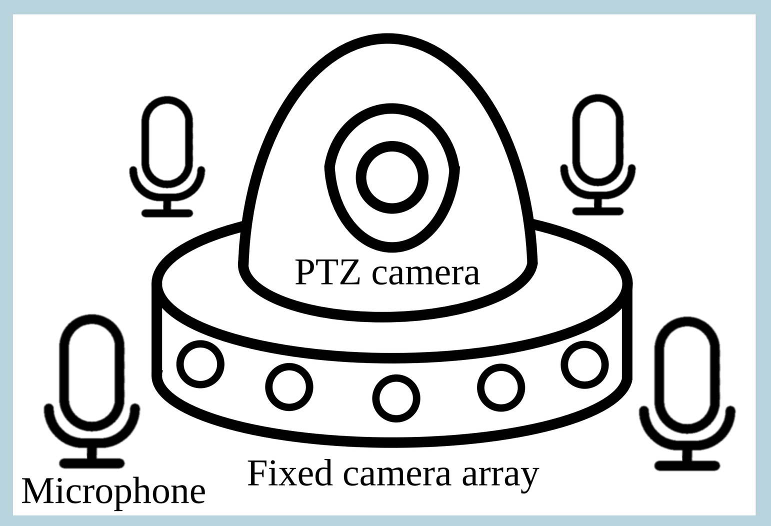
Detection and Localization of Drones and UAVs Using Sound and Vision

Erik Tegler[1], Max Modig[1], Per Skarin[2], Kalle Åström[1], Magnus Oskarsson[1], Gabrielle Flood[1]
[1]Lund University, [2]Saab Kockums

Contribution

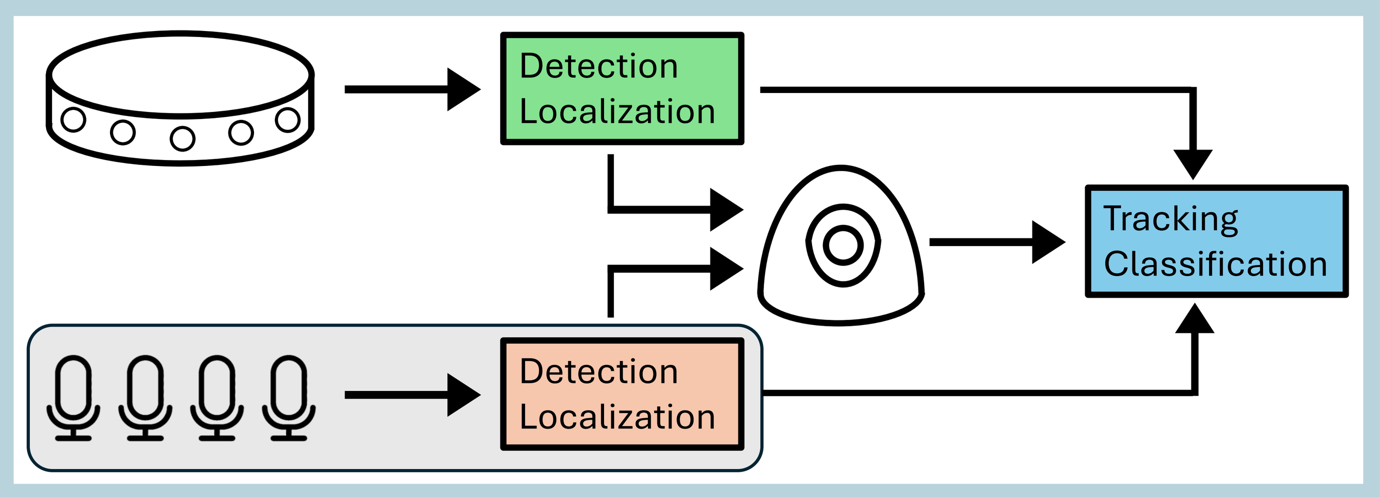
- A robust multi-channel algorithm for locating drones relative to an ad-hoc microphone array.
- A new dataset with multi-channel audio recordings, with ground truth position of drones.
- Evaluation both against ground truth and against localizations by human observers.

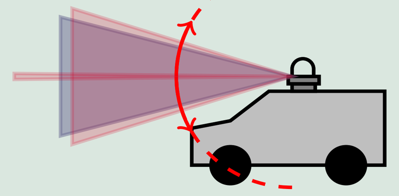


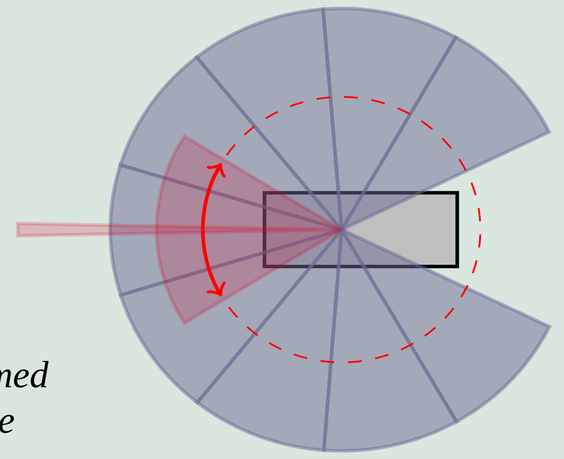
Background

- Detecting drones is important in many settings. Audio and video are both passive detection technologies, which can be advantageous.
- Our system works with a combination of spatially distributed microphones, an array of fixed cameras and a high rolution PTZ camera.
- Here, we focus on detection and coarse localization using audio.
- The localization can then be refined using the PTZ camera.

