Tampere University

A Dataset of Dynamic Reverberant Sound Scenes with Directional Interferers for Sound Event Localization and Detection

Archontis Politis[1], Sharath Adavanne[1], Daniel Krause[1], Antoine Deleforge[2], Prerak Srivastava[2], Tuomas Virtanen[1]

[1]Audio Research Group, Tampere University, Finland, [2]Universite de Lorraine, CNRS, Inria, Nancy, France

Sound Event Localization and Detection

Sound Event Localization and Detection (**SELD**) attempts to simultaneously detect, classify, and localize sound events, aiming at a more holistic spatiotemporal analysis of the sound scene than sound event detection or sound source localization separately.

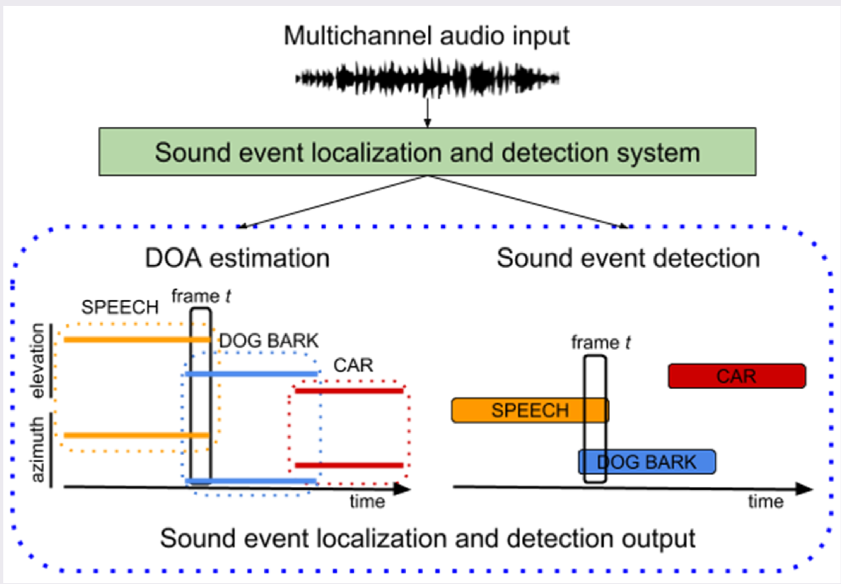

Figure 1: Sound Event Detection and Localization system output.

SELD datasets

	DCASE2019	DCASE2020	DCASE2021
# rooms	5 rooms	13 rooms	13 rooms
# spatial RIRs/positions	504 discrete positions	~200 spatial trajectories (continuously captured SRIRs)	~200 spatial trajectories (continuously captured SRIRs)
Source-to-receiver distances	1m-2m	1m-5m	1m-5m
Spatial ambient noise	30dB SNR	6-30dB SNR	0-30dB SNR
Moving sources	No	Yes	Yes
Non-target interfering events	No	No	Yes
# polyphony/overlapping events	≤2	≤2	≤3 (+ ≤1 interf. event)
% same-class overlapping events	low	low	high
# target classes	11	14	12
# event samples	220	~700	~500 (target events) ~400 (interferer events)

Figure 2: Comparison of SELD datasets created for DCASE Challenges.

Data based on:

- measured spatial room impulse responses from multiple rooms
- recorded spatial ambient noise from the same rooms
- two spatial formats derived from a spherical microphone array

