Blinkies: Sound-to-light conversion sensors and their application to speech enhancement and sound source localization

Robin Scheibler, Daiki Horiike, and Nobutaka Ono
Tokyo Metropolitan University, Tokyo, Japan

TOKYO METROPOLITAN UNIVERSITY
首都大学東京

Blinkies: Sound-to-light sensors

Abstract – We introduce the use of **blinkies** for acoustic sensing and audio processing applications. Blinkies are low-power sensors with a microphone and an LED that can be easily distributed over a large area. The power of the LED is modulated by the sound intensity and the signals from all devices can be captured by a regular video camera. The usefulness of such a system is demonstrated with two applications: First, beamforming informed by a high-quality voice activity signal obtained from a blinky, and second, sound source localization.

Seeing sound with light!

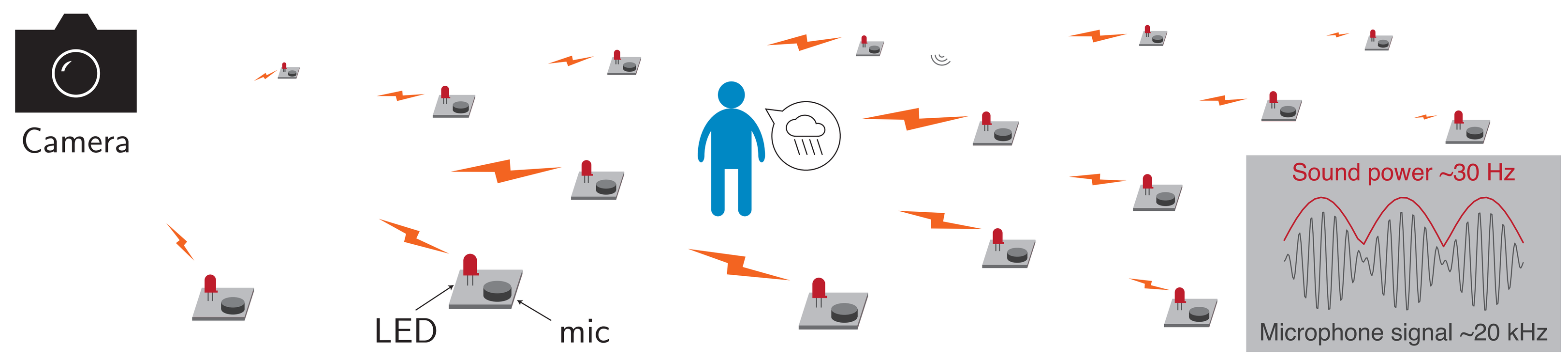
Acoustic sensing with blinkies

Blinkies are compact devices that

- transform sound power to light
- are easy to spread over a large area
- scale to a **very large** number of channels
- are synchronously sampled with a video camera

No Cables | No Wifi | Synchronized | Low-power | Programmable

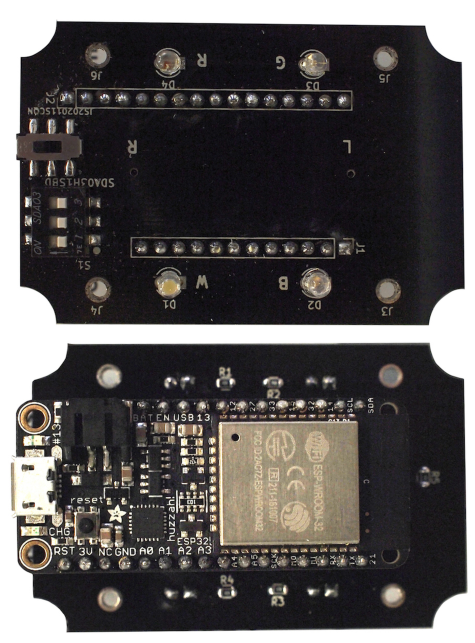
A typical deployment of blinkies around a sound source:



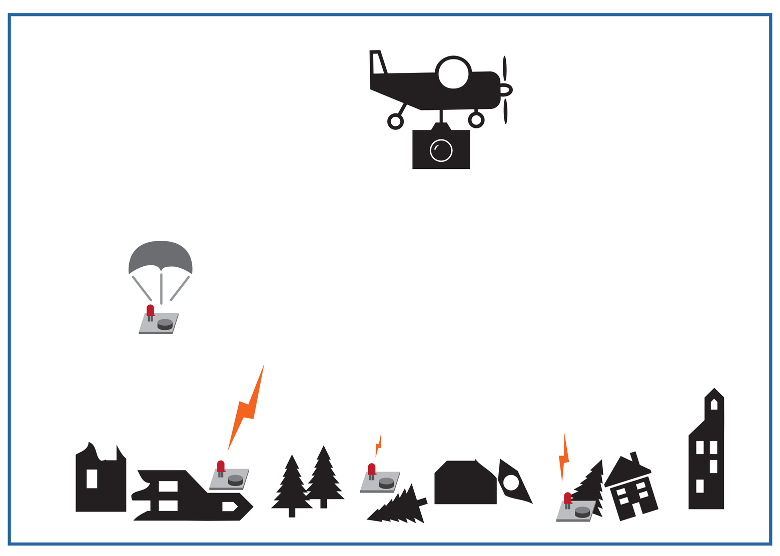
Blinkies: The hardware

Features of our **custom design** [1]

