Parametric Analysis of Ambisonic Audio

Contributions to methods, applications and data generation

Andrés Pérez López
23-10-2020

Supervision
Dr. Emilia Gómez
Dr. Adan Garriga

President Dr. Xavier Serra
Secretary Dr. Máximo Cobos
Vocal Dr. Tuomas Virtanen

Outline

I. INTRODUCTION

- 1. Introduction

II. SCIENTIFIC CONTRIBUTIONS

- 2. Blind Reverberation Time Estimation
- 3. Coherence Estimation
- 4. Sound Event Localization and Detection
- 5. Data Generation and Storage

III. CONCLUSIONS

- 6. Conclusions

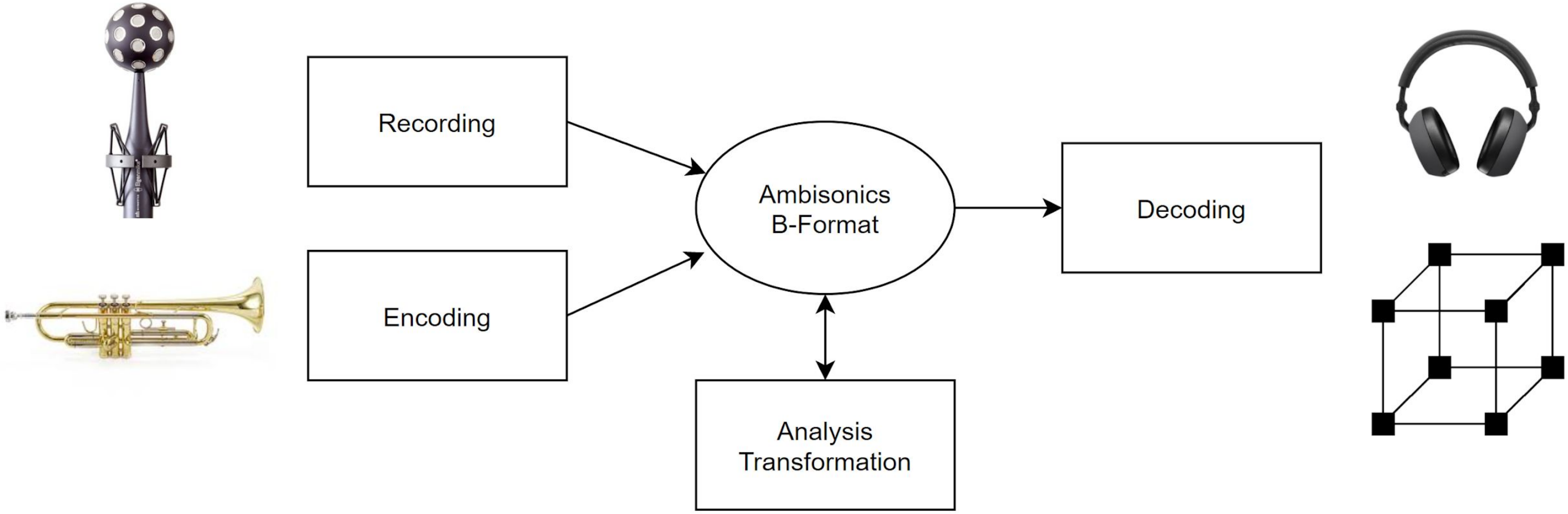
I - INTRODUCTION

1. Introduction

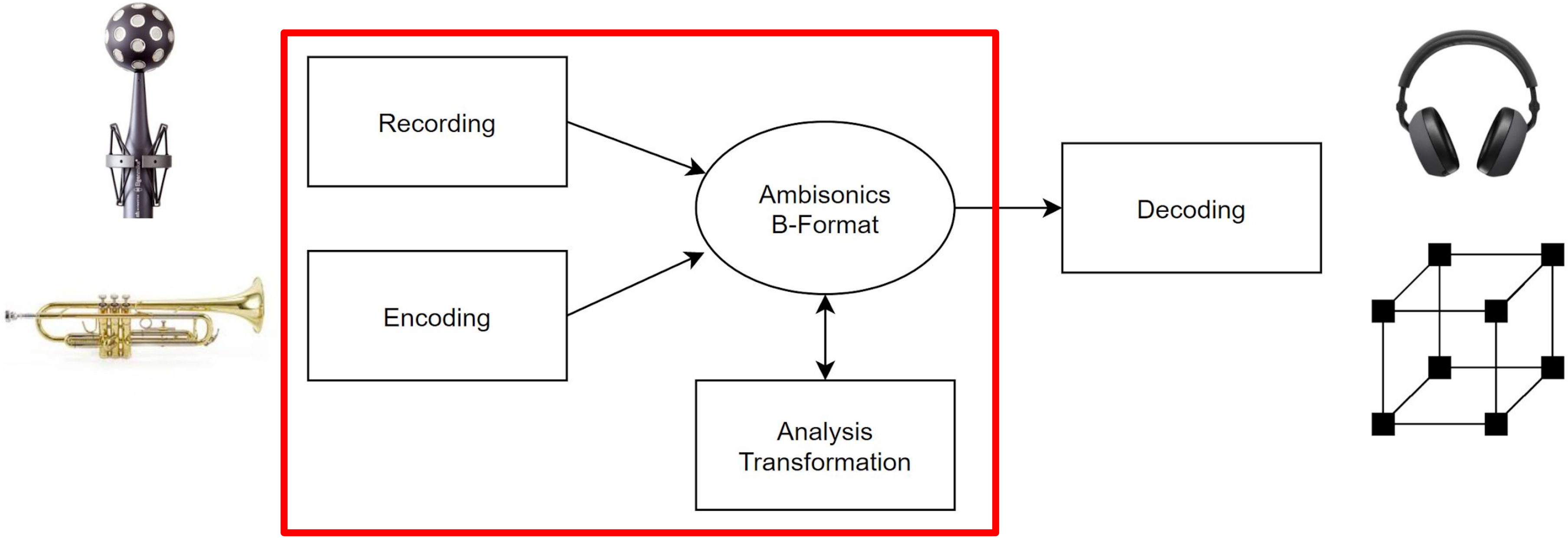
Motivation

- Popularization of VR/AR
- Affordability of Ambisonic (VR) microphones
- Standardization of Ambisonics as spatial format
- New challenges and opportunities

