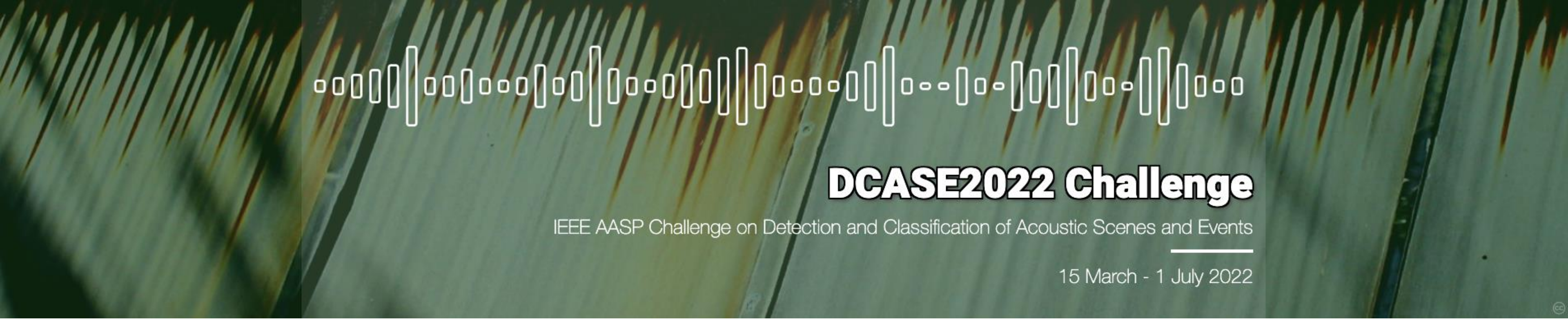


DCASE2022 Challenge

IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events

15 March - 1 July 2022

Task 3

Sound Event Localization and Detection evaluated in real spatial sound scenes

Localization

Task 3

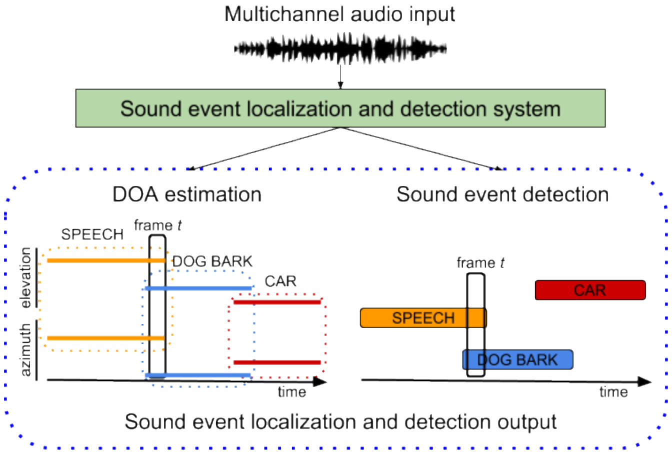
Archontis Politis, Parthasaarathy Sudarsanam, Daniel A. Krause, Sharath Adavanne, Tuomas Virtanen
Kazuki Shimada, Yuichiro Koyama, Naoya Takahashi, Shusuke Takahashi, Yuki Mitsufuji

Tampere University

SONY

Sound Event Localization and Detection

Joint **classification** of sound events, class-wise **activity detection**, and event **localization.**



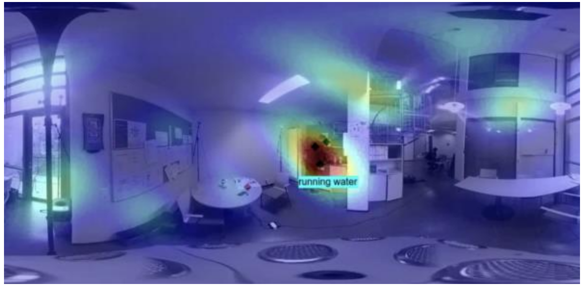
Dataset

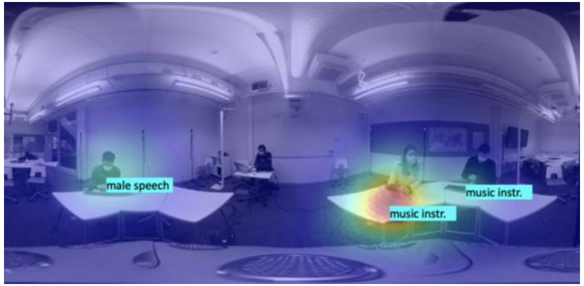
Recordings of naturally acted scenes with multiple human agents in rooms interacting between them and with the environment.