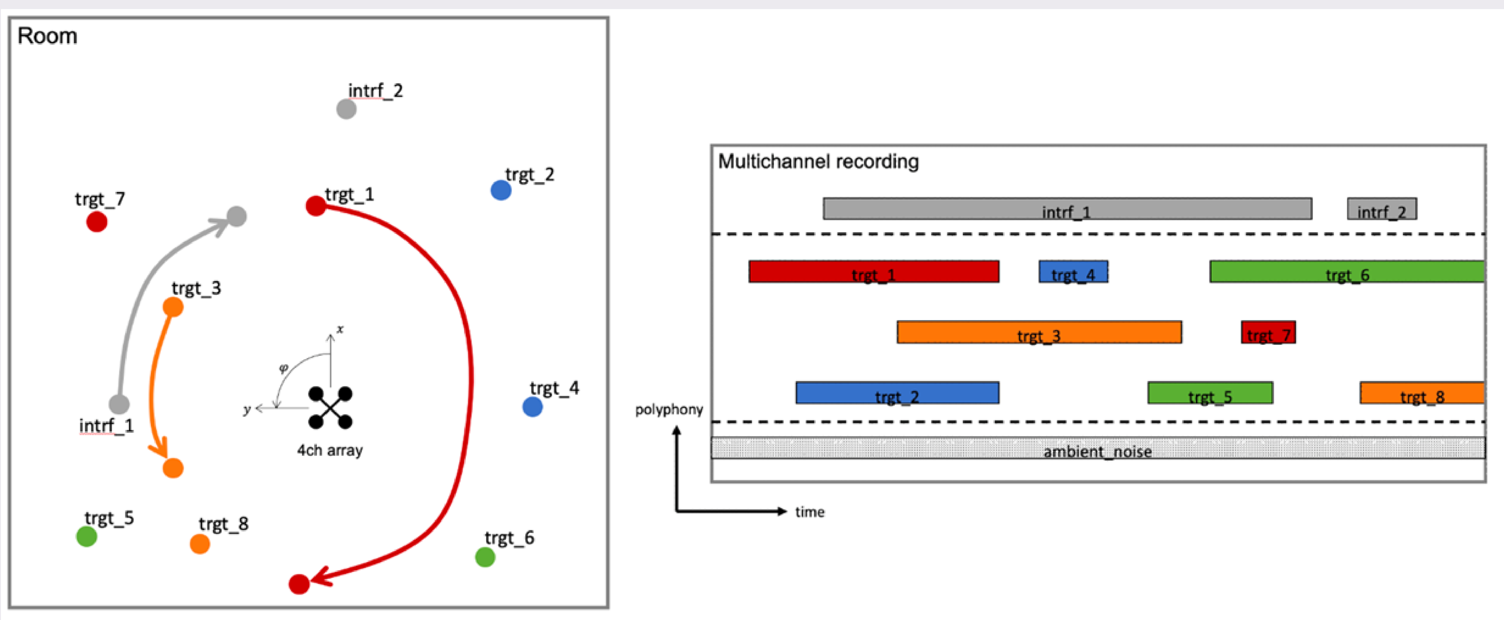

Figure 3: Exemplary depiction of an emulated recording in the dataset.

DCASE2021 baseline system

The new baseline system uses the single ACCDOA output representation, instead of the earlier multi-task output.

	FOA		MIC	
	LE_{CD}	$F1_{20°}$	LE_{CD}	$F1_{20°}$
DCASE2020 dataset				
multi-task	24.3°	44.4%	25.4°	40.4%
ACCDOA	17.9°	51.9%	19.3°	48.5%
DCASE2021 dataset				
multi-task	32.1°	24.7%	41.6°	19.1%
ACCDOA	24.5°	30.7%	30.6°	23.4%

Figure 4: Results of the baseline with different output representations.

Experiments for DCASE2021 dataset

- **Problem description:** to study the influence of the different elements of the spatial sound scenes on the models' performance, we conducted additional tests for different versions of the TAU-NIGENS Spatial Sound Events 2021 dataset.