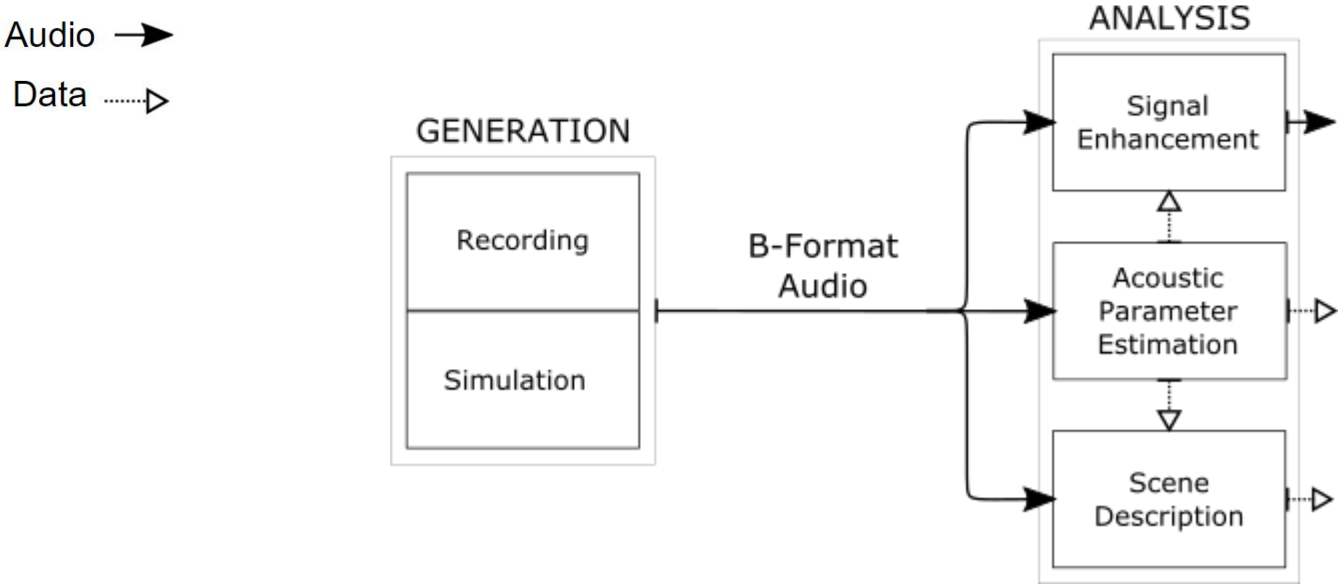
Motivation


Recording
Encoding
Ambisonics
B-Format
Decoding
Analysis
Transformation

Motivation

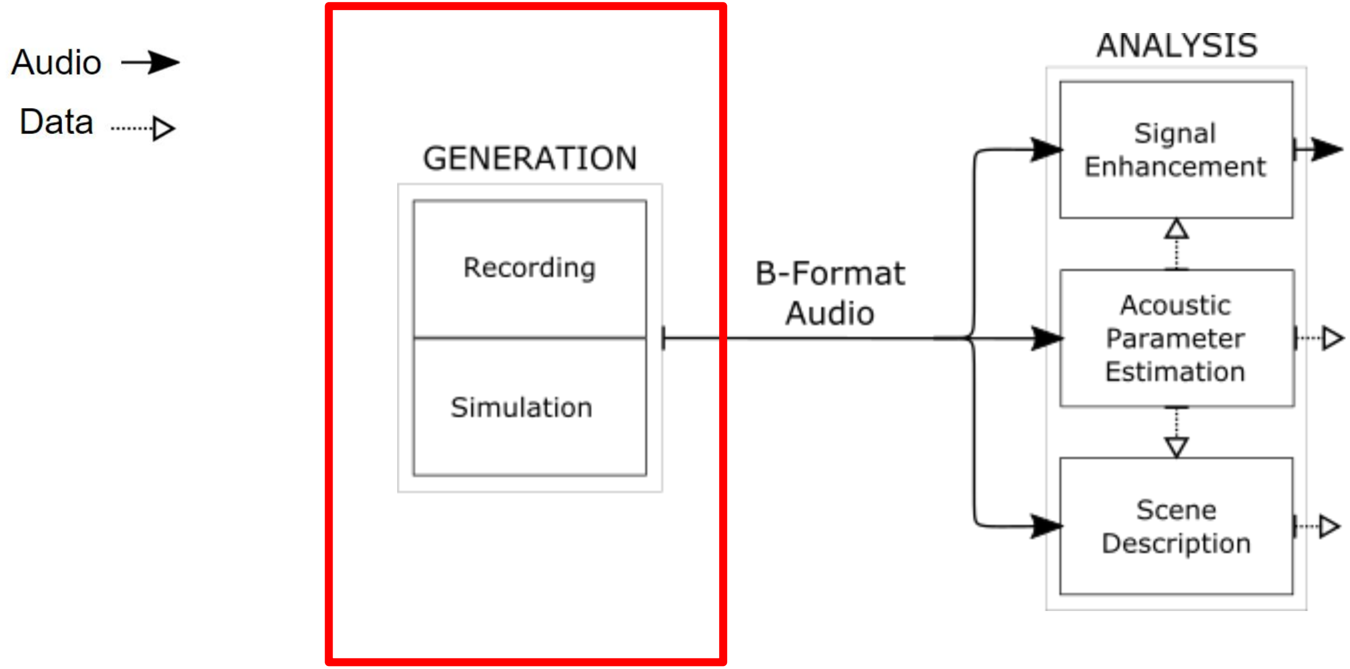

Recording
Encoding
Ambisonics
B-Format
Analysis
Transformation
Decoding

Problem description

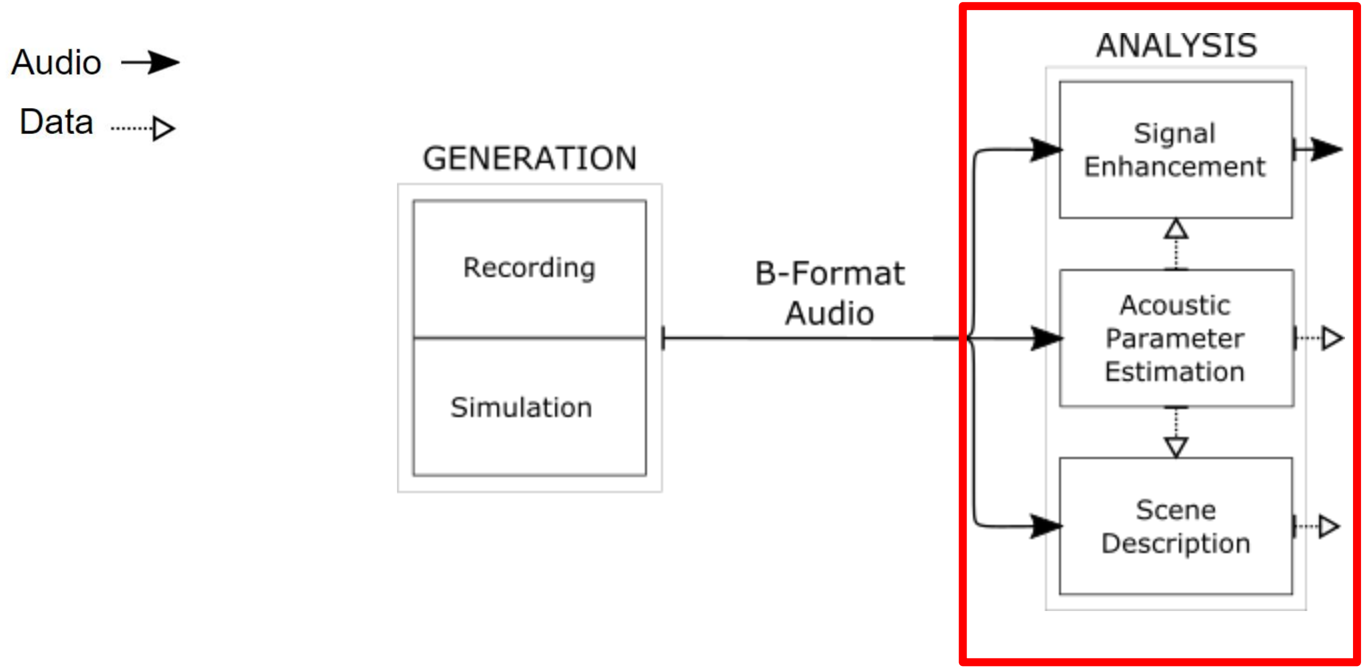


[1] Jarrett, D., Habets, E., and Naylor, P. (2017). *Theory and applications of spherical microphone array processing*, volume 9. Springer.

