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

INTRODUCTION

Sound Event Localization and Detection (SELD)

- Detection of sound event occurrences with their spatial, temporal and class information.

Discretizing spatial targets for SELD

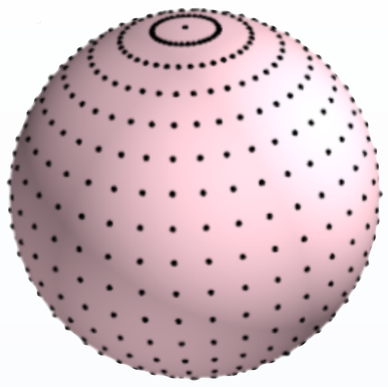
- Datasets typically use equal-angle discretization:
  $\theta = [-90, 90] \in \mathbb{Z}$ and
  $\phi = [-180, 180) \in \mathbb{Z}$ (in degrees).
- **Problem:** this results in non-uniform spherical grid of points along the elevation axis
- An alternate discretization is Fibonacci [3], which results in a uniform spherical grid.

$$r_n = [cos(\phi_n)sin(\theta_n), sin(\phi_n)sin(\theta_n), cos(\theta_n)] \quad n = 1, ..., N$$
$$where,\ \phi_n = 2\pi n(1 - \frac{2}{1+\sqrt{5}}) \ \&\ \theta_n = \arccos(1 - \frac{2n}{N})$$

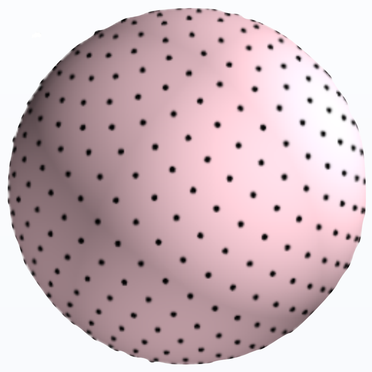
SELD Modeling

- DCASE-22 SELD baseline model is optimized with the ADPIT loss [2], which jointly minimizes detection and localization error using MSE.
- **Hypothesis:** With equal-angle data, ADPIT will give larger localization errors at the equator.

Spatial discretization of sound event targets



Equal-angle



Fibonacci

Research questions

- Does equal-angle discretization affect model localization performance along the elevation axis?
- How can the model's localization performance be robust given the non-uniform grids used to discretize sound events in common SELD datasets?

ANALYSIS OF EQUAL-ANGLE VS UNIFORM DISCRETIZATION

Model

- SELD DCASE 2022 baseline model.

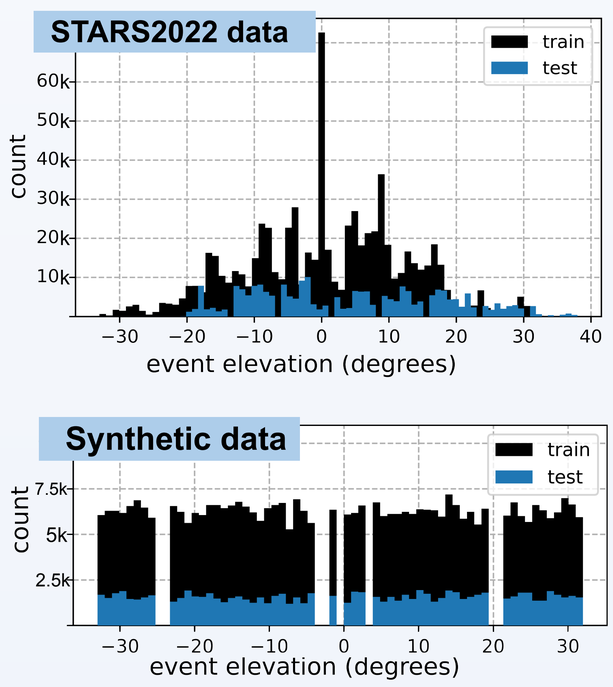
Data

- The DCASE STARSS2022[1] dataset features equal-angle targets (non-uniform grid along the elevation axis).
- In this dataset, more events occur around the equator.
- To control density of events, we generated uniformly-spatialized synthetic data along the elevation axis.

