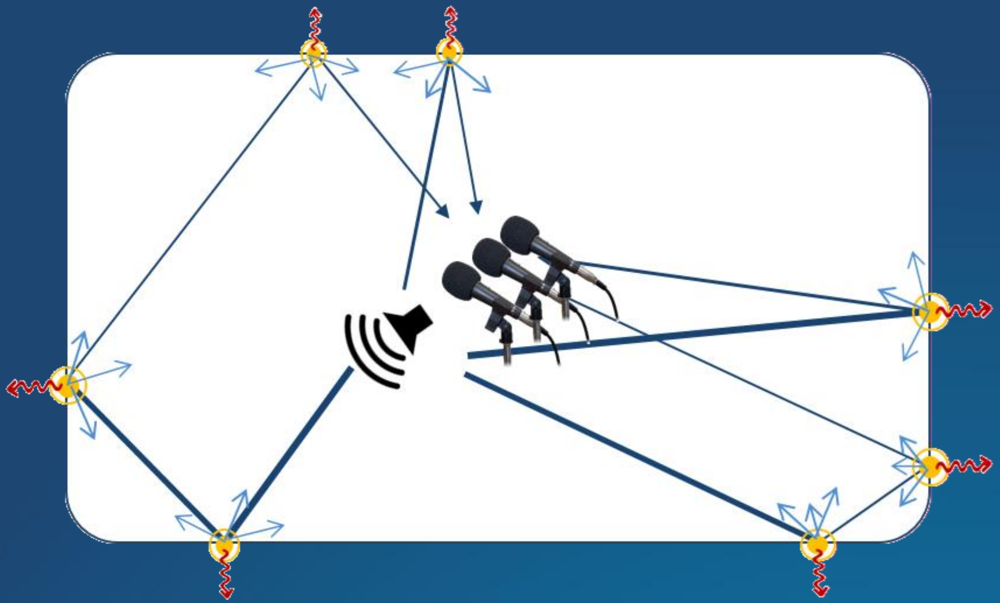
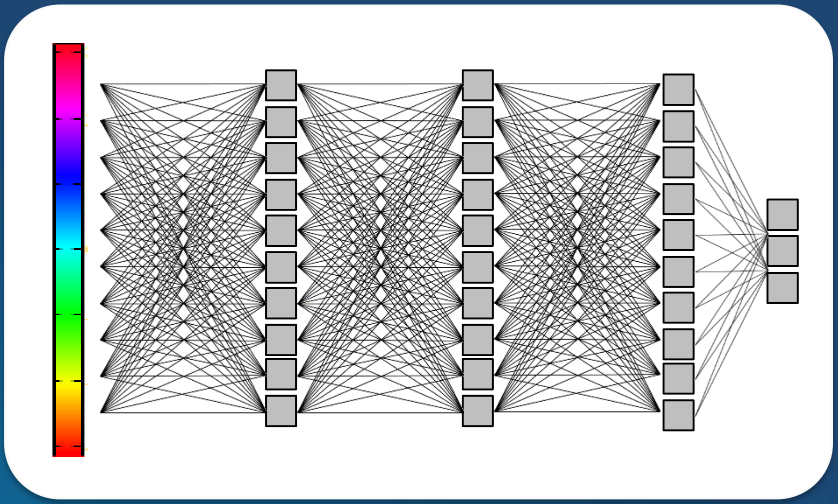
# Machine Learning for Indoor Acoustics

Antoine Deleforge
Inria (Nancy - Grand Est)





👏

*« What is the shape of the room? »*

*« What is the shape of the room? »*

*« Is the floor made of tiles or carpet? »*

# OUTLINE

1) Intro & Background
2) Virtually-Supervised Learning
3) Examples and Results
4) Conclusions and Outlook

# OUTLINE

**1) Intro & Background**

2) Virtually-Supervised Learning

3) Examples and Results

4) Conclusions and Outlook

# Sound Propagation

- **What is sound?**

# Sound Propagation

- **What is sound?**
  - A Mechanical Vibration

# Sound Propagation

- **What is sound?**
  - A Mechanical Vibration
  - A Variation of Air Pressure


Antoine.Deleforge@inria.fr

# Sound Propagation

- **What is sound?**
  - A Mechanical Vibration
  - A Variation of Air Pressure
  - A 3D Wave

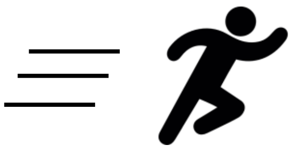

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

# Sound Propagation

- **What is sound?**
  - A Mechanical Vibration
  - A Variation of Air Pressure
  - A 3D Wave


$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

- Sound has a **speed:** $c \approx 343$ m/sec

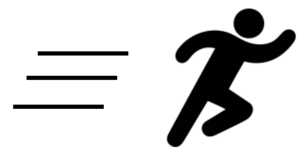


# Sound Propagation

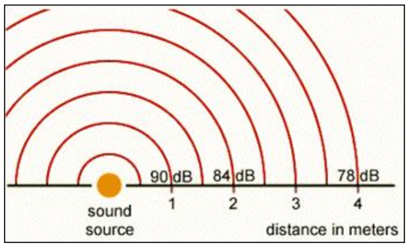
- **What is sound?**
  - A Mechanical Vibration
  - A Variation of Air Pressure
  - A 3D Wave


$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

- Sound has a **speed:** $c \approx 343$ m/sec



- Sound **dissipates:** ≈ -6 dB every doubling of distance

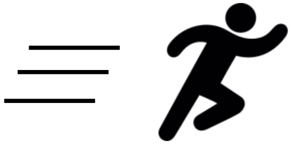


# Sound Propagation

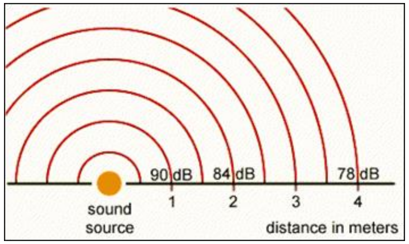
- **What is sound?**
  - A Mechanical Vibration
  - A Variation of Air Pressure
  - A 3D Wave

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

- Sound has a **speed:** $c \approx 343$ m/sec



- Sound **dissipates:** ≈ -6 dB every doubling of distance



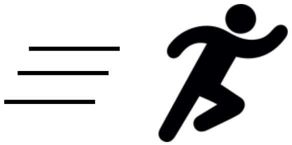

- Sound **Interacts:**

# Sound Propagation

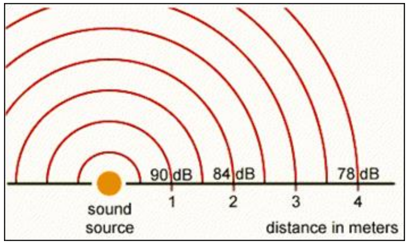
- **What is sound?**
  - A Mechanical Vibration
  - A Variation of Air Pressure
  - A 3D Wave

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

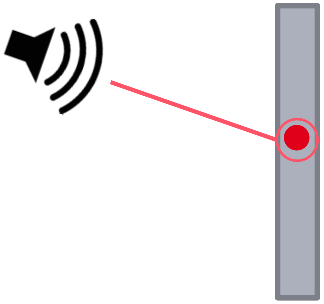
- Sound has a **speed:** $c \approx 343$ m/sec



- Sound **dissipates:** ≈ -6 dB every doubling of distance




- Sound **Interacts:**



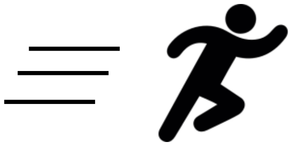
Absorption

# Sound Propagation

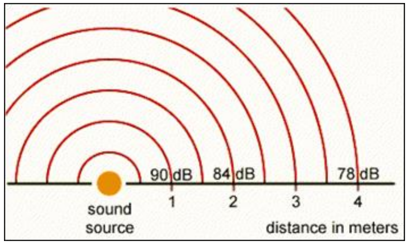
- **What is sound?**
  - A Mechanical Vibration
  - A Variation of Air Pressure
  - A 3D Wave

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

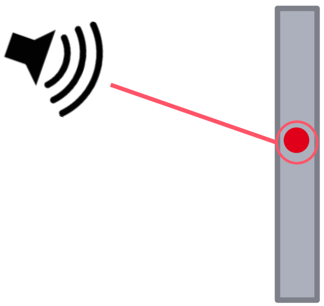
- Sound has a **speed:** $c \approx 343$ m/sec



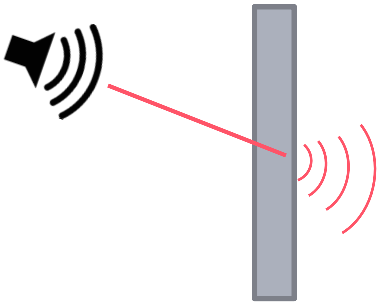
- Sound **dissipates:** ≈ -6 dB every doubling of distance




- Sound **Interacts:**



Absorption



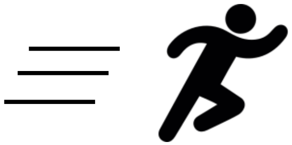
Transmission

# Sound Propagation

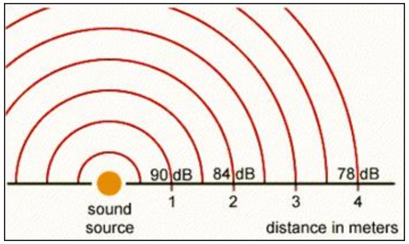
- **What is sound?**
  - A Mechanical Vibration
  - A Variation of Air Pressure
  - A 3D Wave


$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

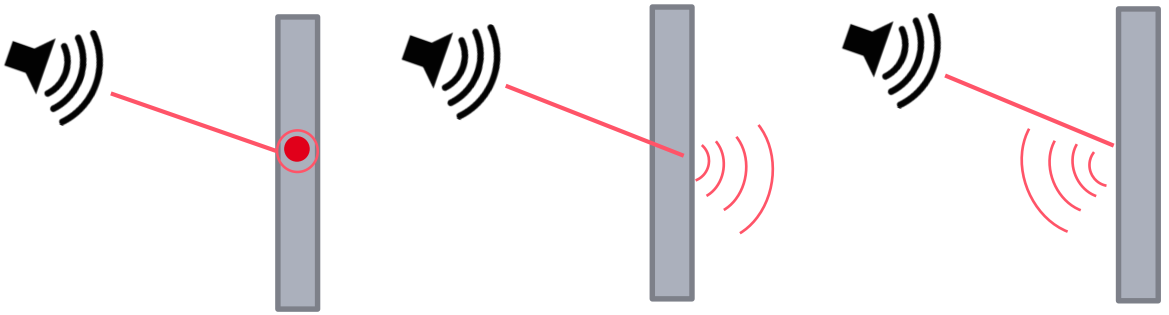
- Sound has a **speed:** $c \approx 343$ m/sec



- Sound **dissipates:** ≈ -6 dB every doubling of distance




- Sound **Interacts:**



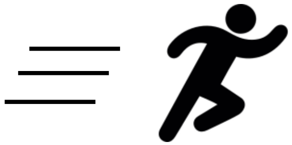
Absorption Transmission Reflexion

# Sound Propagation

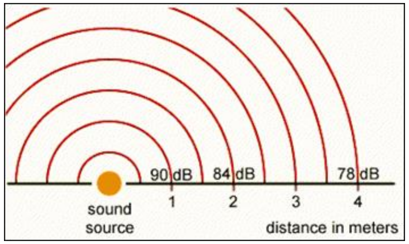
- **What is sound?**
  - A Mechanical Vibration
  - A Variation of Air Pressure
  - A 3D Wave

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

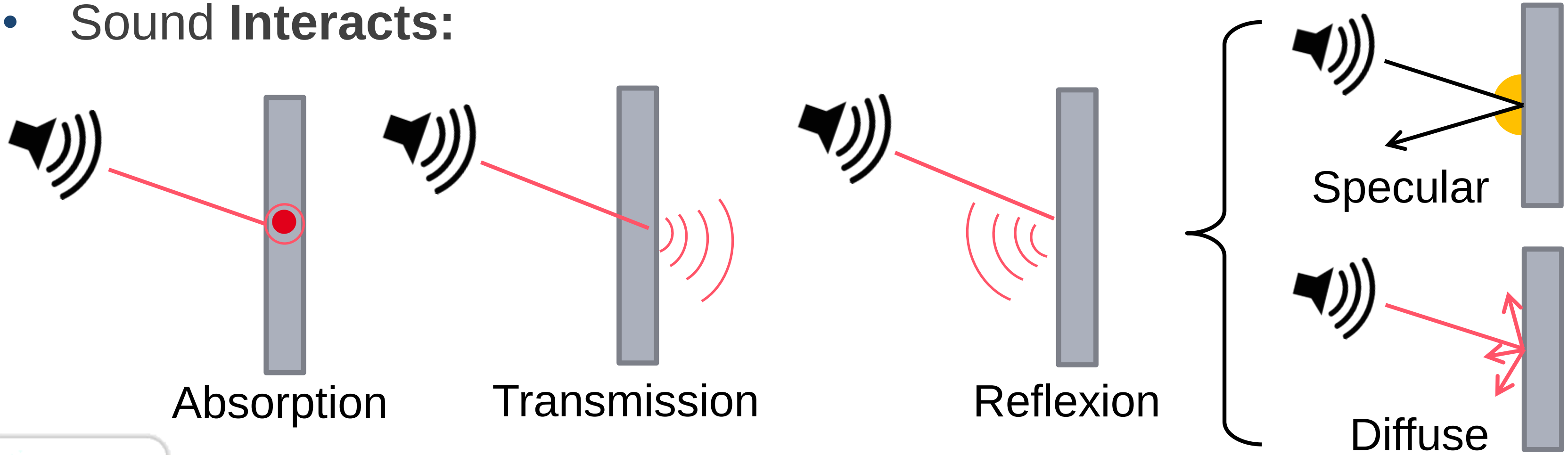
- Sound has a **speed:** $c \approx 343$ m/sec



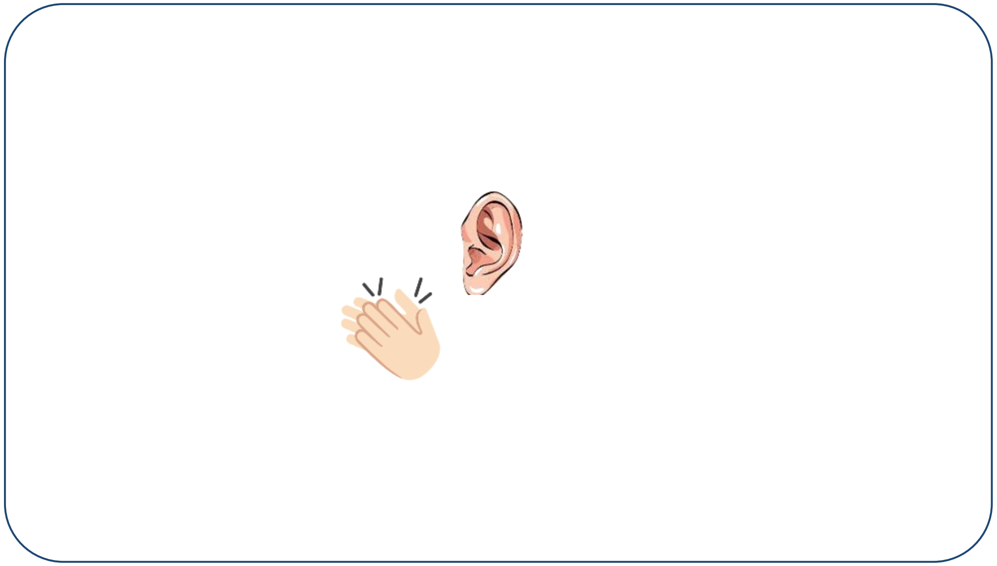
- Sound **dissipates:** ≈ -6 dB every doubling of distance



