

Background

Multimodal source separation

Informing a music source separation system with **additional information** (text, score, video features) was proven to benefit the separation quality, especially when limited or no training data is available.

*What if this additional information comes from the **user**?*

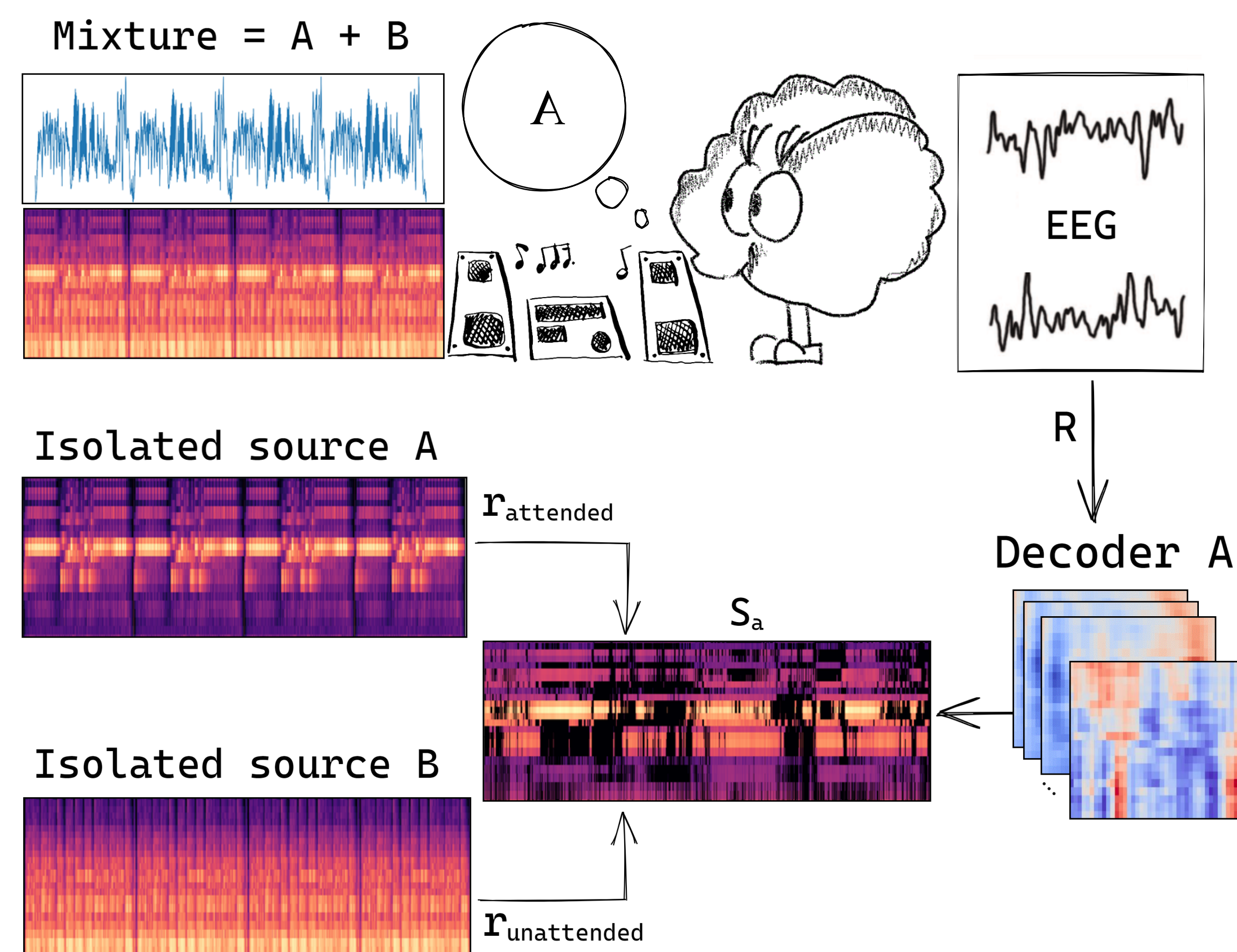
Selective auditory attention

Cognitive mechanism that allows us to focus on a sound source of interest:

- can be tracked in the **neural activity** (EEG, ECoG, MEG)
- the attended source's neural encoding is **stronger** than the other ones

EEG-based Auditory Attention Decoding

[Cantisani et al., 2019]: the EEG tracks musically relevant features highly correlated with the attended source and weakly correlated with the unattended one.