- ESP32, 240 MHz, 512 KB SRAM
- 2 microphones
- 4 LEDs
- Li-Po battery + USB charger
- C++, Arduino, Micropython
- WiFi + BT (reuse as wireless mic!)
- Active dev community (ESP32)



Applications scenarios



Examples:

- Search and rescue in difficult terrain
- Factory monitoring
- Smart home/meeting spaces

Application I: Blinky-informed Beamforming

Goal
Compute an effective beamformer using a microphone array and a blinky

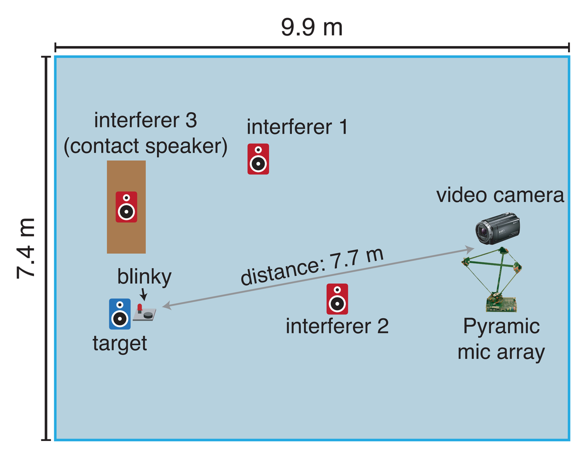
A single blinky is located close to target source and provides **high-quality voice activity detection**. Given the VAD, the beamformer with maximum signal-to-interference-and-noise ratio (SINR) is

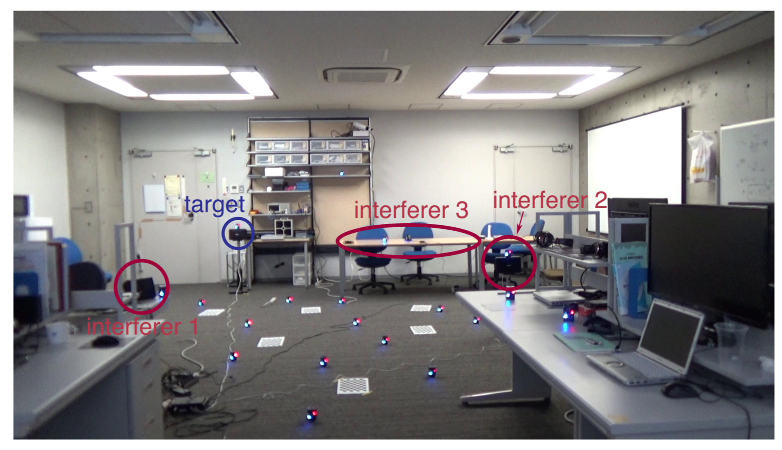
$$\mathbf{w}_{\text{Max-SINR}} = \arg\max_{\mathbf{w}} \frac{\mathbf{w}^H \mathbf{R}_x \mathbf{w}}{\mathbf{w}^H \mathbf{R}_{\text{noise}} \mathbf{w}} \qquad (1)$$

where $\mathbf{R}_x$ is the covariance matrix of the whole signal and

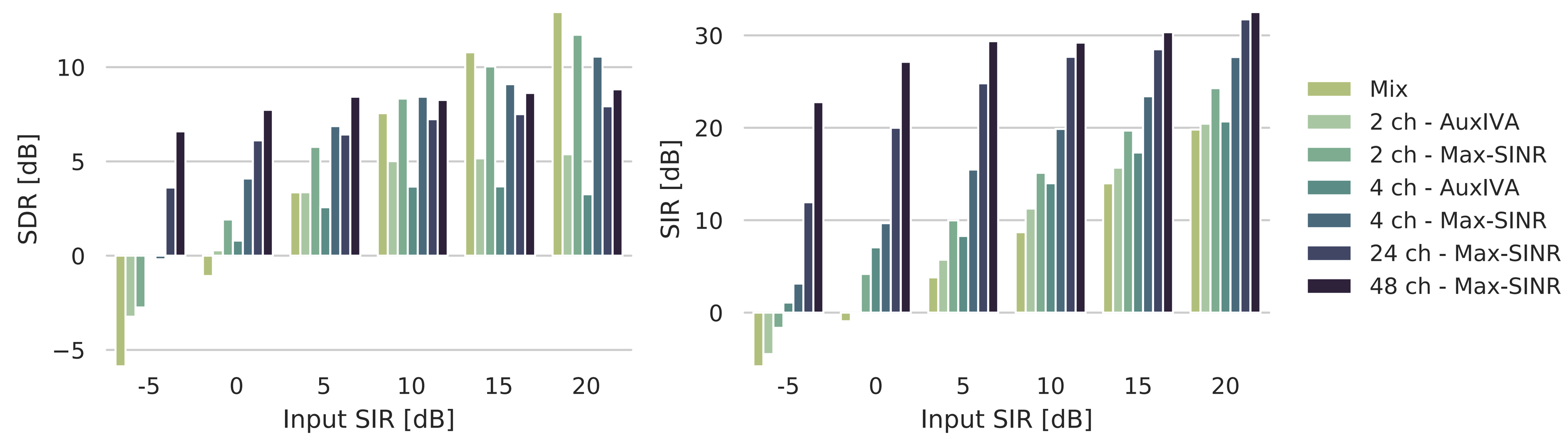
$$\mathbf{R}_{\text{noise}} = \sum_t (1 - \mathsf{VAD}(t)) \mathbf{x}_t \mathbf{x}_t^H. \qquad (2)$$

