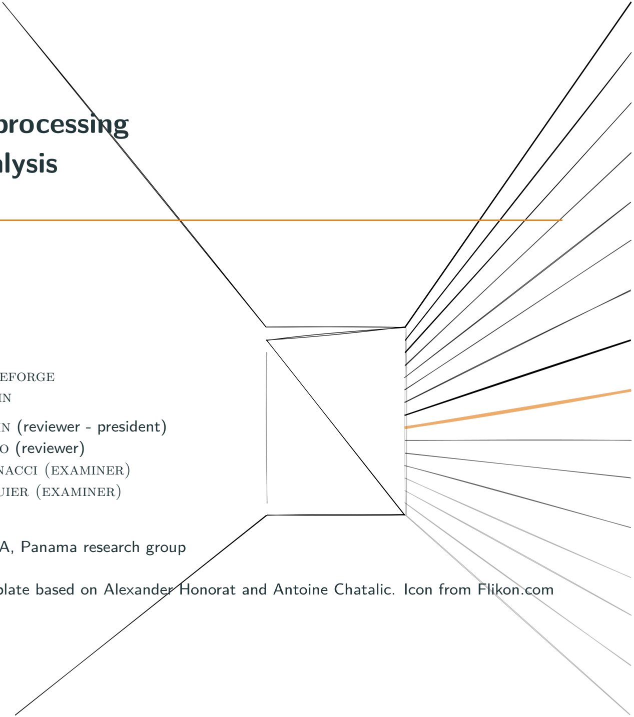
Echo-aware signal processing for audio scene analysis



Diego Di Carlo

November 29, 2020

PhD supervisors:	Antoine Deleforge Nancy Bertin
Jury members:	Laurent Girin (reviewer - president) Simon Doclo (reviewer) Fabio Antonacci (examiner) Renaud Seguier (examiner)

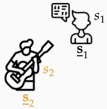
Université de Rennes 1, IRISA/INRIA, Panama research group

Colophon: Metropolis Beamer Template based on Alexander Honorat and Antoine Chatalic. Icon from Flikon.com

Introduction

Echo-aware signal processing for audio scene analysis

Current Scenario

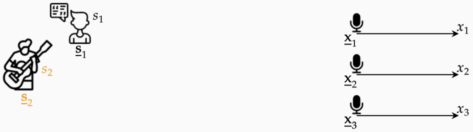


Sound

- produced by sources

Echo-aware signal processing for audio scene analysis

Current Scenario

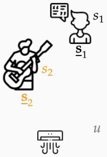

s_1
$\underline{\mathbf{s}}_1$
s_2
$\underline{\mathbf{s}}_2$
$\underline{\mathbf{x}}_1$
x_1
$\underline{\mathbf{x}}_2$
x_2
$\underline{\mathbf{x}}_3$
x_3

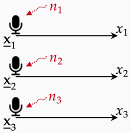
Sound

- produced by sources
- recorded by (array of) microphones

Echo-aware signal processing for audio scene analysis

Current Scenario



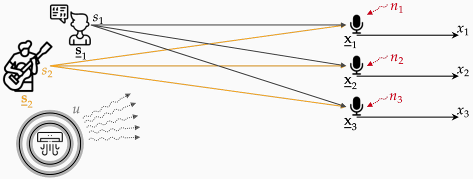

n_1
$\underline{\mathbf{x}}_1$
x_1
n_2
$\underline{\mathbf{x}}_2$
x_2
n_3
$\underline{\mathbf{x}}_3$
x_3

Sound

- produced by sources
- recorded by (array of) microphones
- corrupted by noise

Echo-aware signal processing for audio scene analysis

Current Scenario

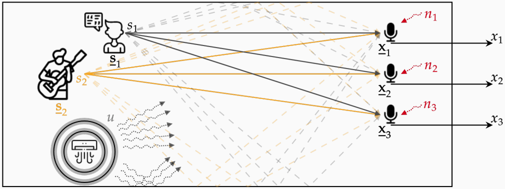

s_1
$\underline{\mathbf{s}}_1$
s_2
$\underline{\mathbf{s}}_2$
u
n_1
n_2
n_3
$\underline{\mathbf{x}}_1$
$\underline{\mathbf{x}}_2$
$\underline{\mathbf{x}}_3$
x_1
x_2
x_3

Sound

- produced by sources
- recorded by (array of) microphones
- corrupted by noise
- propagates in the space

