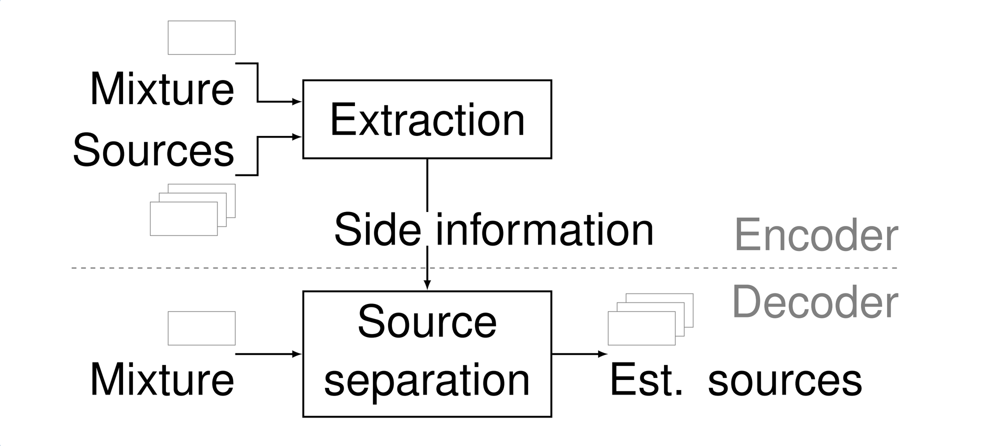
NMF-BASED INFORMED SOURCE SEPARATION

Christian Rohlfing, Julian M. Becker, Mathias Wien

Institut für Nachrichtentechnik, RWTH Aachen University

Informed Source Separation

- Active listening and remixing of music (e.g. karaoke) requires audio objects
- Informed source separation (ISS) uses source separation for audio object coding
  - Encoder extracts compact set of side information with knowledge of sources
  - Side information assists source separation step in decoder to yield estimated sources

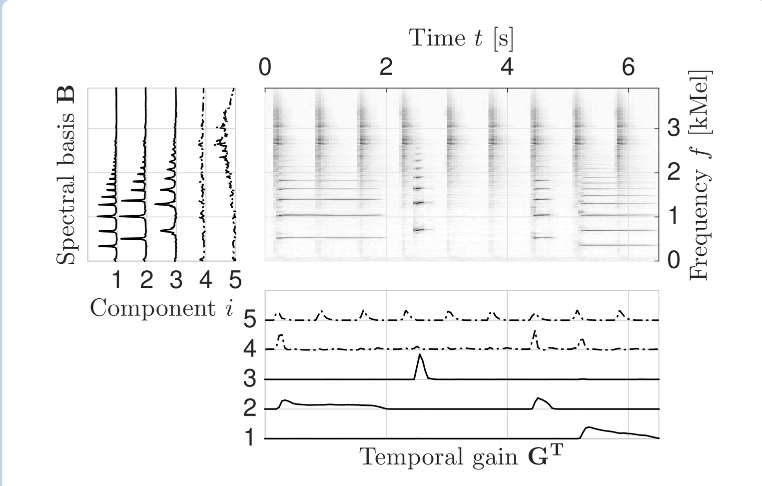


Main Approach

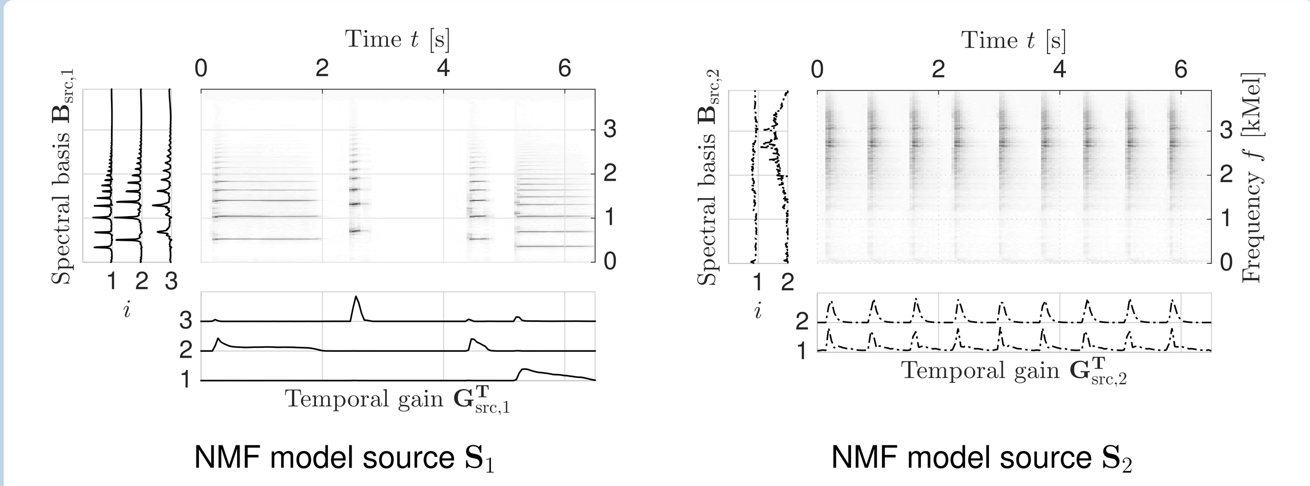
- Proposed decoder uses semi-blind *source separation* (SBSS) algorithm which models mixture spectrogram with non-negative matrix factorization (NMF) to obtain source estimates by time-frequency masking
- Encoder calculates interference-free NMF model of the sources to *extract* side information:
  - compact initialization for mixture NMF in decoder
  - residuals between mixture NMF and source NMF model

