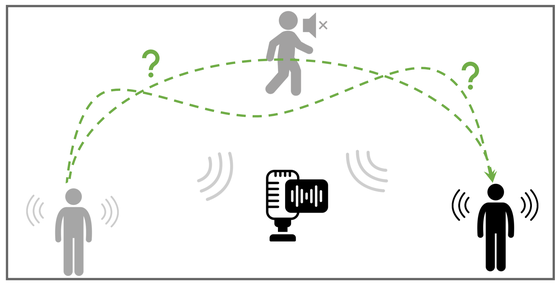
SOUND SOURCE LOCALISATION AND TRACKING IN COMPLEX ACOUSTIC SCENES

Taous Iatariene[12], Alexandre Guérin[1], Romain Serizel[2]
[1]Orange Innovation, 2 Av. de Belle Fontaine, 35512 Cesson-Sévigné, France
[2]Université de Lorraine, CNRS, Inria, Loria 54000 Nancy, France

Goal

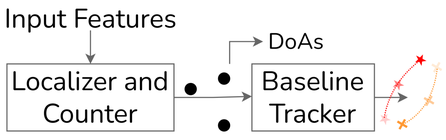
Reconstruct the trajectories of the sound sources present in an acoustic scene.

Focus on **speech sources** and **indoor acoustic scenes.**



Recording made with a **spherical microphone antenna.**

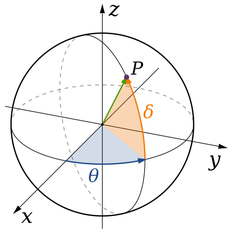
Overview



1. A **Localizer and Counter** [1] provides estimations of the **Directions of Arrival** (DoAs) and **Number of active Sources** (NoS) **for each timestep**.

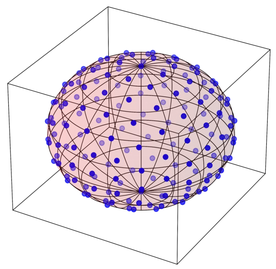
2. A **Baseline Tracker** [2] links the DoAs through time.

DoA : spherical coordinates (θ, δ)

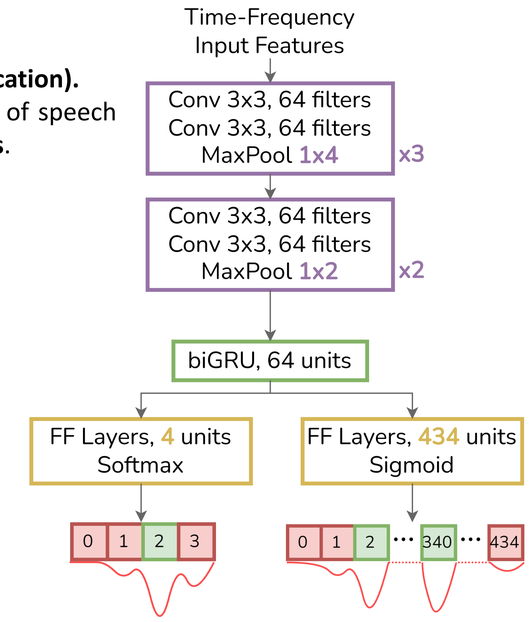


Localization and Counting

- **Multitask neural network.**
- Supervised learning **(classification).**
- Training dataset : 60 hours of speech in **simulated acoustic scenes**.

