Echo-aware signal processing for audio scene analysis

Diego Di Carlo

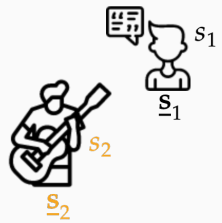
November 27, 2020

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

Introduction

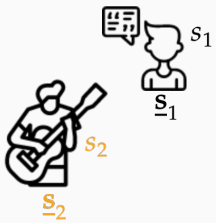
Scenario

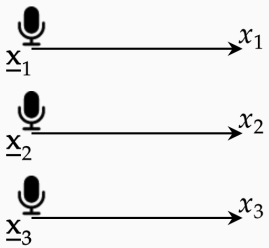


Sound

- produced by sources

Scenario

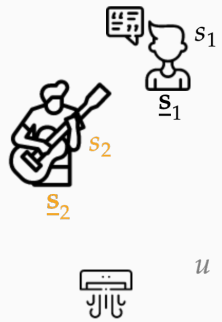


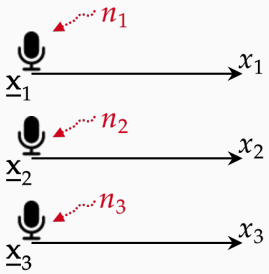


Sound

- produced by sources
- recorded by microphones

Scenario

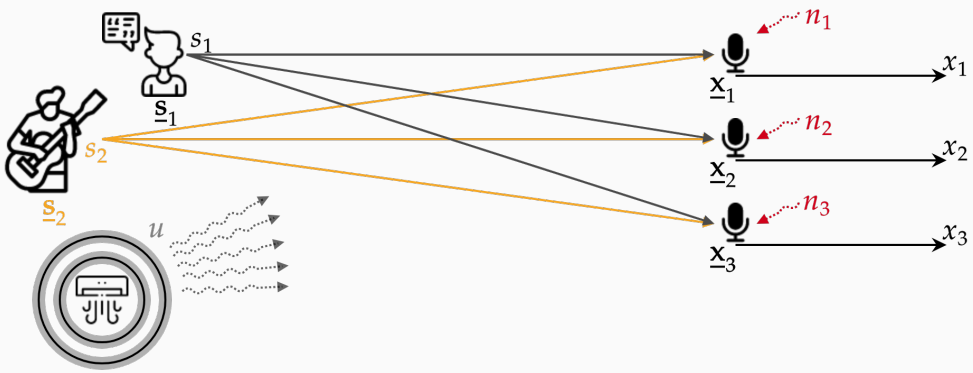




Sound

- produced by sources
- recorded by microphones
- corrupted by noise

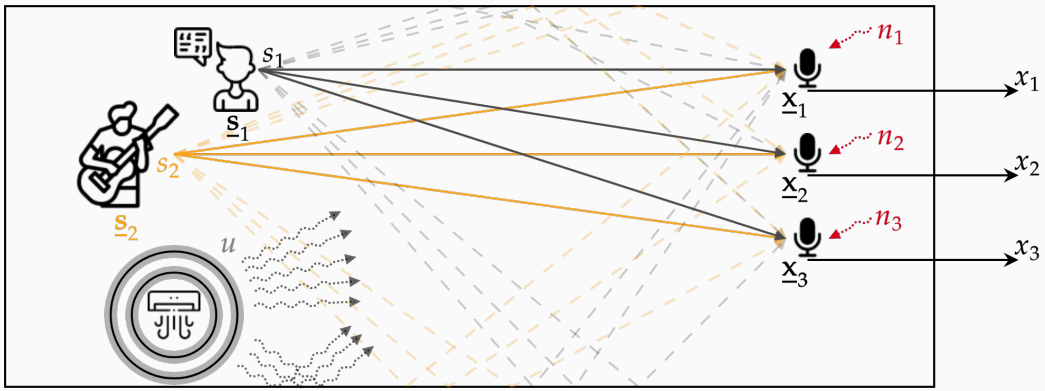
Scenario



Sound

- produced by sources
- recorded by microphones
- corrupted by noise
- propagates in the space

Scenario



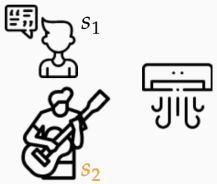
Sound

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

1

Echo-aware signal processing for audio scene analysis

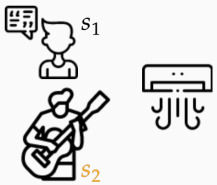
Semantic information



on nature and content

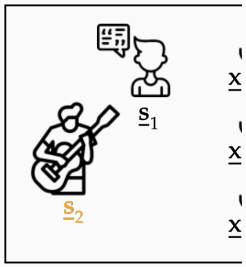
Echo-aware signal processing for audio scene analysis

