Multi-modal Blind Source Separation with Microphones and Blinkies

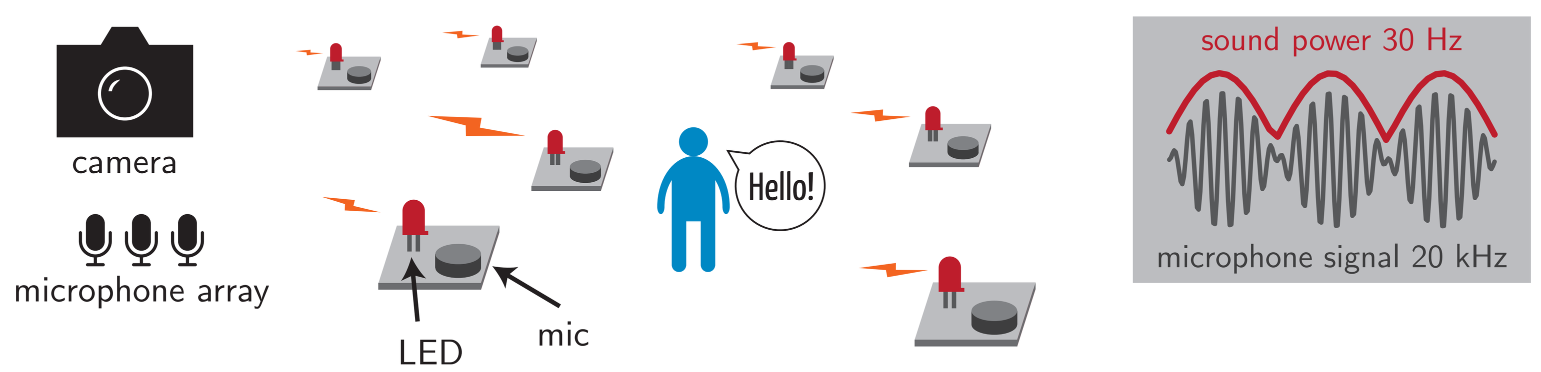
Robin Scheibler and Nobutaka Ono – Tokyo Metropolitan University, Tokyo, Japan

Microphones and Blinkies

Abstract – We propose an algorithm for blind source separation that jointly uses measurements from a microphone array and an ad hoc array of sound power sensors called **blinkies**. These sensors are low-rate and provide less information but circumvent some difficulties of microphones, e.g., deployment. A joint probabilistic model of both types of measurements and efficient update rules minimizing its negative log-likelihood are proposed. The proposed algorithm is validated via numerical experiments.

Sound power sensors

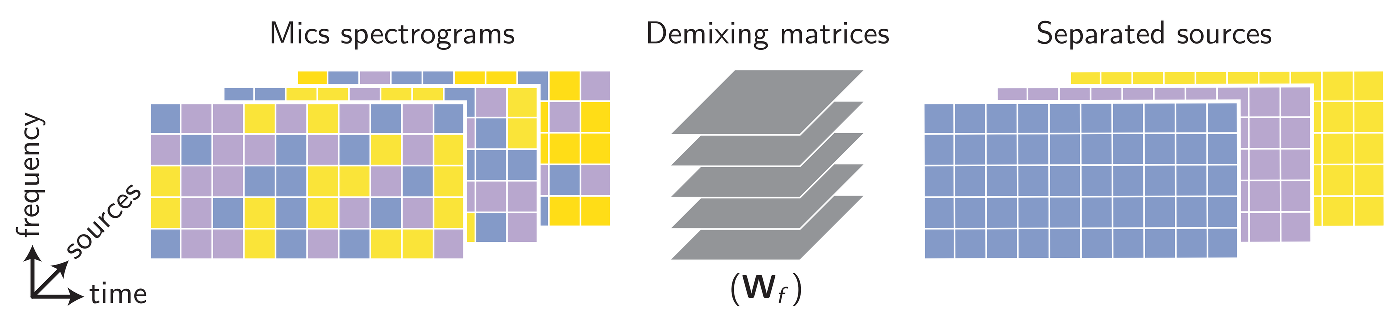
- Easy to spread over a large area
- Scale to a **very large** number of channels
- Low-rate (∼30 Hz)
- **Blinkies**: Transform sound power to light [1, 7]