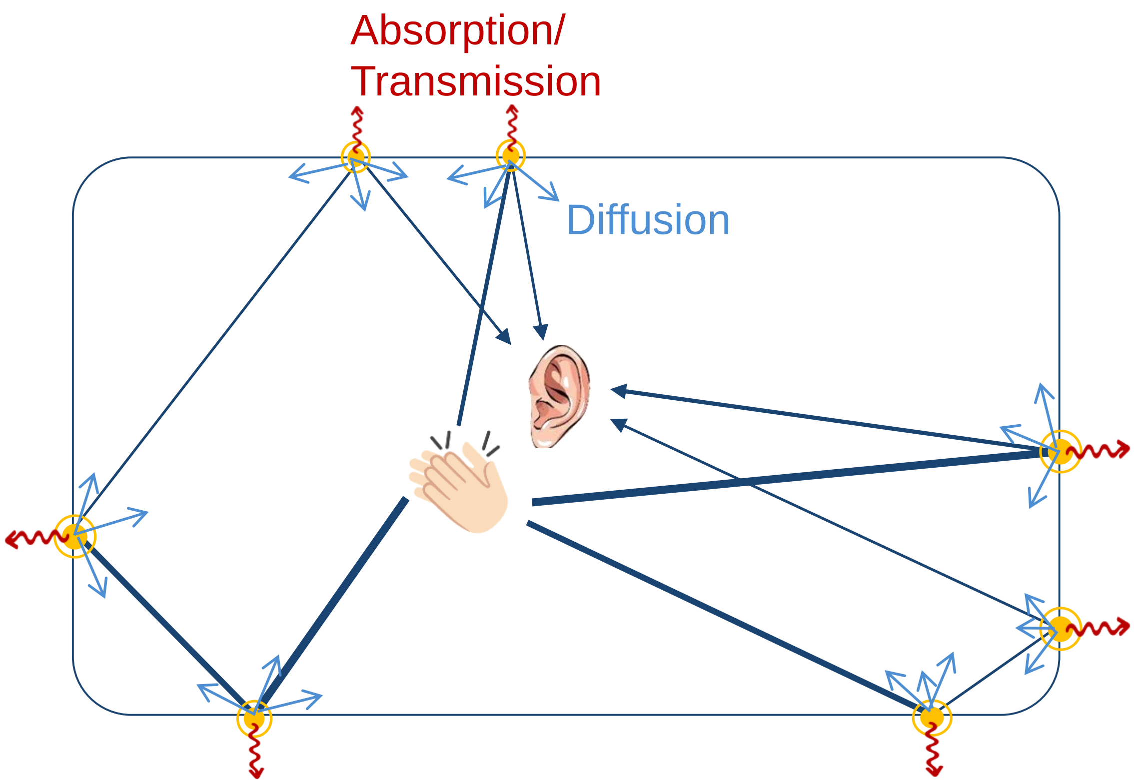

- Sound **Interacts:**



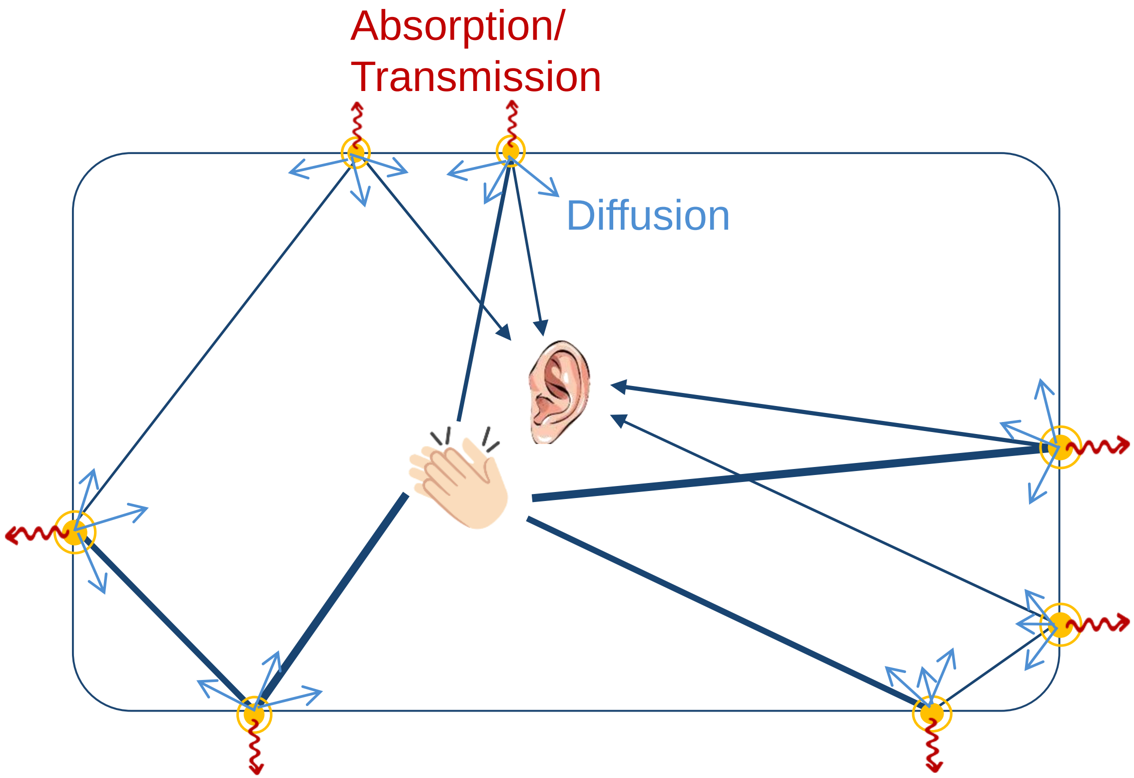
# 1) Introduction



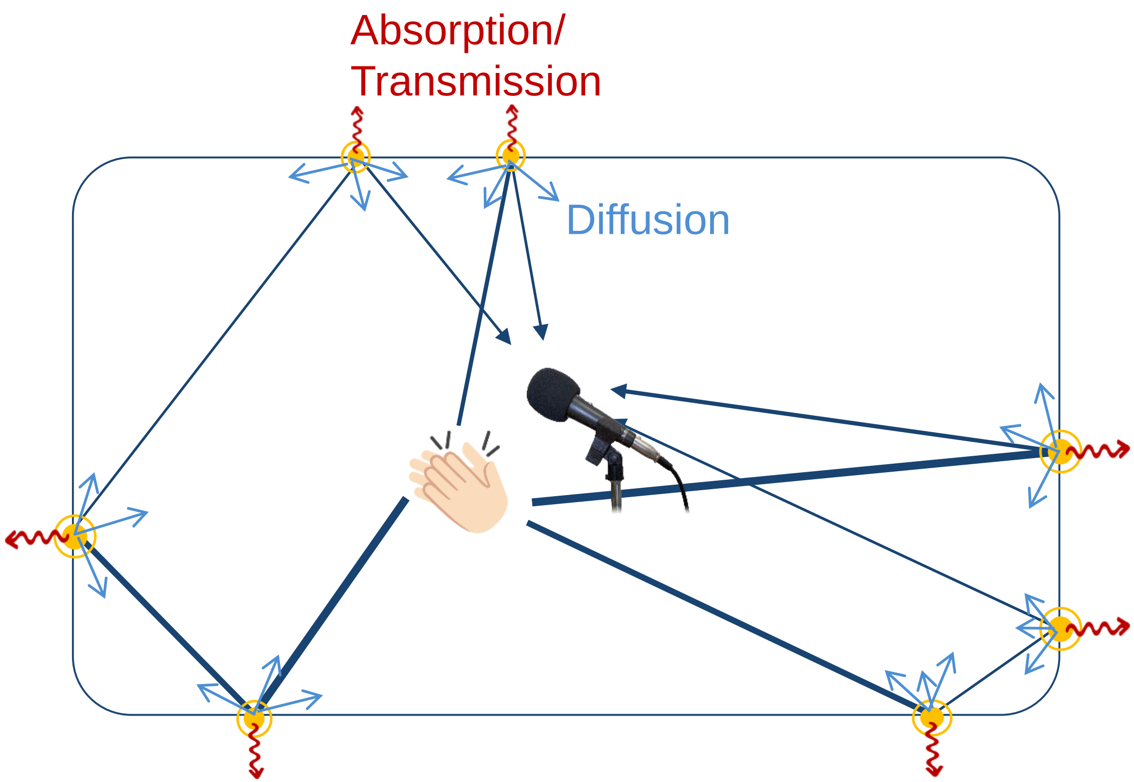
# 1) Introduction



# 1) Introduction




« Reverberation »




« Reverberation »

A signal model of reverberation?

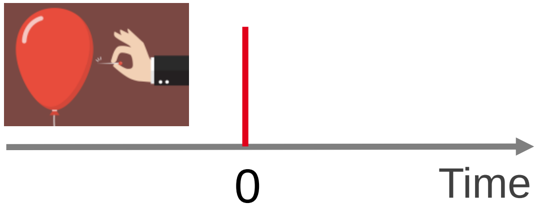
# The Room Impulse Response

- **Impulse response:** The response of an LTI system to a perfect impulse (*Dirac*).
- **Room Impulse response (RIR):** Captures the linear filtering effect due to the propagation of sound from a **point source** to a **microphone** inside a room

# The Room Impulse Response

- **Impulse response:** The response of an LTI system to a perfect impulse (*Dirac*).
- **Room Impulse response (RIR):** Captures the linear filtering effect due to the propagation of sound from a **point source** to a **microphone** inside a room

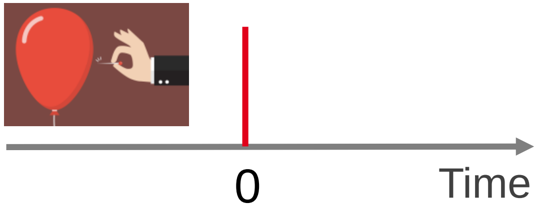
**Input:** $\delta(t)$



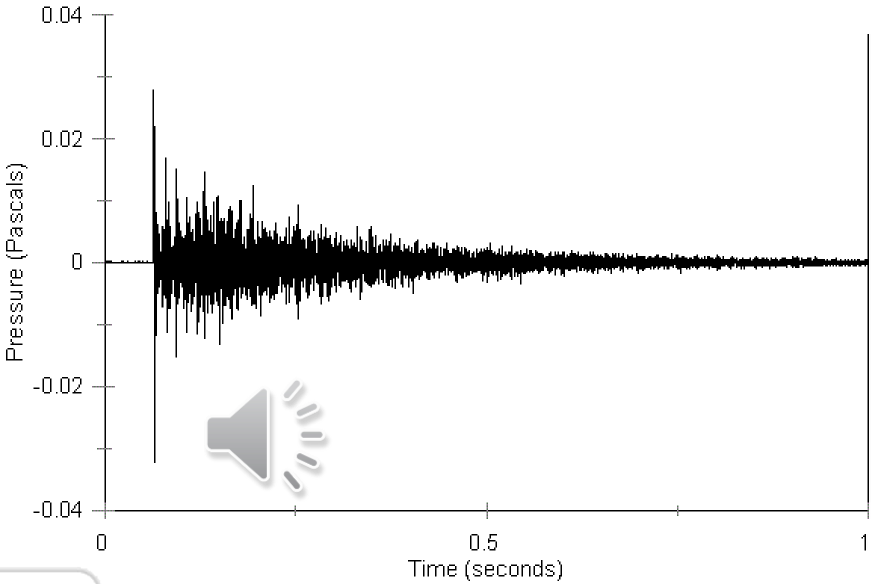
# The Room Impulse Response

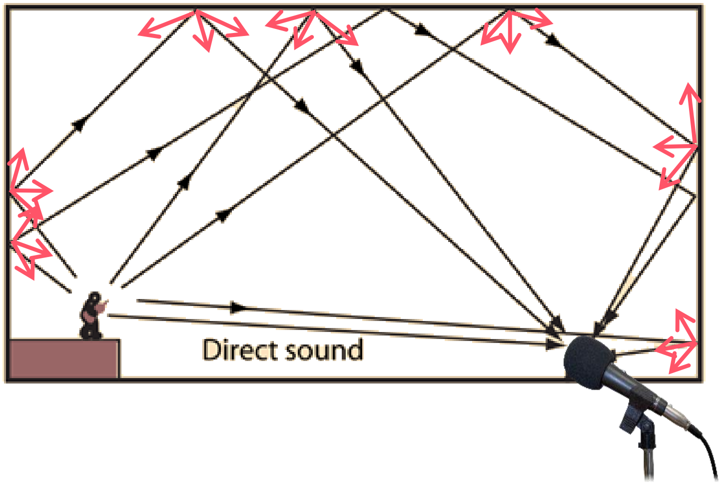
- **Impulse response:** The response of an LTI system to a perfect impulse (*Dirac*).
- **Room Impulse response (RIR):** Captures the linear filtering effect due to the propagation of sound from a **point source** to a **microphone** inside a room

**Input:** $\delta(t)$




**Output:** $h(t)$

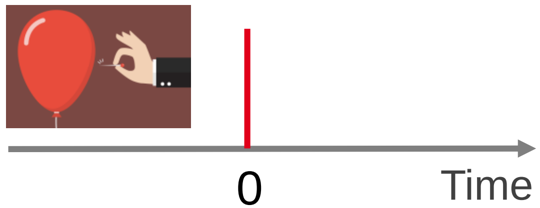


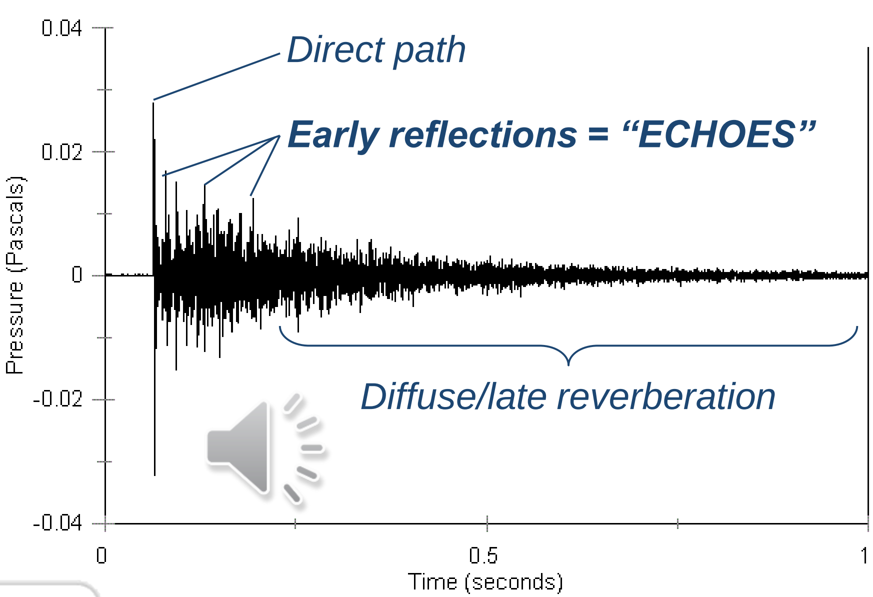
# The Room Impulse Response

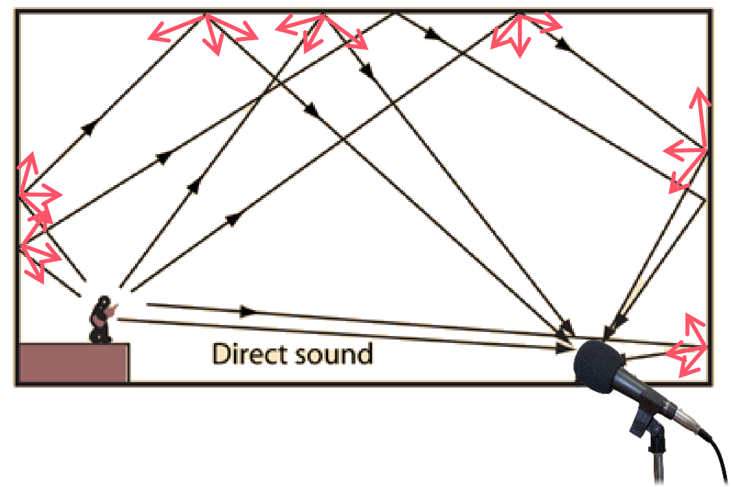
- **Impulse response:** The response of an LTI system to a perfect impulse (*Dirac*).
- **Room Impulse response (RIR):** Captures the linear filtering effect due to the propagation of sound from a **point source** to a **microphone** inside a room