The recordings have been captured with multiple types of sensors and these have been used to annotate them spatiotemporally.

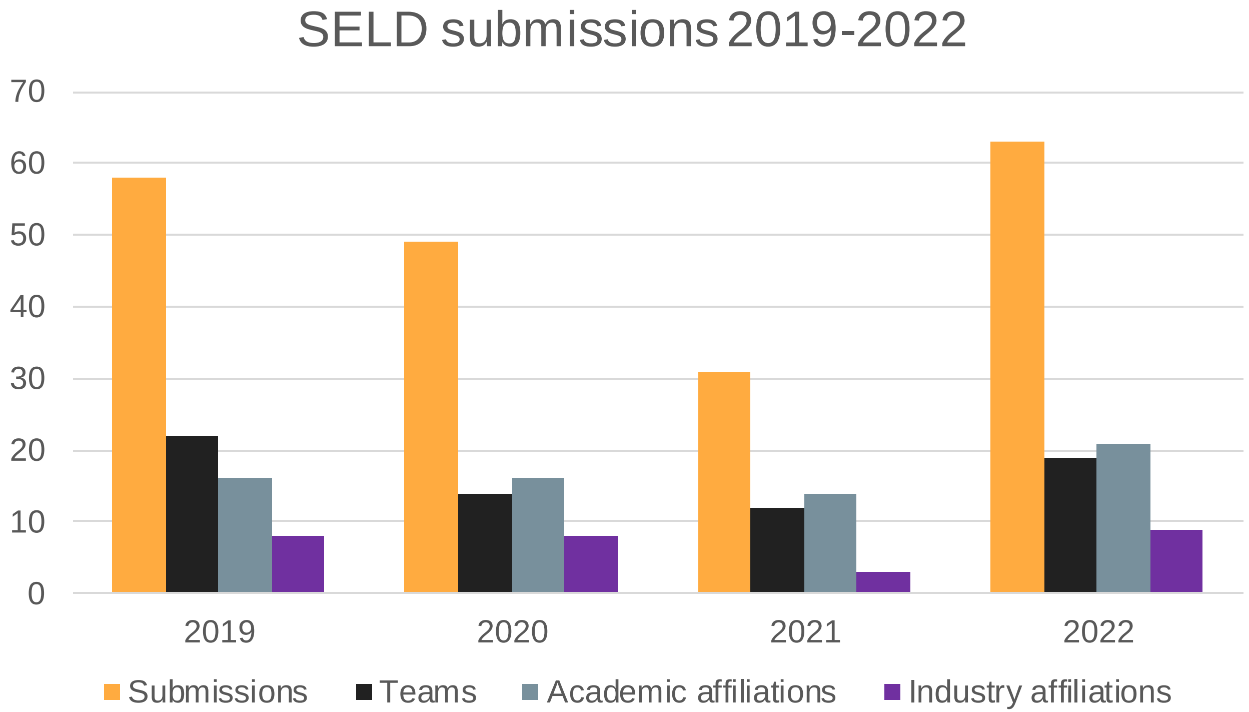
- ~7hrs of recordings captured in Tampere, FI, and Tokyo, JP
- semi improvised scenes of 1-4 actors
- 11 different rooms
- 13 annotated sound classes

- natural composition of classes, class presence, event occurences and co-occurences, and spatial distribution