Blind Source Separation

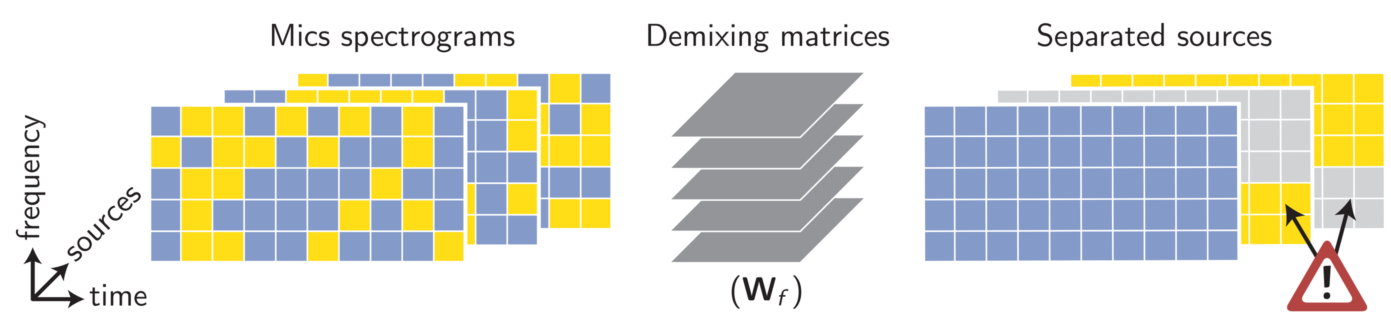
Goal: Find demixing matrices $\mathbf{W}_f$ such that outputs are independent

- Frequency domain
- Independent Vector Analysis [2, 3, 4]



- Ideally: More microphones ⇒ Better performance
- **However**, when channels > sources, mistakes may happen!

Example:



Joint Probabilistic Model

Our model is made up of 3 hypothesis

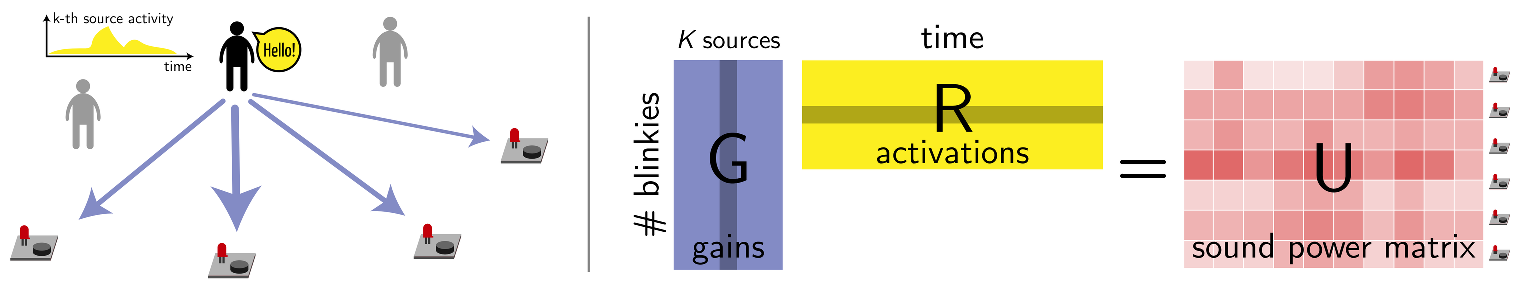
1. There are M independent source signals
2. The source spectra are complex Normal random vectors

$$\mathbf{y}_{kn} \sim \mathcal{CN}(0, r_{kn}\mathbf{I}),$$

   where r_{kn} is the time-varying activation
3. The blinky signals are norms of complex Normal random vectors

$$u_{bn} = \|\mathbf{x}_{bn}\|^2, \quad \mathbf{x}_{bn} \sim \mathcal{CN}\left(0, \mathbf{I} \sum_k g_{bk} r_{kn}\right),$$

   namely, the sound power matrix is rank-K and shares activations with K of the sources:



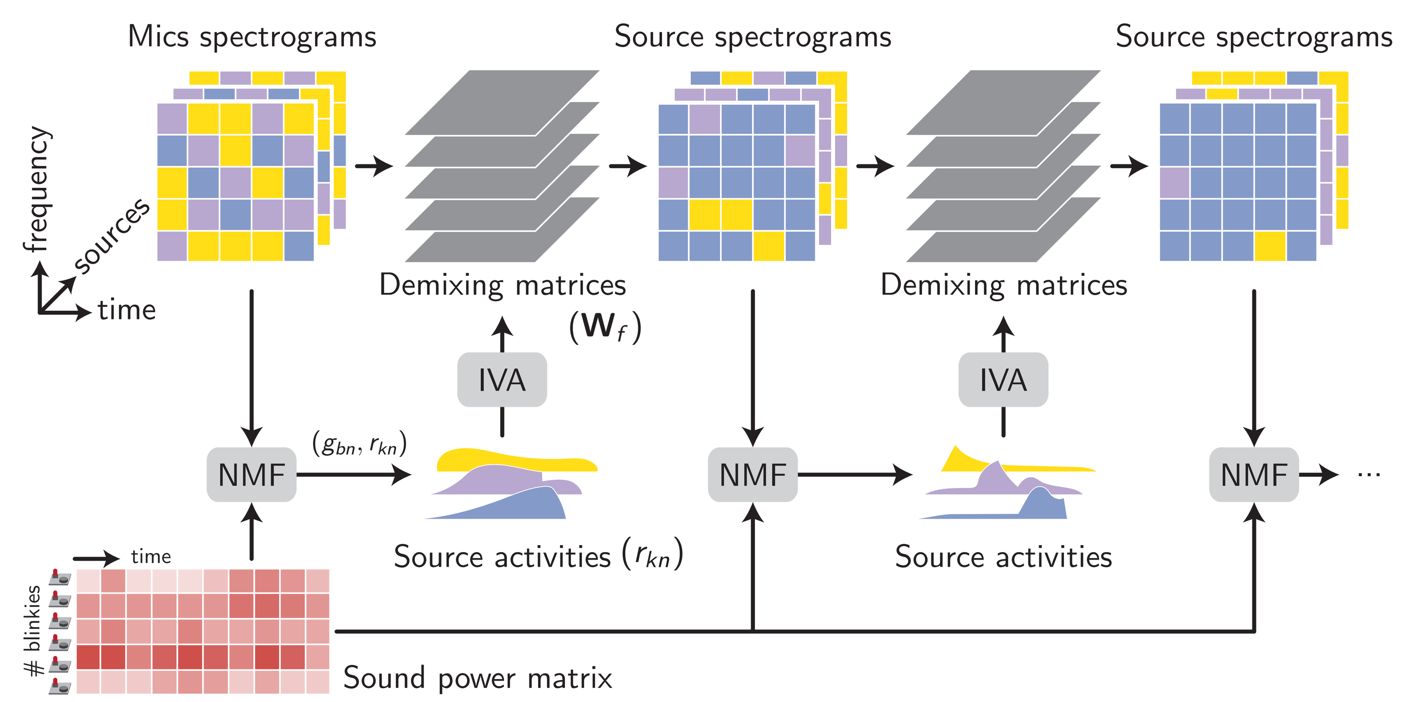
Algorithm

The **negative log-likelihood** (known/unknown quantities)

$$\ell = -2N \sum_f \log|\det \mathbf{W}_f| + \sum_{n=1}^{N}\sum_{k=1}^{M}\left(\frac{\sum_f |\mathbf{w}_{fk}^H \mathbf{x}_{fn}|^2}{r_{kn}} + F\log r_{kn}\right)$$
$$+ \sum_{n=1}^{N}\sum_{b=1}^{B}\left(F \log \sum_{k=1}^{K} g_{bk} r_{kn} + \frac{u_{bn}}{2\sum_{k=1}^{K} g_{bk} r_{kn}}\right)$$

can be minimized by **majorization-minimization**:

- $\mathbf{W}_f$: AuxIVA update [4]
- g_{bk}, r_{kn}: NMF-like multiplicative updates