$\Rightarrow$ Instead of transmitting source model directly, use SBSS and transmit only initialization and residuals

Mixture NMF Model



Source NMF Model



Decoder

Decoder uses SBSS algorithm [1] to estimate M sources $\underline{\mathbf{S}}_m$ out of their linear mixture $\underline{\mathbf{X}} = \sum_{m=1}^{M} \underline{\mathbf{S}}_m$ in time-frequency domain:

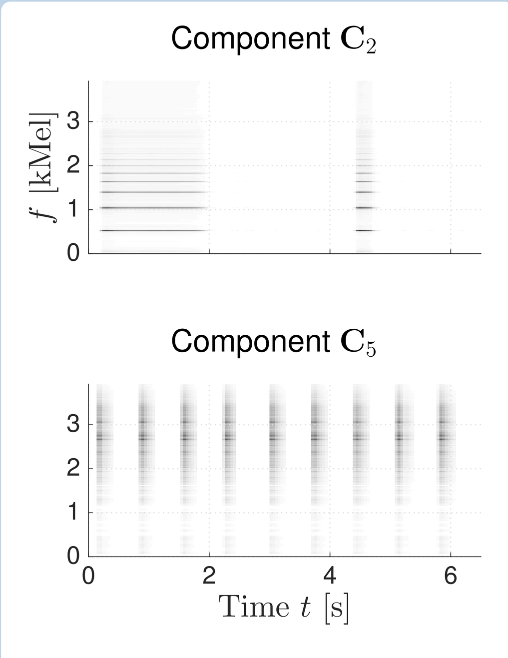
1) **Non-negative matrix factorization (NMF):** NMF of $\mathbf{X} = |\underline{\mathbf{X}}|$ yields spectral basis $\mathbf{B}$ and temporal gain $\mathbf{G}$ with I component spectrograms $\mathbf{C}_i$

$$\mathbf{X}(f,t) \approx \sum_{i=1}^{I} \mathbf{B}(f,i)\mathbf{G}(t,i) = \sum_{i=1}^{I} \mathbf{C}_i(f,t)$$

2) **Wiener filter:** Time-frequency masking of complex mixture $\underline{\mathbf{X}}$ obtains estimated sources $\tilde{\underline{\mathbf{S}}}_m$

$$\tilde{\underline{\mathbf{S}}}_m(f,t) = \underline{\mathbf{X}}(f,t) \sum_{\mathbf{r}(i)=m} \mathbf{C}_i(f,t) / \sum_j \mathbf{C}_j(f,t)$$

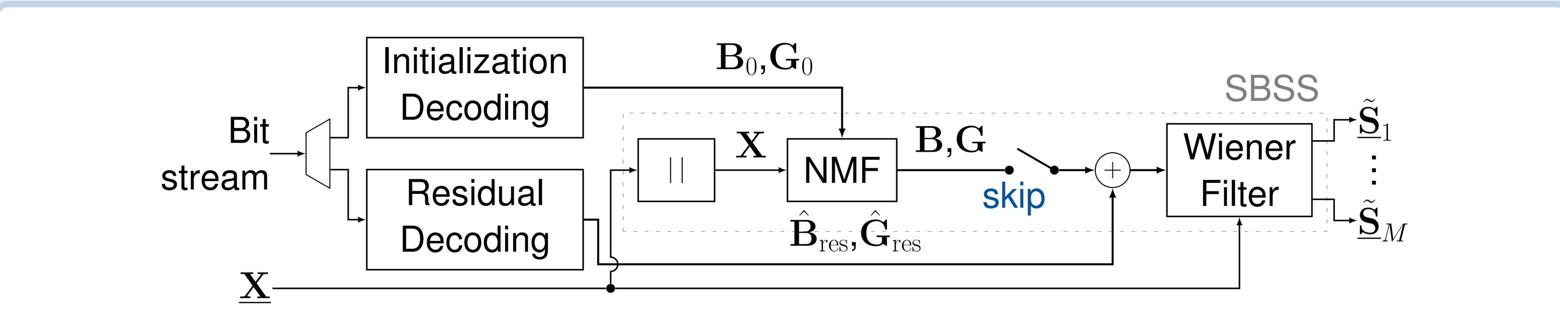
with grouping vector $\mathbf{r}$ linking components to estimated sources