Echo-aware signal processing for audio scene analysis

Current Scenario

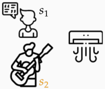

s_1
$\underline{\mathbf{s}}_1$
s_2
$\underline{\mathbf{s}}_2$
u
n_1
n_2
n_3
$\underline{\mathbf{x}}_1$
$\underline{\mathbf{x}}_2$
$\underline{\mathbf{x}}_3$
x_1
x_2
x_3

Sound

- produced by sources
- recorded by (array of) microphones
- corrupted by noise
- propagates in the space
- interacts with the room
  ↪ reverberation

Echo-aware signal processing for audio scene analysis

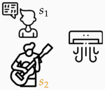
Semantic information



on nature and content

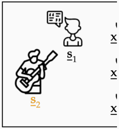
Echo-aware signal processing for audio scene analysis

Semantic information



on nature and content

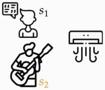
Spatial information



on position and geometry

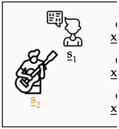
Echo-aware signal processing for audio scene analysis

Semantic information



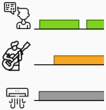
on nature and content

Spatial information



on position and geometry

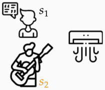
Temporal information



on events activity

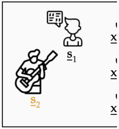
Echo-aware signal processing for audio scene analysis

Semantic information



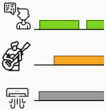
on nature and content

Spatial information



on position and geometry

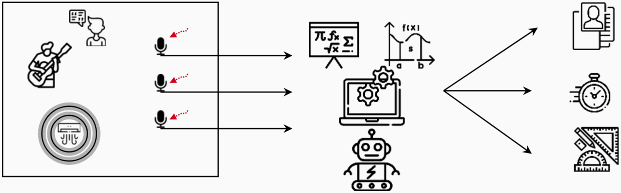
Temporal information



on events activity

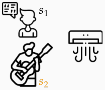
Audio Scene Analysis

Extraction and organization of all the information in the sound



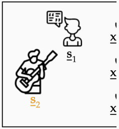
Echo-aware signal processing for audio scene analysis

Semantic information



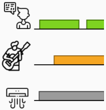
on nature and content

Spatial information



on position and geometry

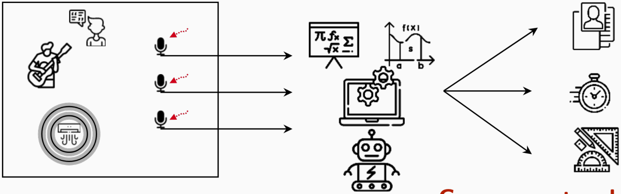
Temporal information



on events activity

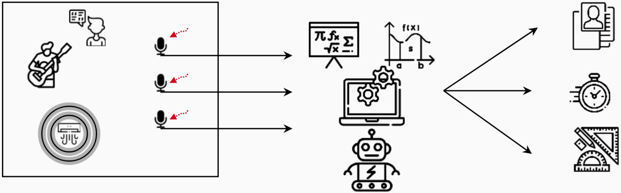
Audio Scene Analysis

Extraction and organization of all the information in the sound

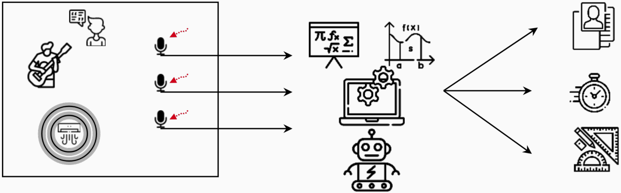


Can computer do it?

Echo-aware signal processing for audio scene analysis



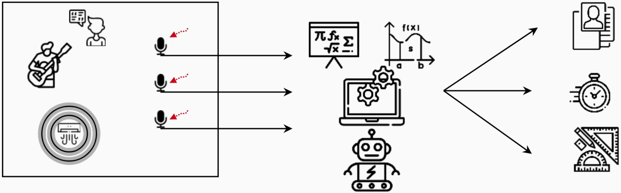
Echo-aware signal processing for audio scene analysis



Signal Processing

Mathematical models, frameworks and tools to tackle and solve such problems

Echo-aware signal processing for audio scene analysis



Signal Processing

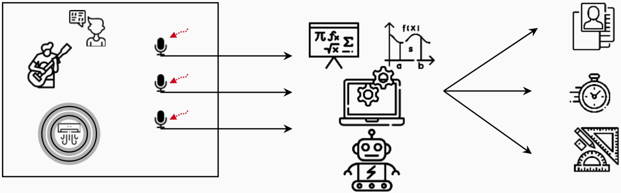
Mathematical models, frameworks and tools to tackle and solve such problems