Experiments

- Model trained with targets that were discretized with either:
  - Equal-angle
  - Fibonacci

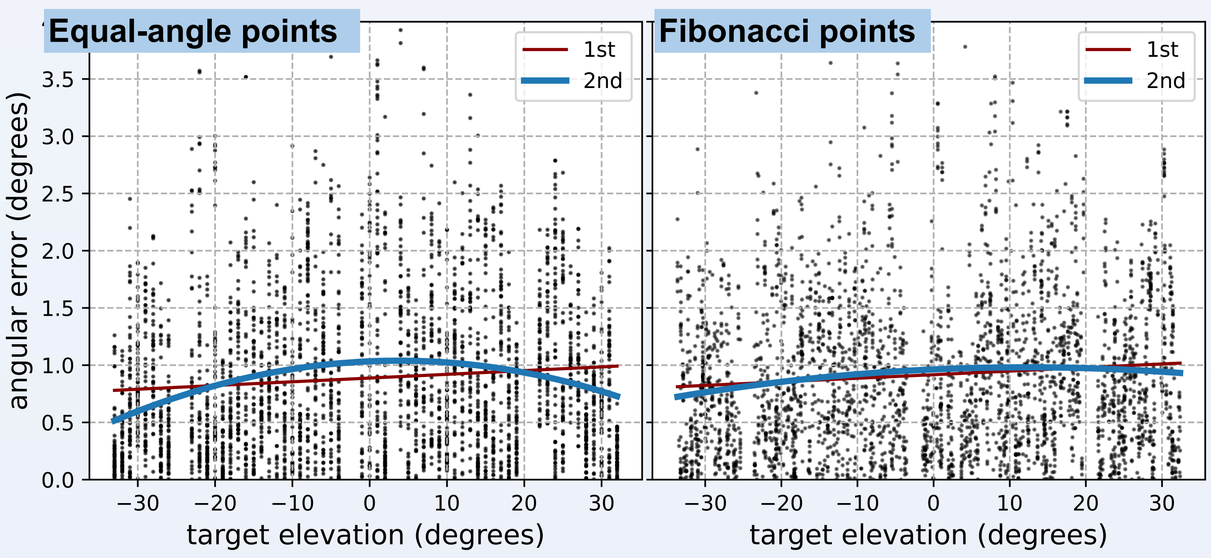
Histogram of target spatializations along the elevation axis



Results

- Equal-angle targets results in larger localization error around equator than pole
- Fibonacci targets show uniform localization error along the elevation axis

SELD angular localization error with:



Thresholded Angular error ADPIT (TAEADPIT) loss

The TAEADPIT Loss:

- Resampling with uniform discretization is not always feasible.
- TAEADPIT penalizes angular localization error, across azimuth and elevation.

$$l^{ALE}_{\alpha,ct} = \max(\boldsymbol{a}_{\alpha,nct} ALE_{\alpha,nct}, H)$$
$$ALE_{\alpha,nct} = \angle(p(\boldsymbol{P}_{\alpha,nct}), p(\hat{\boldsymbol{P}})_{nct})$$
$$L^{TAEADPIT} = \frac{1}{CT} \sum_c^C \sum_t^T \min_{\alpha \in \text{Perm}(ct)} l^{ACCDOATAE}_{\alpha,ct}$$
$$l^{ACCDOATAE}_{\alpha,ct} = \frac{1}{N} \sum_n^N MSE(\boldsymbol{P}^*_{\alpha,nct}, \hat{\boldsymbol{P}}_{nct}) + \beta(l^{ALE}_{\alpha,ct} - H)$$

Experiment 1

Research question

- Does fine-tuning with the TAEDPIT loss reduce equal-angle impact?