Semantic information



on nature and content

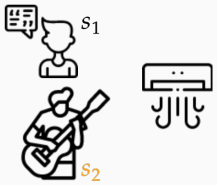
Spatial information



on position and geometry

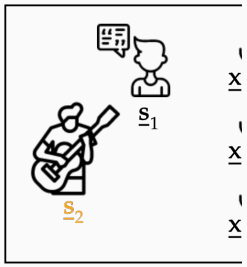
Echo-aware signal processing for audio scene analysis

Semantic information



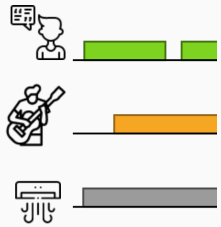
on nature and content

Spatial information



on position and geometry

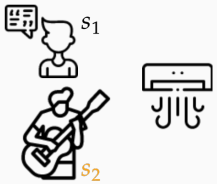
Temporal information



on events activity

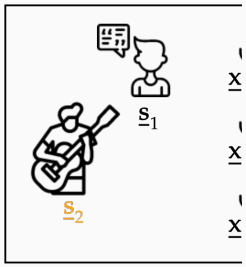
Echo-aware signal processing for audio scene analysis

Semantic information



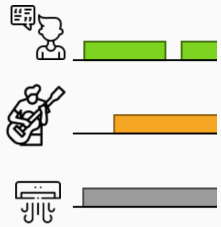
on nature and content

Spatial information



on position and geometry

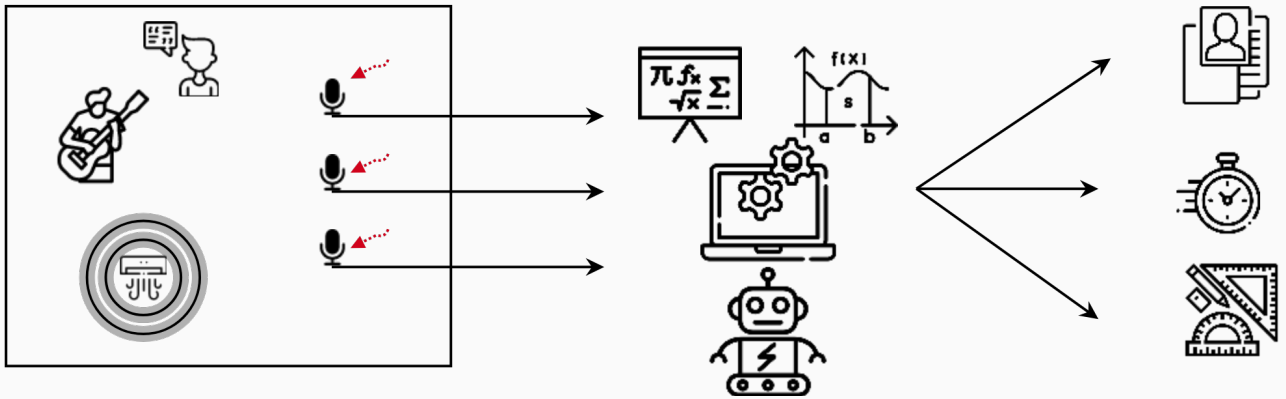
Temporal information



on events activity

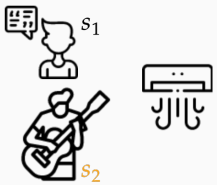
Audio Scene Analysis

Extraction and organization of all the information in the sound



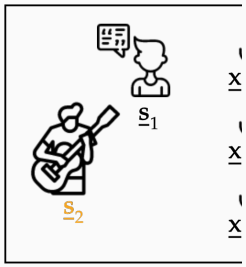
Echo-aware signal processing for audio scene analysis