Problem description



Problem description

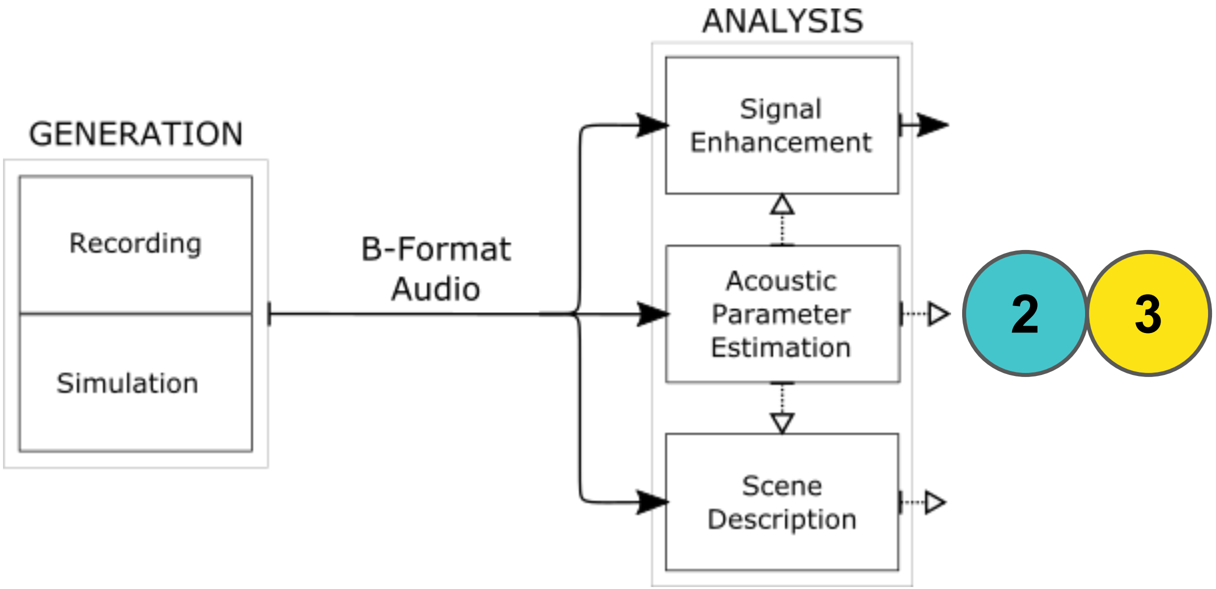


Scientific Objectives

1. **Acoustic Parameter Estimation**

Scientific Objectives

1. Acoustic Parameter Estimation

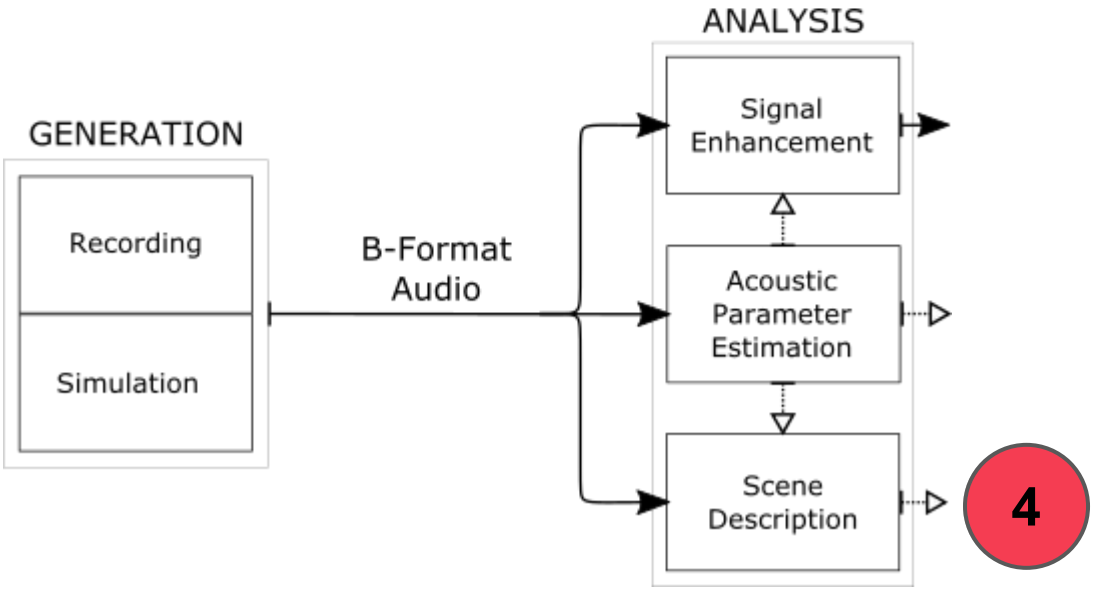


Scientific Objectives

2. Sound Event Localization and Detection

Scientific Objectives

2. Sound Event Localization and Detection

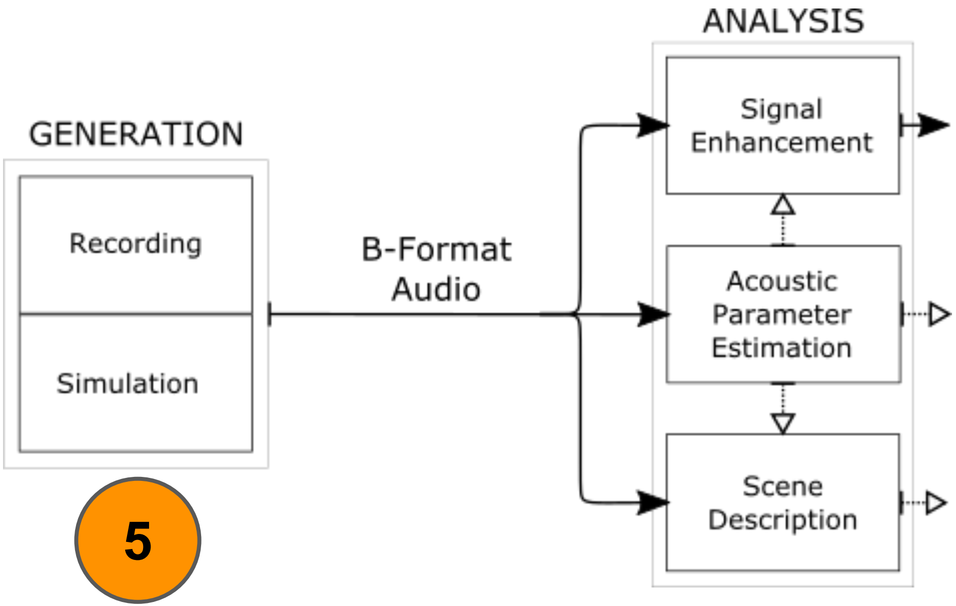


Scientific Objectives

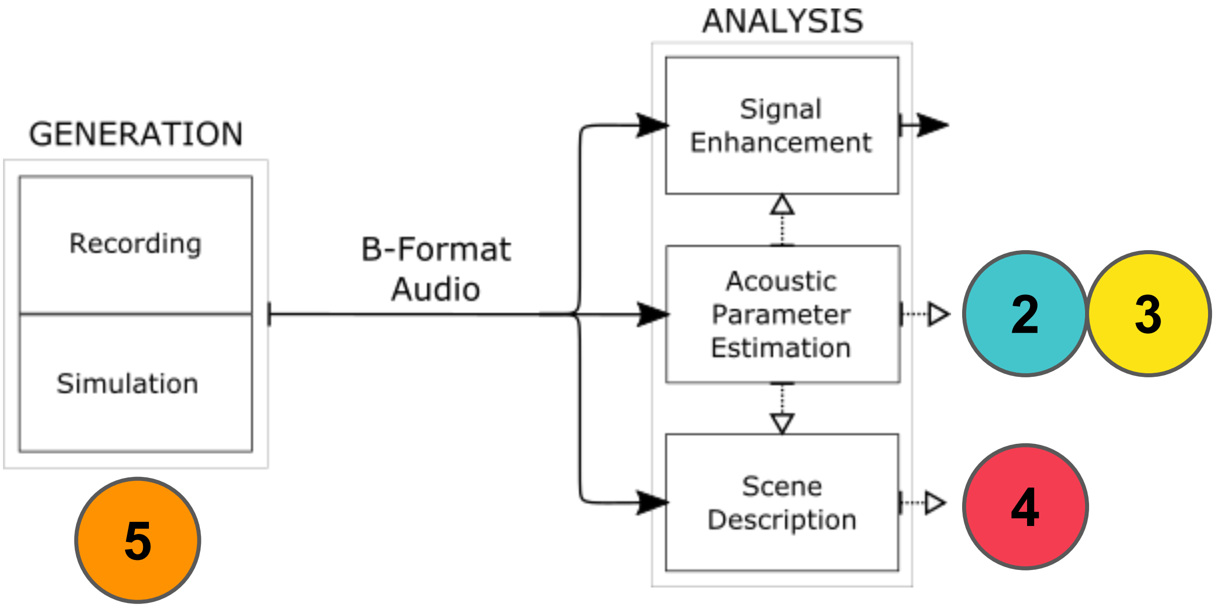
3. Data Generation and Storage

Scientific Objectives

3. Data Generation and Storage

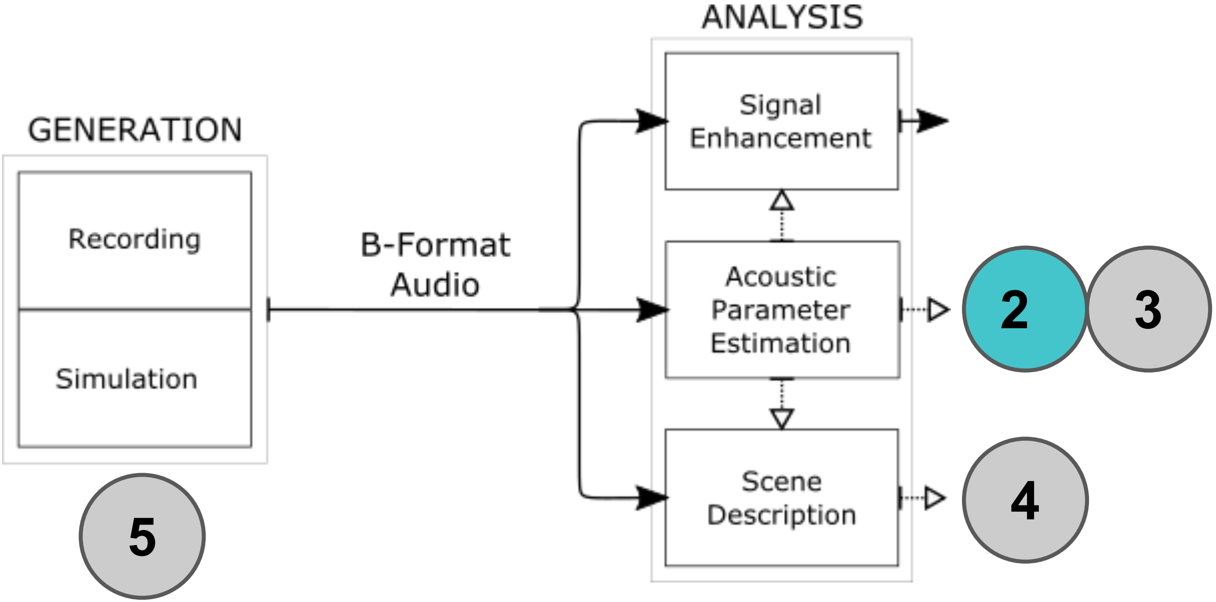


Outline


GENERATION
Recording
Simulation
5
B-Format
Audio
ANALYSIS
Signal
Enhancement
Acoustic
Parameter
Estimation
Scene
Description
2
3
4

II - SCIENTIFIC CONTRIBUTIONS

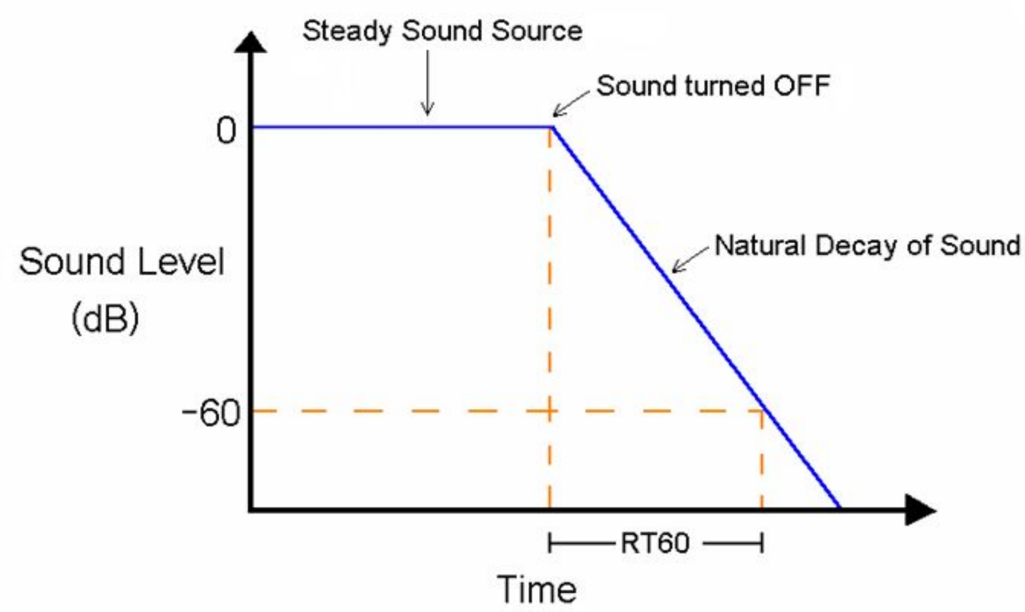
2. Blind Reverberation Time Estimation



GENERATION
Recording
Simulation
5
B-Format
Audio
ANALYSIS
Signal
Enhancement
Acoustic
Parameter
Estimation
Scene
Description
2
3
4

Introduction

Reverberation Time (*RT60)*



[2] Audioforo.com. *Curso de medidas acústicas.*
https://audioforo.com/2015/03/18/curso-de-medidas-acusticas-3-como-mido-el-tr/

Introduction

- State of the art: ACE Challenge 2015 [3]
- Focus on single-channel

[3] J. Eaton, N. D. Gaubitch, A. H. Moore, and P. A. Naylor, "Estimation of room acoustic parameters: The ACE challenge" IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 10, pp. 1681– 1693, 2016

Introduction

- State of the art: ACE Challenge 2015 [3]
- Focus on single-channel

Proposal:

novel *RT60* estimation method for first order ambisonics

[3] J. Eaton, N. D. Gaubitch, A. H. Moore, and P. A. Naylor, “Estimation of room acoustic parameters: The ACE challenge” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 10, pp. 1681– 1693, 2016

Proposed Method

1. **Dereverberation**

Obtain a clean
signal estimate

2. System Identification

Compute the RIR
by inverse filtering

Proposed Method

Dereverberation

- Split early and late reverberation
- Multichannel Autoregressive Model (MAR)
- Offline optimization based on group sparsity [4]

System Identification

- Estimate impulse response from dry and recorded
- Compute *RT60* from Schroeder integral

[4] A. Jukic, T. van Waterschoot, T. Gerkmann, and S. Doclo, “Group sparsity for mimo speech dereverberation,” in 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA). IEEE, 2015, pp. 1–5.

Experimental Setup

Baseline Method [5]

- 1st classified in ACE regarding Pearson correlation coefficient
- Subband detection of energy decays in signal offsets

[5] T. d. M. Prego, A. A. de Lima, S. L. Netto, B. Lee, A. Said, R. W. Schafer, and T. Kalker, "A blind algorithm for reverberation-time estimation using subband decomposition of speech signals," The Journal of the Acoustical Society of America, vol. 131, no. 4, pp. 2811–2816, 2012.

Experimental Setup

Datasets

1. Speech

LibriSpeech [6]

test-clean

2. Drums

DSD100 [7]

test-drums

[6] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: an asr corpus based on public domain audio books," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2015, pp. 5206–5210.

[7] A. Liutkus, F.-R. Stoter, Z. Rafii, D. Kitamura, B. Rivet, N. Ito, N. Ono, and J. Fontecave, "The 2016 signal separation evaluation campaign," in Latent Variable Analysis and Signal Separation - 12th International Conference, LVA/ICA 2015, Liberec, Czech Republic, August 25-28, 2015. Springer International Publishing, 2017, pp. 323–332.

Experimental Setup

Datasets

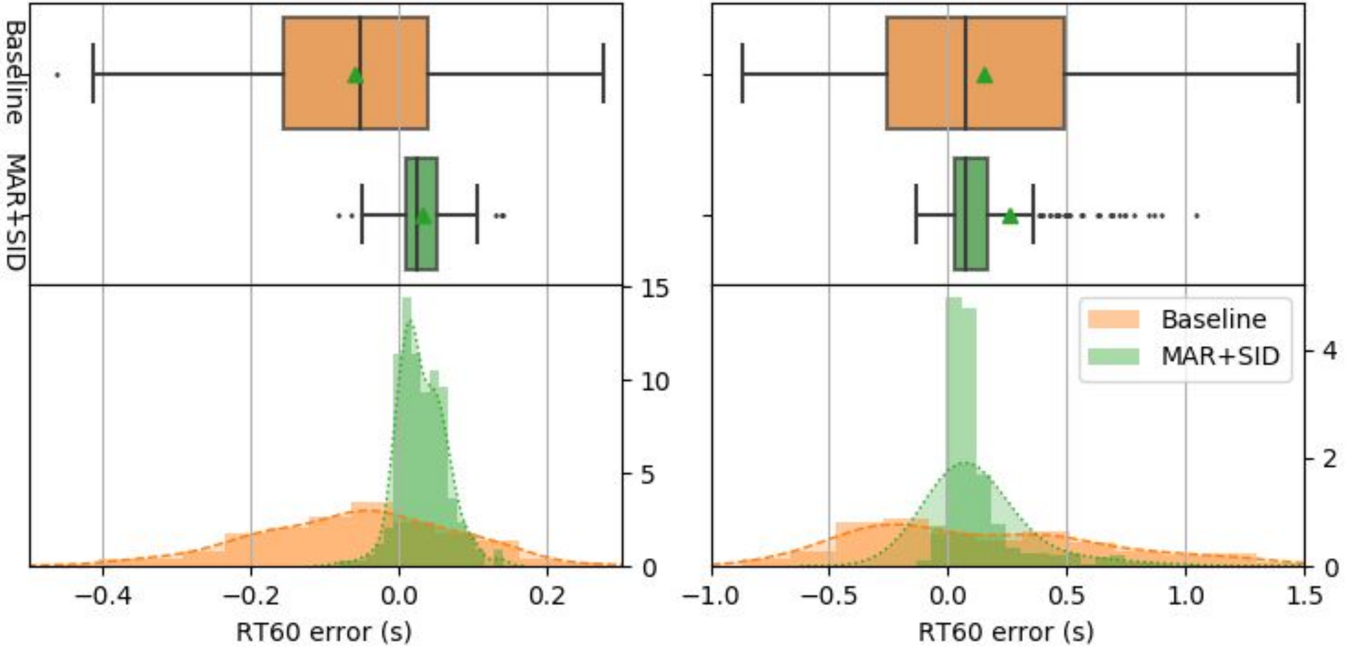
- 9 simulated First Order Ambisonic RIRs
- *RT60* between 0.4 and 1.1 s (@1 kHz)
- Random (uniform) Direction of Arrival