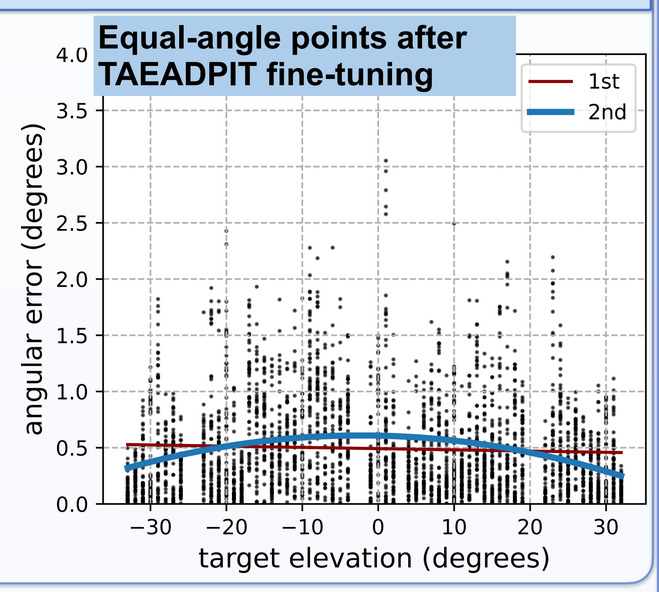
Methodology

- Use TAEADPIT to fine-tune the model previously trained with our synthetic equal-angle data.

Results

- Finetuning makes the localization error along the elevation axis more uniform.

Equal-angle points after TAEADPIT fine-tuning



Experiment 2

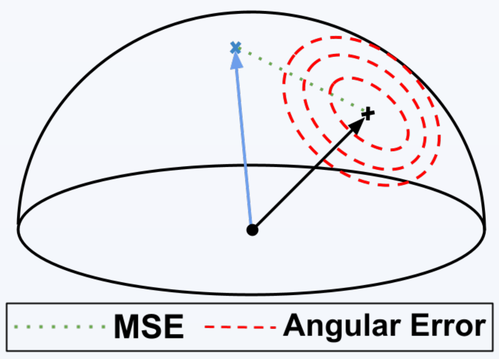
Research question

- Does using the TAEADPIT loss improve metrics on the DCASE 2022 SELD baseline model?

Methodology

- DCASE22 SELD baseline model trained on STARS2022 (and supplementary synthetic) dataset with:
  - original ADPIT,
  - TAEADPIT fine-tune and from scratch
  - ADPIT with fibonacci targets.

Components of TAEADPIT loss



Results

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

SELD model performance with different discretization and losses.
Training with TAEADPIT reduces localization error, similar to Fibonacci

CONCLUSIONS AND FUTURE WORK

- The non-uniform localization error along the elevation axis can be mitigated by
  - Resampling with uniform spatial discretized targets
  - Finetuning with TAEADPIT
- In the future we would like to extend our study to other:
  - audio formats
  - target discretizations.

REFERENCES

1. A. Politis, K. Shimada, P. Sudarsanam, S. Adavanne, D. Krause, Y. Koyama, N. Takahashi, S. Takahashi, Y. Mit- sufuji, and T. Virtanen, "Starss22: A dataset of spatial record- ings of real scenes with spatiotemporal annotations of sound events," *arXiv preprint arXiv:2206.01948*, 2022.
2. K. Shimada, Y. Koyama, S. Takahashi, N. Takahashi, E. Tsunoo, and Y. Mitsufuji, "Multi-accdoa: Localizing and detecting overlapping sounds from the same class with aux- iliary duplicating permutation invariant training," in *IEEE In- ternational Conference on Acoustics, Speech and Signal Pro- cessing (ICASSP)*, Singapore, Singapore, May 2022.
3. B. Keinert, M. Innmann, M. Sa¨nger, and M. Stamminger, "Spherical fibonacci mapping," *ACM Transactions on Graph- ics (TOG)*, vol. 34, no. 6, pp. 1–7, 2015.