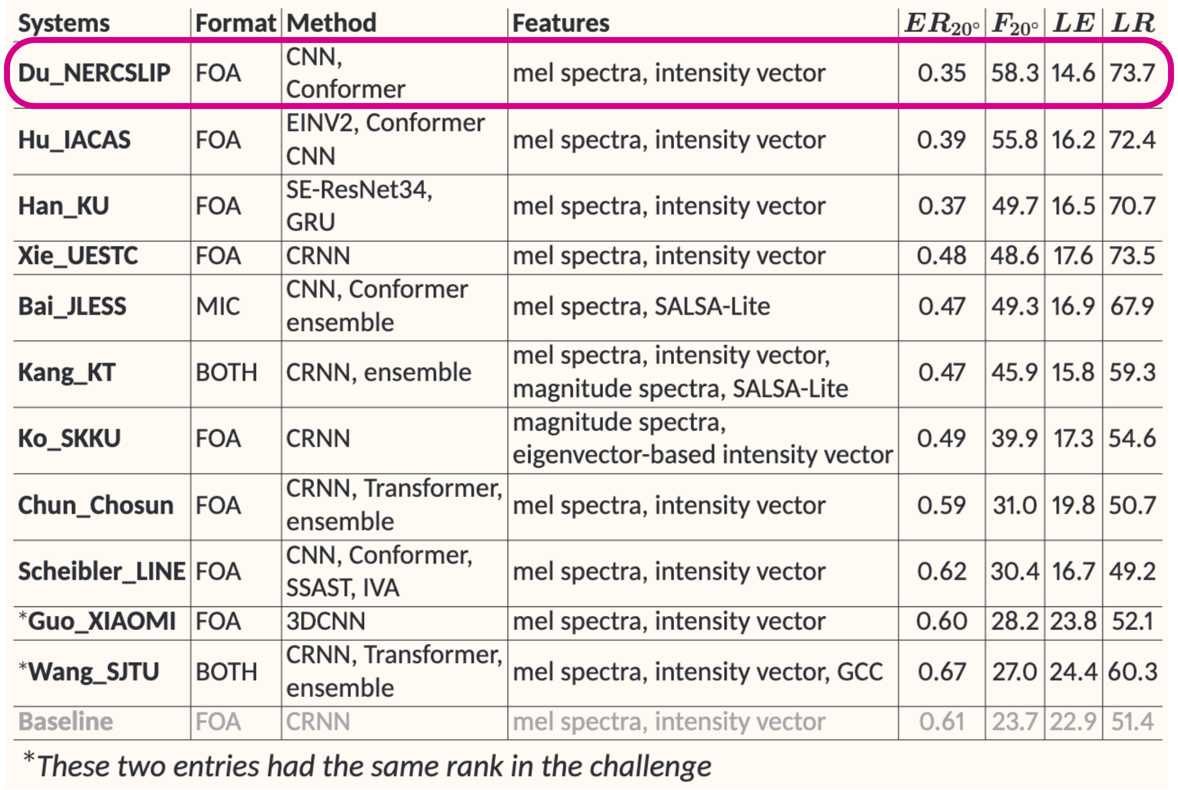




Submissions



Results



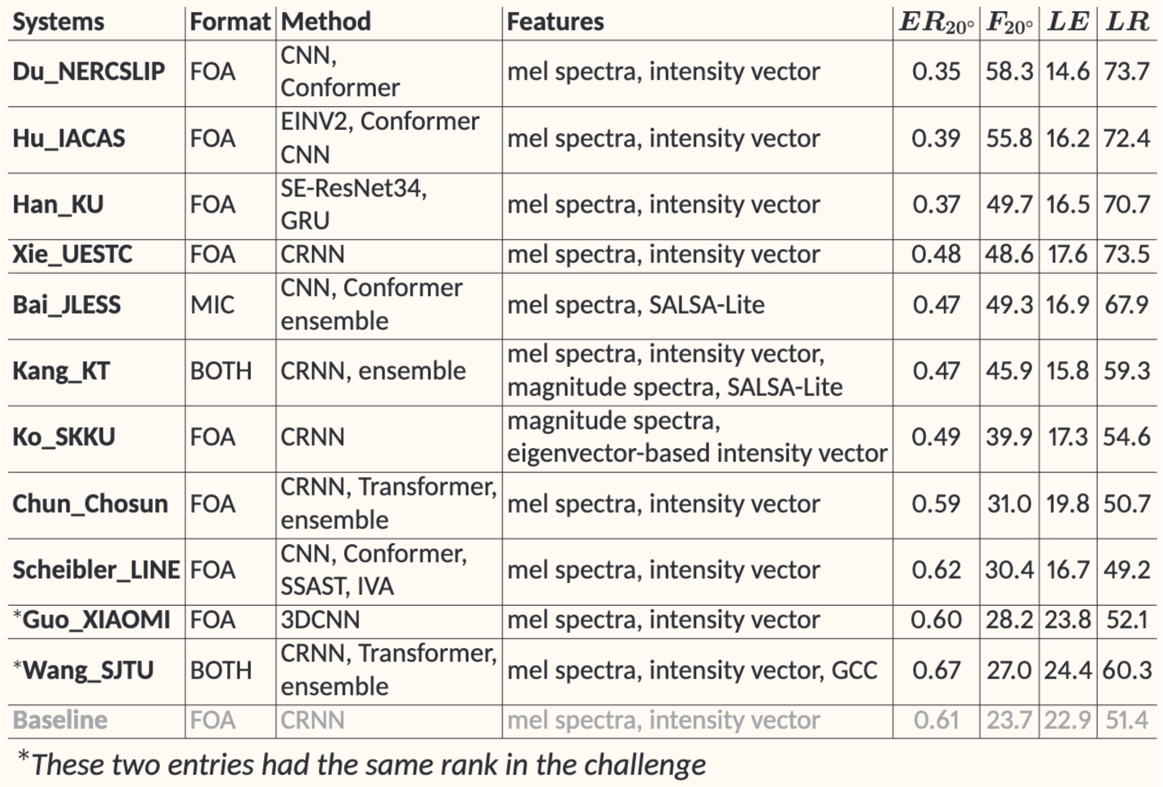
Systems	Format	Method	Features	ER_{20°	F_{20°	LE	LR
Du_NERCSLIP	FOA	CNN, Conformer	mel spectra, intensity vector	0.35	58.3	14.6	73.7
Hu_IACAS	FOA	EINV2, Conformer CNN	mel spectra, intensity vector	0.39	55.8	16.2	72.4
Han_KU	FOA	SE-ResNet34, GRU	mel spectra, intensity vector	0.37	49.7	16.5	70.7
Xie_UESTC	FOA	CRNN	mel spectra, intensity vector	0.48	48.6	17.6	73.5
Bai_JLESS	MIC	CNN, Conformer ensemble	mel spectra, SALSA-Lite	0.47	49.3	16.9	67.9
Kang_KT	BOTH	CRNN, ensemble	mel spectra, intensity vector, magnitude spectra, SALSA-Lite	0.47	45.9	15.8	59.3
Ko_SKKU	FOA	CRNN	magnitude spectra, eigenvector-based intensity vector	0.49	39.9	17.3	54.6
Chun_Chosun	FOA	CRNN, Transformer, ensemble	mel spectra, intensity vector	0.59	31.0	19.8	50.7
Scheibler_LINE	FOA	CNN, Conformer, SSAST, IVA	mel spectra, intensity vector	0.62	30.4	16.7	49.2
***Guo_XIAOMI**	FOA	3DCNN	mel spectra, intensity vector	0.60	28.2	23.8	52.1
***Wang_SJTU**	BOTH	CRNN, Transformer, ensemble	mel spectra, intensity vector, GCC	0.67	27.0	24.4	60.3
Baseline	FOA	CRNN	mel spectra, intensity vector	0.61	23.7	22.9	51.4

**These two entries had the same rank in the challenge*

- 12/19 systems did better than the baseline

- Top system *Du_NERCSLIP* had 145% improvement in spatial F-score and 36% improvement in localization error.

