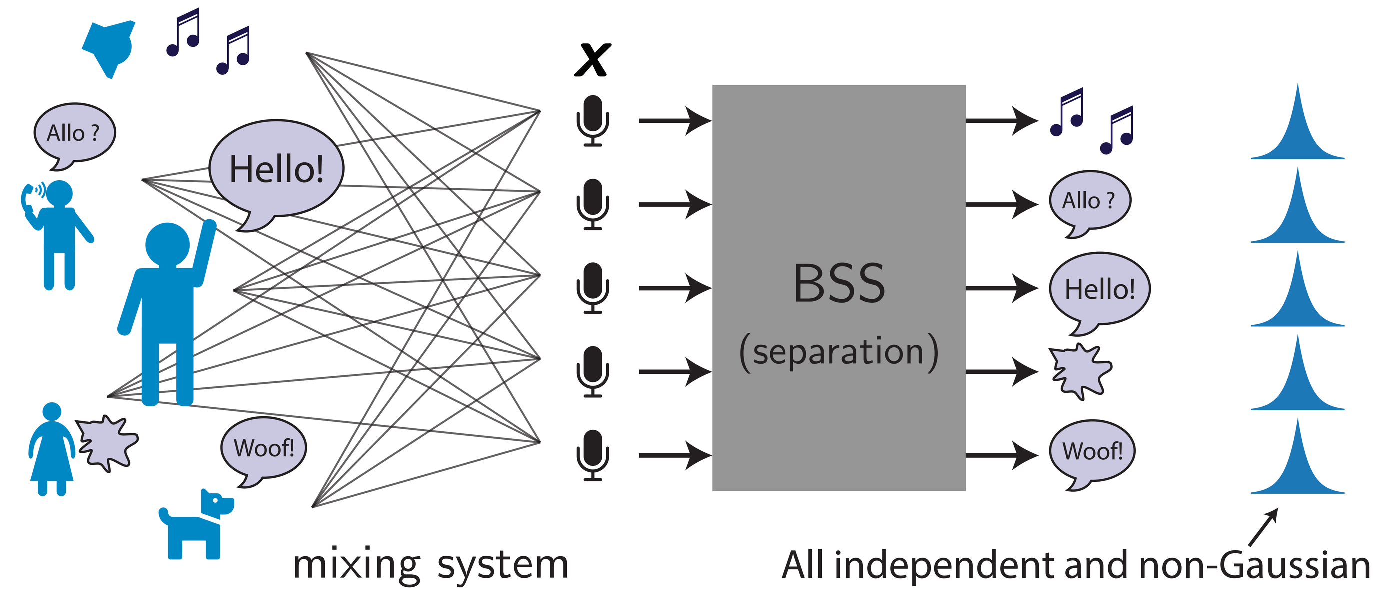
Surrogate Source Model Learning for Determined Source Separation