Here: $\mathbf{r} = (1,1,1,2,2)^{\mathrm{T}}$

Additional steps to enhance separation quality:

- *Initialization:* Decoder uses $\mathbf{B}_0,\mathbf{G}_0$ provided by encoder to initialize NMF
- *Residuals:* Residual model $\hat{\mathbf{B}}_{\mathrm{res}},\hat{\mathbf{G}}_{\mathrm{res}}$ is added to mixture NMF model



[1] Martin Spiertz, *Underdetermined Blind Source Separation for Audio Signals*, vol. 10 of Aachen Series on Multimedia and Communications Engineering. Aachen: Shaker Verlag, July 2012.

Encoder

Encoder steers decoder by initializing and refining mixture NMF:

1) **Source NMF model:** Factorize sources with NMF to yield $\mathbf{B}_{\mathrm{src}},\mathbf{G}_{\mathrm{src}}$
2) **Initialization:** Construct $\mathbf{B}_0$ out of frames of $\mathbf{X}$ with frame indices $\mathbf{d}$ and $\mathbf{G}_0(t,i) \in \{0,1\}$ with $\mathbf{G}_{\mathrm{src}}$ in dB and threshold τ

$$\mathbf{B}_0(f,i) = \mathbf{X}(f,\mathbf{d}(i)) \quad \text{and} \quad \mathbf{G}_0 = \mathbf{G}'_{\mathrm{src}} > \tau$$

Binary initial gain matrix is run-length encoded

$\Rightarrow$ Needs only small amount of bit rate

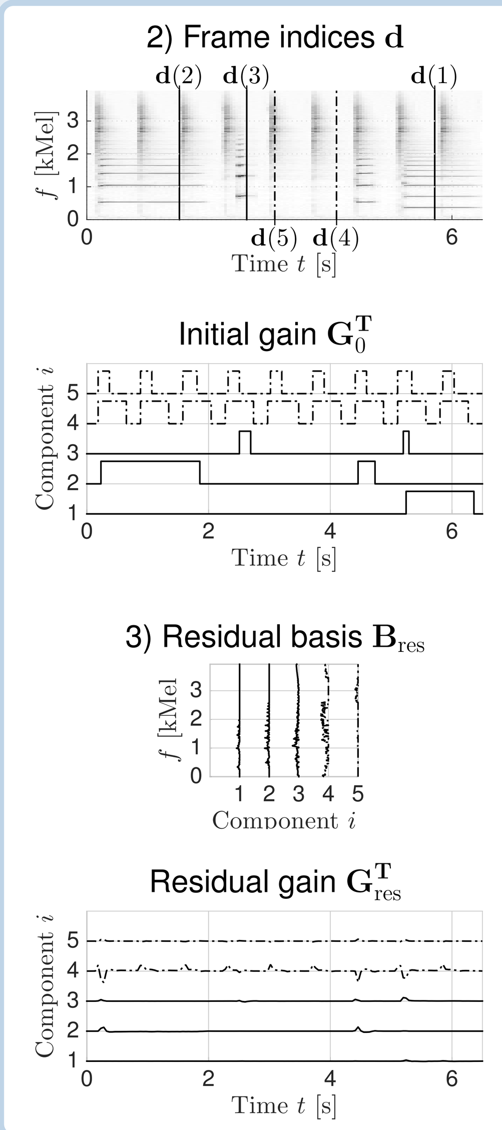
3) **Residuals:** Remaining differences between mixture and source NMF after normalization