Results: General trends



Systems	Format	Method	Features	ER_{20°	F_{20°	LE	LR
Du_NERCSLIP	FOA	CNN, Conformer	mel spectra, intensity vector	0.35	58.3	14.6	73.7
Hu_IACAS	FOA	EINV2, Conformer CNN	mel spectra, intensity vector	0.39	55.8	16.2	72.4
Han_KU	FOA	SE-ResNet34, GRU	mel spectra, intensity vector	0.37	49.7	16.5	70.7
Xie_UESTC	FOA	CRNN	mel spectra, intensity vector	0.48	48.6	17.6	73.5
Bai_JLESS	MIC	CNN, Conformer ensemble	mel spectra, SALSA-Lite	0.47	49.3	16.9	67.9
Kang_KT	BOTH	CRNN, ensemble	mel spectra, intensity vector, magnitude spectra, SALSA-Lite	0.47	45.9	15.8	59.3
Ko_SKKU	FOA	CRNN	magnitude spectra, eigenvector-based intensity vector	0.49	39.9	17.3	54.6
Chun_Chosun	FOA	CRNN, Transformer, ensemble	mel spectra, intensity vector	0.59	31.0	19.8	50.7
Scheibler_LINE	FOA	CNN, Conformer, SSAST, IVA	mel spectra, intensity vector	0.62	30.4	16.7	49.2
*__Guo_XIAOMI__	FOA	3DCNN	mel spectra, intensity vector	0.60	28.2	23.8	52.1
*__Wang_SJTU__	BOTH	CRNN, Transformer, ensemble	mel spectra, intensity vector, GCC	0.67	27.0	24.4	60.3
Baseline	FOA	CRNN	mel spectra, intensity vector	0.61	23.7	22.9	51.4

**These two entries had the same rank in the challenge*

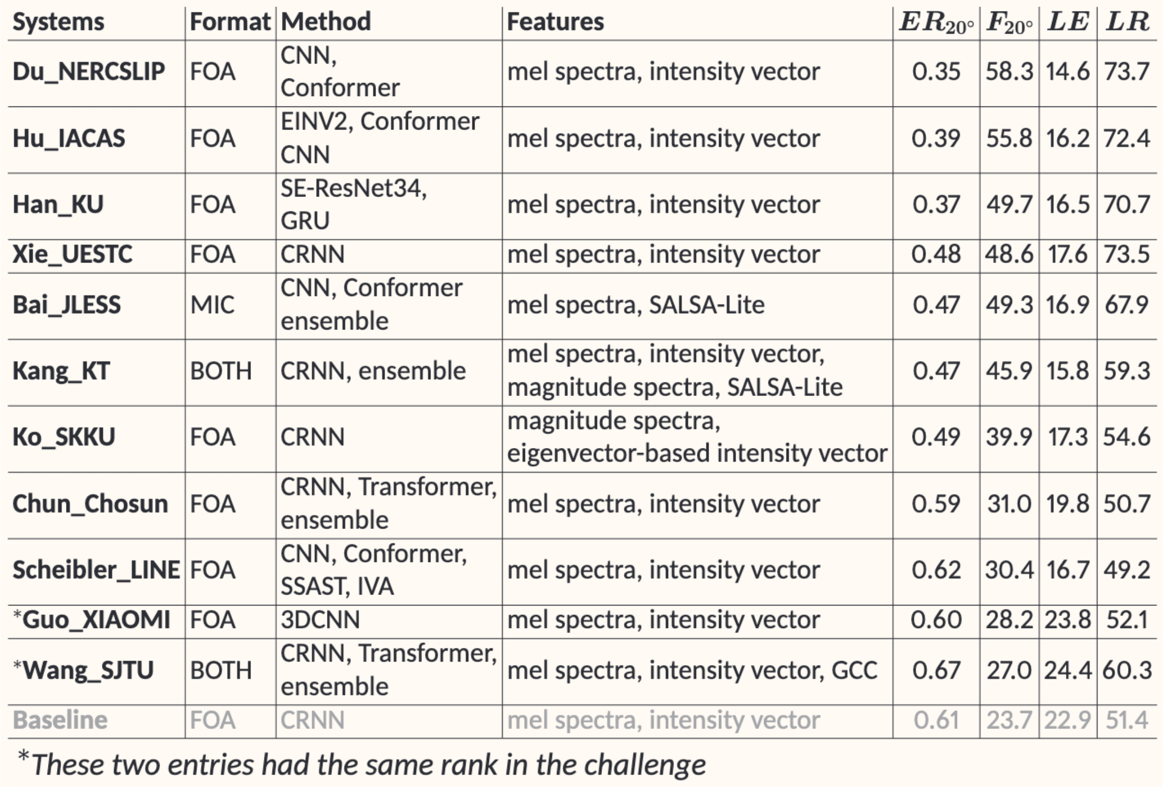
- Model: Baseline CRNN is widely used, and many teams upgrade the model with CNN, Transformer, or Conformer.

- Feature: Most teams keep the feature of the baseline, mel spectra and intensity vector, while a few teams take SALSA-Lite or others.