**Input:** $\delta(t)$




**Output:** $h(t)$

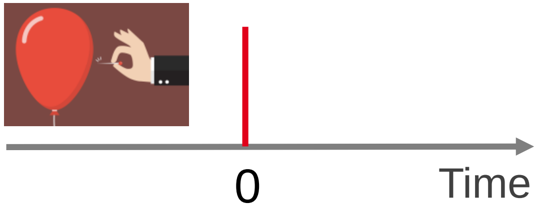


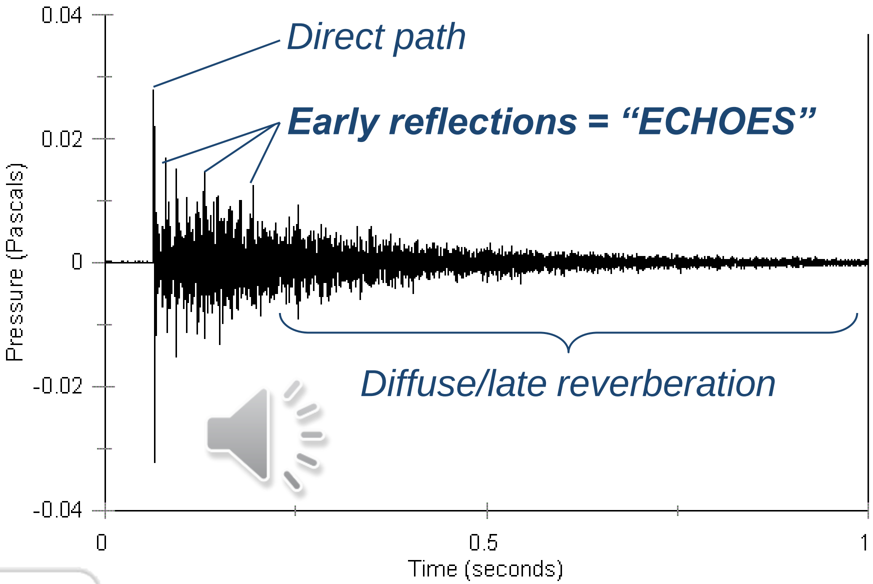
# The Room Impulse Response

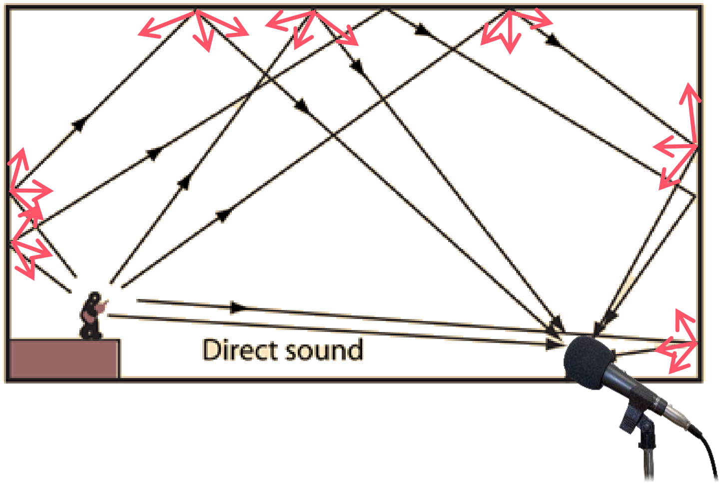
- **Impulse response:** The response of an LTI system to a perfect impulse (*Dirac*).
- **Room Impulse response (RIR):** Captures the linear filtering effect due to the propagation of sound from a **point source** to a **microphone** inside a room

**Input:** $\delta(t)$




**Output:** $h(t)$




The Fourier transform $\widetilde{h}(\omega)$ of a RIR is called ***Room Transfer Function***

It captures the effect of the room in different ***frequency bands***.

Antoine.Deleforge@inria.fr

# The Room Impulse Response

- Can be used to « reverberate » any dry sound source signal $s(t)$ :

$$x(t) = (h * s)(t) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} h(u)s(t-u)du \quad \overset{\text{Fourier}}{\longleftrightarrow} \quad \widetilde{x}(\omega) = \widetilde{h}(\omega)\widetilde{s}(\omega)$$

# The Room Impulse Response

- Can be used to « reverberate » any dry sound source signal $s(t)$ :

$$x(t) = (h * s)(t) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} h(u)s(t-u)du \quad \xrightarrow{\text{Fourier}} \quad \widetilde{x}(\omega) = \widetilde{h}(\omega)\widetilde{s}(\omega)$$

www.openair.hosted.york.ac.uk/

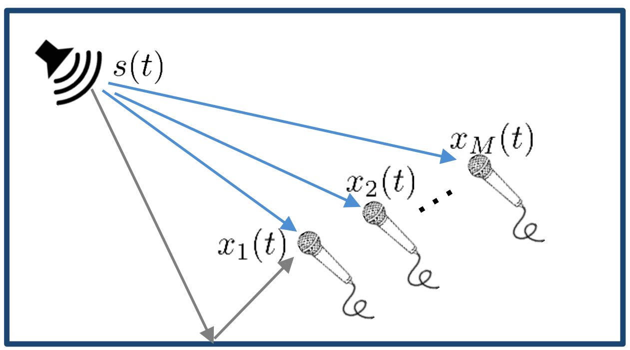
# The Room Impulse Response

- Can be used to « reverberate » any dry sound source signal $s(t)$ :

$$x(t) = (h * s)(t) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} h(u)s(t-u)du \quad \xrightarrow{\text{Fourier}} \quad \widetilde{x}(\omega) = \widetilde{h}(\omega)\widetilde{s}(\omega)$$

www.openair.hosted.york.ac.uk/

- Generalization to multiple microphones:



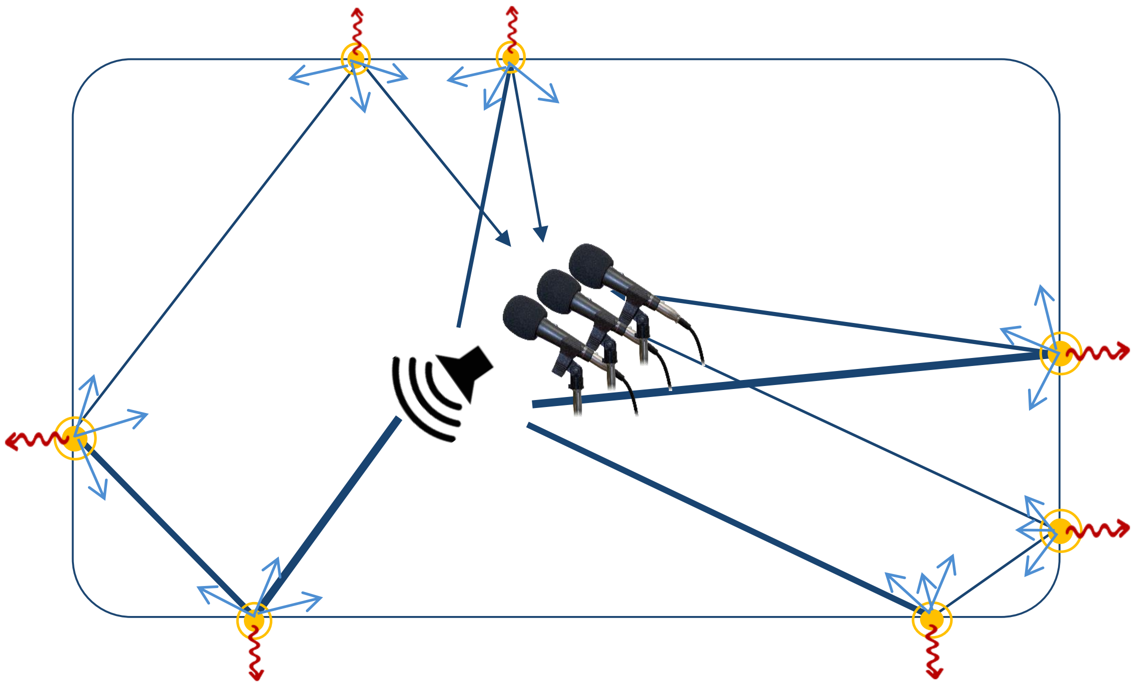

$$\begin{cases} x_1(t) = (h_1 * s)(t) + n_2(t) \\ x_2(t) = (h_2 * s)(t) + n_2(t) \\ \quad\vdots \qquad\qquad \vdots \\ x_M(t) = (h_M * s)(t) + n_M(t) \end{cases}$$