Neuro-steered Audio Source Separation

Typically the separation and the decoding tasks are tackled **sequentially**:

- a separation system provides the reference sources for the decoding;
- and the decoding system selects the source which needs to be enhanced.

Acknowledgements

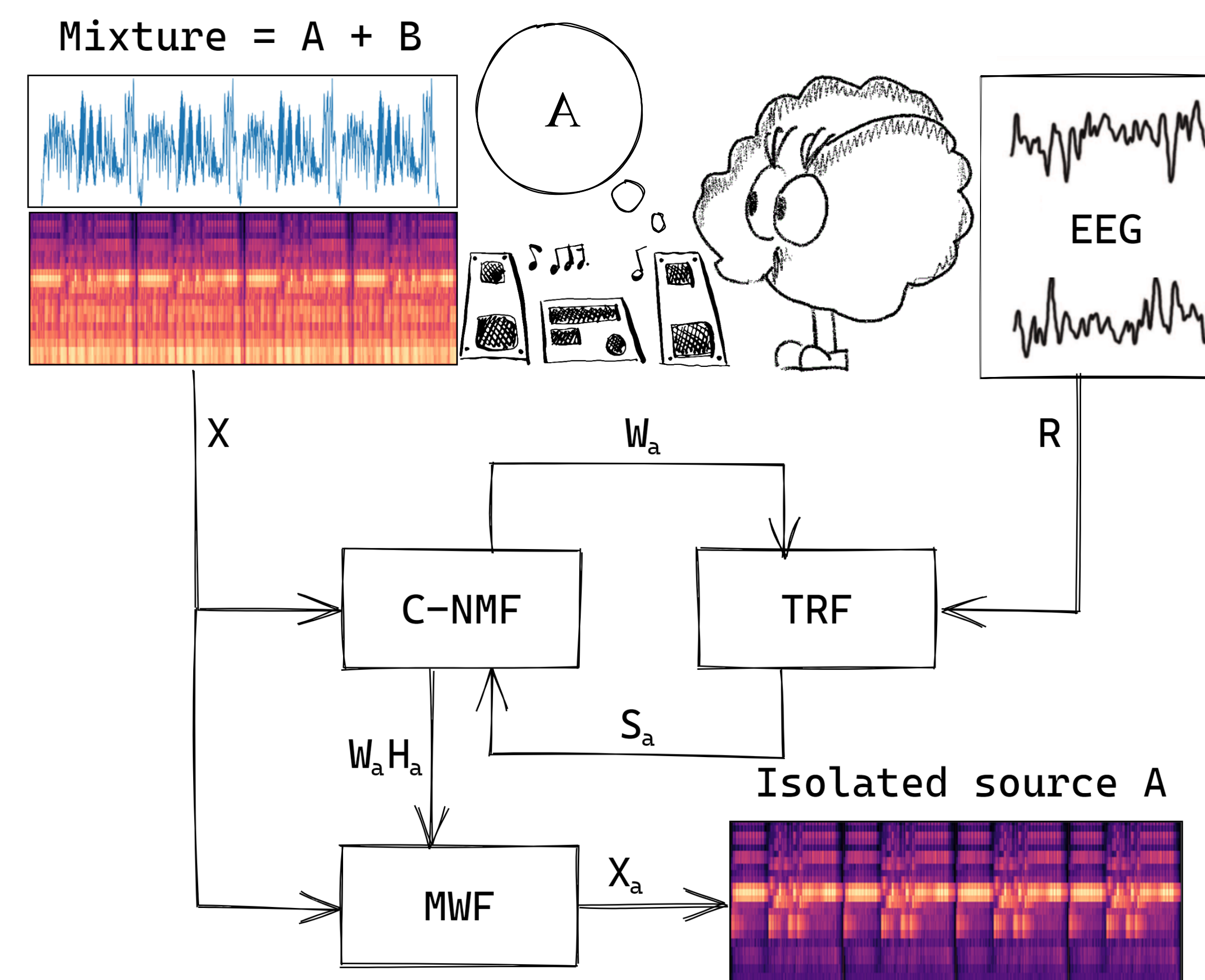


This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 765068

Framework

Proposed model: a joint approach

Unsupervised nonnegative matrix factorisation variant, named **Contrastive-NMF**, that separates a target instrument, guided by the user's selective auditory attention to that instrument, which is tracked in his/her EEG response to music.