- SELD method: More than half teams follow the baseline to use Multi-ACCDOA while some teams use ACCDOA, EINV2, or others.

- Data augmentation:
  * Multichannel data simulation
  * Audio channel swapping (Rotation)
  * Mixup
  * SpecAugment
  * Band-pass filter
  * Perturbation of gain/frequency/frame/pitch
  * Angle noise to label

Results: Comments on several systems

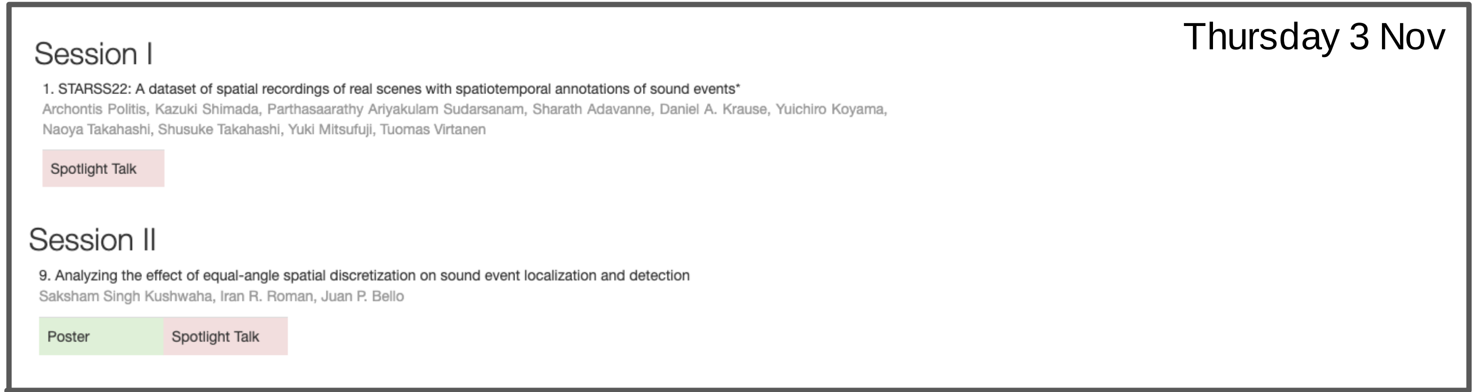


Systems	Format	Method	Features	ER_{20°	F_{20°	LE	LR
Du_NERCSLIP	FOA	CNN, Conformer	mel spectra, intensity vector	0.35	58.3	14.6	73.7
Hu_IACAS	FOA	EINV2, Conformer CNN	mel spectra, intensity vector	0.39	55.8	16.2	72.4
Han_KU	FOA	SE-ResNet34, GRU	mel spectra, intensity vector	0.37	49.7	16.5	70.7
Xie_UESTC	FOA	CRNN	mel spectra, intensity vector	0.48	48.6	17.6	73.5
Bai_JLESS	MIC	CNN, Conformer ensemble	mel spectra, SALSA-Lite	0.47	49.3	16.9	67.9
Kang_KT	BOTH	CRNN, ensemble	mel spectra, intensity vector, magnitude spectra, SALSA-Lite	0.47	45.9	15.8	59.3
Ko_SKKU	FOA	CRNN	magnitude spectra, eigenvector-based intensity vector	0.49	39.9	17.3	54.6
Chun_Chosun	FOA	CRNN, Transformer, ensemble	mel spectra, intensity vector	0.59	31.0	19.8	50.7
Scheibler_LINE	FOA	CNN, Conformer, SSAST, IVA	mel spectra, intensity vector	0.62	30.4	16.7	49.2
*__Guo_XIAOMI__	FOA	3DCNN	mel spectra, intensity vector	0.60	28.2	23.8	52.1
*__Wang_SJTU__	BOTH	CRNN, Transformer, ensemble	mel spectra, intensity vector, GCC	0.67	27.0	24.4	60.3
Baseline	FOA	CRNN	mel spectra, intensity vector	0.61	23.7	22.9	51.4