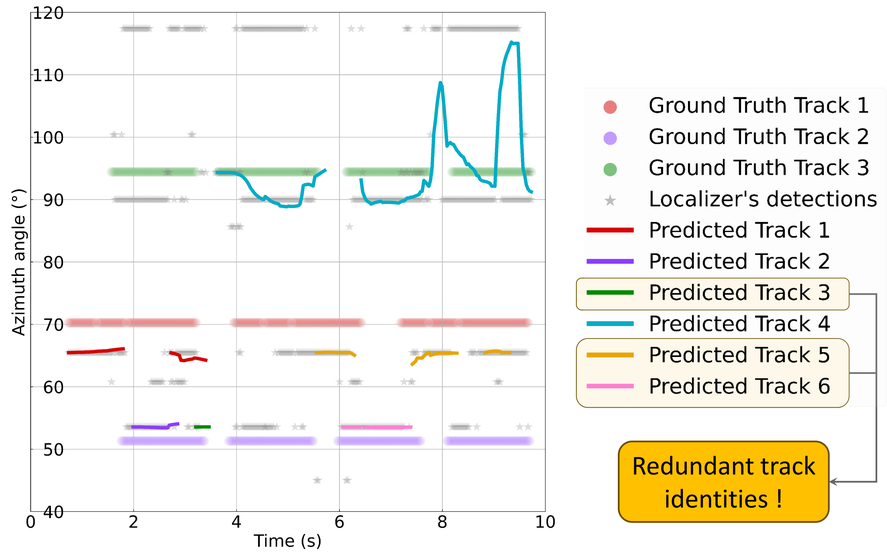


Discretization of the DoA space.



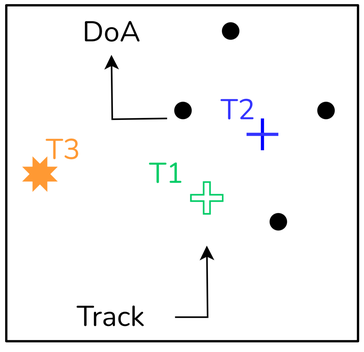
Results

An example of the predicted tracks that the baseline tracker can make given the multitask neural network DoAs.



Redundant track identities !

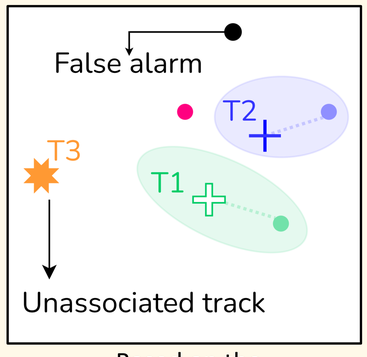
Baseline Tracker



The baseline tracker **receives DoAs as detections** and **updates its track states** accordingly.

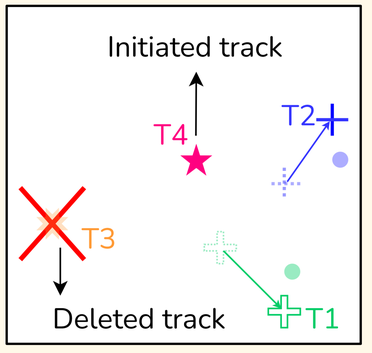
Two main steps

1) Detection-to-Track assignment



Based on the **minimization of the angular distances.**

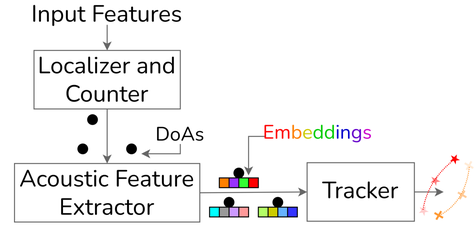
2) Track Update



Track activity management and **track position update** with Bayesian filtering.

Perspectives

Give more than DoA to "feed" the tracker : an **acoustic feature extractor** to generate acoustic detections.



References

[1] P. -A. Grumiaux, S. Kitić, L. Girin and A. Guérin, "Improved feature extraction for CRNN-based multiple sound source localization," 2021 29th European Signal Processing Conference (EUSIPCO), Dublin, Ireland, 2021, pp. 231-235, doi:10.23919/EUSIPCO54536.2021.9616124.
[2] Kitić, S., & Guérin, A. (2018). TRAMP: Tracking by a Real-time AMbisonic-based Particle filter. arXiv preprint arXiv:1810.04080