Analyzing the effect of equal-angle spatial discretization on Sound event localization and detection

Saksham Singh Kushwaha[1,2], Iran R. Roman[2], Juan Pablo Bello[2,3]

[1]Courant Institute of Mathematical Sciences, New York University, NY, USA

[2]Music and Audio Research Lab, New York University, NY, USA

[3]Center for Urban Science and Progress, New York University, NY, USA

NYU

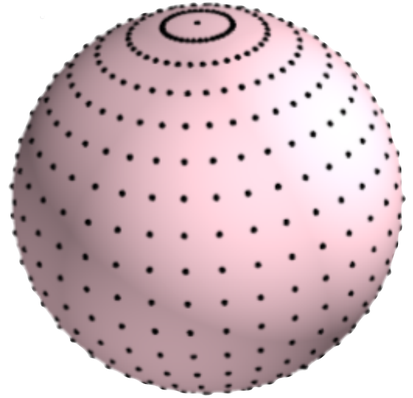
MARL

NYU COURANT INSTITUTE OF MATHEMATICAL SCIENCES

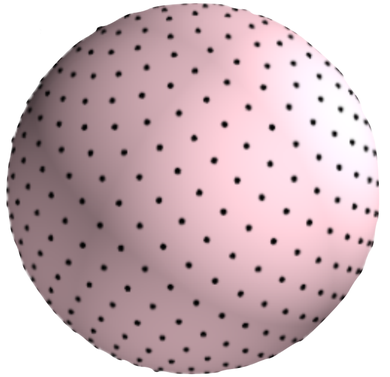
Sound event localization and detection(SELD)

- Detecting spatial, temporal and class information of sound events
- **Discretizing spatial targets**
  - Equal-angle targets: $\theta \in [-90°, 90°], \phi \in [-180°, 180°)$
    - **Problem**: Non-uniform spacing between targets along elevation
  - Alternative:
    - Fibonacci uniform spherical grid
- ADPIT LOSS**:**
  - MSE to minimise detection and localization error(LE)
  - **Hypothesis**: With equal-angle, $LE_{equator} > LE_{poles}$

Spatial discretization of sound event targets



Equal-angle

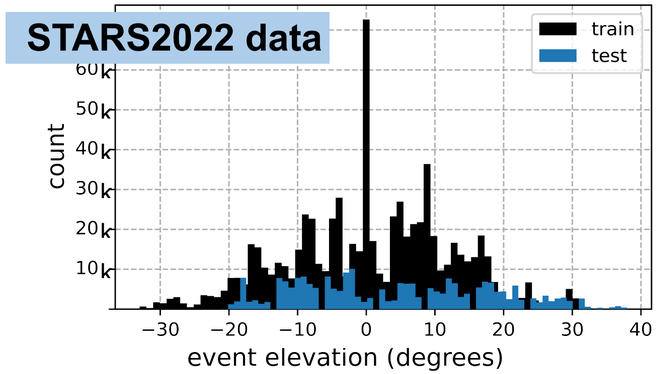


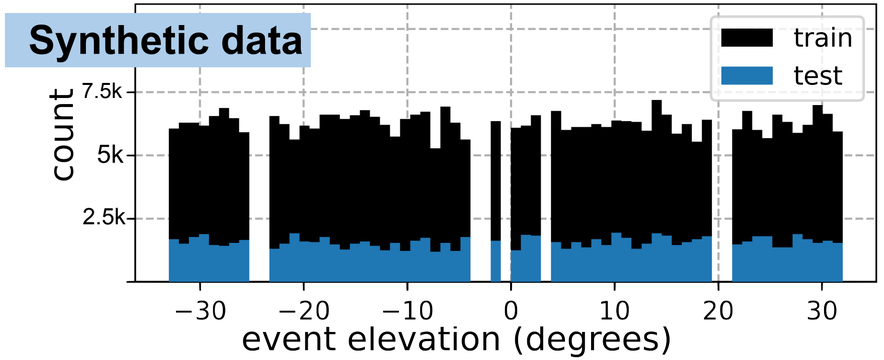
Fibonacci

Question: Does equal-angle discretization affect model localization performance along the elevation axis?