Linear


Circular

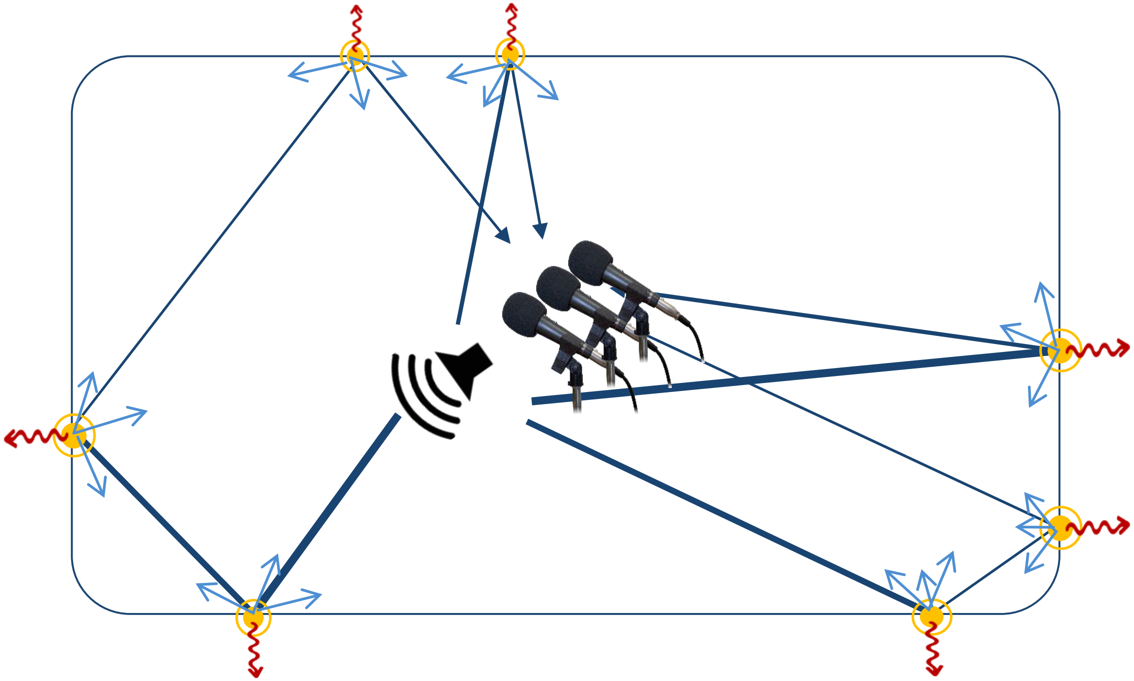
Spherical

# 1) Intro & Background

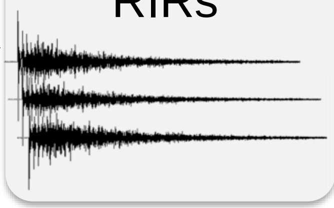


- Source & receivers positions & properties
- Room geometry
- Surface properties

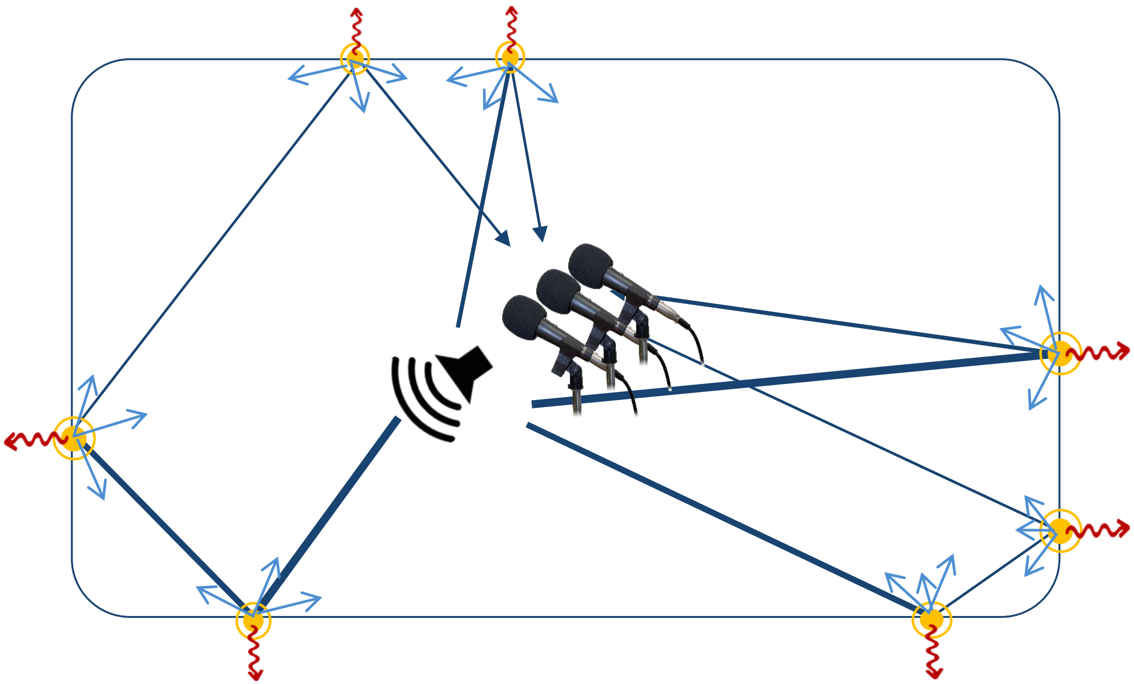
# 1) Intro & Background



- Source & receivers positions & properties
- Room geometry
- Surface properties



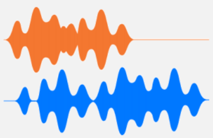
# 1) Intro & Background

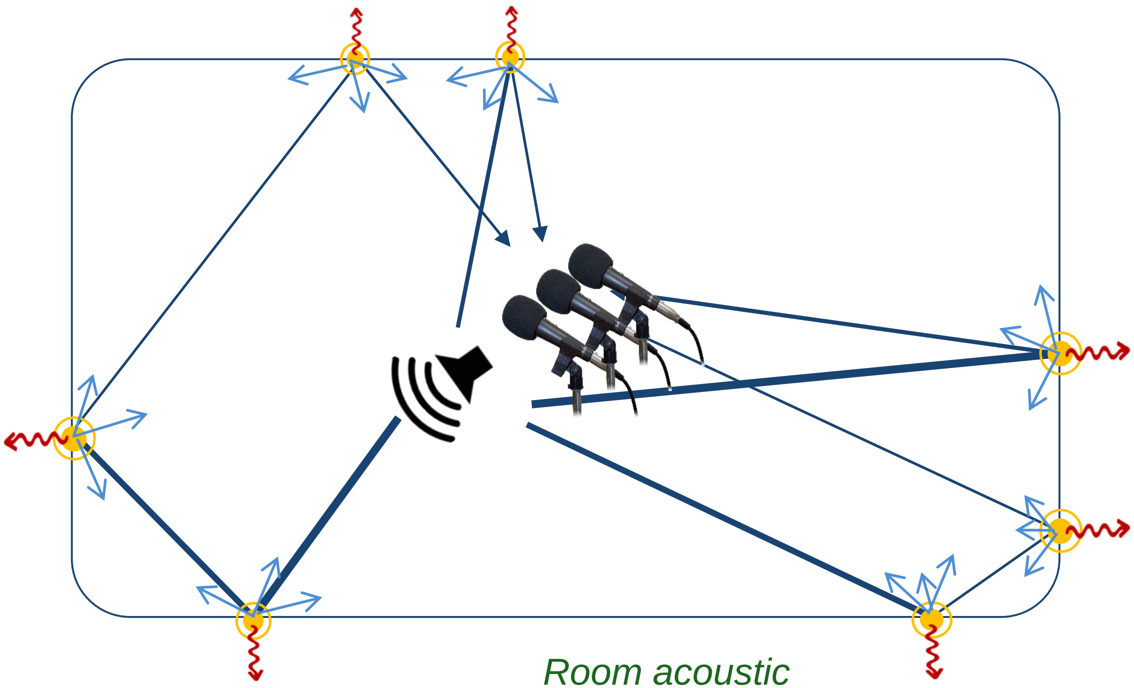


- Source & receivers positions & properties
- Room geometry
- Surface properties

RIRs

Revereberated audio signals

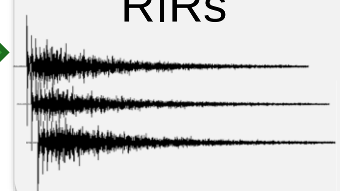




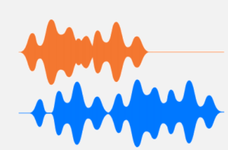
*Room acoustic simulators ☑*

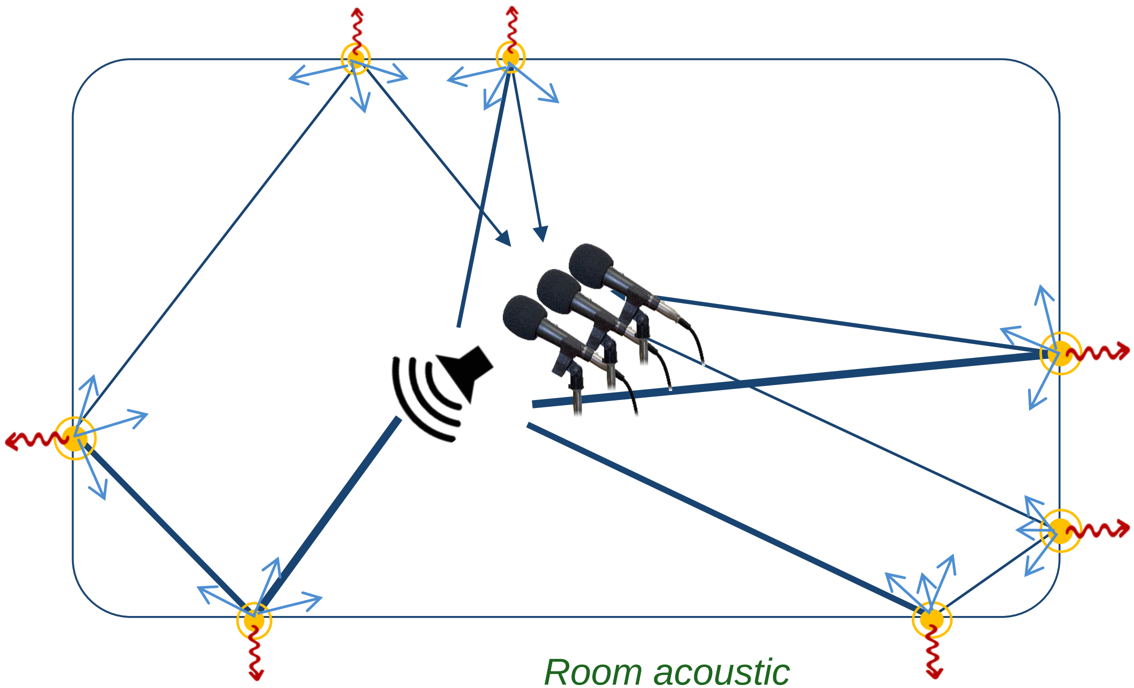
- Source & receivers positions & properties
- Room geometry
- Surface properties

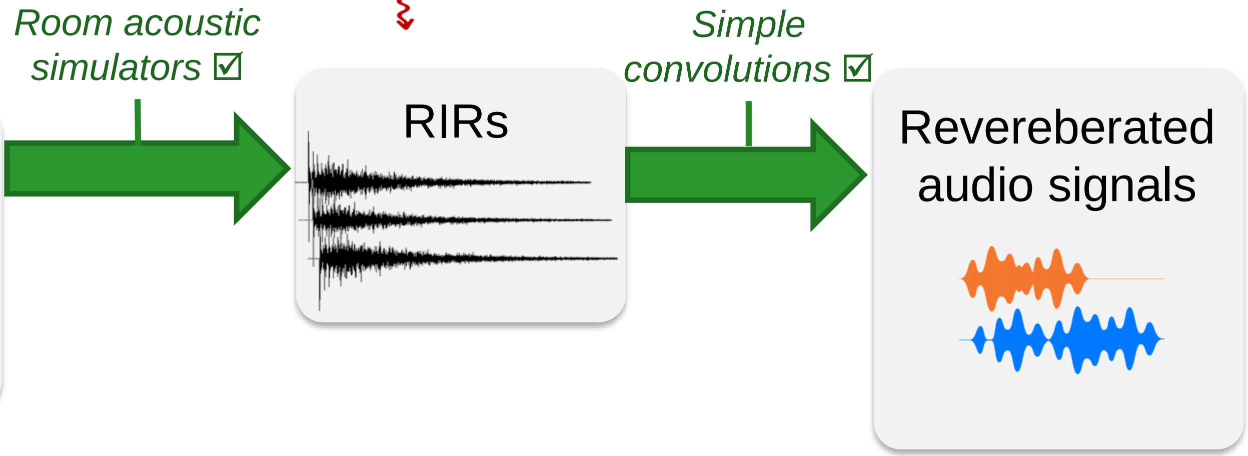
RIRs



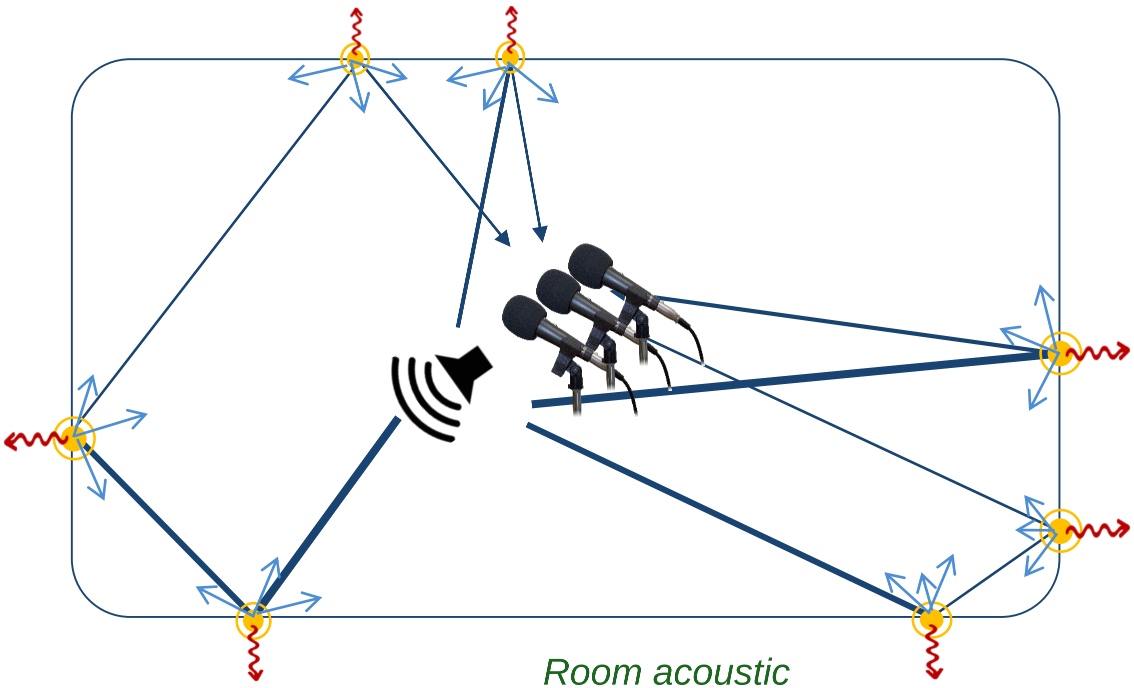
Revereberated audio signals








- Source & receivers positions & properties
- Room geometry
- Surface properties



- Source & receivers positions & properties
- Room geometry
- Surface properties

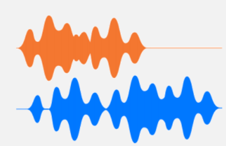
*Room acoustic simulators ☑*

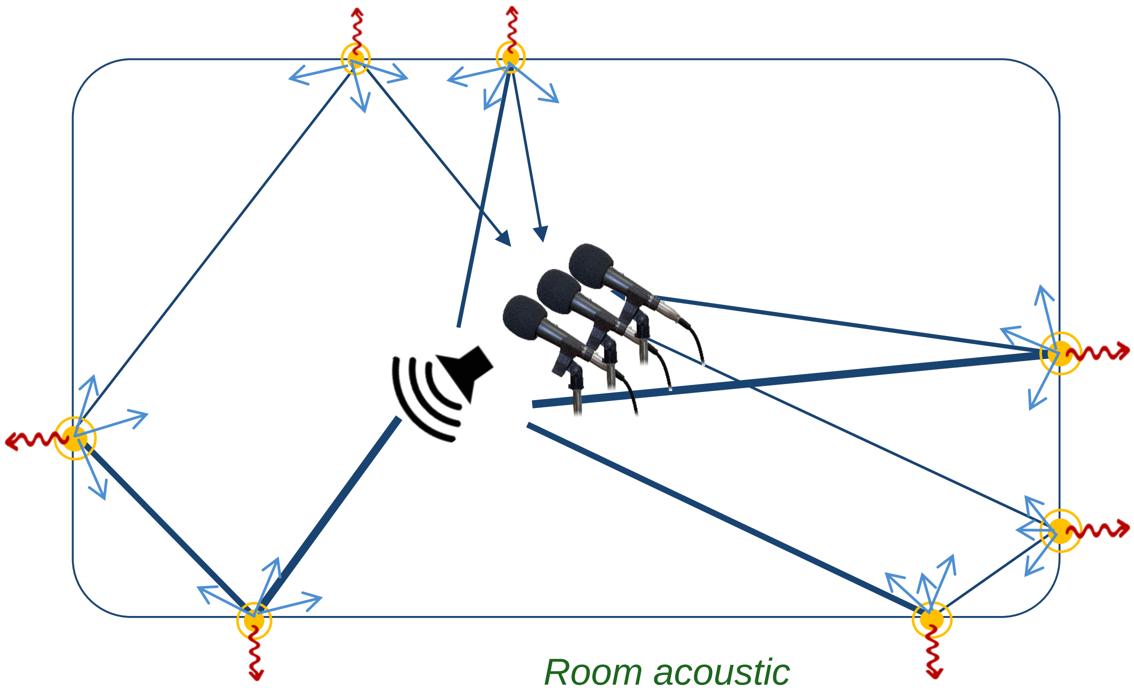
RIRs

?

*Simple convolutions ☑*

Revereberated audio signals





- Source & receivers positions & properties
- Room geometry
- Surface properties

*Room acoustic simulators* ☑

?

RIRs

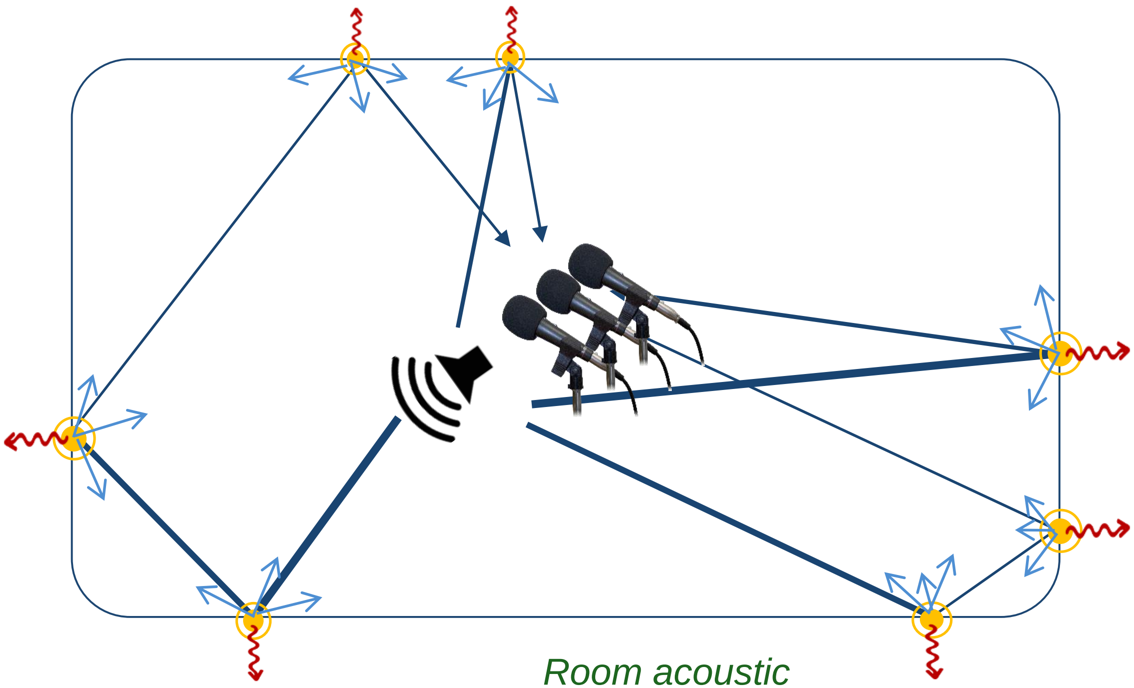
*Simple convolutions* ☑

Revereberated audio signals

*(Blind)*

?

# 1) Intro & Background



*Room acoustic simulators* ☑

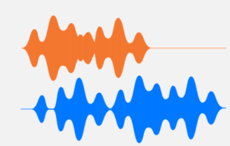
*Simple convolutions* ☑

- Source & receivers positions & properties
- Room geometry
- Surface properties

RIRs

?

Revereberated audio signals



*(Blind)*

?