Contrastive-NMF

$$\begin{cases} C(\mathbf{W}, \mathbf{H}) = \underbrace{D(\mathbf{X}|\mathbf{W}\mathbf{H})}_{\text{audio factorization}} + \underbrace{\mu\|\mathbf{H}\|_1 + \beta\|\mathbf{W}\|_1}_{\text{sparsity}} - \underbrace{\delta(\|\mathbf{H}_a\mathbf{S}_a^T\|_F^2 - \|\mathbf{H}_u\mathbf{S}_u^T\|_F^2)}_{\text{contrast}} \\ \mathbf{W}, \mathbf{H}, \mathbf{S}_a \geq 0 \\ \|\mathbf{h}_{k:}\|_2 = 1, \|\mathbf{s}_{k:}\|_2 = 1. \end{cases}$$

The proposed cost aims at:

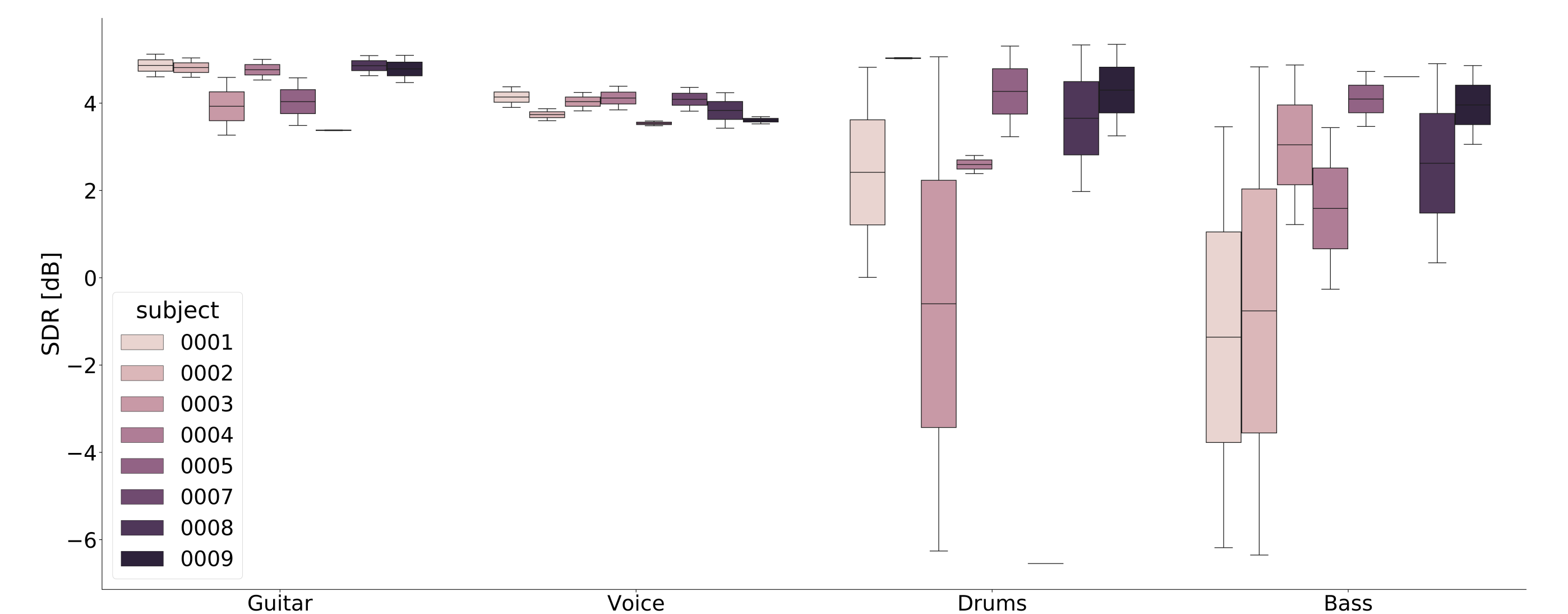
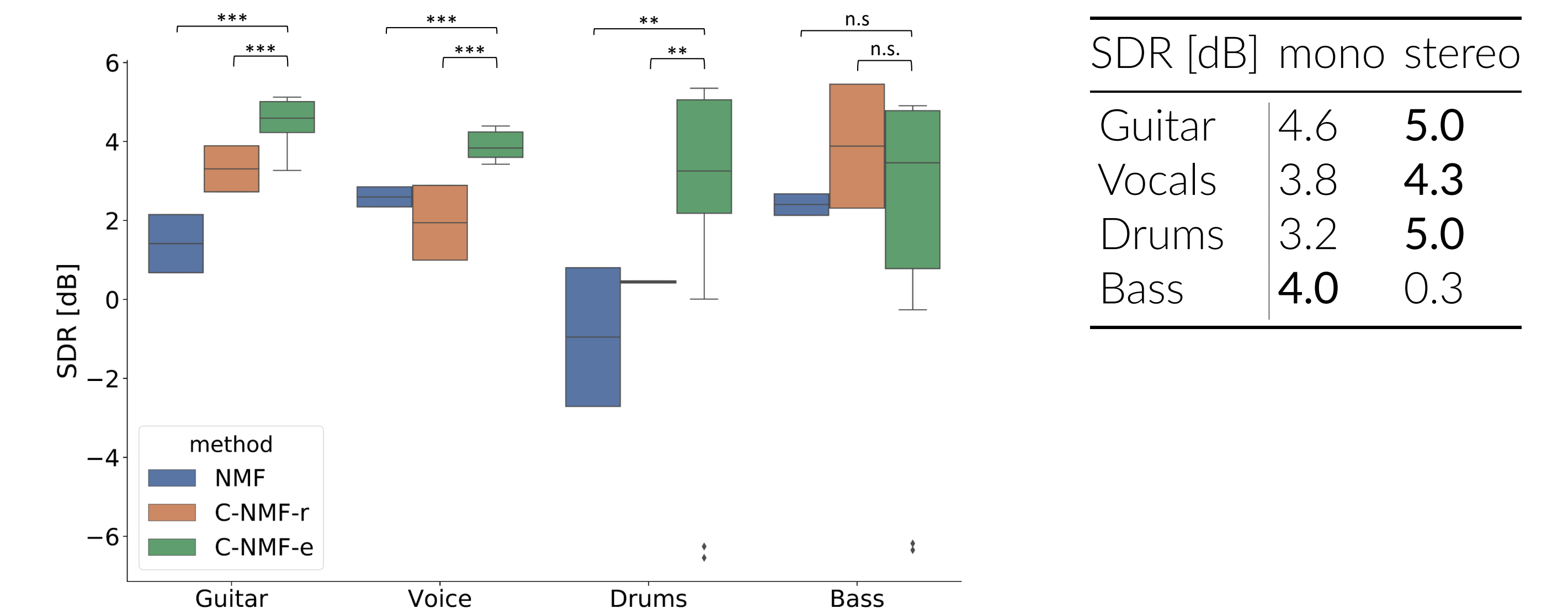
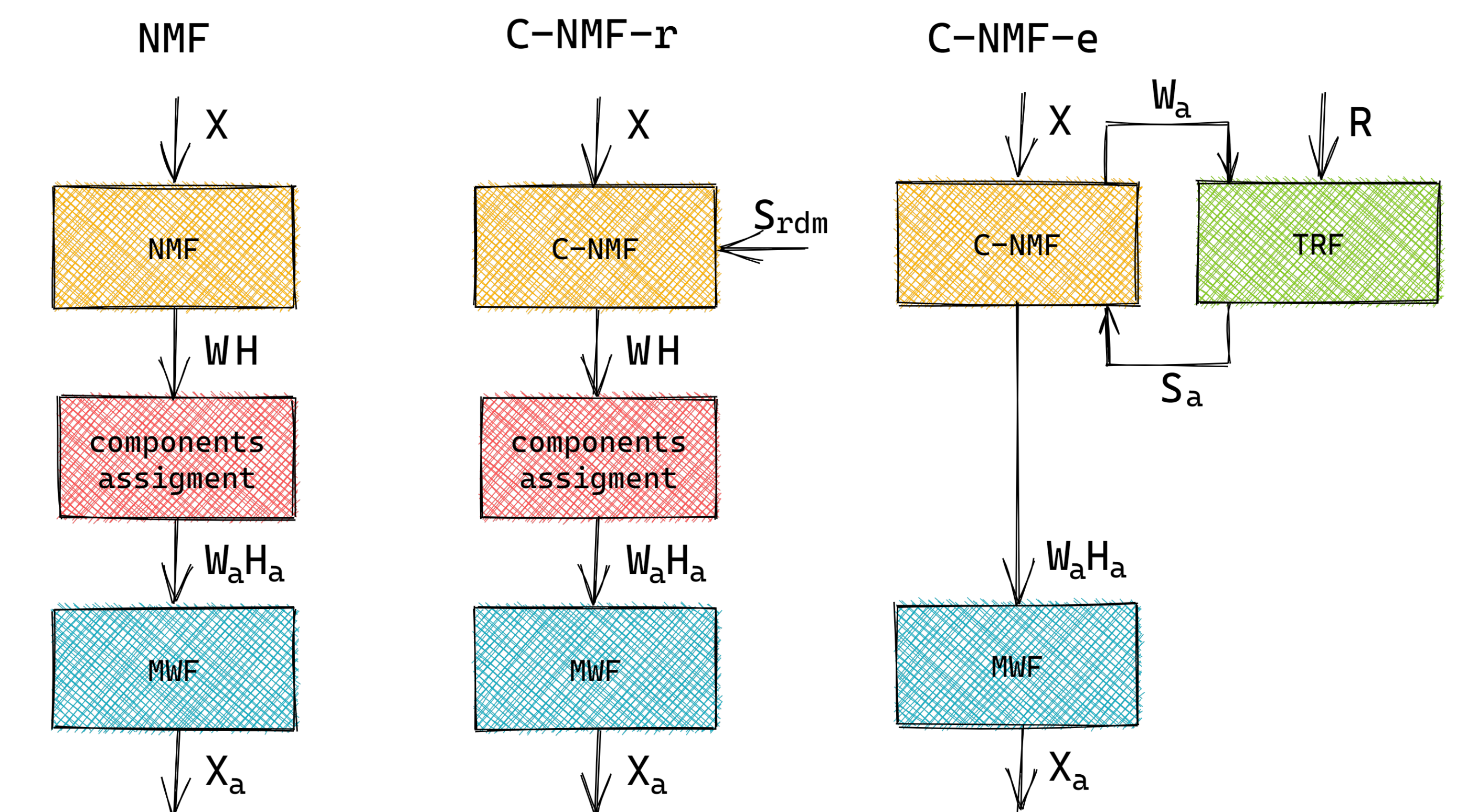
- decomposing** the mixture spectrogram;
- imposing** the sparsity of \mathbf{W} and \mathbf{H} ;
- maximising** the similarity of \mathbf{S}_a with \mathbf{H}_a ;
- minimising** the similarity of \mathbf{S}_a with \mathbf{H}_u .

Conclusions

We propose a new model for **neuro-steered music source separation**:

- improves the separation**, especially in difficult cases;
- automatically separates the attended source**;
- auditory attention decoding **without the reference sources**.

Experiments



Resources

- Code: <https://github.com/giorgiacantisani/Contrastive-NM>
- Demo: <https://giorgiacantisani.github.io/projects/C-NMF>
- Update rule: <https://hal.telecom-paris.fr/hal-02978978v4/file/Update-rule-C-NMF.pdf>