(\omega,t)\,P_j(\omega,t)}{\sum_{t=1}^{N_t}\hat{V}_i(\omega,t)^{\beta-1}\,P_j(\omega,t)}$$

$$V_i(\omega,t) \triangleq |X_i(\omega,t)|^2$$
$$\hat{V}_i(\omega,t) \triangleq \textstyle\sum_j \lambda_{ij}(\omega)\,P_j(\omega,t)$$



$$\hat{Y}_{32} = W_{32} \otimes X_3$$

Evaluation – Listening Test



Cello section

- 3 microphones
- More microphones are beneficial for interference reduction

Singer

- 1 microphone
- More iterations may decrease the overall quality

Kokkinis 2012 [1]
KAMIR ρ = 1, φ(j) = j, no iteration
KAMIR ρ = 0.1, φ_2, 4 iterations
KAMIR ρ = 0.1, φ_2, 4 iterations
KAMIR ρ = 0.1, φ_2, 4 iterations, interference matrix learned
KAMIR ρ = 0.1, φ_2, 1`iteration, interference matrix learned, sigmoid on gains

References

[1] Elias K. Kokkinis, Joshua D. Reiss, and John Mourjopoulos, "A Wiener filter approach to microphone leakage reduction in close-microphone applications," *IEEE Transactions on Audio, Speech & Language Processing*, vol. 20, no. 3, pp. 767–779,2012.

[2] Antoine Liutkus, Derry Fitzgerald, Zafar Rafii, Bryan Pardo, and Laurent Daudet, "Kernel Additive Models for Source Separation," *IEEE Transactions on Signal Processing*, June 2014.

[3] Accompanying website: http://www.audiolabs-erlangen.de/resources/MIR/2015-ICASSP-KAMIR

Acknowledgments

This work has been supported by the BMBF project *Freischütz Digital* (Funding Code 01UG1239A to C). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer IIS.

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG · Inria informatics mathematics · NYU · MARL · Fraunhofer IIS