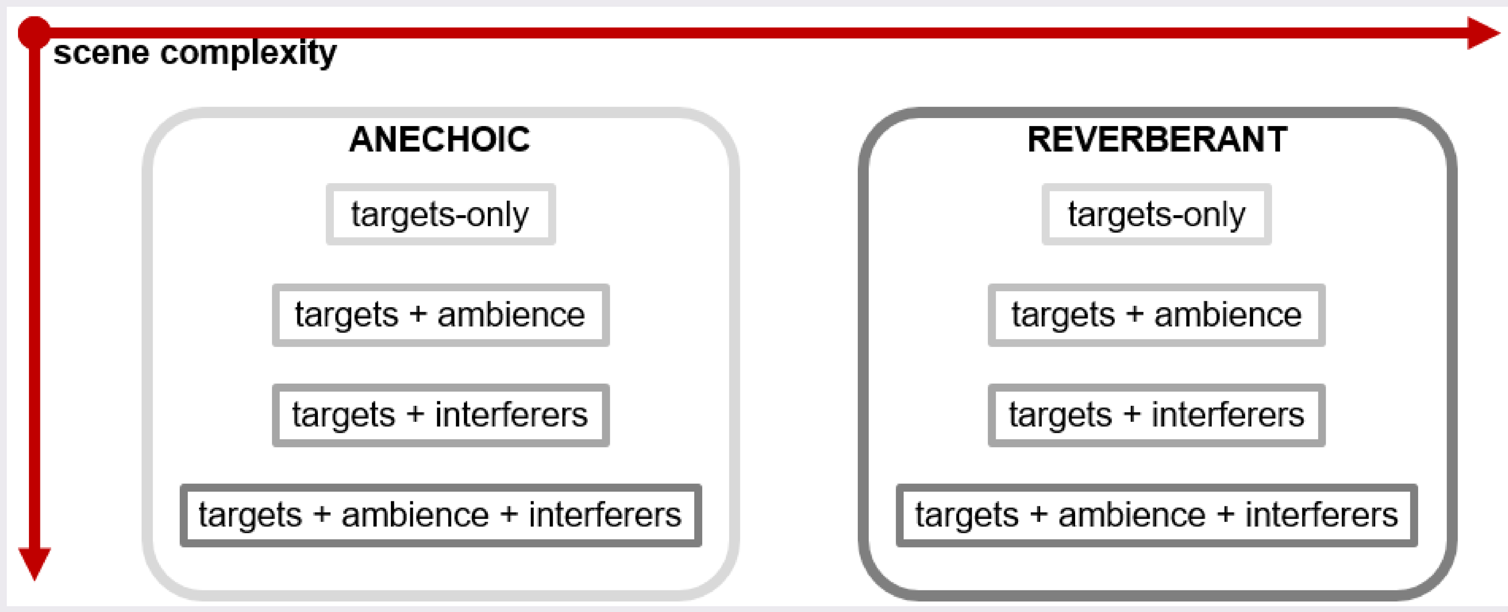

Figure 5: Composition of the different dataset versions.

- **Results**

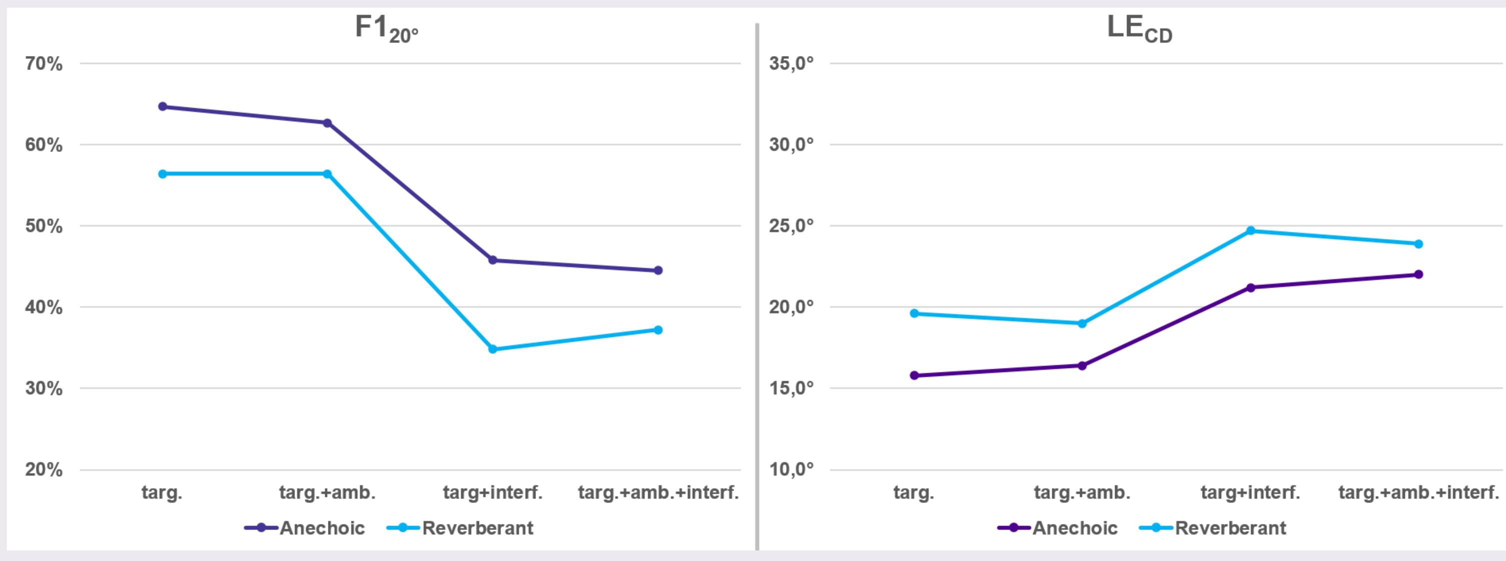

Figure 6: Results obtained for different versions of the dataset using the FOA format.