**These two entries had the same rank in the challenge*

- Top 3 teams used external data and sophisticated data augmentation techniques.
- *Kang_KT* applied AD-PIT to multi-task SELDnet.
- *Ko_SKKU* modified original mixup for ACCDOA.
- *Scheibler_LINE* used IVA to separte sources, while Park_SU used ResUNet.
- *Guo_XIAOMI* proposed a network to consider time alignment.
- Many more SELD-specific innovations proposed (COLOC representation, Spatial Mixup a.o)

Task 3 @ DCASE Workshop



Thursday 3 Nov

Session I

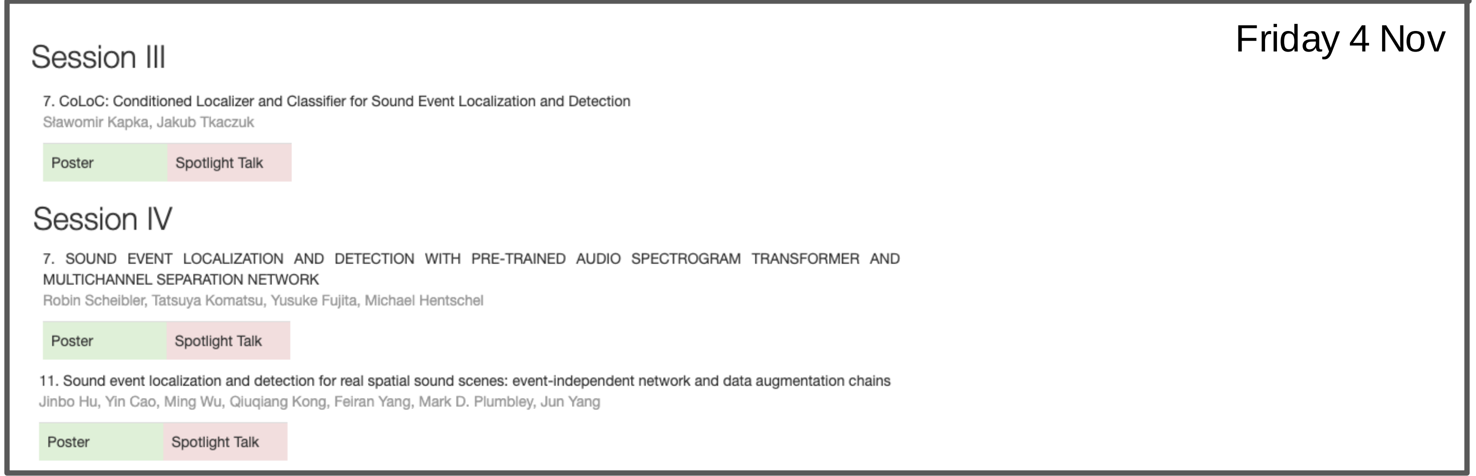
1. STARSS22: A dataset of spatial recordings of real scenes with spatiotemporal annotations of sound events*
Archontis Politis, Kazuki Shimada, Parthasaarathy Ariyakulam Sudarsanam, Sharath Adavanne, Daniel A. Krause, Yuichiro Koyama, Naoya Takahashi, Shusuke Takahashi, Yuki Mitsufuji, Tuomas Virtanen

Spotlight Talk

Session II

9. Analyzing the effect of equal-angle spatial discretization on sound event localization and detection
Saksham Singh Kushwaha, Iran R. Roman, Juan P. Bello

Poster | Spotlight Talk



Friday 4 Nov

Session III

7. CoLoC: Conditioned Localizer and Classifier for Sound Event Localization and Detection
Sławomir Kapka, Jakub Tkaczuk

Poster | Spotlight Talk

Session IV

7. SOUND EVENT LOCALIZATION AND DETECTION WITH PRE-TRAINED AUDIO SPECTROGRAM TRANSFORMER AND MULTICHANNEL SEPARATION NETWORK
Robin Scheibler, Tatsuya Komatsu, Yusuke Fujita, Michael Hentschel

Poster | Spotlight Talk

11. Sound event localization and detection for real spatial sound scenes: event-independent network and data augmentation chains
Jinbo Hu, Yin Cao, Ming Wu, Qiuqiang Kong, Feiran Yang, Mark D. Plumbley, Jun Yang

Poster | Spotlight Talk

Thank you!

DCASE2022 Workshop
Workshop on Detection and Classification of Acoustic Scenes and Events
3-4 November 2022, Nancy, France