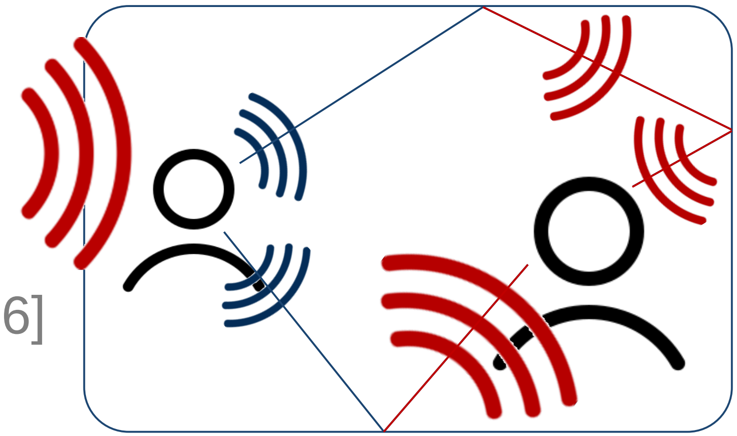
Difficult (interesting) inverse problems!

# Why do we care?

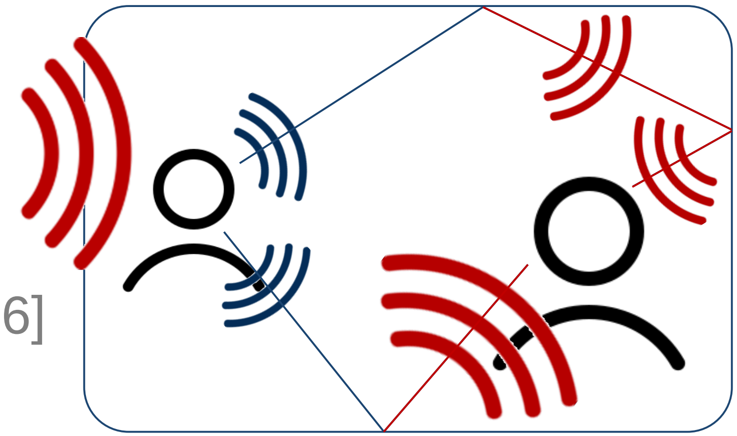
# Why do we care?

1) Indoor noise disturbance

➡ Make acoustic diagnosis faster / better [16]



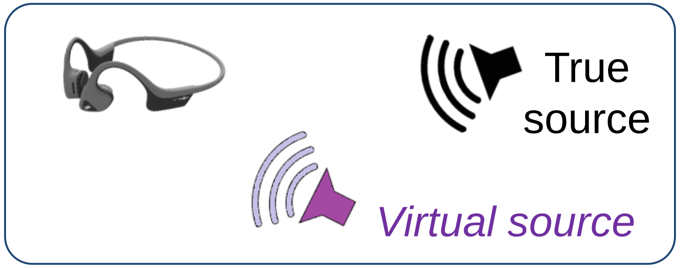
# Why do we care?



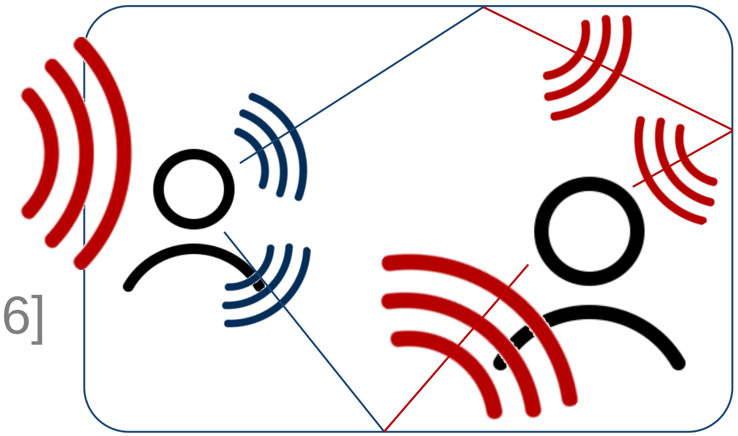
1) Indoor noise disturbance

➡ Make acoustic diagnosis faster / better [16]

2) Audio Augmented Reality [6, 17]



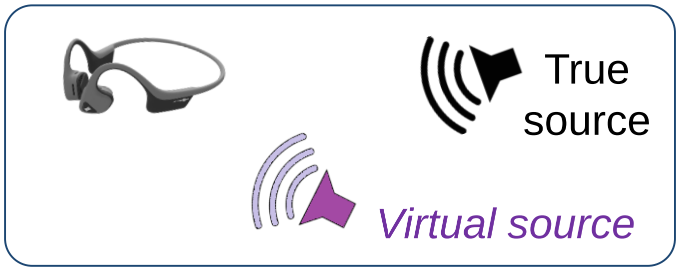
# Why do we care?

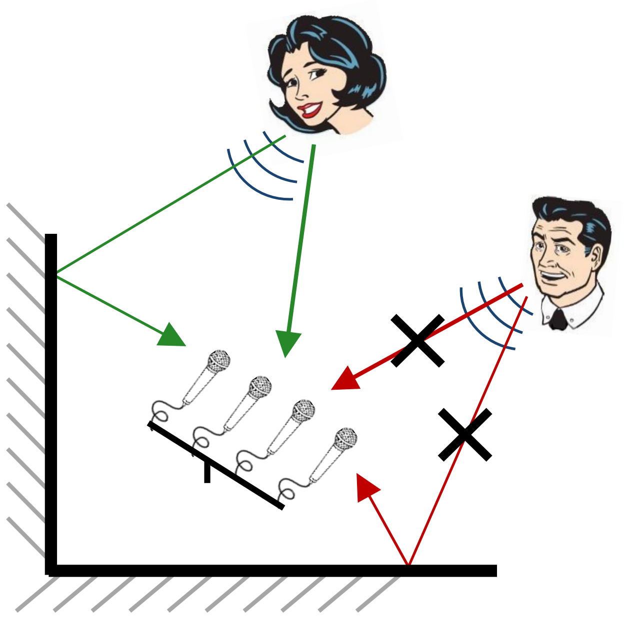


1) Indoor noise disturbance

➡ Make acoustic diagnosis faster / better [16]

2) Audio Augmented Reality [6, 17]



3) « *Echo-Aware* » Audio Signal Processing [7, 8]

- Hearing aids
- Vocal assistant devices
- …

# OUTLINE

**1) Intro & Background**

2) Virtually-Supervised Learning

3) Examples and Results

4) Conclusions and Outlook

# OUTLINE

1) Intro & Background

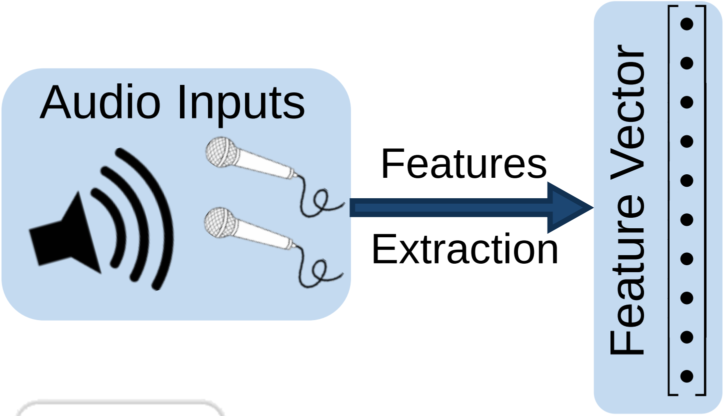
**2) Virtually-Supervised Learning**

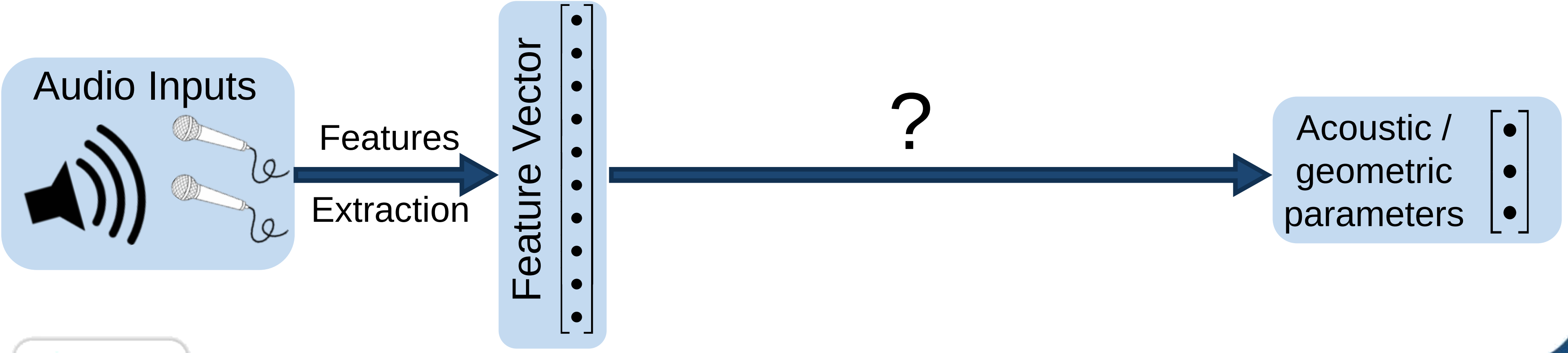
3) Examples and Results

4) Conclusions and Outlook

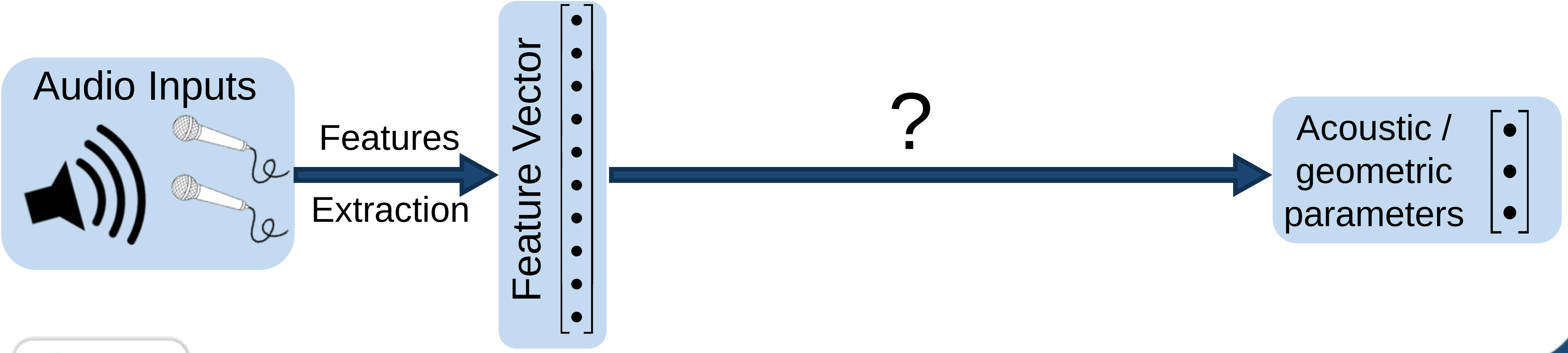
# 2) Virtually Supervised Learning

# 2) Virtually Supervised Learning

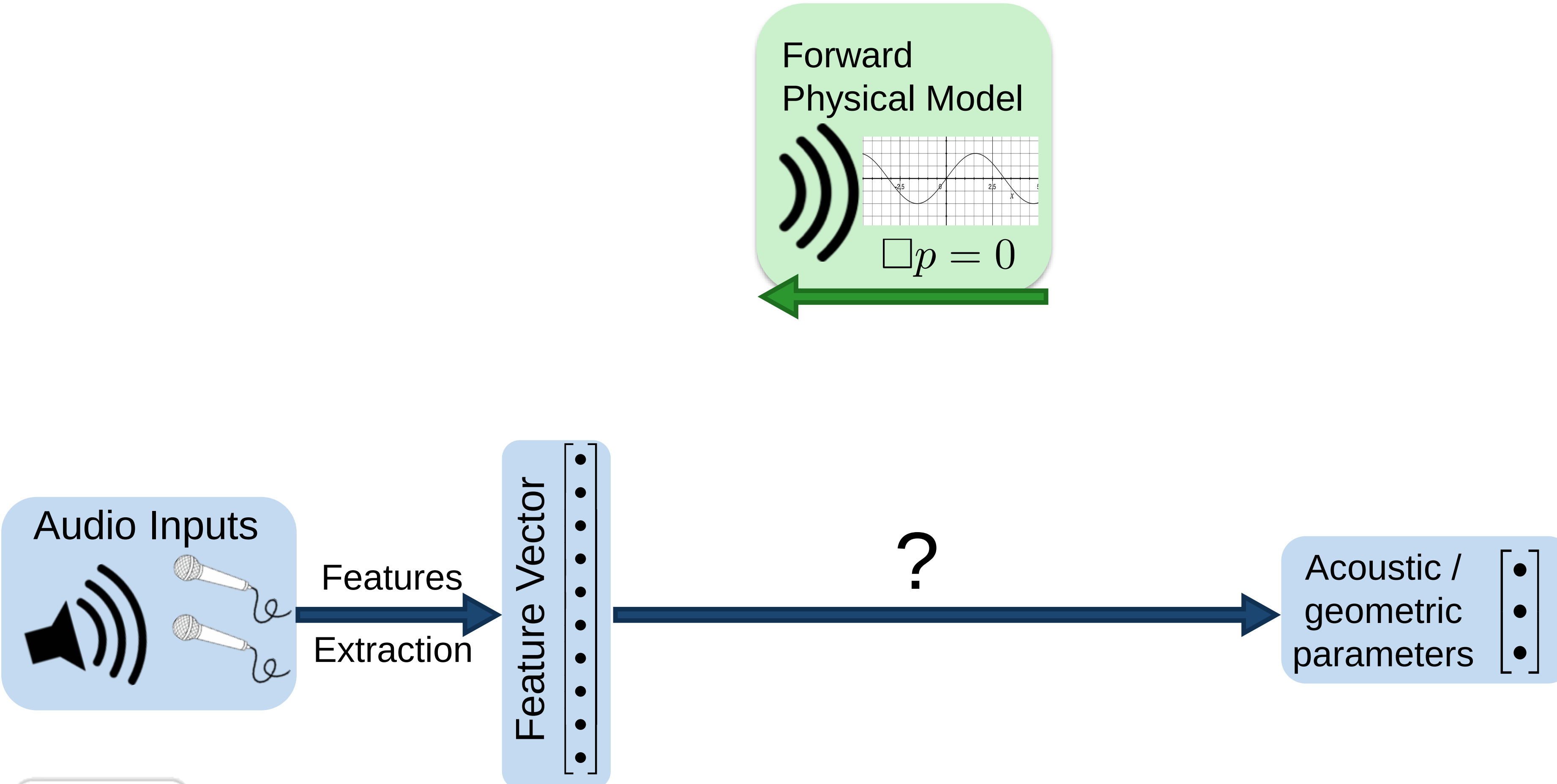



Audio Inputs
Features
Extraction
Feature Vector
?
Acoustic /
geometric
parameters