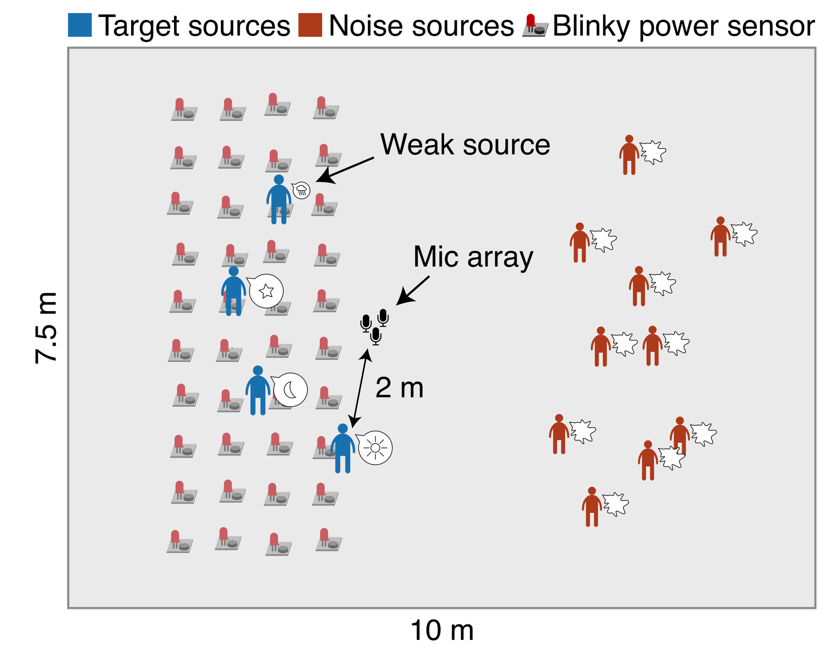


Performance Evaluation

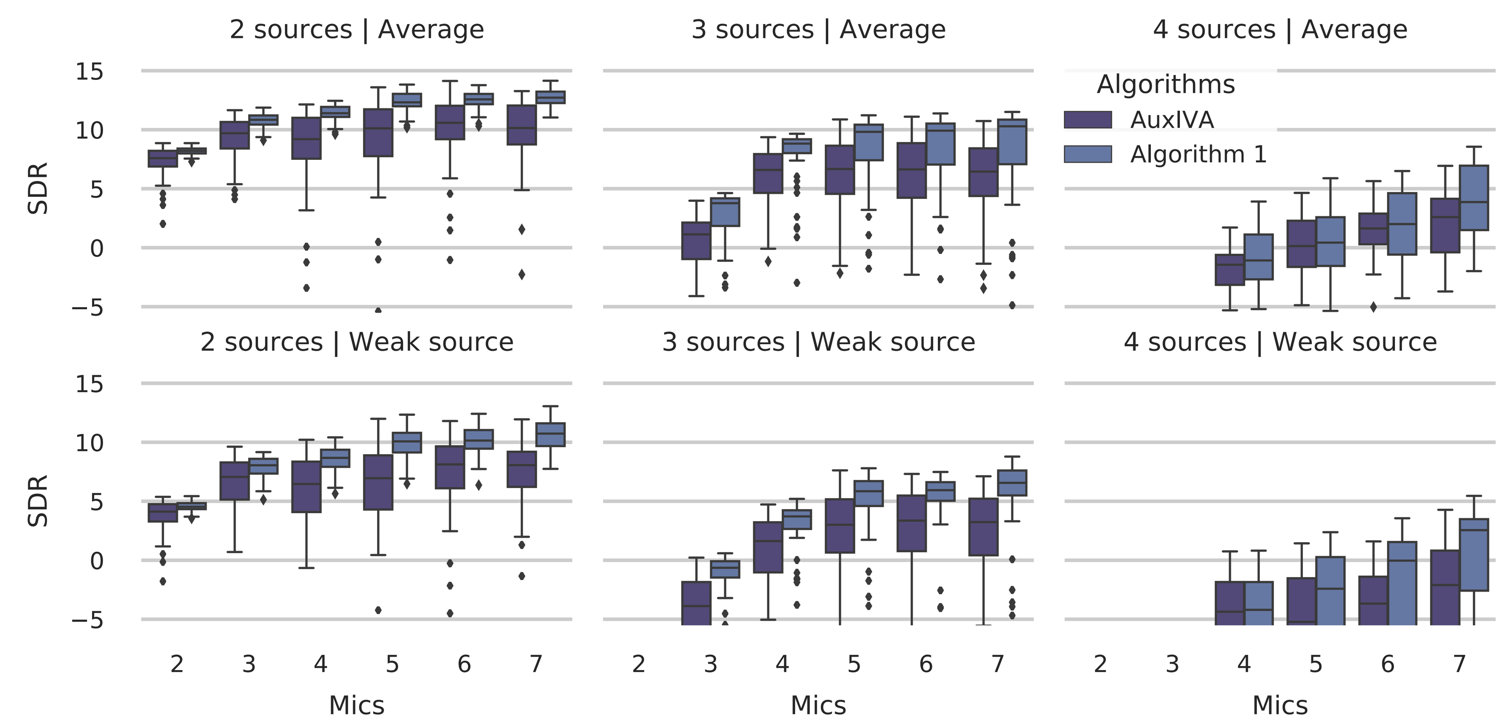
We compare the proposed algorithm to AuxIVA [4] via simulation

Simulation Setup

- `pyroomacoustics` simul. [5]
- fs: 16 kHz
- STFT: 4096, 1/2-o.l., Hann
- 2 to 7 mics, 40 blinkies
- 2 to 4 sources
- 1 weak source (1/4 power)
- 10 interferers
- SINR 10 dB



Results



In the spirit of reproducible research, the code is made available [6]

Highlights

- Microphones + Blinkies ⇒ Higher SDR/SIR
- Stable separation (smaller variance)
- Weak source recovered
- Up to 8 dB improvement over AuxIVA [4]

References

[1] Scheibler et al., in *Proc. APSIPA*, pp. 1899–1904, 2018.
[2] T. Kim et al., IEEE TASLP, vol. 15, no. 1, pp. 70–79, Dec. 2006.
[3] A. Hiroe, in *Proc. ICA*, vol. 3889, no. 2, pp. 601–608, 2006.
[4] N. Ono, in *Proc. IEEE WASPAA*, pp. 189–192, 2011.
[5] `https://github.com/LCAV/pyroomacoustics`
[6] `https://github.com/onolab-tmu/blinky-iva`
[7] `https://github.com/onolab-tmu/blinky`

This work was supported by a JSPS post-doctoral fellowship and grant-in-aid (17F17049), and the SECOM Science and Technology Foundation.