Equal-angle vs uniform targets

- **Datasets**
  - STARSS22: more events at equator
  - **Our Synthetic data**: uniform along elevation axis

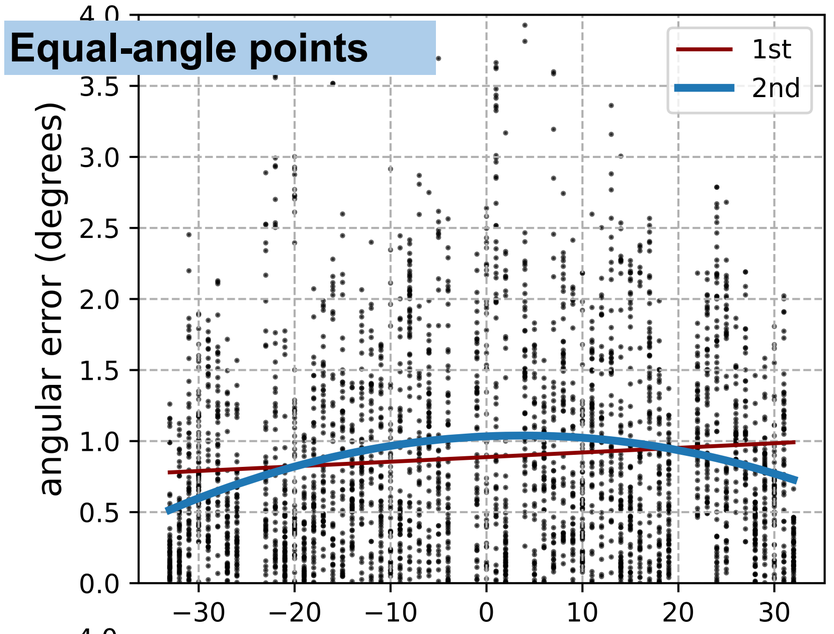


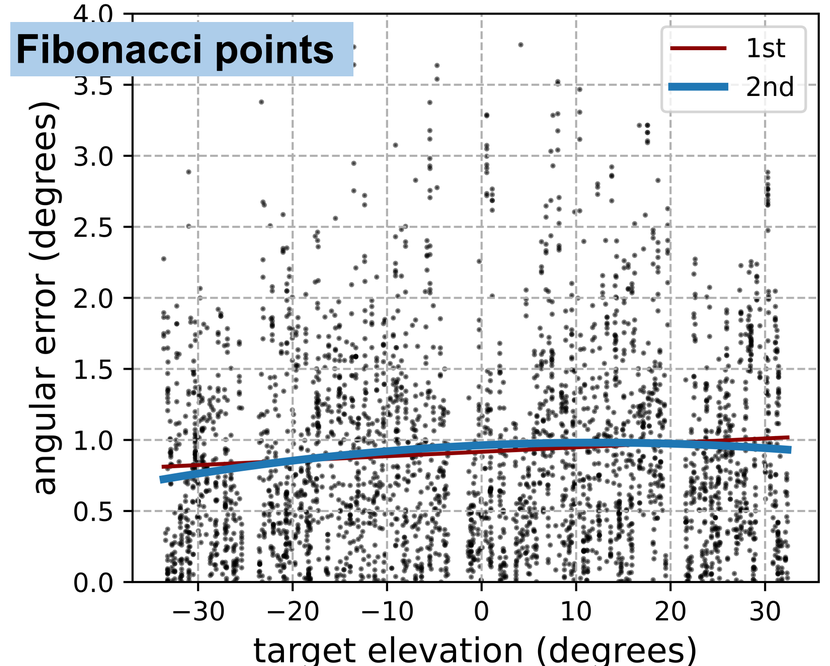


- **Experiment:** DCASE SELD baseline model trained with
  - Equal-angle targets
  - Fibonacci targets

- **Result:** Fibonacci has more uniform LE along elevation axis

Next: Is resampling targets the only way to mitigate bias?