Robin Scheibler and Masahito Togami

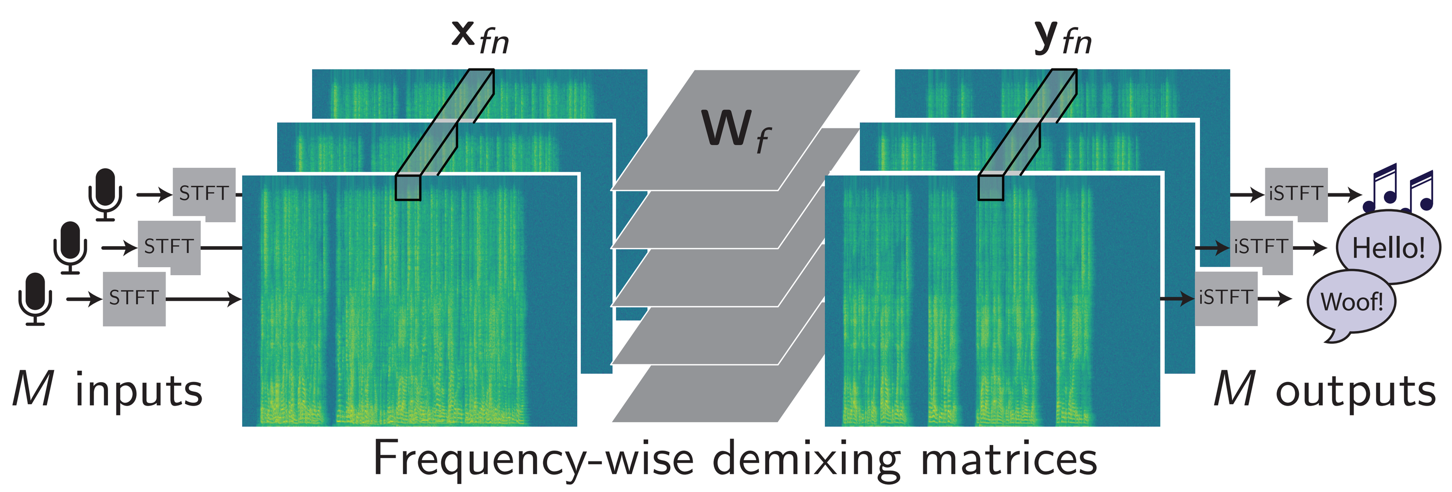
LINE

Blind Source Separation

Abstract —We propose to replace the surrogate function of AuxIVA by a DNN. The model is trained end-to-end and shows superior performance. It generalizes to different number of channels and BSS algorithms.



Frequency-domain BSS



Independent Vector Analysis [1, 2]

1. Sources are independent
2. Source model (joint pdf), $\mathbf{Y}$ is the spectrogram

$$p(\mathbf{Y}) = \frac{1}{c} e^{-G(\mathbf{Y})}$$

Then, the maximum-likelihood estimator is the minimum of

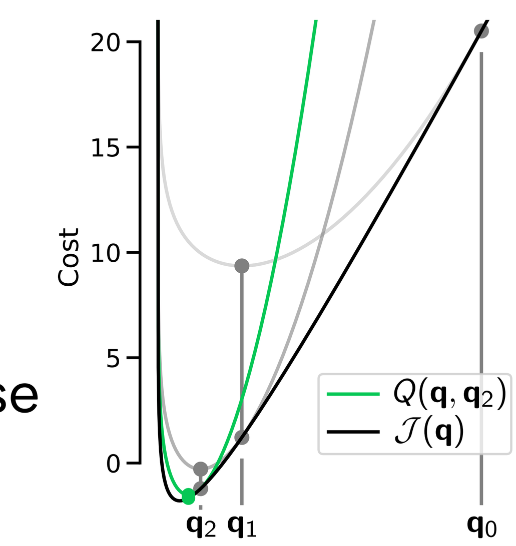
$$\mathcal{L} = \sum_{k=1}^{M} G(\mathbf{Y}_k) - 2N \sum_f \log|\det(\mathbf{W}_f)| + \text{const.}$$

where $y_{kfn} = \mathbf{w}_{kf}^H \mathbf{x}_{fn}$.

Suppose there exists $u_{fn}(\mathbf{Y})$ s.t.

$$G(\mathbf{Y}) \leq \sum_{fn} u_{fn}(\hat{\mathbf{Y}})|y_{fn}|^2 + \text{const.},$$

with equality iff $\mathbf{Y} = \hat{\mathbf{Y}}$. Then, we can use AuxIVA [3], an MM algorithm!



Iterative Source Steering for AuxIVA [5]

ISS is an efficient algorithm to perform AuxIVA. For $k = 1, \ldots, M$, and $m = 1, \ldots, M$, do

$$y_{mfn} \leftarrow y_{mfn} - \left(\frac{\sum_n u_{fn}(\mathbf{Y}_m) y_{mfn} y_{kfn}^*}{\sum_n u_{fn}(\mathbf{Y}_m)|y_{kfn}|^2} \right) y_{kfn},$$

It can be interpreted as

$$\min_{v \in \mathbb{C}} \sum_n u_{fn}(\mathbf{Y})|y_{mfn} - v\, y_{kfn}|^2$$

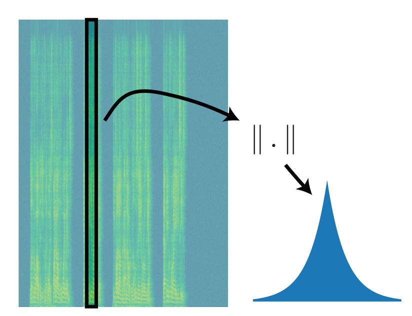
where $u_{fn}(\mathbf{Y}_m)$ is a mask removing the influence of source m.