Setup







This work has been supported by Sweden's Innovation Agency VINNOVA and the Swedish Armed Forces within the programme "Civil-Militärt Innovationsprogram", project nr 2024-03191, the strategic research project ELLIIT and the Wallenberg AI, Autonomous Systems and Software Program (WASP), funded by the Knut and Alice Wallenberg (KAW) Foundation.

Method

- We have $i = 1\ldots12$ known microphone positions $\mathbf{r}_i$ and received sound signals $\mathbf{s}_i^{(k)}(t)$ coming from k unknown sound source positions $\mathbf{x}^{(k)}$.
- To find the DOA of the sound source positions, compare TDOAs estimated from sound signals $\mathbf{s}_i^{(k)}$

$$\tilde{R}^{(k)}_{\mathbf{s}_\kappa \mathbf{s}_i}(T) = \mathcal{F}^{-1}\left(\frac{\mathcal{S}^{(k)}_\kappa(f)\bar{\mathcal{S}}^{(k)}_i(f)}{|\mathcal{S}^{(k)}_\kappa(f)\bar{\mathcal{S}}^{(k)}_i(f)| + \varepsilon}\right)$$

$$\tilde{\tau}^{(k)}_{\kappa i} = \arg\max_T \tilde{R}^{(k)}_{\mathbf{s}_\kappa \mathbf{s}_i}(T)$$

to DOA candidates for different angles

$$\tau_{\kappa i}(\theta,\phi) = \frac{[\cos(\theta)\cos(\phi), \sin(\theta)\cos(\phi), \sin(\phi)](\mathbf{r}_i - \mathbf{r}_\kappa)}{c}$$

- Two models: 1) only azimuth angle, 2) azimuth and elevation angle.
- Done for every sound segment, leading to a track of the drone movement.

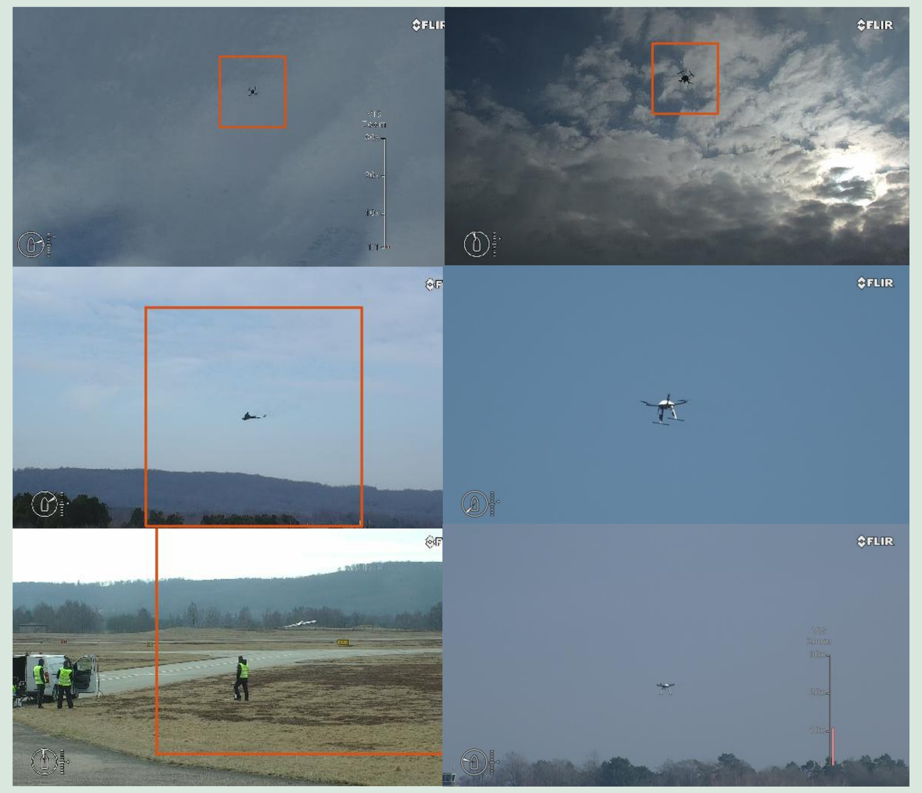
Algorithm 1 Find Direction of Arrival (DOA)