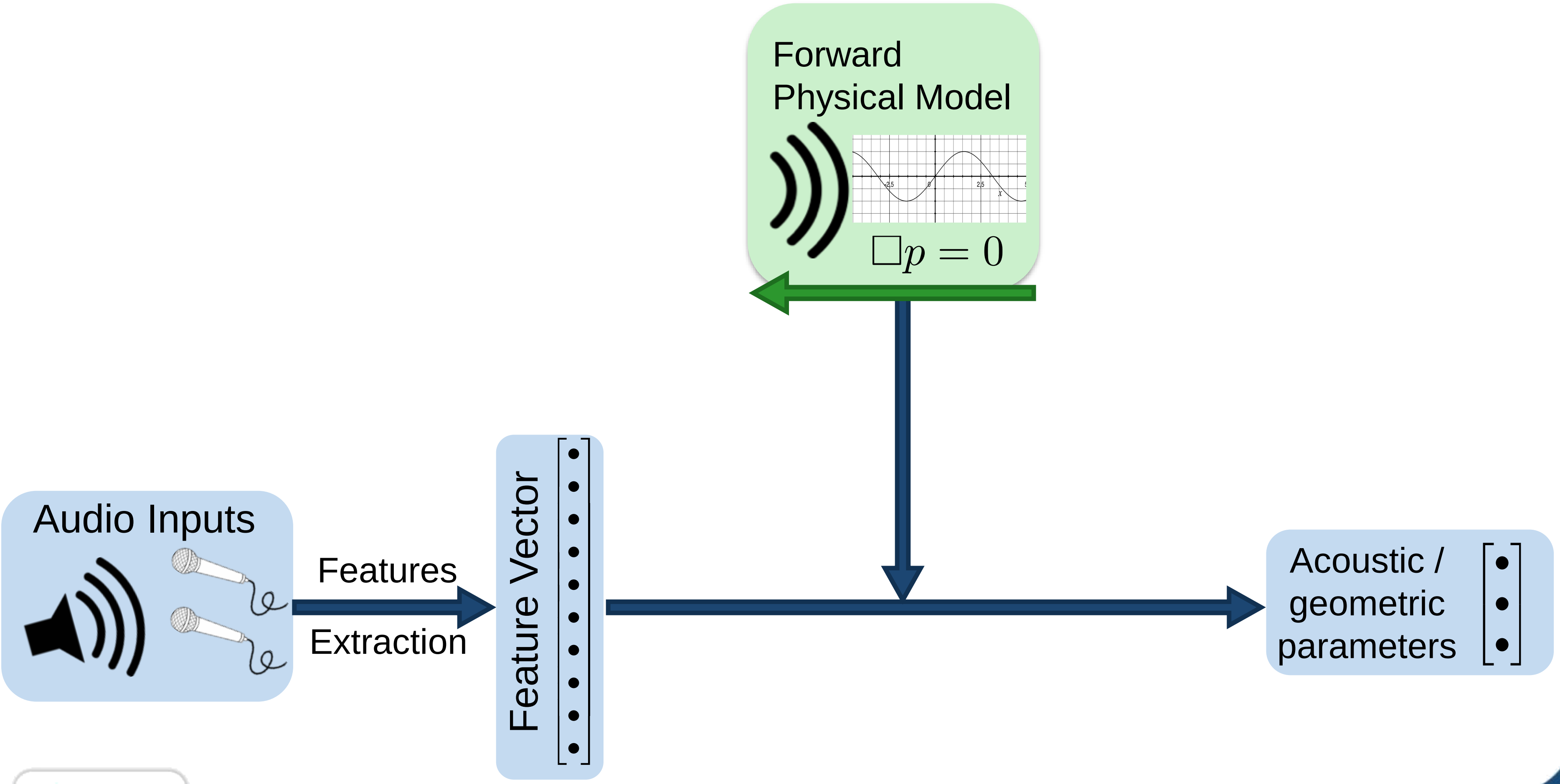
## a) Physics-Driven Approaches



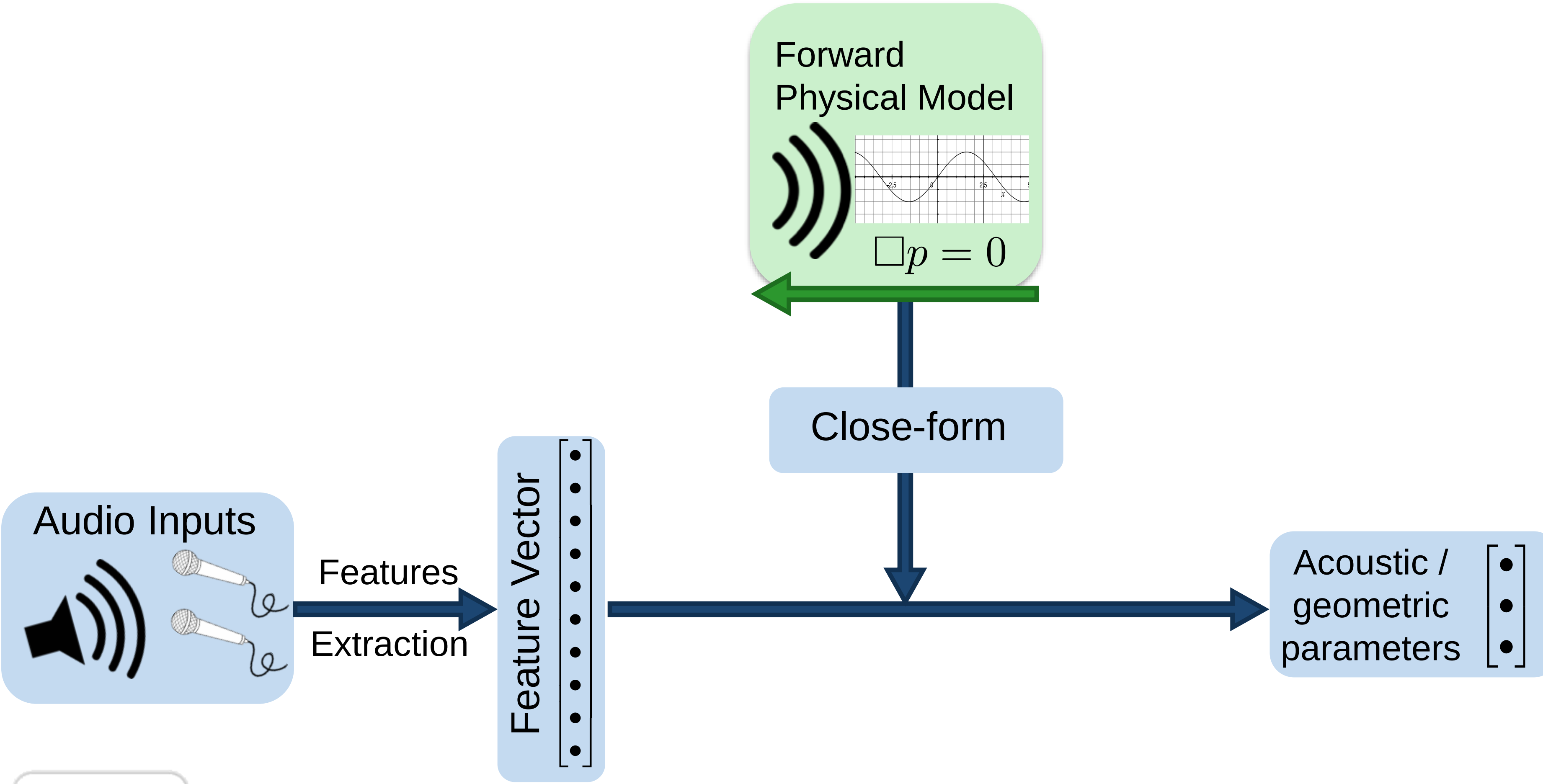
# a) Physics-Driven Approaches



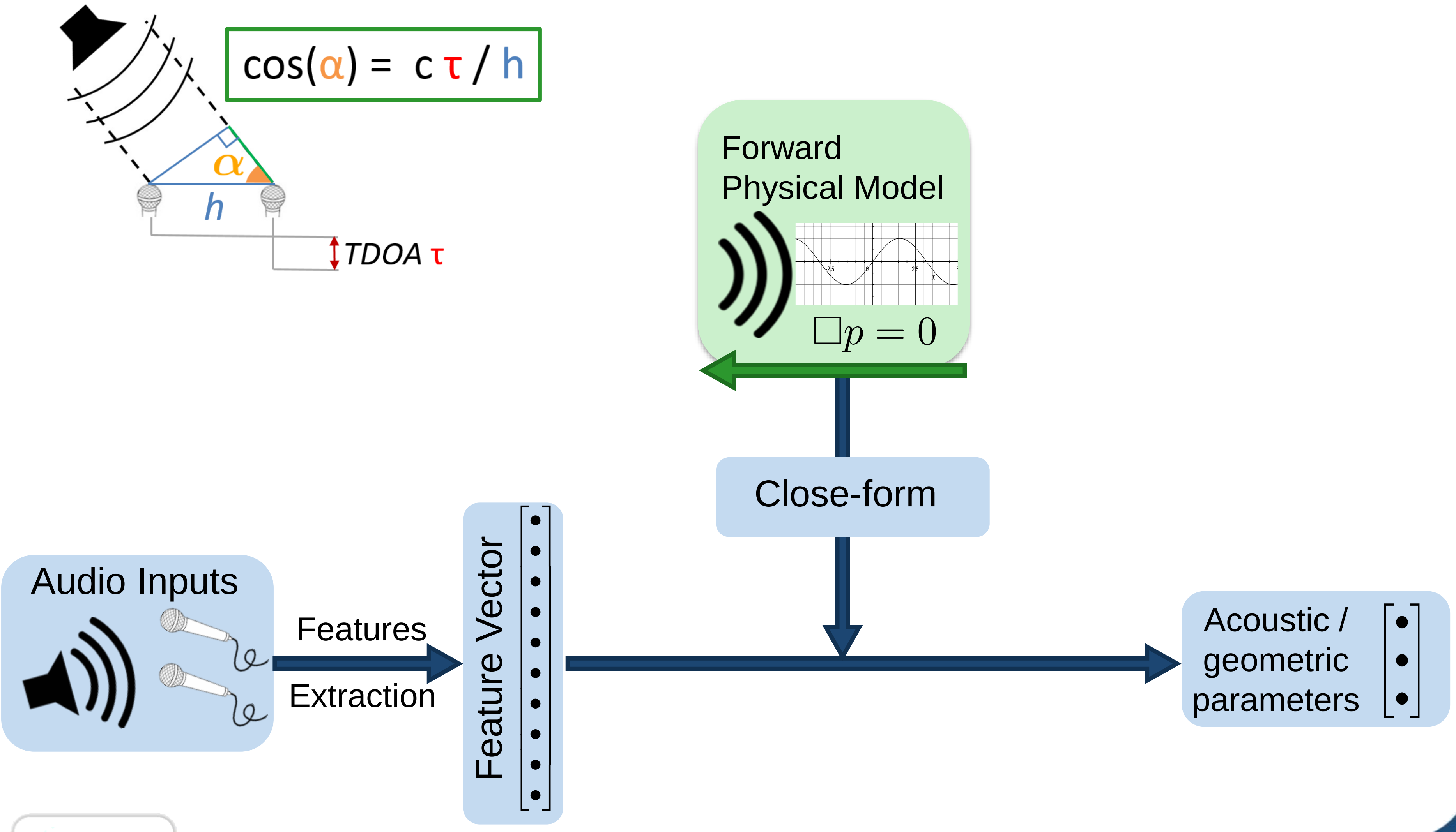
# a) Physics-Driven Approaches



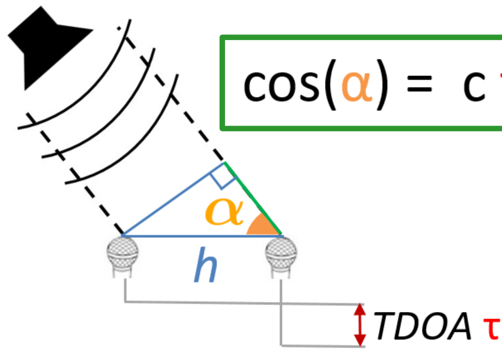
# a) Physics-Driven Approaches



# a) Physics-Driven Approaches



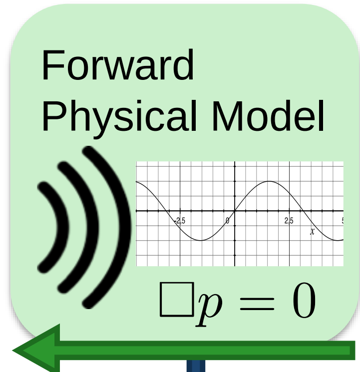
# a) Physics-Driven Approaches

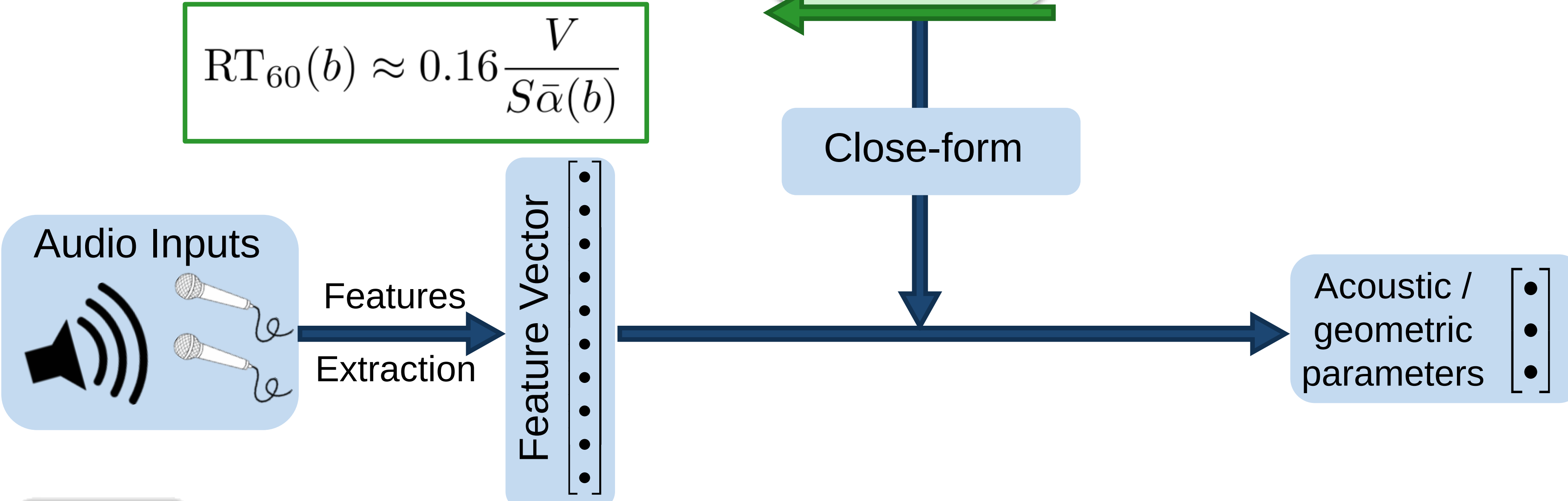



$$\cos(\alpha) = c\,\tau / h$$

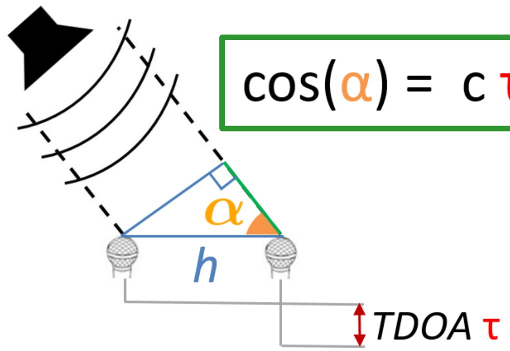
*Sabine's law:*

$$\mathrm{RT}_{60}(b) \approx 0.16 \frac{V}{S\bar{\alpha}(b)}$$



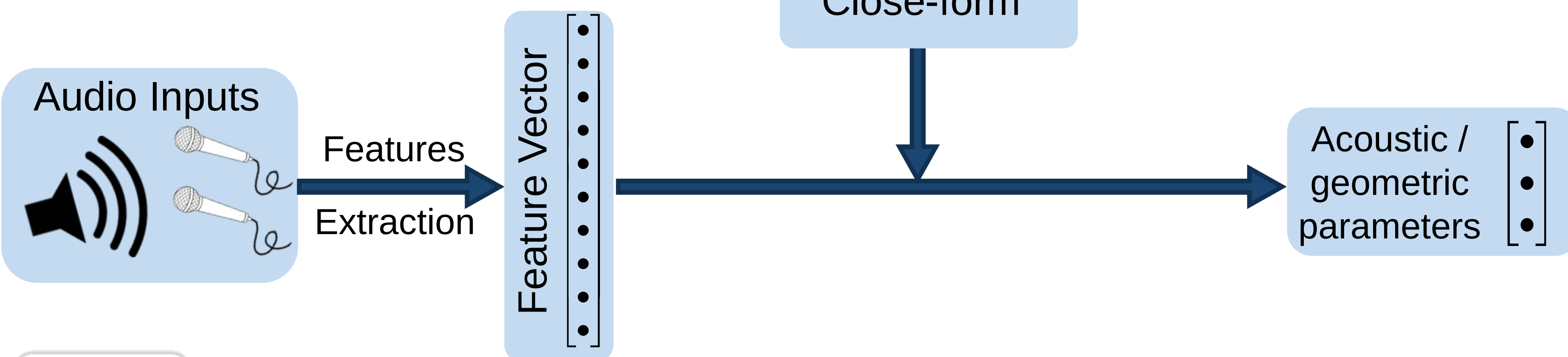
# a) Physics-Driven Approaches




$\cos(\alpha) = c\,\tau / h$

*Sabine's law:*

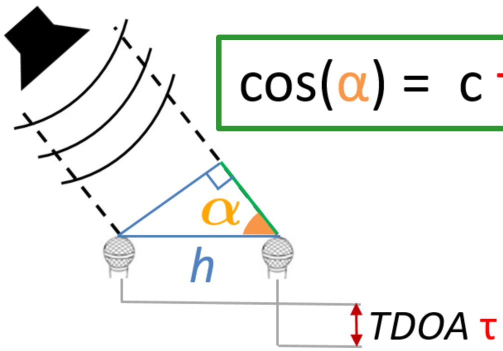
$$\mathrm{RT}_{60}(b) \approx 0.16 \frac{V}{S\bar{\alpha}(b)}$$