Well-suited for DNN: no matrix inv., low-complexity.

Traditional Source Models

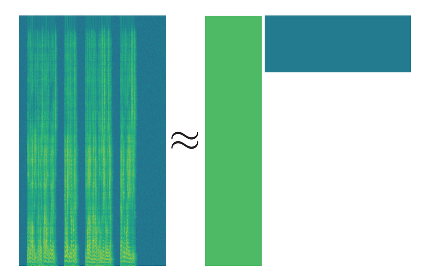
Circularly Symmetric [1, 3]

- No dep. accross time
- All freq. equal



Non-negative Low-rank [4]

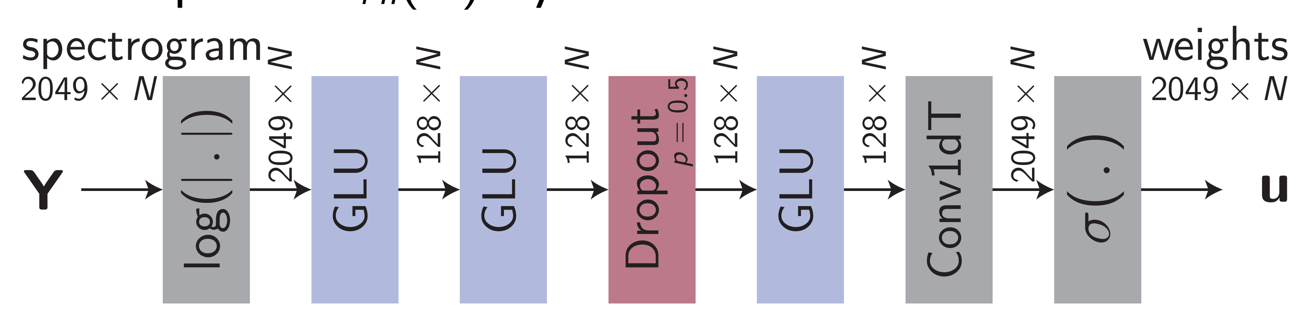
- Extra variables to est.
- Not always appropriate



⇒ Lack of flexibility!

Surrogate Source Model Learning

Key Idea Replace $u_{fn}(\mathbf{Y})$ by a DNN



Train the weights end-to-end using ISS for BSS.

Loss Functions (with PIT)

SI-SDR (time-domain)

$$L_{\text{SDR}}(\hat{\mathbf{y}}, \mathbf{s}) = 10 \log_{10} \left(\frac{\|\alpha \mathbf{s}\|^2}{\|\alpha \mathbf{s} - \hat{\mathbf{y}}\|^2} \right), \quad \alpha = \frac{\hat{\mathbf{y}}^\top \mathbf{s}}{\mathbf{s}^\top \mathbf{s}}.$$

Coherence (time-freq. domain)

$$L_{\text{Coh}}(\hat{\mathbf{Y}}, \mathbf{S}) = \frac{1}{F} \sum_f \frac{|\mathbb{E}[(\hat{\mathbf{Y}})_{fn}(\mathbf{S})_{fn}^*]|}{\sqrt{\mathbb{E}[|(\hat{\mathbf{Y}})_{fn}|^2]\mathbb{E}[|(\mathbf{S})_{fn}|^2]}}.$$

Experimental Validation

Baseline Methods

- AuxIVA-Laplace [3] and ILRMA [4]
- Single channel phase-sensitive masking [6]
- Mask-based generalized eigenvalue beamforming [7]