Input: Microphones $\mathbf{r}_1, \ldots, \mathbf{r}_N$, sounds $\mathbf{s}_1, \ldots, \mathbf{s}_N$
Output: DOA estimates $\tilde{\theta}^{(k)}$ (or $\tilde{\theta}^{(k)}$ and $\tilde{\phi}^{(k)}$)

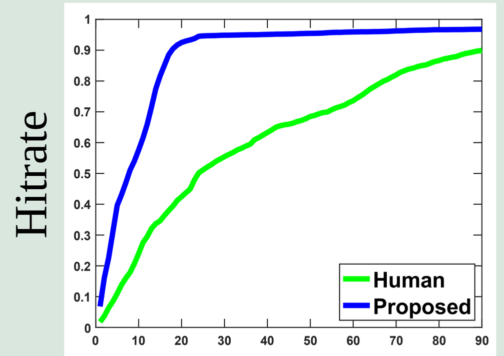
```
1: Divide sound signals s_i in 0.1 s long segments s_i^(k)
2: for each k do
3:     Compute the GCC-PHAT R~^(k)_{s_κ s_i}(T)
4:     Find measured TDOAs τ~^(k)_{κi}
5: end for
6: Update τ~^(k)_{κi} to be a moving median of τ~_{κi}
7: Grid space in azimuth θ for Model 1 (or θ and elevation φ for Model 2)
8: for each angle θ (or angle combination (θ,φ)) do
9:     Compute angle dependent TDOA τ_{κi}(θ) (or τ_{κi}(θ,φ))
10: end for
11: for each k do
12:    Estimate DOA by comparing τ~^(k)_{κi} and τ_{κi}(θ) (or τ_{κi}(θ,φ))
13: end for
```

Results

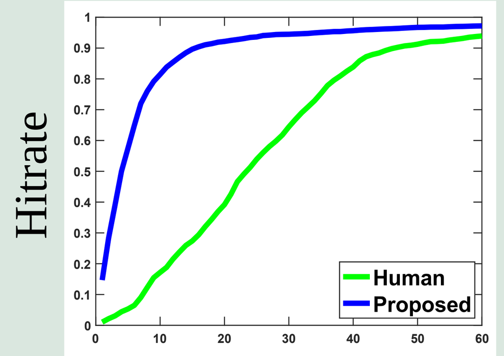
The red rectangles show estimated angles from our system



Comparison between our system and a human listener

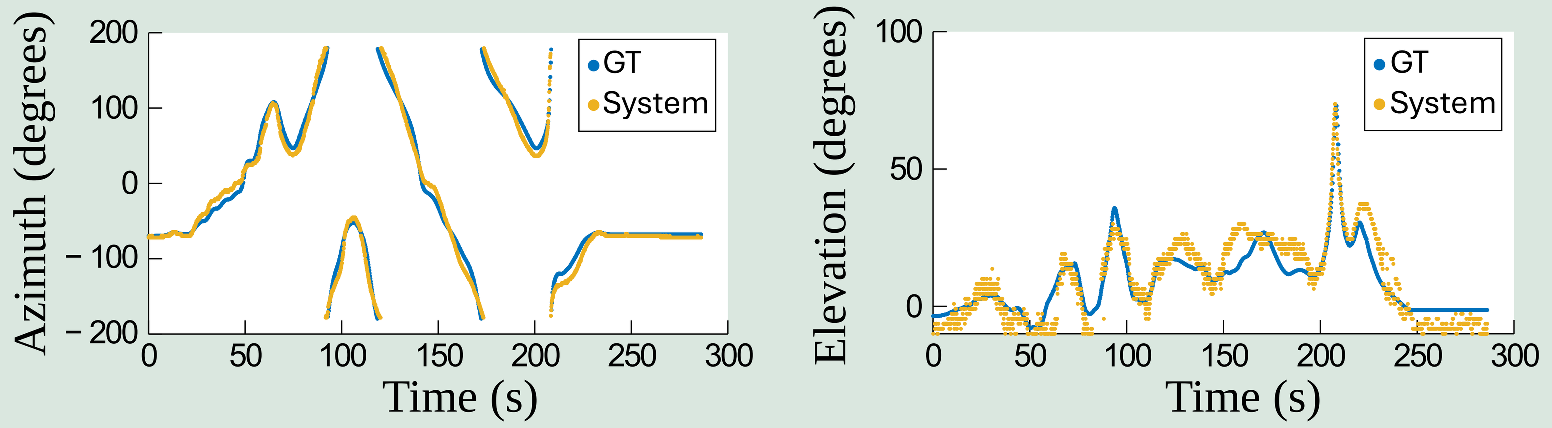


Azimuth error (degrees)



Elevation error (degrees)

Comparison between our system and ground truth on one test sequence



Conclusion

- A drone detection and tracking system based on multimodal sensor input.
- Azimuth and elevation angles can be estimated reasonably well.
- Works up to several hundred meters distance.
- A new dataset with sound and GT positions released.
- Future work: integrate coarse localization from sound with the camera array and then use the PTZ camera for fine localization.

Paper
Code
Dataset