✓ No training data needed

# a) Physics-Driven Approaches

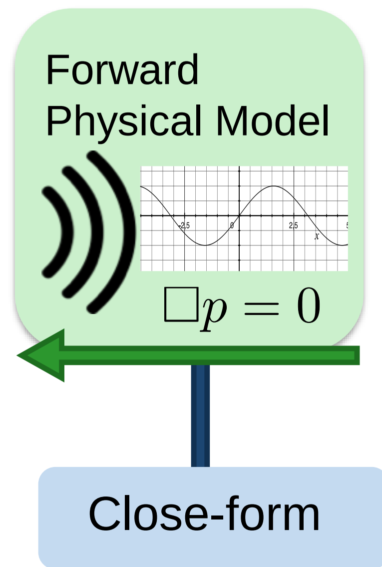
$$\cos(\alpha) = c\,\tau / h$$

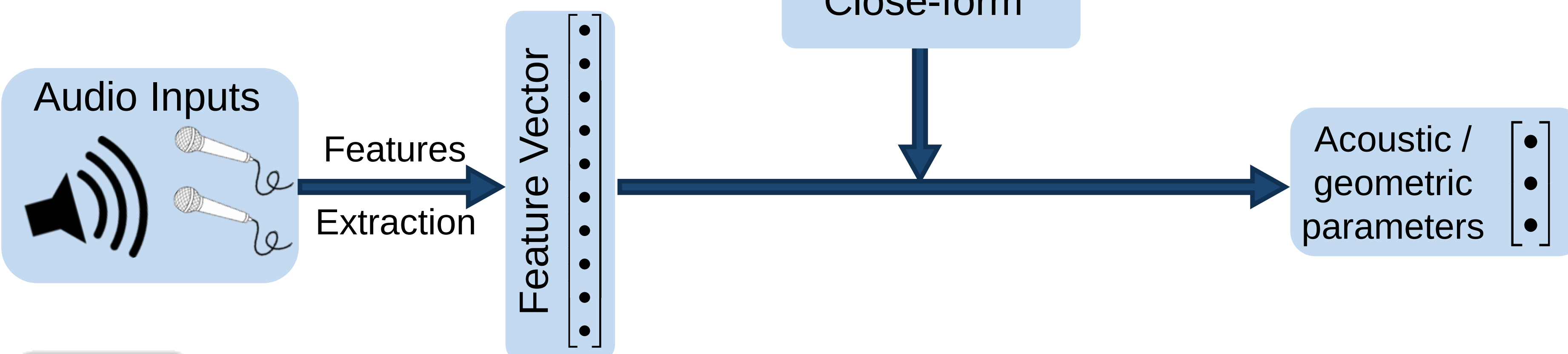



*Sabine's law:*

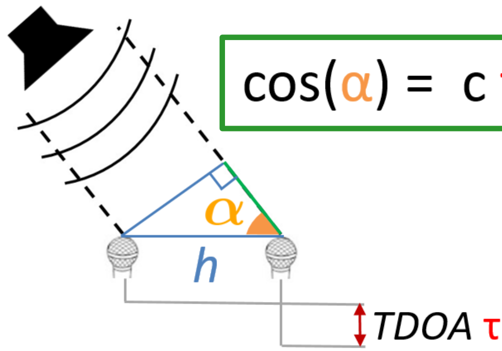
$$\mathrm{RT}_{60}(b) \approx 0.16 \frac{V}{S\bar{\alpha}(b)}$$

- ✓ No training data needed
- ✓ Computationally efficient



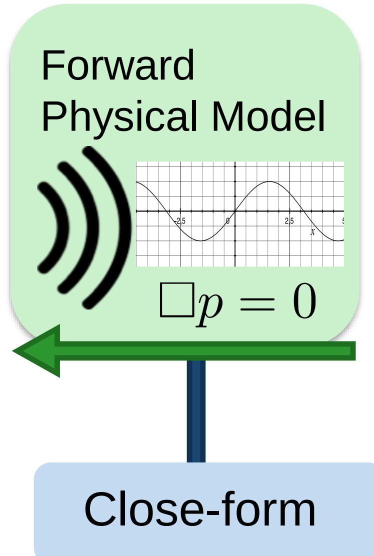
# a) Physics-Driven Approaches



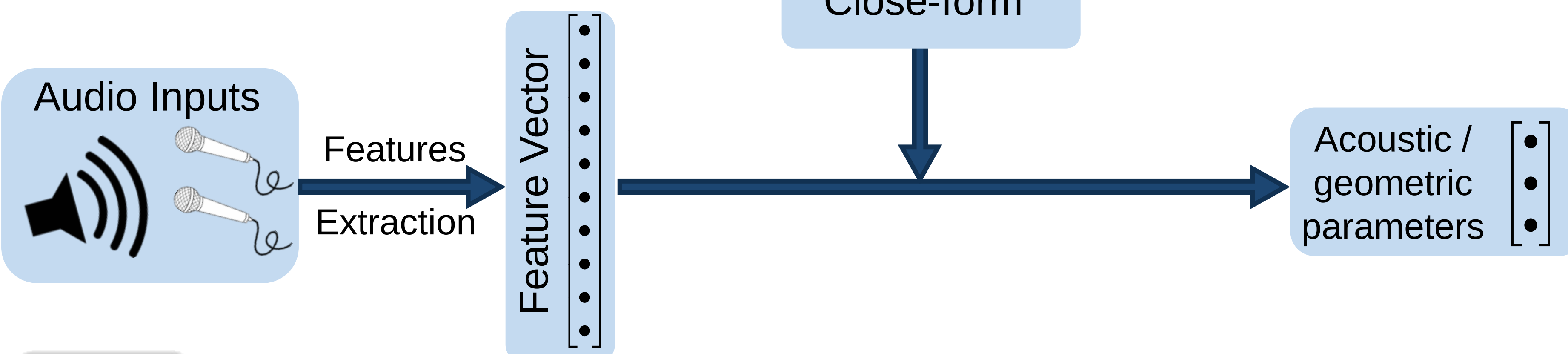

$$\cos(\alpha) = c\,\tau / h$$

*Sabine's law:*

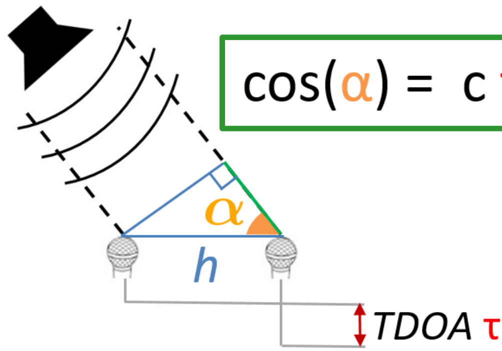
$$\mathrm{RT}_{60}(b) \approx 0.16 \frac{V}{S\bar{\alpha}(b)}$$




- ✓ No training data needed
- ✓ Computationally efficient
- ☒ Suffers in complex conditions



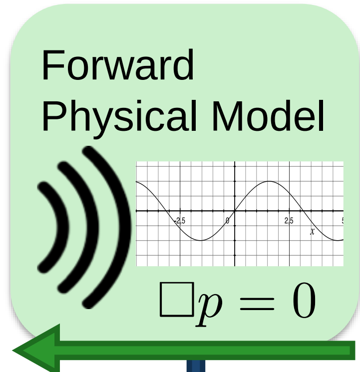
# a) Physics-Driven Approaches




$\cos(\alpha) = c\,\tau / h$

*Sabine's law:*

$$\mathrm{RT}_{60}(b) \approx 0.16 \frac{V}{S\bar{\alpha}(b)}$$




- ✓ No training data needed
- ✓ Computationally efficient
- ☒ Suffers in complex conditions
- ☒ Limited

Close-form

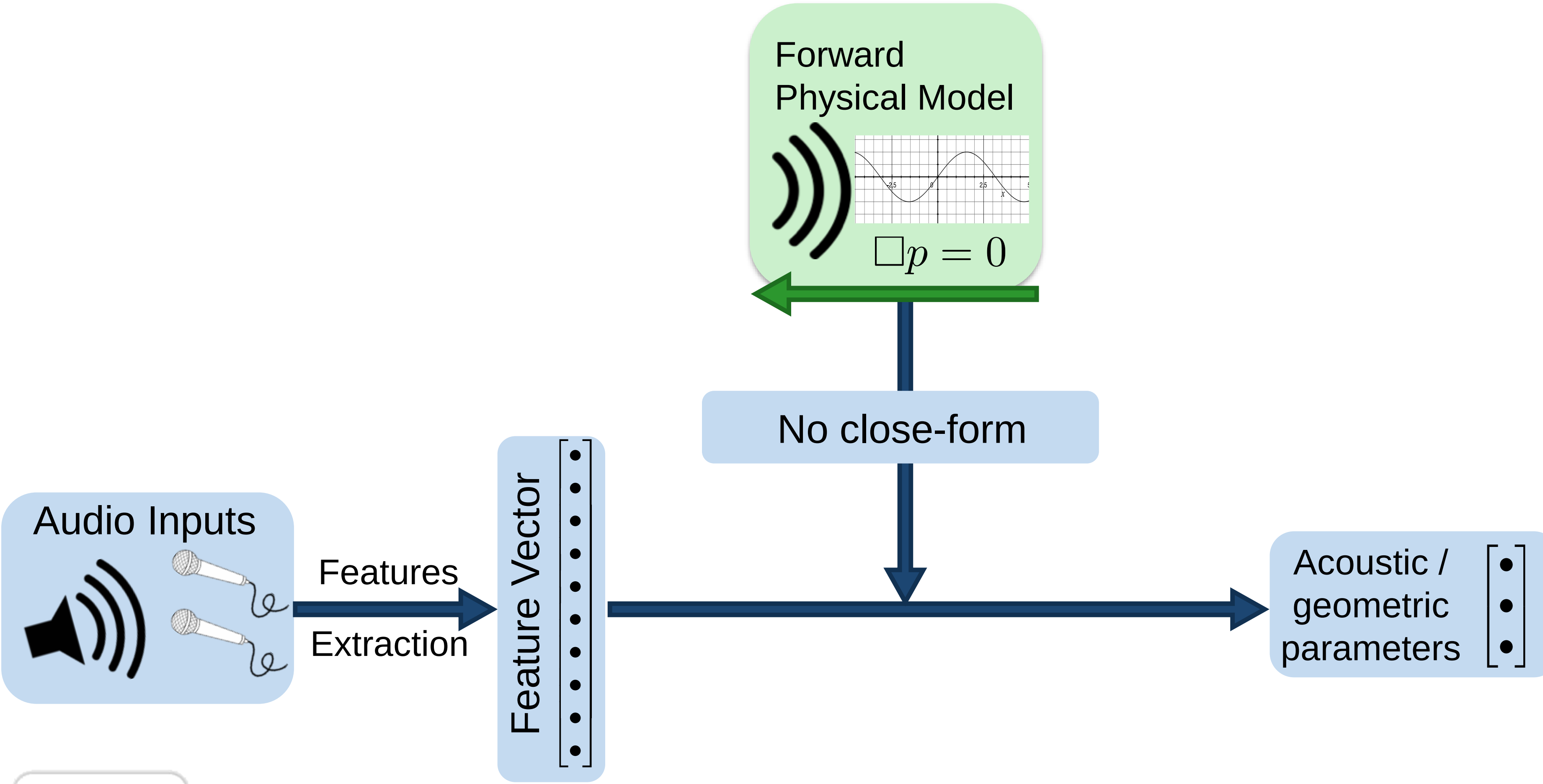
Audio Inputs

Features Extraction

Feature Vector

Acoustic / geometric parameters

## a) Physics-Driven Approaches

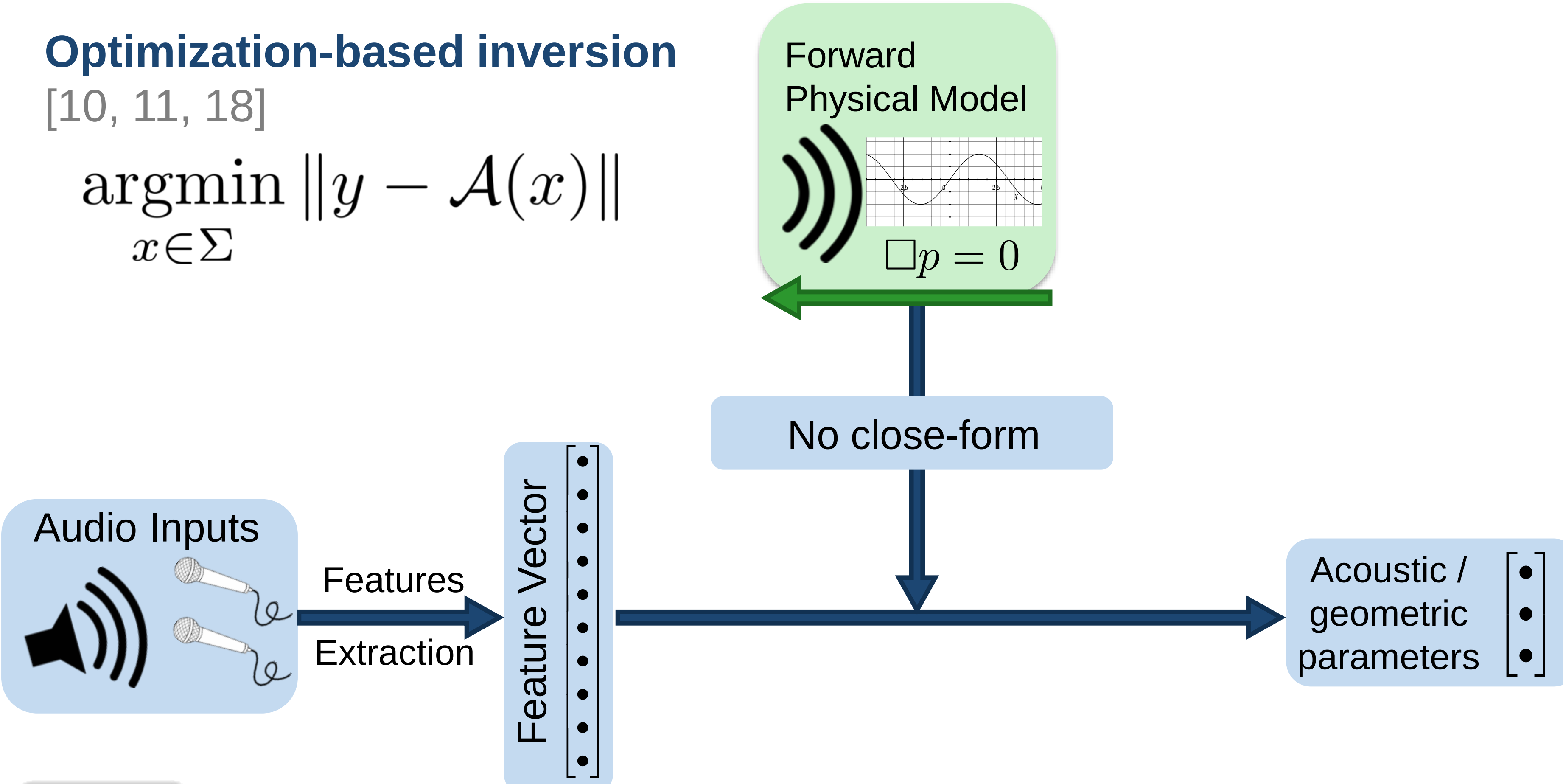


# a) Physics-Driven Approaches

**Optimization-based inversion**

[10, 11, 18]