Semantic information



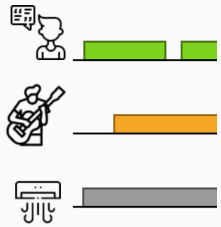
on nature and content

Spatial information



on position and geometry

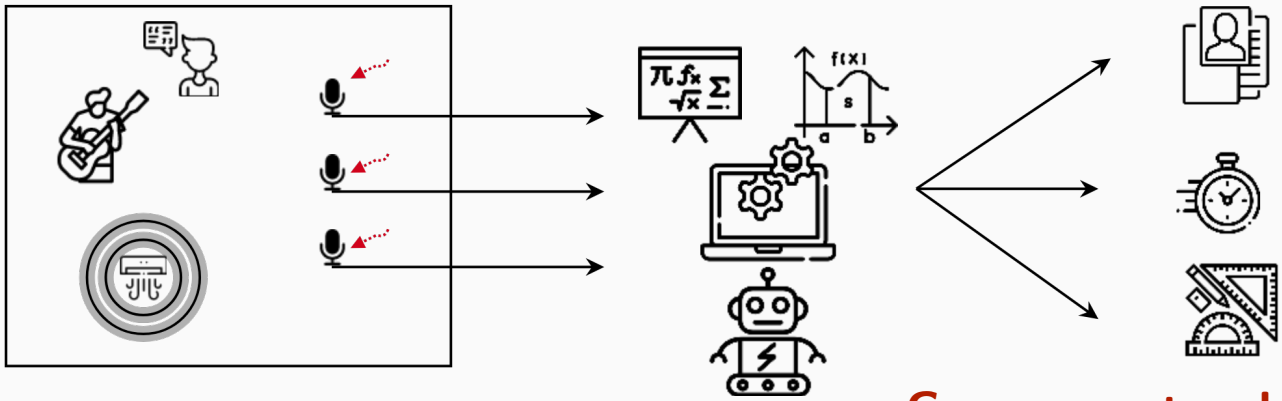
Temporal information



on events activity

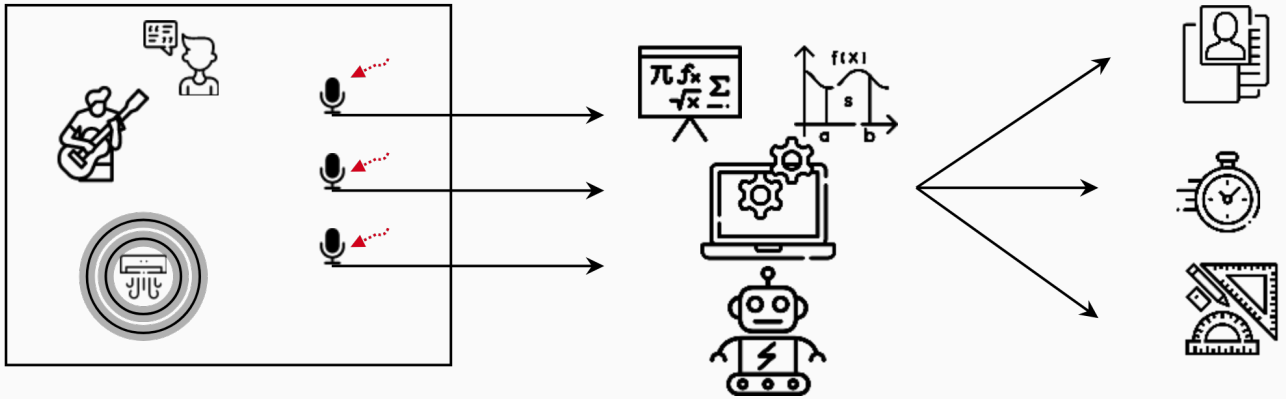
Audio Scene Analysis