Some (inverse) problems

- Speaker Identification
- Sound Source Separation (SSS)
- Speech Enhancement (SE)
- Automatic Speech Recognition (ASR)
- Sound Source Localization (SSL)

- Voice Activity Detection
- Diarization
- RT$_{60}$ estimation
- Acoustic Channel Estimation
- Wall Absorption Estimation

Echo-aware signal processing for audio scene analysis



Signal Processing

Mathematical models, frameworks and tools to tackle and solve such problems

Some (inverse) problems

- Speaker Identification
- Sound Source Separation (SSS)
- Speech Enhancement (SE)
- Automatic Speech Recognition (ASR)
- Sound Source Localization (SSL)

- Voice Activity Detection
- Diarization
- RT_{60} estimation
- Acoustic Channel Estimation
- Wall Absorption Estimation

Echo-aware signal processing for audio scene analysis

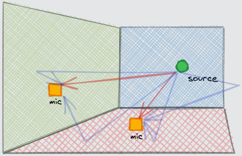
Sound interacts with indoor environment:

it is reflected,
specularly and diffusely
+ it is absorbed,
+ it is transmitted,
+ and other.

= all reverberation

Acoustic Echoes

- Elements of reverberation
- Specular reflection standing out for time and strength
- Repetition of a sound but after
  - same content
  - delay ⇔ distance



4/8

Everyday examples:

Outline and contributions

Thesis title

Audio Scene Analysis

↓

context and problems

Outline and contributions

Thesis title

Audio Scene Analysis	Signal Processing
↓	↓
context and problems	models and frameworks

Outline and contributions

Thesis title

Audio Scene Analysis
↓
context and problems

Signal Processing
↓
models and frameworks

Echo-aware
↓
better processing

Outline and contributions

Thesis title

Audio Scene Analysis	Signal Processing	Echo-aware
↓	↓	↓
context and problems	models and frameworks	better processing

Thesis content:

How to estimate them?

- Analytical method
- Learning-based method

no parameter tuning

no full sound modeling

How to use them?

- Source Localization
- Source Separation
- Speech Enhancement
- Room Geometry Estimation

Where to find them?

Echo-aware database for
estimation and application

Problem Statement

Signal model

Sound propagation process ⇔ Source → Filter → Receiver model

source signal

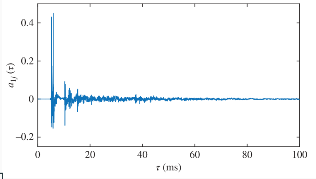
microphone signal ← $\tilde{x}_i(t) = (\tilde{h}_i * \tilde{s})(t) + +\tilde{n}(t)$ → noise term

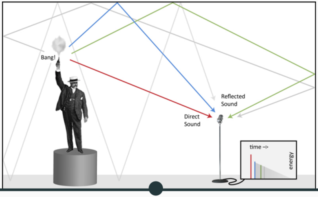
continuous-time convolution

⚠ continuous time

Room Impulse Response (RIR)

- linear filtering effect of the sound
- acoustic response of a room to a (prefect) impulsive sound
- depends on spatial properties (room geometry, mic/src position)

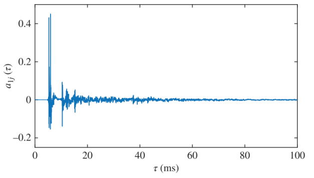




Echoes in the RIR

RIR model

$$\tilde{h}_i(t) = \tilde{h}_i^{\mathsf{d}}(t) + \tilde{h}_i^{\mathsf{e}}(t) + \tilde{h}_i^{\mathsf{lrev}}(t) + \varepsilon_i(t)$$



Echoes can be modeled as sum of Dirac's delta

$$\tilde{h}_i^{\mathsf{echoes}} = \tilde{h}_i^{\mathsf{d}}(t) + \tilde{h}_i^{\mathsf{e}}(t) \approx \sum_{r=0}^{R} \alpha_i^{(r)} \delta(t - \tau_i^{(r)})$$

Goal: estimated the $\tau_{i_{i,r}}$

Challenges:

- α distortion (even if we know it $\implies$ labeling)
- $\alpha \to \alpha(t)$ (sum of diracs → sum of filters)
- h_l reverberation is included in the noise term
- depends on the scene geometry (room, source and mic position)

References i