Spatial Scaper: A library to simulate and augment soundscapes for sound event localization and detection in realistic rooms

Iran R. Roman[1] Christopher Ick[1] Sivan Ding[1] Adrian S. Roman[2] Brian McFee[1] Juan P. Bello[1]

[1] Music and Audio Research Laboratory, New York University, New York, USA
[2] Viterbi School of Engineering, University of Southern California, California, USA

Summary

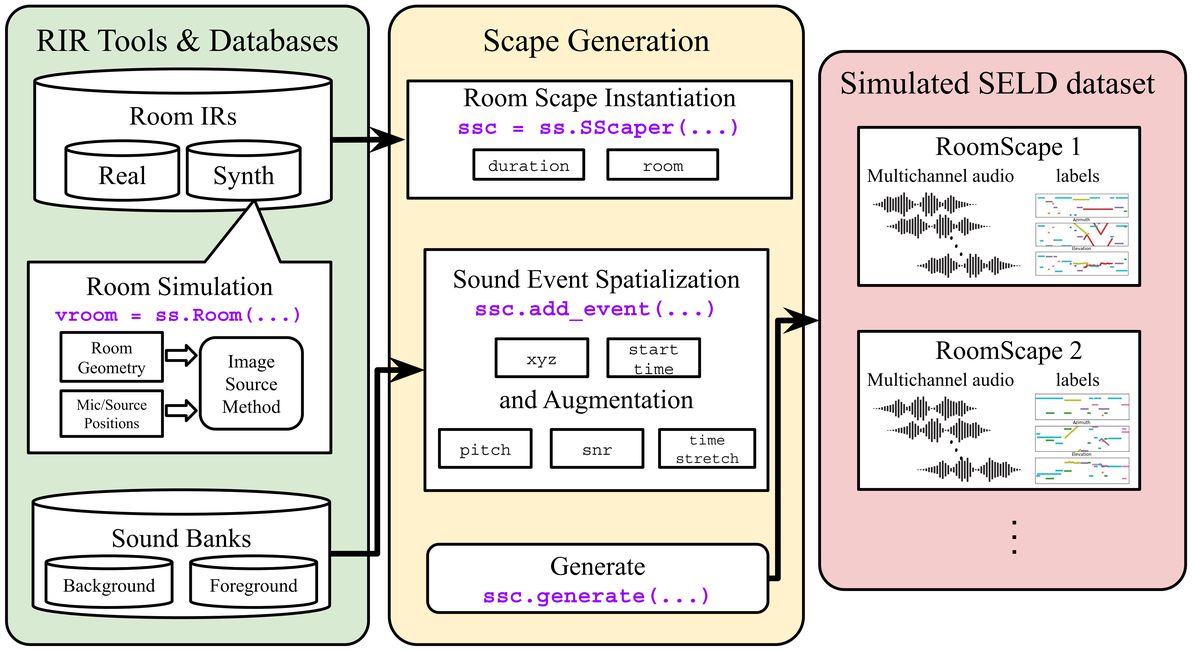
- We introduce `spatial_scaper`, a library for SELD data simulation and augmentation via real and simulated room impulse responses
- We demonstrate comparative benefits for various forms of data augmentation for training SELD models
- We show progressive performance improvements for increasing room variety when training SELD models

Sound Event Localization and Detection (SELD)

SELD is the the joint task of identifying the time and class of an event, and estimating it's direction-of-arrival. SELD sees use in many fields, ranging from bioacoustics, to urban monitoring, machine fault detection, and assistive devices for the hard-of-hearing. As modern methods for SELD trend towards highly parametrized models, i.e. deep neural networks, the data requirements for strongly-labeled spatial audio data are higher than ever. Real data is costly to produce and annotate, and is further limited by microphone configuration, room geometry, and acoustic class diversity. An alternative to field recordings is convolving spatial room impulse responses (SRIRs) with mono audio, but limited SRIR datsets similarly limit our data scale. To ameliorate this, we present `spatial_scaper`, a library for generating high volume spatial audio from prerecorded SRIRs, simulated SRIRs, and augmenting these datasets using audio deformation and rotation.

MARL NYU Music and Audio Research Laboratory
USC Viterbi School of Engineering
NEW YORK UNIVERSITY

Data Generation Pipeline



Key Contributions / Features

- Spatial audio generation from **any pre-recorded SRIR datasets**
- Simulated SRIR generation for **user-chosen room and microphone array specs**
- Augmenting SRIR/spatial audio datasets via **audio deformation and rotation**

Comparison of Features with Other Acoustic Simulators

	External SRIR Support	Custom Mic Arrays	Custom Rooms	Dataset Augmentation	Required Data
`spatial_scaper`	✓	✓	✓	✓	SRIRs or Room Specs
DCASE Generator [4]	✗	✗	✗	✗	SRIRs
SoundSpaces 2.0 [2]	✗	✓	✗	✗	RGB-D (Matterport/Replica)
ThreeDWorld [3]	✗	✗	✓	✗	Room Specs

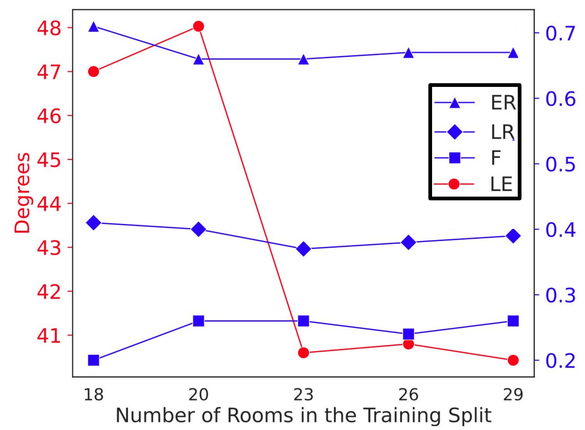
Spatialize some audio with `spatial_scaper` today:

Data Augmentation

Data	ER	F	LE	LR
Original DCASE2020	0.71	20.2	47.0	**40.5**
`spatial_scaper`	0.71	19.8	46.5	34.0
+Channel Swapping	**0.59**	**31.7**	**29.5**	31.2
+Pitch Shifting	0.67	17.8	48.4	36.9

SELDnet[1] performance on the STARSS23[5] "dev-test-sony" as a test split for different versions of training data used.

Room Diversity via Simulation



Performance as a function of adding rooms (i.e. increasing acoustic diversity) to the training split.

References

[1] S. Adavanne, A. Politis, J. Nikunen, and T. Virtanen.
Seld of overlapping sources using convolutional recurrent neural networks.
CoRR, abs/1807.00129, 2018.

[2] C. Chen et al.
Soundspaces 2.0: A simulation platform for visual-acoustic learning.
In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 8896–8911. Curran Associates, Inc., 2022.

[3] C. Gan et al.
ThreeDWorld: A platform for interactive multi-modal physical simulation.
In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1)*, 2021.

[4] D. Krause and A. Politis.
github.com/danielkrause/dcase2022-data-generator.

[5] A. Politis, K. Shimada, et al.
STARSS22: A dataset of spatial recordings of real scenes with spatiotemporal annotations of sound events.
In *Proceedings of the 8th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022)*, pages 125–129, Nancy, France, November 2022.

[6] J. Salamon, D. MacConnell, M. Cartwright, P. Li, and J. P. Bello.
Scaper: A library for soundscape synthesis and augmentation.
In *WASPAA*, pages 344–348, 2017.

This material is based upon work partly supported by the National Science Foundation under NSF Award 1922658.