- Total clips: 270 (speech), 450 (drums)

Results

Metric	**speech**		**drums**	
	Baseline	***MAR+SID***	***Baseline***	***MAR+SID***
Bias	-0.0599	**0.0305**	**0.1521**	0.2568
MSE	0.6366	**0.0594**	**13.9376**	16.5261
ρ	0.8212	**0.9848**	0.3705	**0.7552**

Results



Baseline
MAR+SID
15
10
5
0
−0.4
−0.2
0.0
0.2
RT60 error (s)
Baseline
MAR+SID
4
2
0
−1.0
−0.5
0.0
0.5
1.0
1.5
RT60 error (s)

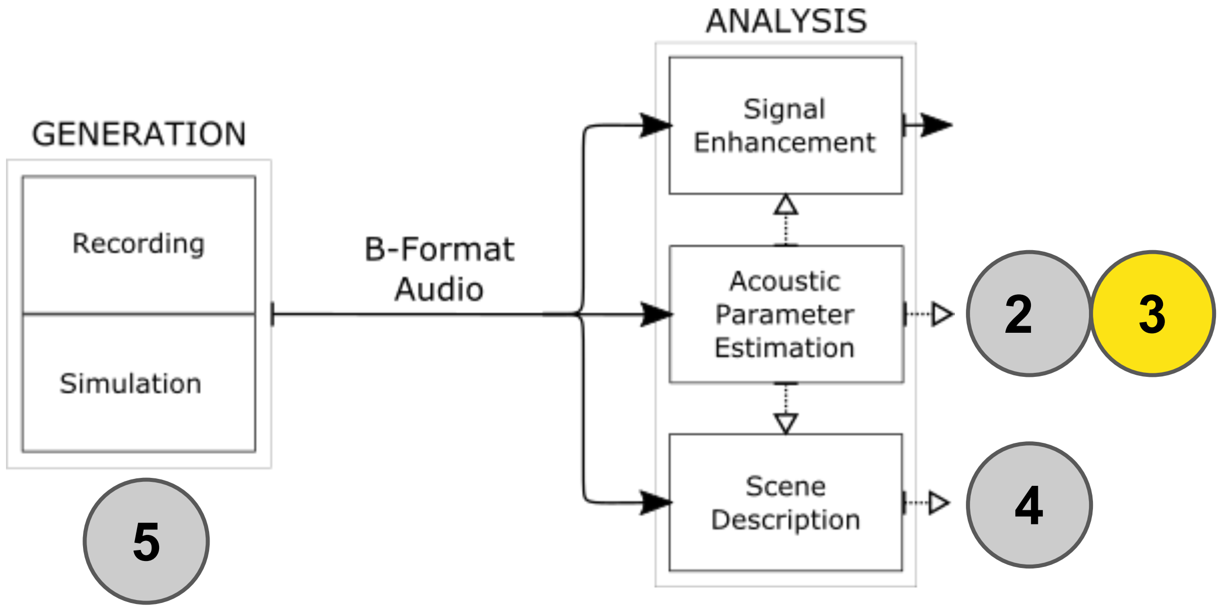
Summary of contributions

- **Novel method** for blind *RT60* estimation
  - Dereverberation using auto-recursive model
  - System identification

- **Outperforms state-of-the-art** baseline method in most metrics

- Improved **consistency** and **robustness**

3. Coherence Estimation


GENERATION
Recording
Simulation
5
B-Format
Audio
ANALYSIS
Signal
Enhancement
Acoustic
Parameter
Estimation
Scene
Description
2
3
4

Introduction

- Coherence/diffuseness as sound field parameter
  - Spatial domain: Magnitude Squared Coherence (MSC) [8]
  - Ambisonic domain: sound field diffuseness (e.g. DirAC [9])

[8] Elko, G. W. (2001). *Spatial coherence functions for differential microphones in isotropic noise fields.* In Microphone Arrays, pages 61–85. Springer, New York.
[9] Pulkki, V. (2007). *Spatial sound reproduction with directional audio coding*. Journal of the Audio Engineering Society, 55(6):503–516

Introduction

- A-Format microphones:
  - Spatial domain: known curve [8]
  - Ambisonic domain: **theoretically incoherent** [8]

[8] Elko, G. W. (2001). *Spatial coherence functions for differential microphones in isotropic noise fields.* In Microphone Arrays, pages 61–85. Springer, New York.

Introduction

Problem definition

Model non-ideal behavior of coherence estimators:

- A-Format microphones
- Spherical isotropic noise
- Recorded and simulated audio
- Spatial and spherical harmonic domain

Methods

Simulation

Isotropic noise using the *geometrical method* [10, 11]

- 1024 plane waves
- Ideal open-sphere A-Format microphone
  - Radius 0.015 m
  - Directivity factor 0.5 (cardioid)

[10] Habets, E. A. P. and Gannot, S. (2007). Generating sensor signals in isotropic noise fields. The Journal of the Acoustical Society of America, 122(6):3464–3470.
[11] Habets, E. A. P. and Gannot, S. (2010). Comments on generating sensor signals in isotropic noise fields.

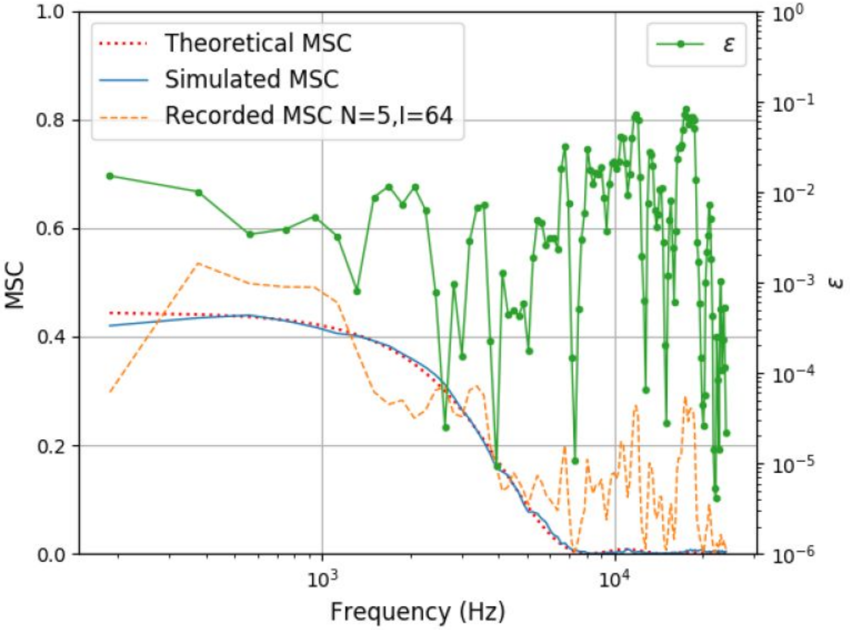
Methods

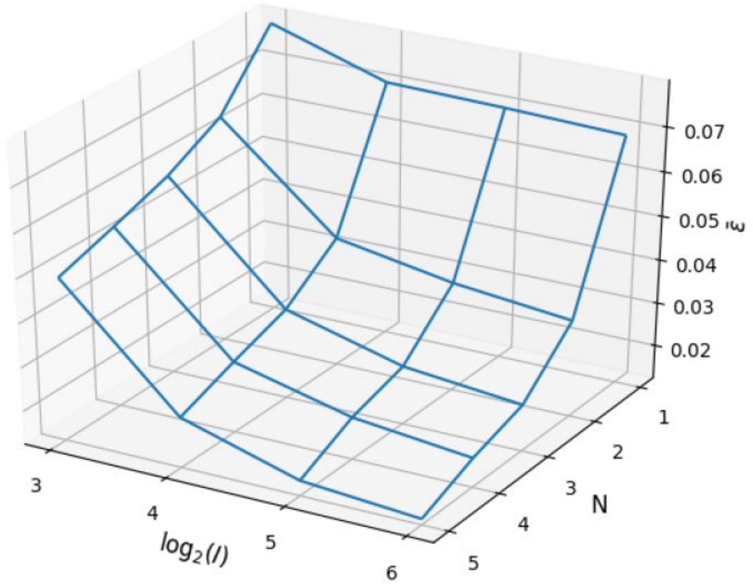
Recording

- Spherical loudspeaker layout: 25 *Genelec 8040*
- Plane-wave ambisonic encoding of gaussian noise sources
  - Ambisonic orders $N \in [1, 5]$
  - Number of sources $I = [8, 16, 32, 64]$