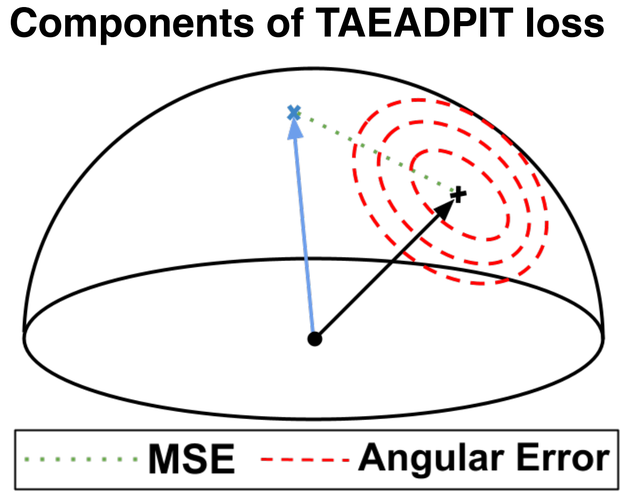


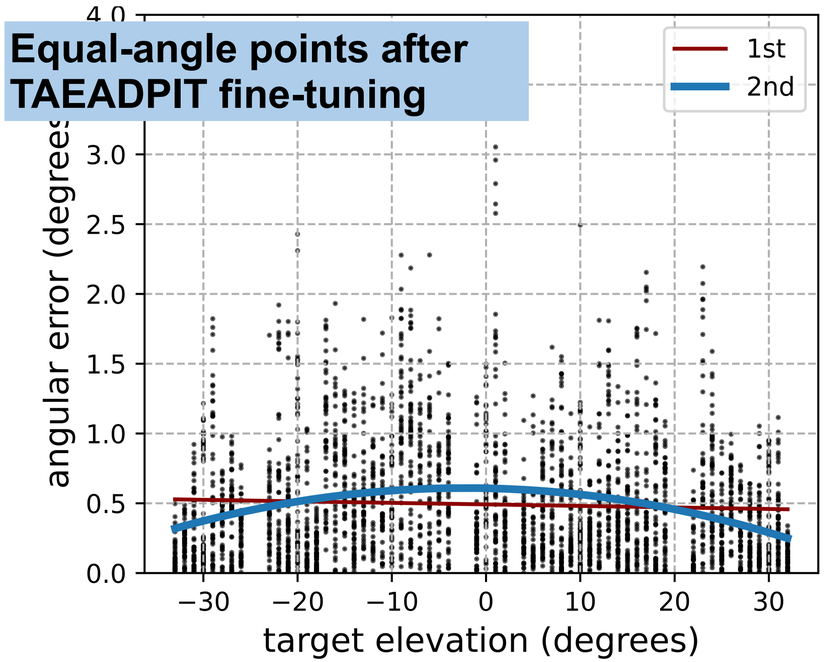


Thresholded Angular error ADPIT(TAEADPIT) loss

- Penalizes angular LE across azimuth and elevation
- **Experiment 1:** fine-tuning with the TAEDPIT loss
  Result: reduce equal-angle impact
- **Experiment 2:** DCASE22 SELD baseline model performance with TAEADPIT loss
  Result: improves localization metrics

Loss	ER_{20^o}	F_{20^o}	LE_{CD}	LR_{CD}	SELD
ADPIT-base	0.69	0.24	30.43	0.43	0.55
TAEDPIT-tune	0.71	0.23	28.86	0.47	0.54
TAEDPIT	0.71	0.20	26.42	0.41	0.56
ADPIT-Fib	0.68	0.22	26.11	0.46	0.54





Conclusion and Future Work

- Equal-angle results in non-uniform Localization error along the elevation axis

- To mitigate this bias, we:
  - Resampled with Fibonacci targets
  - Finetuned with TAEADPIT Loss

- In future, we would like to extend our study to
  - Other audio formats (eg. FOA, stereo)
  - Other spatial discretizations