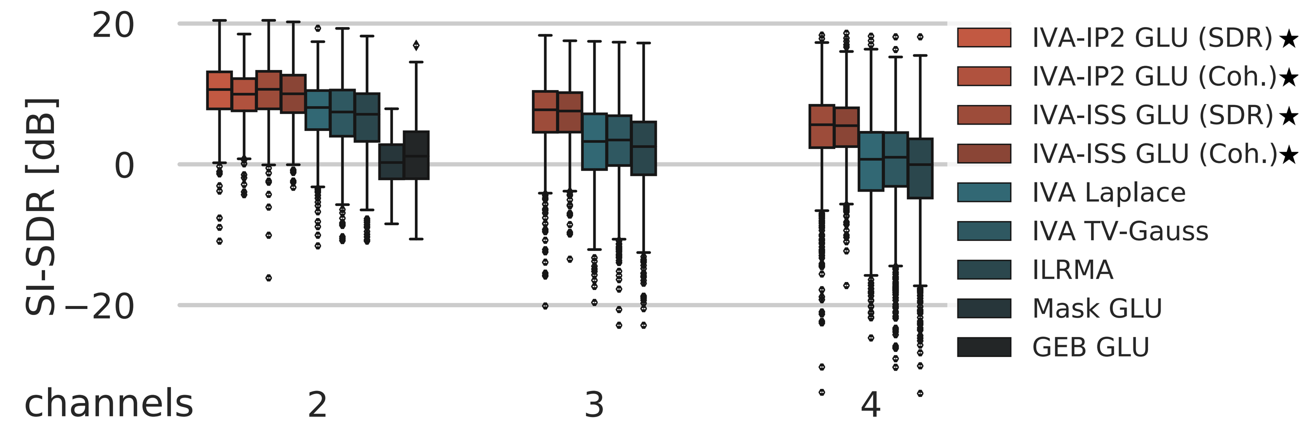
Dataset

- WSJ0
- Reverberation
- Noise from CHIME3

Training

- 2 channels
- 20 iterations of ISS

Results



Ch.	New	Algo.	Model	Loss	SDR (↑)	SIR (↑)	WER (↓)	CER (↓)
2		GEB	GLU	PSM	1.2	9.2	95.0%	60.5%
		IVA	Laplace	—	8.1	21.9	54.5%	31.6%
	⋆	IVA	GLU	SDR	**10.7**	24.1	33.5%	18.0%
	⋆	IVA	GLU	Coh.	10.0	24.9	**33.0%**	17.8%
3		IVA	Laplace	—	3.2	13.6	80.0%	50.3%
	⋆	IVA	GLU	SDR	**7.7**	20.1	47.1%	27.3%
	⋆	IVA	GLU	Coh.	7.6	21.1	**43.5%**	25.2%
4		IVA	Laplace	—	0.7	10.2	91.2%	58.6%
	⋆	IVA	GLU	SDR	**5.6**	17.4	58.3%	35.0%
	⋆	IVA	GLU	Coh.	5.5	18.4	**55.3%**	32.5%

Conclusion

- High performance and flexible
- Generalizes to unseen number of channels / BSS algorithms

References

[1] T. Kim et al., IEEE TASLP, vol. 15, no. 1, pp. 70–79, 2006.
[2] A. Hiroe, in Proc. ICA, vol. 3889, no. 2, pp. 601–608, 2006.
[3] N. Ono, in Proc. IEEE WASPAA, pp. 189–192, 2011.
[4] D. Kitamura et al., IEEE/ACM TASLP, vol. 24, no. 9, pp. 1626–1641, 2016.
[5] R. Scheibler and N. Ono, in Proc. IEEE ICASSP, pp. 236–240, 2020.
[6] H. Erdogan et al., in Proc. IEEE ICASSP, pp. 708–712, 2015.
[7] J. Heymann et al., in Proc. IEEE ICASSP, pp. 196–200, 2016.
[8] `https://github.com/fakufaku/auxiva-iss-dnn`