Results

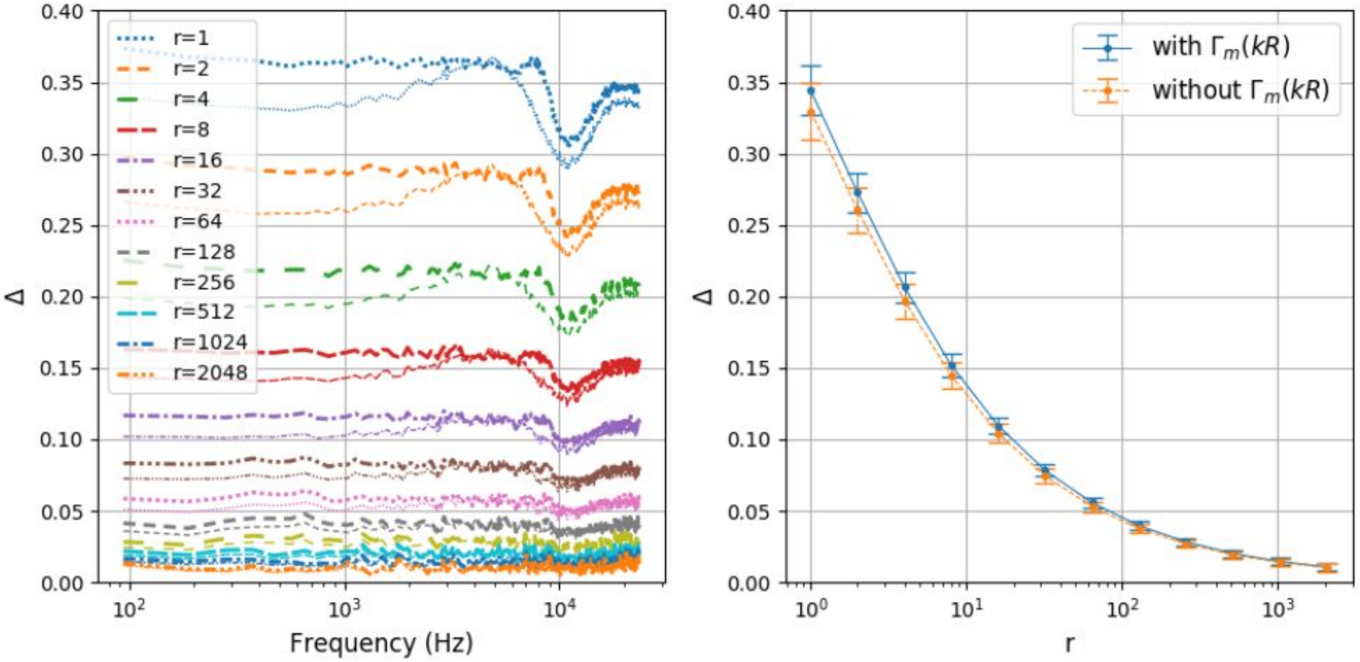
Spatial domain - simulated vs. recorded





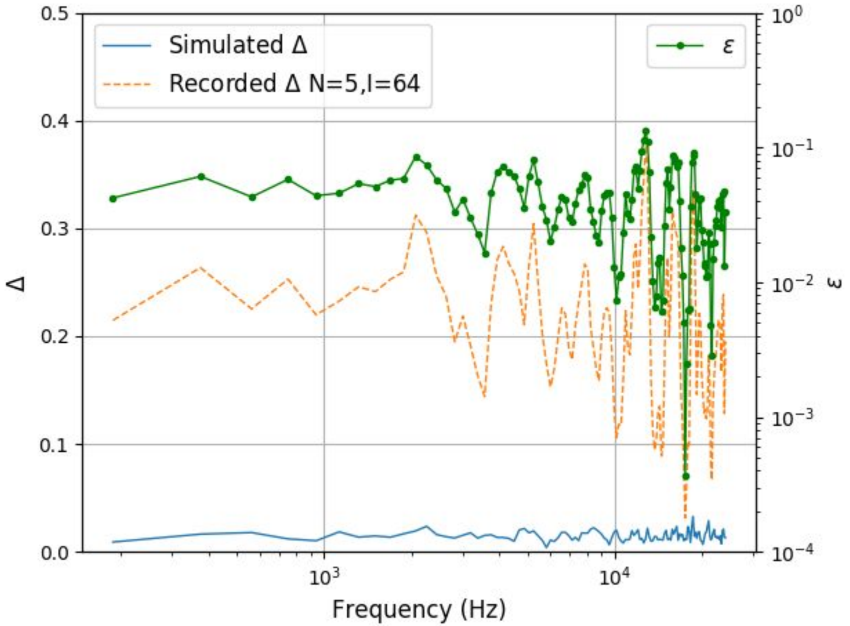
Results

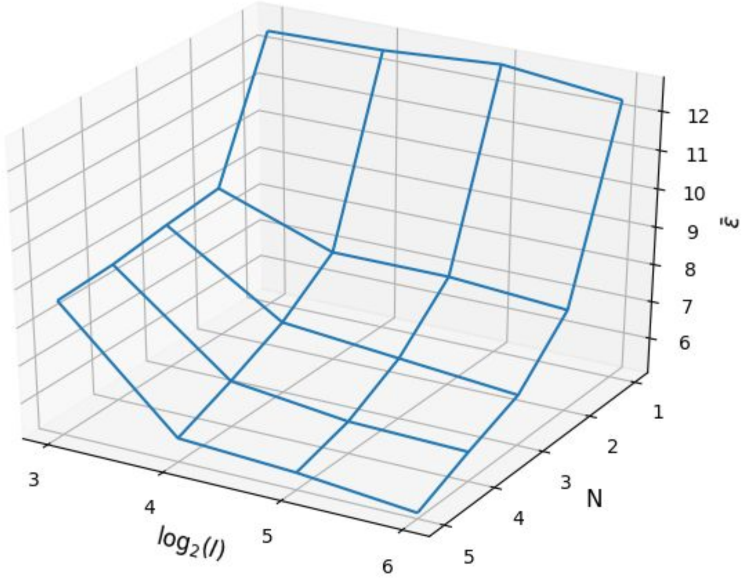
Ambisonic domain - simulated



Results

Ambisonic domain - simulated vs. recorded

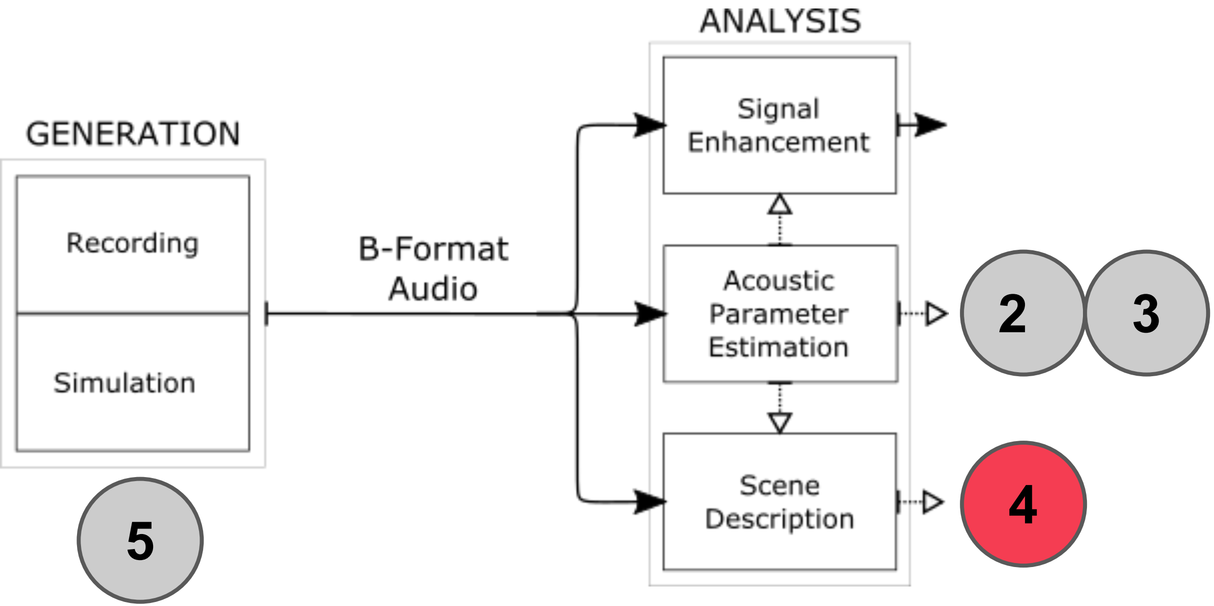




Summary of contributions

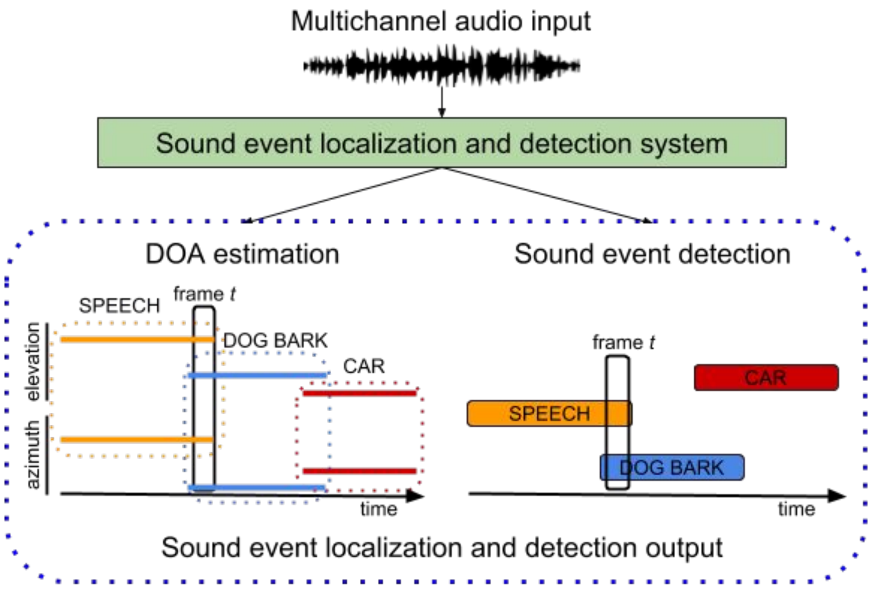
- **Novel diffuse field characterization**
  - Most common ambisonic microphone
  - Two different metrics / domains
- **Quantify** high impact of diffuseness time-averaging window
- Study feasibility of **diffuse sound reproduction** in ambisonics

4. Sound Event Localization and Detection


GENERATION
Recording
Simulation
5
B-Format
Audio
ANALYSIS
Signal
Enhancement
Acoustic
Parameter
Estimation
Scene
Description
2
3
4

Introduction

Sound Event Localization and Detection (SELD)



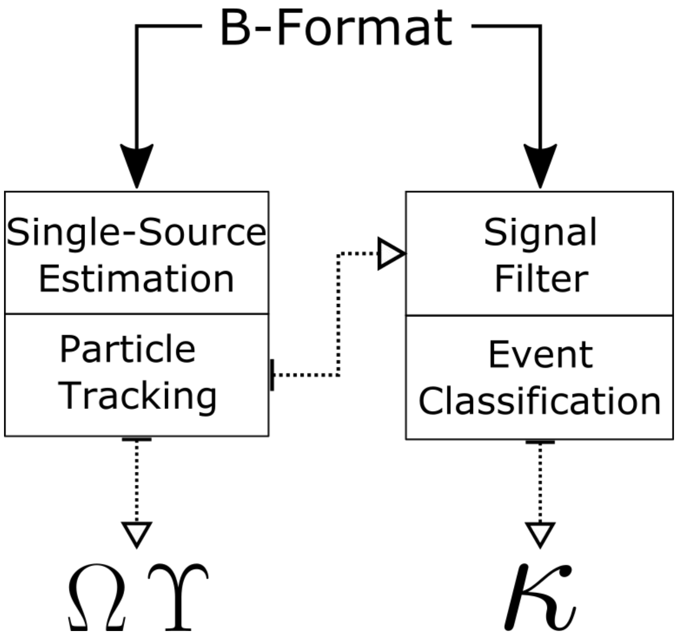
[12] DCASE Workshop, Task 3. http://dcase.community/challenge2020/task-sound-event-localization-and-detection

Introduction

Proposed method

- ***PAPAFIL***: PArametric PArticle Filter
  - Focus on low-complexity
  - Uses *traditional* machine learning
  - *First Localization, then Classification*
- Presented to DCASE Challenge 2020