Extraction and organization of all the information in the sound

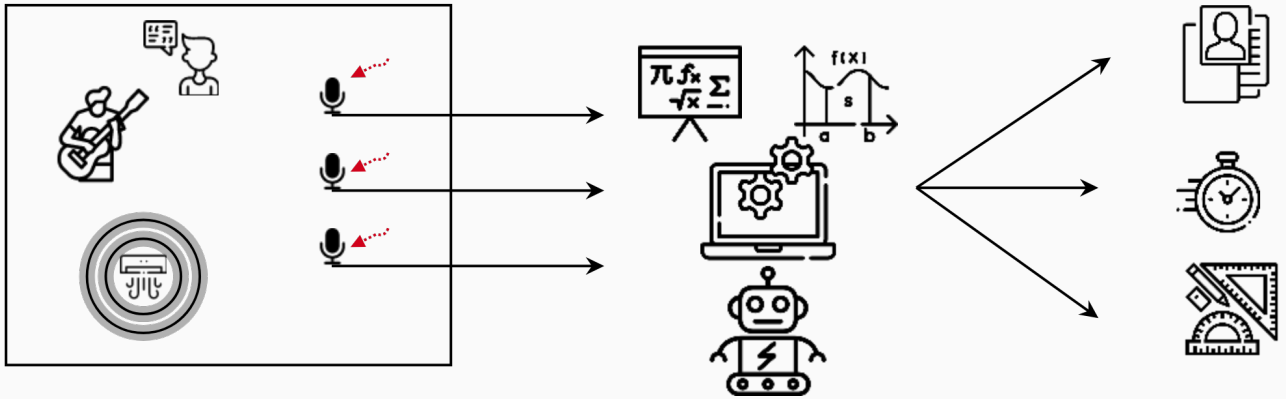


Can computer do it?

Echo-aware signal processing for audio scene analysis



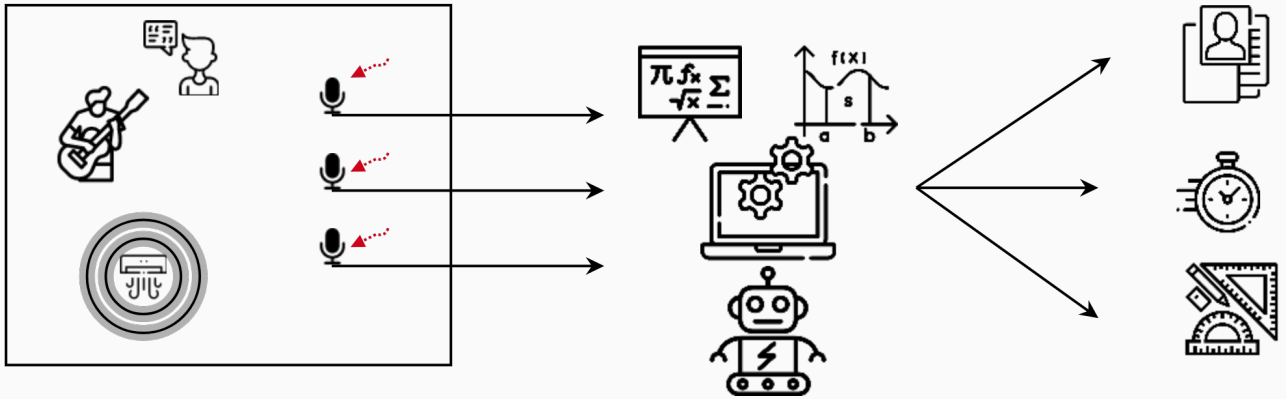
Echo-aware signal processing for audio scene analysis