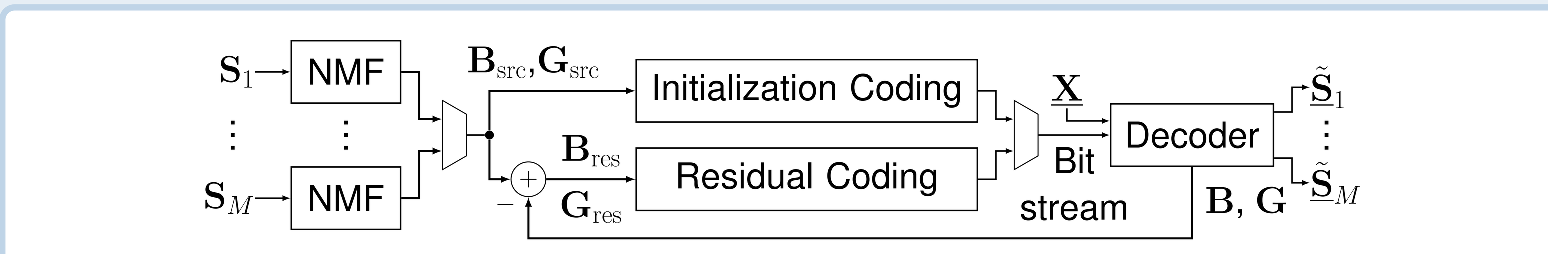
$$\mathbf{B}_{\mathrm{res}} = \mathbf{B}''_{\mathrm{src}} - \mathbf{B}'' \quad \text{and} \quad \mathbf{G}_{\mathrm{res}} = \mathbf{G}''_{\mathrm{src}} - \mathbf{G}''$$

Residuals are scalar quantized and encoded with adaptive arithmetic coding

4) **Parameter optimization:** Encoder tests decoder with different parameter combinations to choose optimal configuration

Skip Mode: For direct transmission of source model, mixture NMF in decoder can be skipped (skip) such that $\mathbf{B}_{\mathrm{res}} = \mathbf{B}_{\mathrm{src}}$ and $\mathbf{G}_{\mathrm{res}} = \mathbf{G}_{\mathrm{src}}$





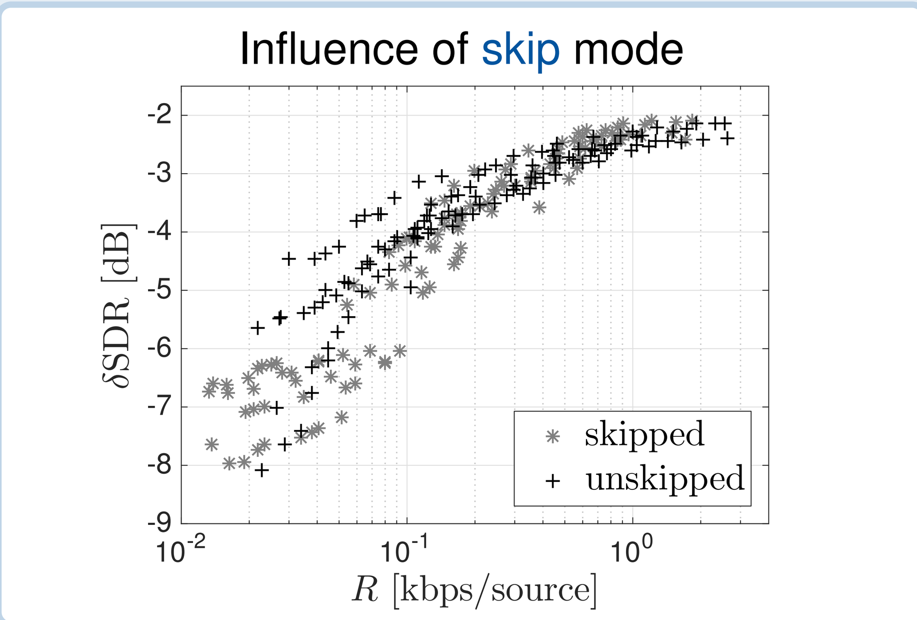
Results and Conclusions

Setup:

- Five mixtures (3 to 6 sources, taken from QUASI database [2])
- STFT and Mel-filtering for spectral dimension reduction
- NMF: Minimize β-divergence with $\beta \in \{0,1,2\}$
- $(\delta\mathrm{SDR}, R)$ points for number of components $I/M \in [2,30]$ and different number of quantization bins for residual quantization

Results: Compared to direct transmission of source NMF model, using SBSS improves quality significantly at lower bit rates.

[2] http://www.tsi.telecom-paristech.fr/aao/en/2012/03/12/quasi/



Conclusions

- Proposed ISS decoder uses semi-blind source separation (SBSS)
- Encoder transmits compact SBSS initialization and residuals
- Using SBSS enhances separation quality at lower bit rates
- Proposed scheme works even blindly without any side information

rohlfing@ient.rwth-aachen.de

www.ient.rwth-aachen.de

Institut für Nachrichtentechnik, Melatener Str. 23, 52074 Aachen

ICASSP 2016