System Description



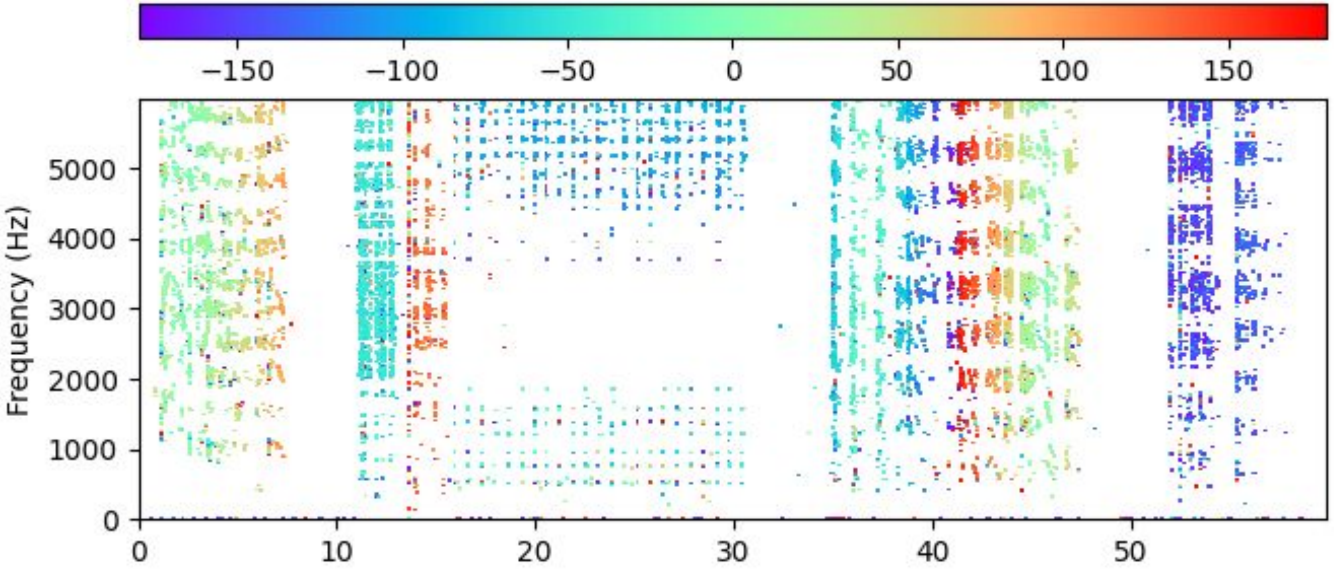
Ω Spatial location

Υ Temporal activity

κ Sound class

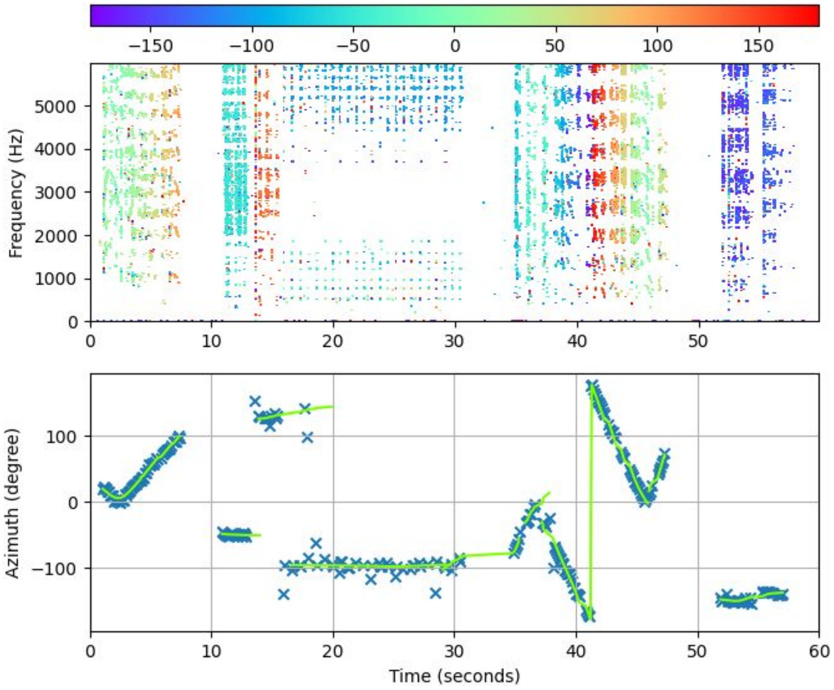
System Description

1. SINGLE SOURCE ESTIMATION



System Description

2. PARTICLE TRACKING



System Description

3. SIGNAL FILTER

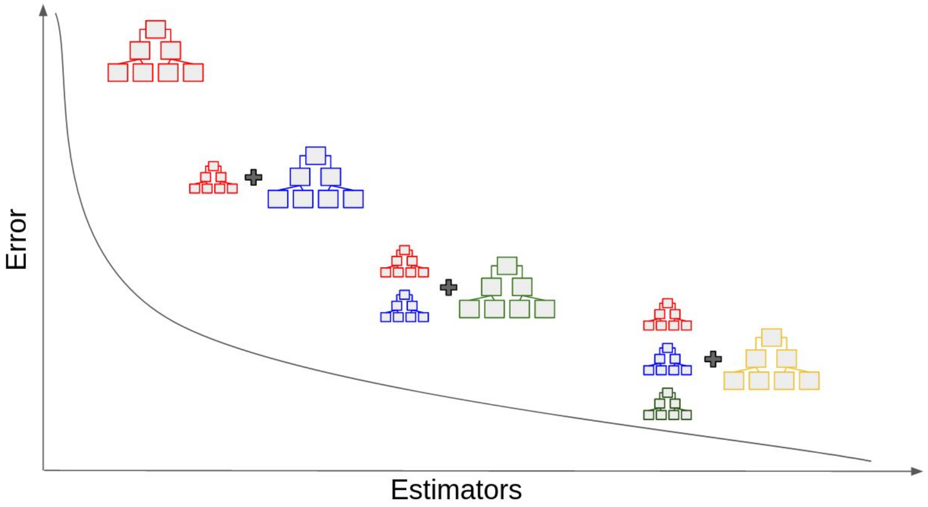
Steer a first-order virtual cardioid towards the estimated DOAs:

- Improve SNR of target sound event
- Mono downmix
- Sound event estimation (with temporal segmentation)

System Description

4. EVENT CLASSIFICATION

Gradient Boosting Machine (GBM) [13]



[13] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," Annals of statistics, pp. 1189–1232, 2001.

System Description

4. EVENT CLASSIFICATION

Sound features obtained from *Essentia* [14]*:*

- Frame-based: first order statistics
- Whole event

Type	Features	Number
Low-level	Mel bands	24
	MFCC	13
	Spectral Features	26
SFX	Duration	2
	Harmonic	4
	Sound envelope	11
	Pitch envelope	4

[14] D. Bogdanov, N. Wack, E. Gomez Gutierrez, S. Gulati, H. Boyer, O. Mayor, G. Roma Trepat, J. Salamon, J. R. Zapata Gonzalez, X. Serra, et al., "Essentia: An audio analysis library for music information retrieval," in 14th Conference of the International Society for Music Information Retrieval (ISMIR); p. 493-8., Curitiba, Brazil, November 2013.

Experiments

DATASET

TAU-NIGENS Spatial Sound Events 2020 - FOA [15]

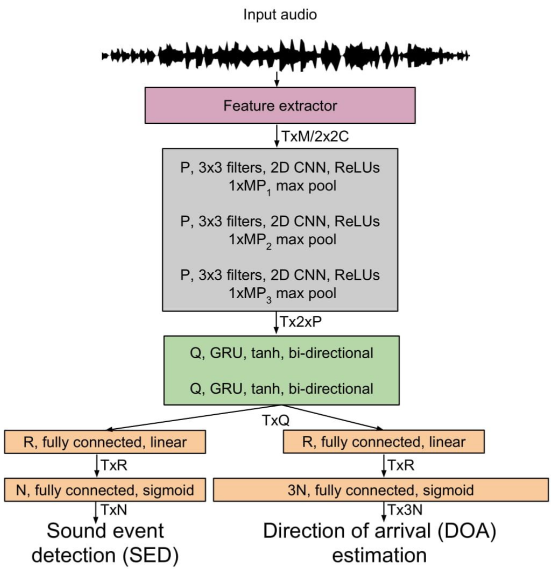
[15] I. Trowitzsch, J. Taghia, Y. Kashef, and K. Obermayer, "The nigens general sound events database,"

Experiments

BASELINE

Based on *SELD-net* [16]*:*

- CRNN architecture
- Joint localization and classification
- Improved after DCASE2019 Challenge



[16] S. Adavanne, A. Politis, J. Nikunen, and T. Virtanen, “Sound event localization and detection of overlapping sources using convolutional recurrent neural networks,” IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 1, pp. 34–48, March 2018

Experiments

PAPAFIL 1

Oracle space-time information (parsing annotation files)

PAPAFIL 2

True parametric particle filtering method

Results

6-fold cross-validation (development set)

Method	ER_{20}	F_{20}	LE_{CD}	LR_{CD}	SELD
BASELINE	0.72	37.4 %	22.8°	**60.7** %	0.47
PAPAFIL1	0.60	49.8 %	**13.4**°	54.4 %	0.41
PAPAFIL2	**0.57**	**54.0** %	13.8°	59.7 %	**0.38**
PAPAFIL1-O	0.37	67.0 %	2.0°	68.6 %	0.26
PAPAFIL2-O	0.32	79.6 %	8.5°	82.4%	0.19

Results

6-fold cross-validation (development set)

Method	ER_{20}	F_{20}	LE_{CD}	LR_{CD}	SELD
BASELINE	0.72	37.4 %	22.8°	**60.7** %	0.47
PAPAFIL1	0.60	49.8 %	**13.4**°	54.4 %	0.41
PAPAFIL2	**0.57**	**54.0** %	13.8°	59.7 %	**0.38**
PAPAFIL1-O	0.37	67.0 %	2.0°	68.6 %	0.26
PAPAFIL2-O	0.32	79.6 %	8.5°	82.4%	0.19

Results

Method	ER_{20}	F_{20}	LE_{CD}	LR_{CD}	SELD
BASELINE-DEV	0.72	37.4 %	22.8°	60.7 %	0.47
BASELINE-EVAL	0.70	39.5 %	23.2°	62.1 %	0.45
PAPAFIL2-DEV	0.57	54.0 %	13.8°	59.7 %	0.38
PAPAFIL2-EVAL	**0.51**	**60.1** %	**12.4**°	**65.1** %	**0.33**