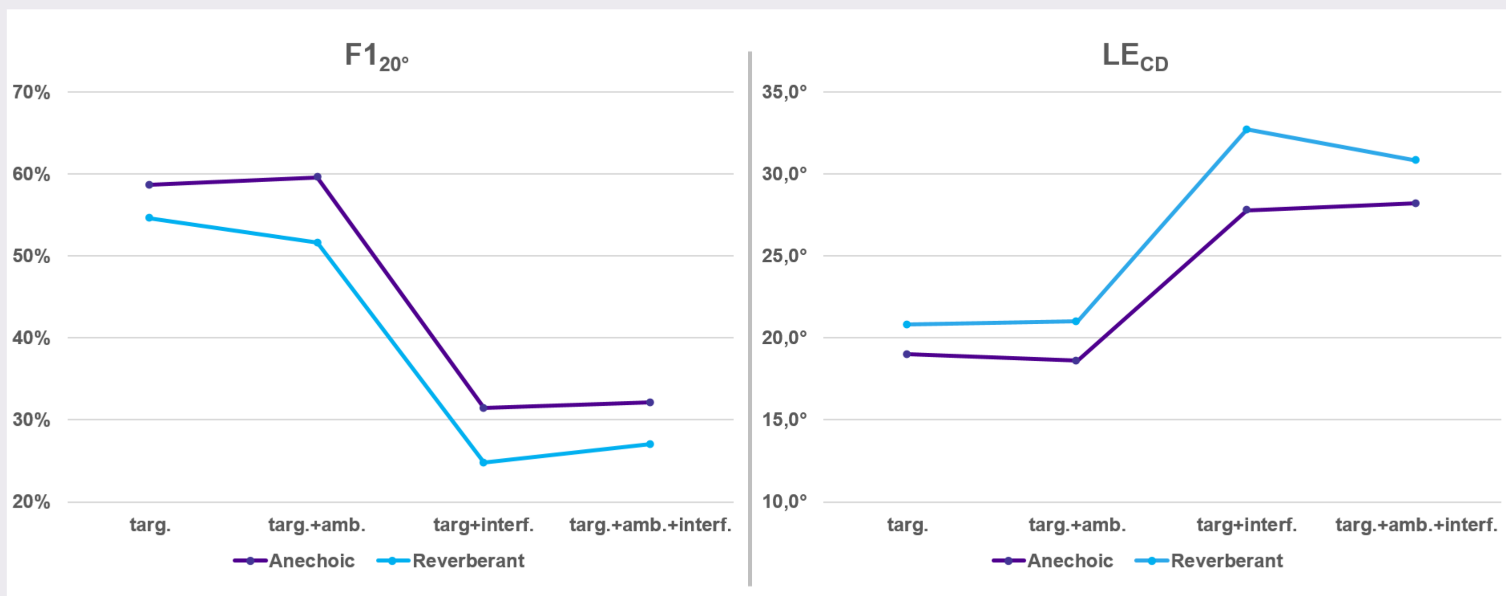

Figure 7: Results obtained for different versions of the dataset using the microphone array format.

Summary and conclusions

- This work introduces a new dataset for SELD including polyphony of up to 3 sound events, directional interfering events, and significant number of overlapping same-class events.
- Reverberation affects negatively all investigated SELD metrics.
- Ambient noise does not seem to significantly impact the results.
- Directional interfering events cause the most severe effect in SELD performance.
- The microphone array format with GCC spatial features achieve significantly worse results than the FOA format with acoustic intensity spatial features.

TAU-NIGENS Spatial Sound Events 2021
https://doi.org/10.5281/zenodo.5476980
DCASE2021 baseline system
https://github.com/sharathadavanne/seld-dcase2021

https://www.tuni.fi/en/archontis-politis/ archontis.politis@tuni.fi