Experiment [2] (4 sound sources, 1 blinky + 48-ch mic array)





Results (BSS-AuxIVA vs Blinky-informed beamforming)



Conclusion
The light signal enables us to compute a powerful beamformer!

Application II: Energy-based Localization

Goal
Locate multiple sound sources using only a few blinkies

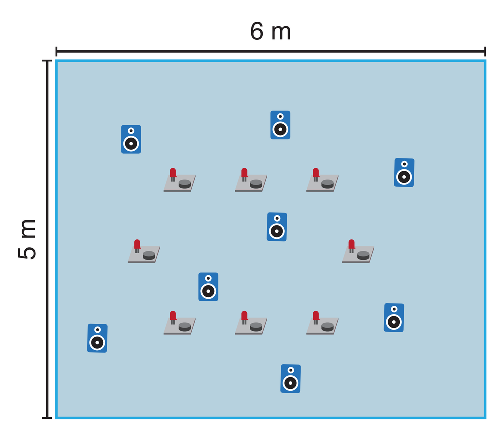
We adapt an existing algorithm [3]. The energy-decay from source to blinky is modeled as follows

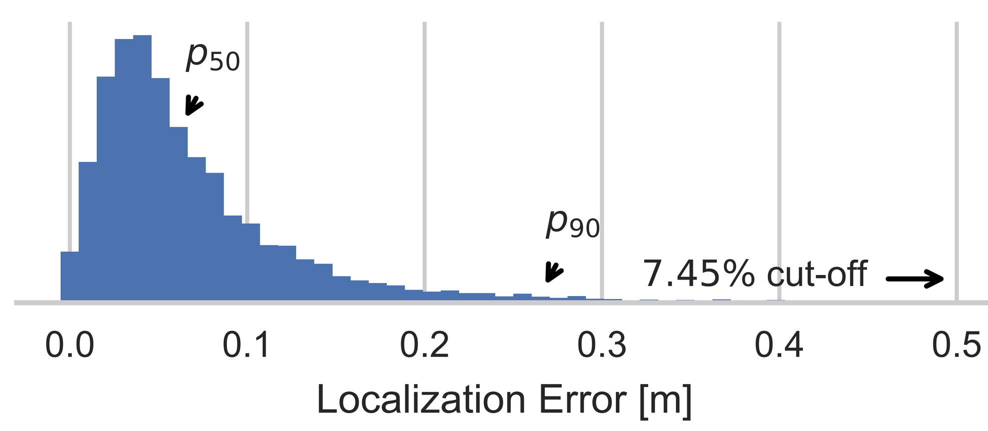
$$a_{mk} = \frac{g_m p_k}{\|\mathbf{r}_m - \mathbf{s}_k\|^{2\alpha}}, \qquad (3)$$

with g_m m-th blinky gain, p_k k-th source power, and the **distance** in the denominator. The propagation loss is characterized by α. Locations of the sources $\mathbf{s}_k$ are recovered via a **non-linear least-squares** problem

$$\min_{\alpha, \tilde{g}_m, \tilde{p}_k, \mathbf{s}_k} \sum_{m=1}^{M} \sum_{k=1}^{K} (\tilde{a}_{mk} - \tilde{g}_m + \alpha \log \|\mathbf{r}_m - \mathbf{s}_k\|^2 - \tilde{p}_k)^2. \qquad (4)$$

Numerical Experiments [2]





Conclusion
We successfully localize ∼ 90% of sources!

References

[1] https://github.com/onolab-tmu/blinky
[2] https://github.com/onolab-tmu/otohikari
[3] M. Chen, Z. Liu, L.-W. He, P. Chou, and Z. Zhang, "Energy-based position estimation of microphones and speakers for ad hoc microphone arrays" in *Proc. IEEE WASPAA*, 2007.

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