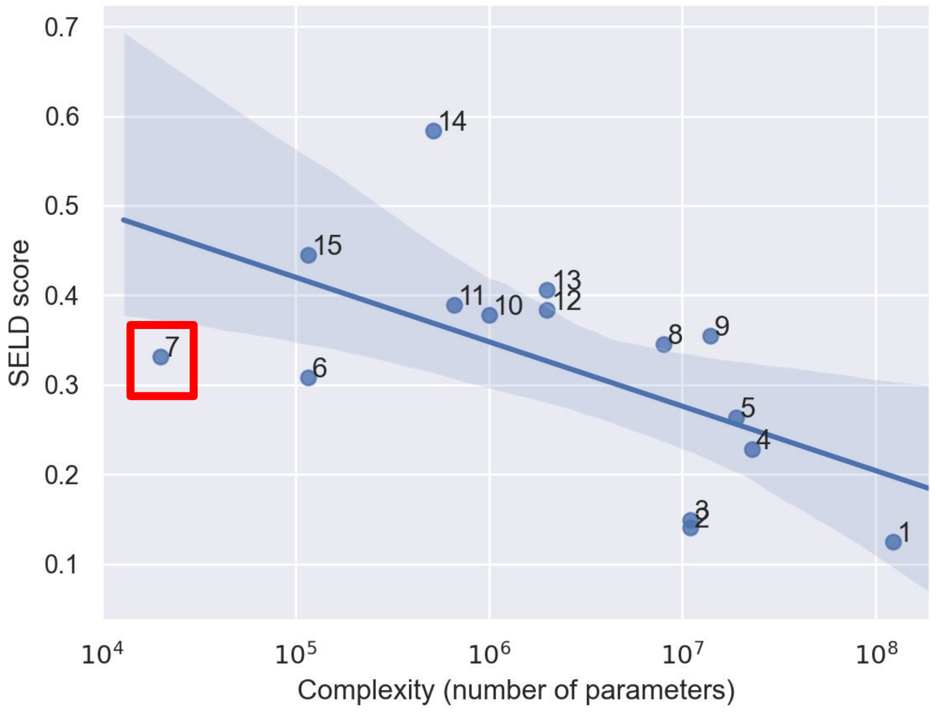
Results

Method	ER_{20}	F_{20}	LE_{CD}	LR_{CD}	SELD
BASELINE-DEV	0.72	37.4 %	22.8°	60.7 %	0.47
BASELINE-EVAL	0.70	39.5 %	23.2°	62.1 %	0.45
PAPAFIL2-DEV	0.57	54.0 %	13.8°	59.7 %	0.38
PAPAFIL2-EVAL	**0.51**	**60.1** %	**12.4**°	**65.1** %	**0.33**

Results

	Submission Information		Evaluation dataset				
Rank	Submission name	Technical Report	Best official system rank	Error Rate (20°)	F-score (20°)	Localization error (°)	Localization recall
1	Du_USTC_task3_4		1	0.20	84.9 %	6.0	88.5 %
2	Nguyen_NTU_task3_2		4	0.23	82.0 %	9.3	90.0 %
3	Shimada_SONY_task3_4		5	0.25	83.2 %	7.0	86.2 %
4	Cao_Surrey_task3_4		11	0.36	71.2 %	13.3	81.1 %
5	Park_ETRI_task3_4		13	0.43	65.2 %	16.8	81.9 %
6	Phan_QMUL_task3_3		15	0.49	61.7 %	15.2	72.4 %
7	PerezLopez_UPF_task3_2		16	0.51	60.1 %	12.4	65.1 %
8	Sampathkumar_TUC_task3_1		20	0.53	56.6 %	14.8	66.5 %
9	Patel_MST_task3_4		22	0.55	55.5 %	14.4	65.5 %
10	Ronchini_UPF_task3_2		28	0.58	50.8 %	16.9	65.5 %
11	Naranjo-Alcazar_VFY_task3_2		30	0.61	49.1 %	19.5	67.1 %
12	Song_LGE_task3_3		31	0.57	50.4 %	20.0	64.3 %
13	Tian_PKU_task3_1		36	0.64	47.6 %	24.5	67.5 %
14	Singla_SRIB_task3_2		38	0.88	18.0 %	53.4	66.2 %
15	DCASE2020_MIC_baseline		39	0.69	41.3 %	23.1	62.4 %

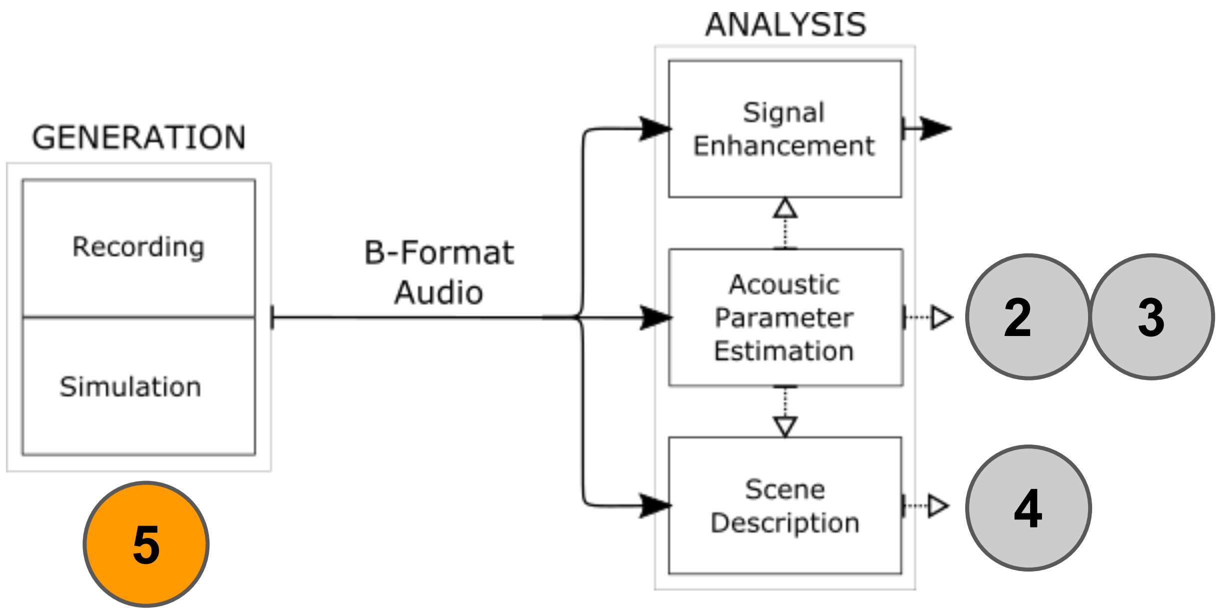
Results


0.7
0.6
0.5
0.4
0.3
0.2
0.1
SELD score
14
15
11
10
13
12
8
9
7
6
5
4
3
2
1
10^4
10^5
10^6
10^7
10^8
Complexity (number of parameters)

Summary of contributions

- **Novel** low-complexity method for SELD
  - Localization with parametric particle filtering
  - Classification by GBM
- **Very low complexity**
- Results **outperform baseline**

5. Dataset Generation and Storage


GENERATION
Recording
Simulation
5
B-Format
Audio
ANALYSIS
Signal
Enhancement
Acoustic
Parameter
Estimation
Scene
Description
2
3
4

Introduction

Support generation of parametrizable ambisonic datasets

- Synthetic and recorded
- Stress on Room Impulse Responses
- Using python

MASP

MASP: **M**ultichannel **A**coustic **S**ignal **P**rocessing Library

- Port of several Matlab libraries by A. Politis
- Acoustic simulation and microphone arrays
- Special focus on spherical configurations

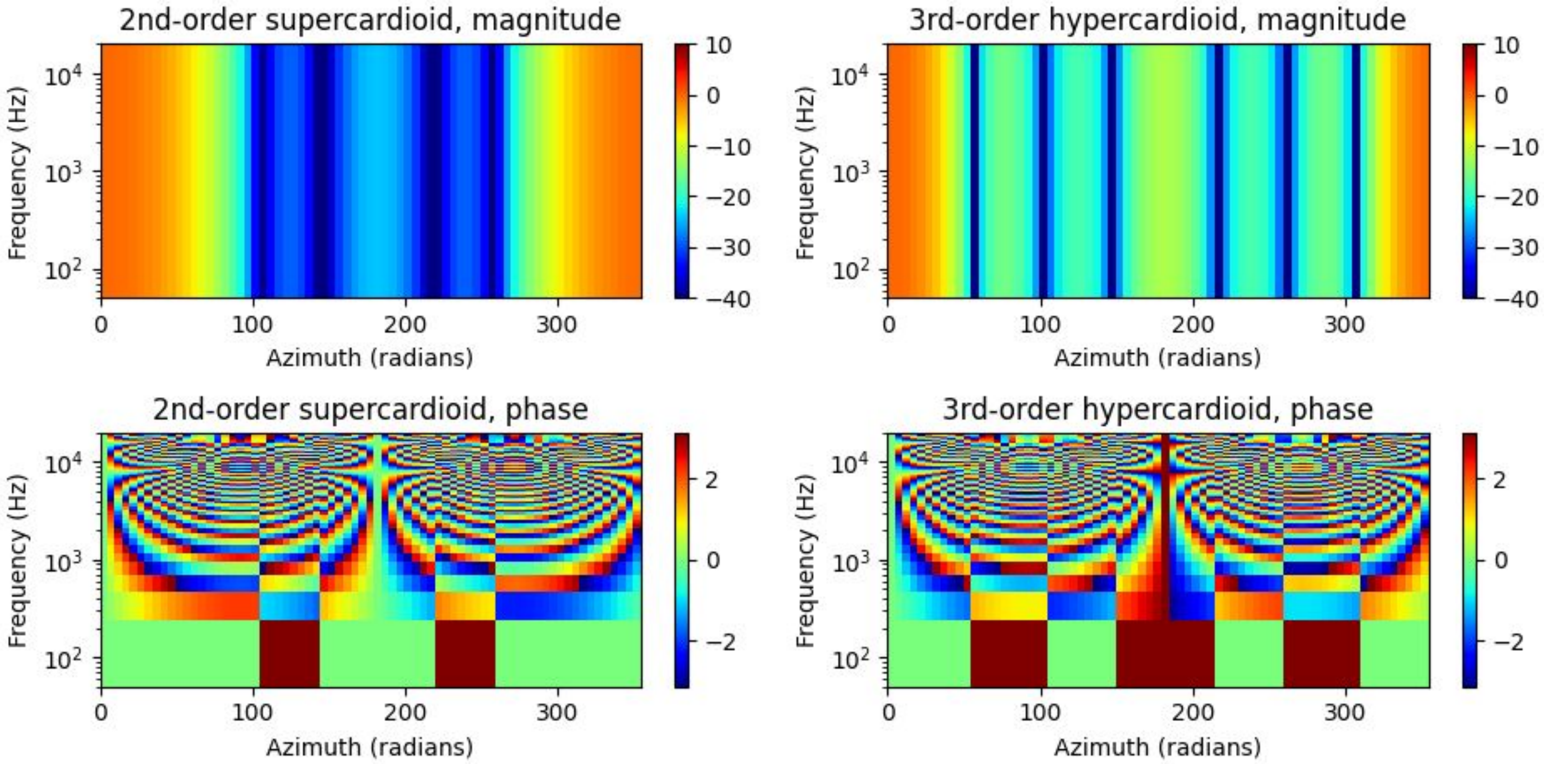
MASP

Components:

- **Array Response Simulator** Simulation of spherical microphones
- **Shoebox Room Model** Image Source Method [17]:
- **Spherical Array Processing** Analysis and transformation tools
- **Spherical Harmonic Transform** Mathematical convenience tools

[17] Allen, J. B., & Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, *65*(4), 943-950.