Signal Processing

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

Echo-aware signal processing for audio scene analysis



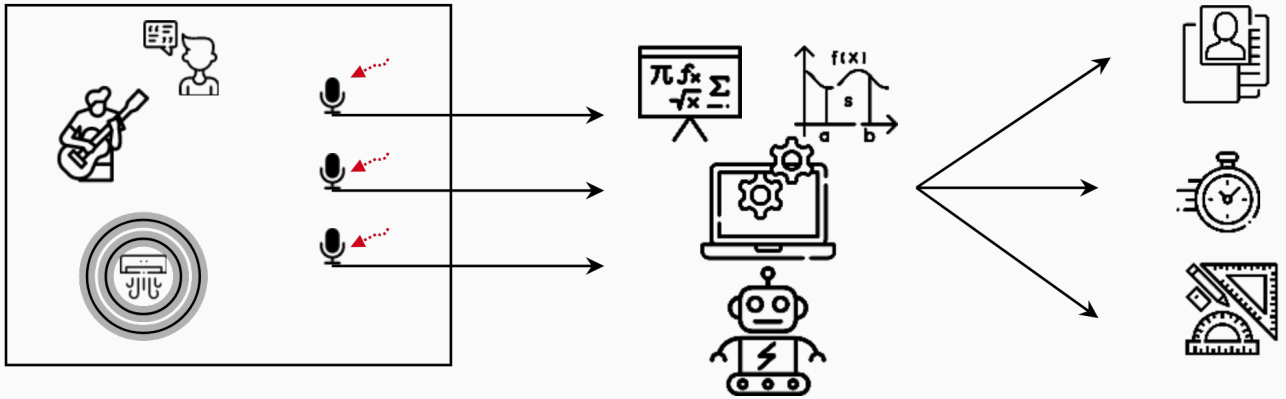
Signal Processing

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

Echo-aware signal processing for audio scene analysis

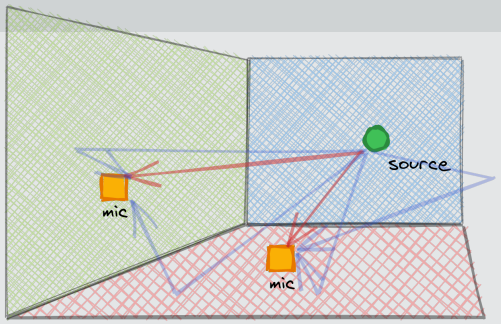
Sound interacts with environment

it is reflected (specularly and diffusely)
+ it is diffracted
+ it is absorbed and transmitted
+ other physical interaction

} = all sound propagation

Acoustic Echoes

- **Elements** of **reverberation**
- **Standing out** for time and strength
- **Repetition** of a sound but **after**
  - time $\Leftrightarrow$ distance
  - same content



Everyday examples:

Echo points Bat Dolphins Room coloration

Typically sound propagation is

- ignored $\Longrightarrow$ simple processing

but reverberation = noise

Echo-aware methods

Outline and contributions

Thesis title:

Audio Scene Analysis

↓

context and problems

Outline and contributions

Thesis title:

Audio Scene Analysis ↓ context and problems

Signal Processing ↓ models and frameworks

Outline and contributions

Thesis title:

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

How to use them?

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

Where to find them?

Echo-aware database for estimation and application

Problem Statement

Signal model

Sound propagation process $\Leftrightarrow$ Source $\rightarrow$ Filter $\rightarrow$ Receiver model

source signal

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

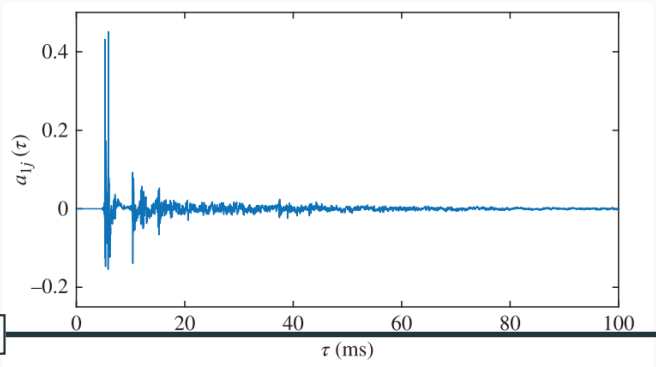
continuous-time convolution

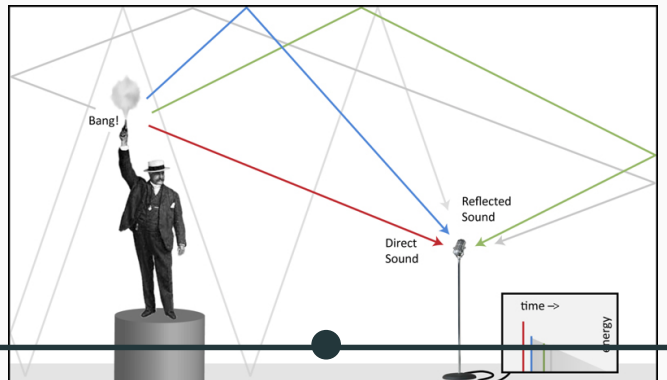
-3mm

⚠ 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)

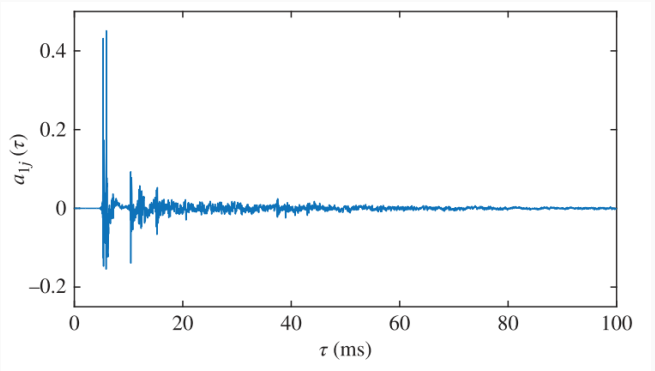




Echoes in the RIR

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 $\to$ sum of filters)
- h_l reverberation is included in the noise term
- depends on the scene geometry (room, source and mic position)

References i