MASP


2nd-order supercardioid, magnitude
3rd-order hypercardioid, magnitude
2nd-order supercardioid, phase
3rd-order hypercardioid, phase
Frequency (Hz)
Azimuth (radians)
10^2
10^3
10^4
0
100
200
300
10
0
−10
−20
−30
−40
2
0
−2

SOFA

SOFA: Spatially **O**riented **F**ormat for **A**coustics [18]

- File format for storage of HRTFs and other IRs
- Widely used, standard (AES-69)
- Solves the interoperatibility problem between datasets

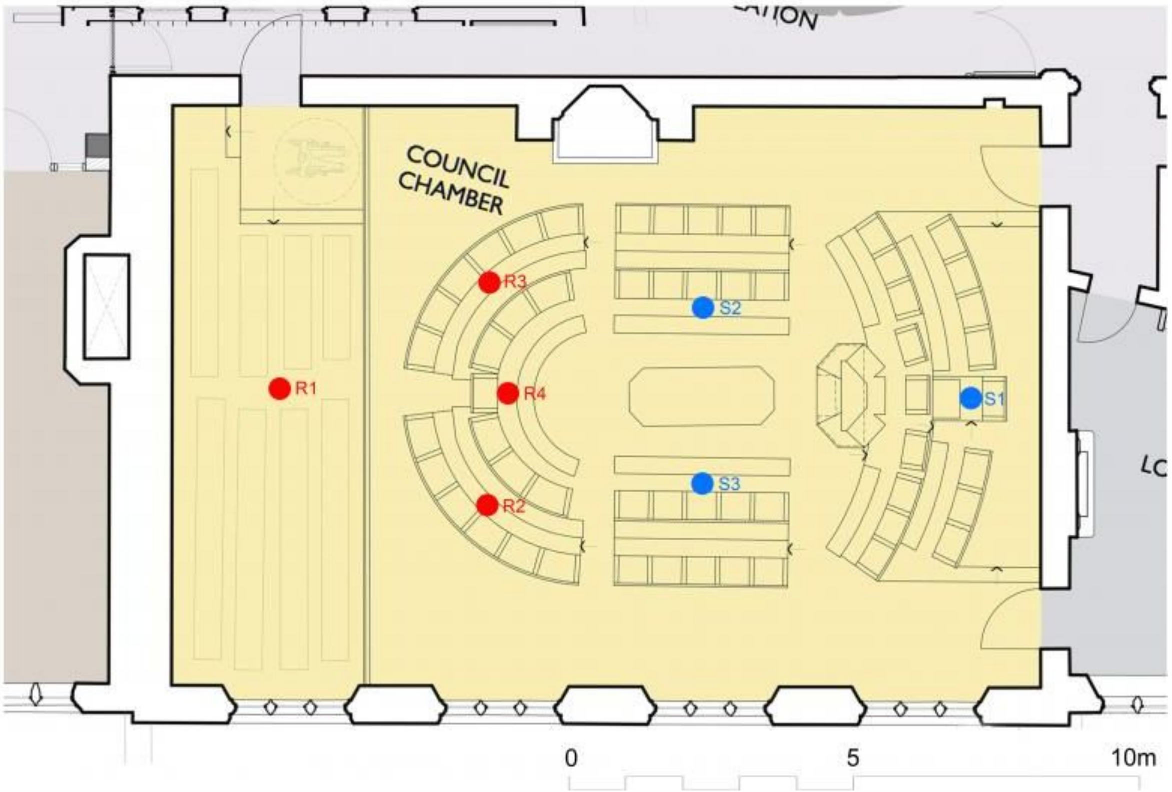
[18] Majdak, P., Iwaya, Y., Carpentier, T., Nicol, R., Parmentier, M., Roginska, A., Suzuki, Y., Watanabe, K., Wierstorf, H., Ziegelwanger, H., et al. (2013). *Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions*. In Audio Engineering Society Convention 134. Audio Engineering Society.

SOFA

Ambisonics Directional Room Impulse Response for SOFA

- Proposed extension of SOFA for Ambisonic RIRs
- Support multi-perspective measurements
- Standard **integration currently discussed** in the AES-69 steering group

SOFA



SOFA

pysofaconventions

- Python implementation of SOFA 1.0
- Adaptation from C++ library
- Used in several external projects
    - Facebook / Chalmers University
    - Several fixes / pull requests

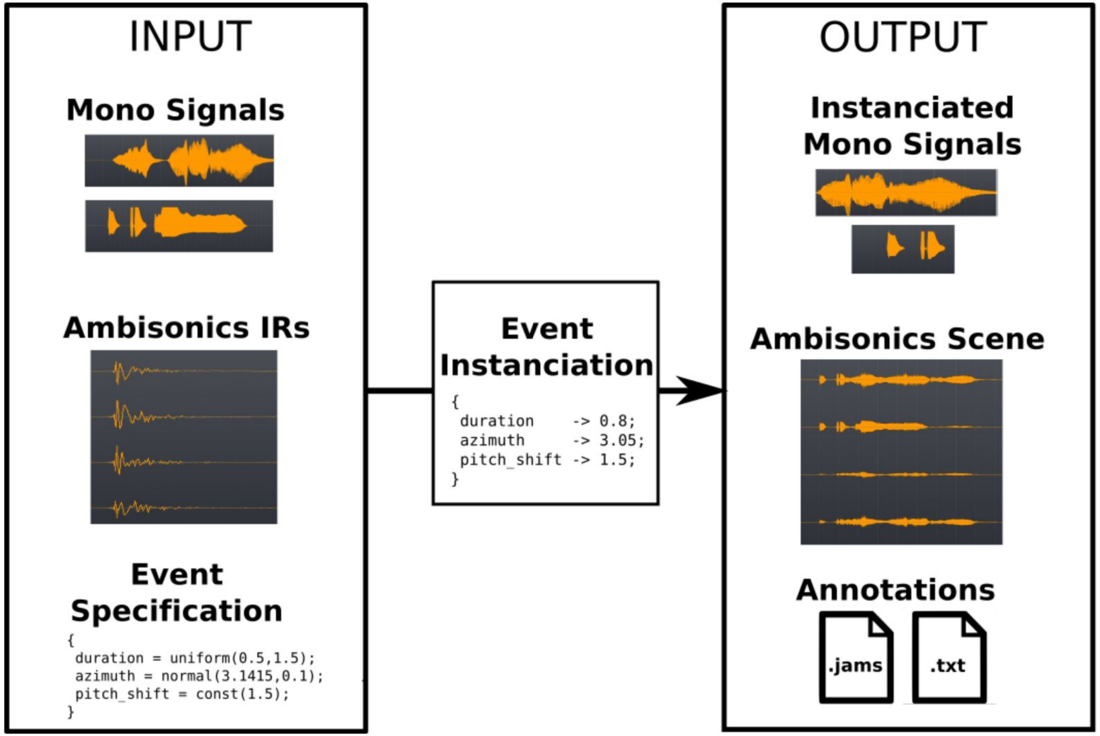
Ambiscaper

Ambiscaper

- Based on J. Salomon's *Scaper* [19]
- Tool for automatic generation of ambisonic datasets
- Statistical rule-based generation
- Synthetic / recorded IRs

[19] Salamon, J., MacConnell, D., Cartwright, M., Li, P., and Bello, J. P. (2017). *Scaper: A library for soundscape synthesis and augmentation*. In Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pages 344–348. IEEE.

Ambiscaper


INPUT
Mono Signals
Ambisonics IRs
Event Specification
{
duration = uniform(0.5,1.5);
azimuth = normal(3.1415,0.1);
pitch_shift = const(1.5);
}
Event Instanciation
{
duration -> 0.8;
azimuth -> 3.05;
pitch_shift -> 1.5;
}
OUTPUT
Instanciated Mono Signals
Ambisonics Scene
Annotations
.jams
.txt

Ambiscaper

- Low adoption
  - Developed in 2018 (before LOCATA, DCASE)
  - Much more datasets available nowadays
  - Maintenance cost
- Dataset reproducibility premises are still valid

III - CONCLUSIONS

6. Conclusions

Summary of Contributions

Contribution to different components of an ambisonics analysis and generation framework

- Making use of parametric signal representations
- Focusing on applied problems and research reproducibility

Summary of Contributions

Main objectives:

1. To develop methods for the **characterisation of acoustic parameters** from recordings originated from ambisonic microphones.
2. To propose methodologies for **sound event localization and detection** in ambisonic domain which are grounded on spatial parametric analysis.
3. To contribute to the **generation and storage** of ambisonic sound scenes, for their usage in controlled experimental environments.

Summary of Contributions

Main contributions:

1.) Blind reverberation time estimation

- First attempt in literature to address the problem
- Performance comparable to state-of-the-art
- Improved result consistency
- **IEEE MMSP 2020 Best Paper Runner-Up Award**

Summary of Contributions

Main contributions:

2.) Coherence estimation

- Characterization of A-Format microphone under spherical noise
- Rendering capabilities of spherical arrays using ambisonics

Summary of Contributions

Main contributions:

3.) Sound Event Localization and Detection

- Novel approach with parametric particle filter
- Very low complexity
- **Q2 results in DCASE 2020 Challenge (7/15)**

Summary of Contributions

Main contributions:

4.) Data generation and management

- Acoustic simulation library
- SOFA library and **standard proposal**
- Dataset generation software

Future work

Technology transfer: convert prototypes into products

List of Contributions

- 1 peer-reviewed journal article
  - **"Analysis of spherical isotropic noise fields with an A-Format tetrahedral microphone"**. A. Pérez-López and N.Stefanakis. The Journal of the Acoustical Society of America 146.4 (2019): EL329-EL334.

List of Contributions

- 4 peer-reviewed conference articles
  - **"Blind reverberation time estimation from ambisonic recordings"**. A. Pérez-López, A. Politis and E. Gómez. IEEE MMSP 2020.
  - **"PAPAFIL: a low complexity sound event localization and detection method with parametric particle filtering and gradient boosting"**. A. Pérez-López and R. Ibañez-Usach. Submitted to Detection and Classification of Acoustic Scenes and Events 2020 Workshop.
  - **"A hybrid parametric-deep learning approach for sound event localization and detection"**. A. Pérez-López, E. Fonseca and X. Serra. In Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop.
  - **"Ambiscaper: A tool for automatic generation and annotation of reverberant ambisonics sound scenes"**. A. Pérez-López. In Proceedings of the 16th International Workshop on Acoustic Signal Enhancement, (IWAENC). IEEE, 2018.

List of Contributions

- 3 conference engineering briefs (abstract-reviewed)
  - **"Ambisonics directional room impulse response as a new convention of the spatially oriented format for acoustics".** A. Pérez-López and J. De Muynke. In Proceedings of the 144th Audio Engineering Society Convention. Audio Engineering Society 2018.
  - **"pysofaconventions, a Python API for SOFA".** A. Pérez-López. In Proceedings of the 148th Audio Engineering Society Convention. Audio Engineering Society, 2020.
  - **"A Python library for multichannel acoustic signal processing".** A. Pérez-López and A. Politis. In Proceedings of the 148th Audio Engineering Society Convention. Audio Engineering Society, 2020.

List of Contributions

- 1 peer-reviewed conference article (as supervisor)
  - **"Sound event localization and detection based on CRNN using dense rectangular filters and channel rotation data augmentation".** F. Ronchini, D. Arteaga and A. Pérez-López. Submitted to Detection and Classification of Acoustic Scenes and Events 2020 Workshop (DCASE2020).

List of Contributions

Software contributions

- 3 open-source libraries
  - masp https://github.com/andresperezlopez/masp
  - pysofaconventions https://github.com/andresperezlopez/pysofaconventions
  - Ambiscaper https://github.com/andresperezlopez/ambiscaper
- 3 open implementations of papers
  - RT60 estimation https://github.com/andresperezlopez/ambisonic_rt_estimation
  - DCASE2020 https://github.com/andresperezlopez/DCASE2020
  - DCASE2019 https://github.com/andresperezlopez/DCASE2019_task3

Thanks