$$\underset{x\in\Sigma}{\operatorname{argmin}} \|y - \mathcal{A}(x)\|$$

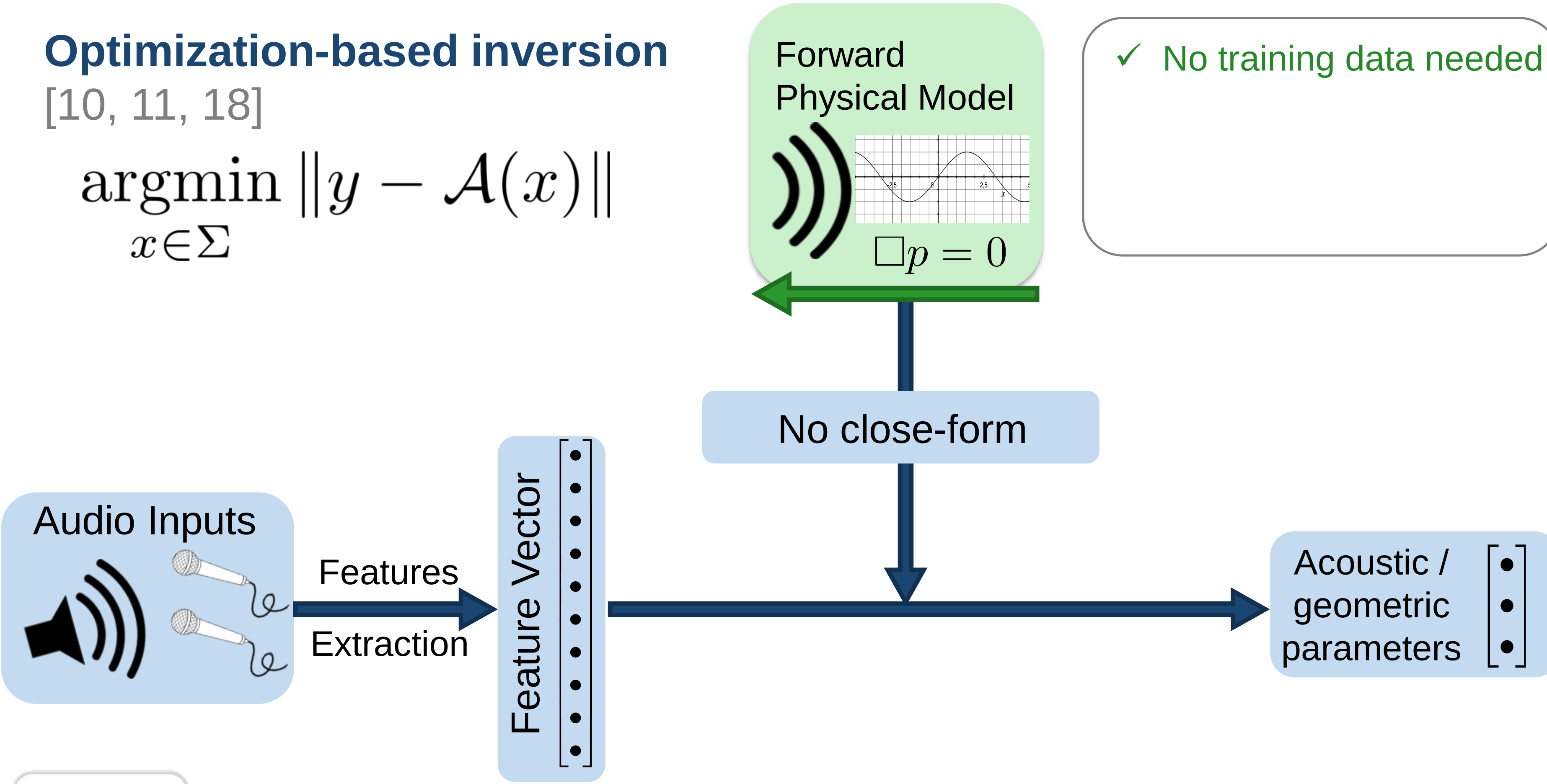


# a) Physics-Driven Approaches

**Optimization-based inversion**
[10, 11, 18]

$$\underset{x \in \Sigma}{\operatorname{argmin}} \|y - \mathcal{A}(x)\|$$

✓ No training data needed

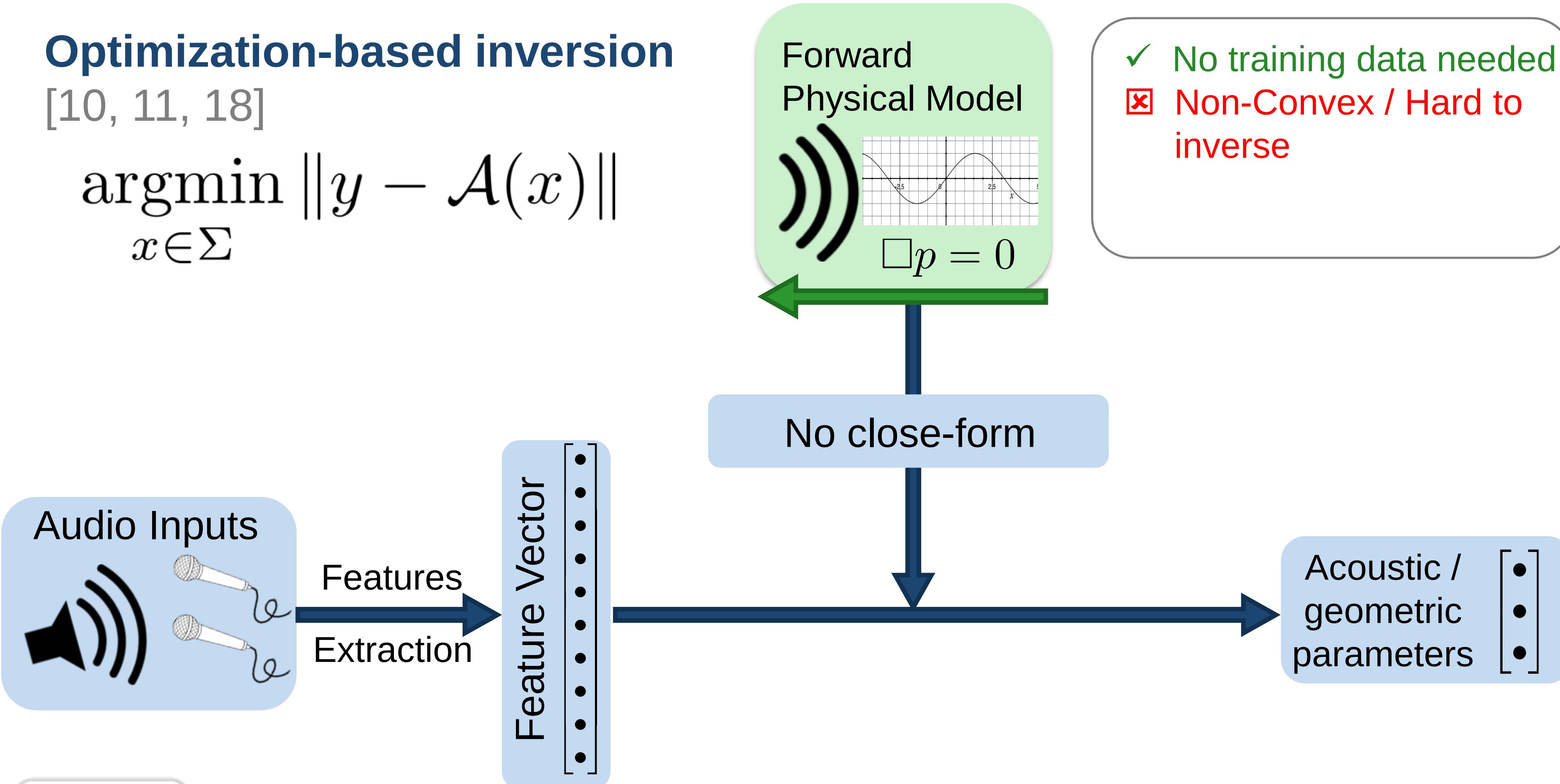


# a) Physics-Driven Approaches

**Optimization-based inversion**
[10, 11, 18]

$$\underset{x \in \Sigma}{\operatorname{argmin}} \|y - \mathcal{A}(x)\|$$

- ✓ No training data needed
- ☒ Non-Convex / Hard to inverse



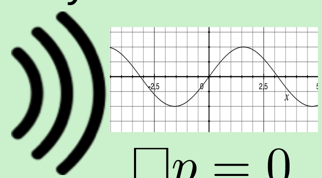
# a) Physics-Driven Approaches

**Optimization-based inversion**

[10, 11, 18]

$$\underset{x \in \Sigma}{\operatorname{argmin}} \|y - \mathcal{A}(x)\|$$

Forward Physical Model



$\Box p = 0$

- ✓ No training data needed
- ☒ Non-Convex / Hard to inverse
- ☒ Sensitive to model mismatch

No close-form

Audio Inputs

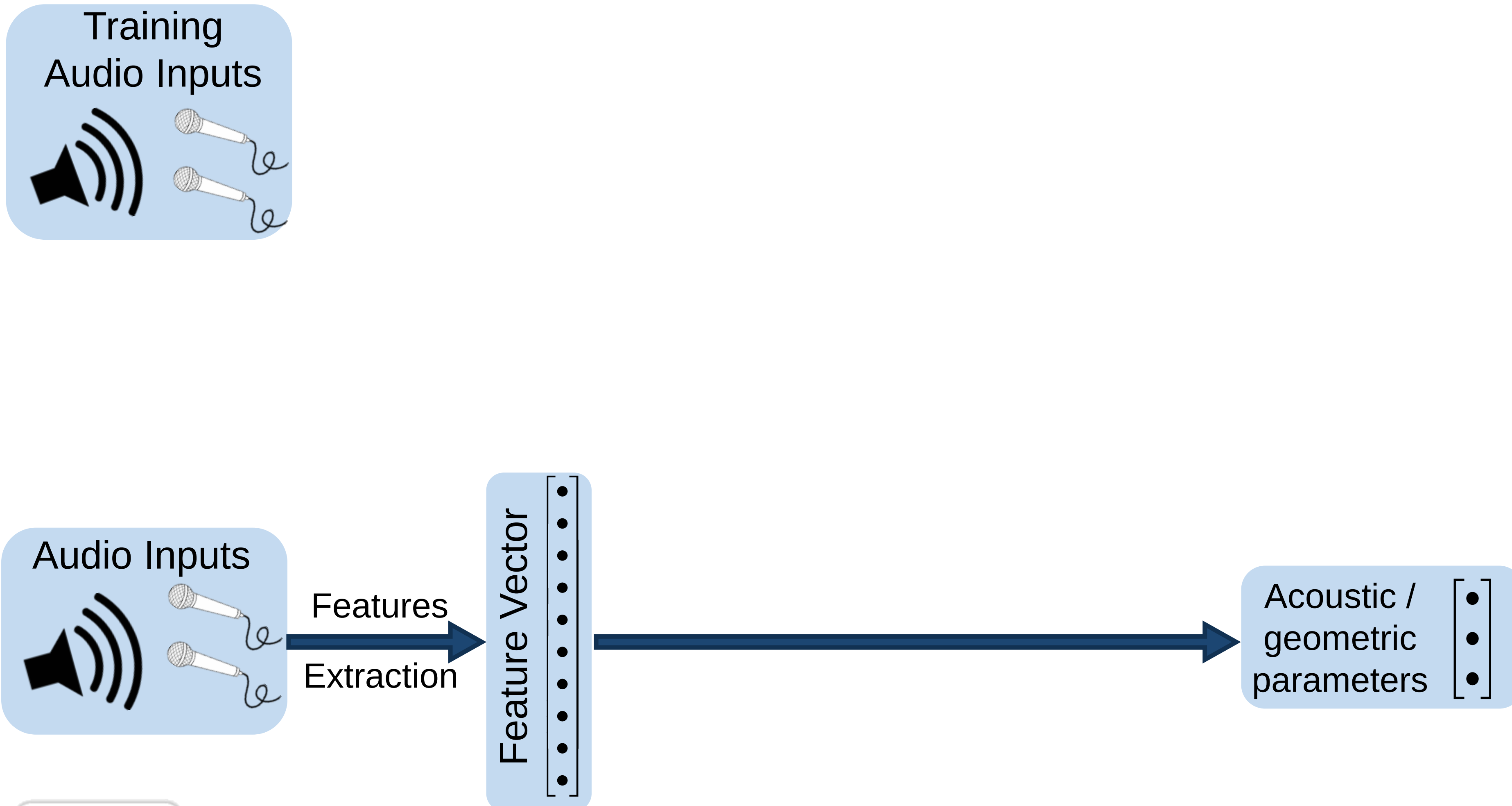
Features Extraction

Feature Vector

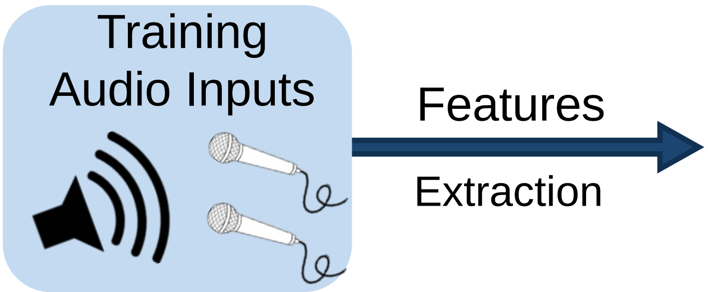
Acoustic / geometric parameters

## b) Real-Data-Driven Approaches [1, 2, 3, 6]

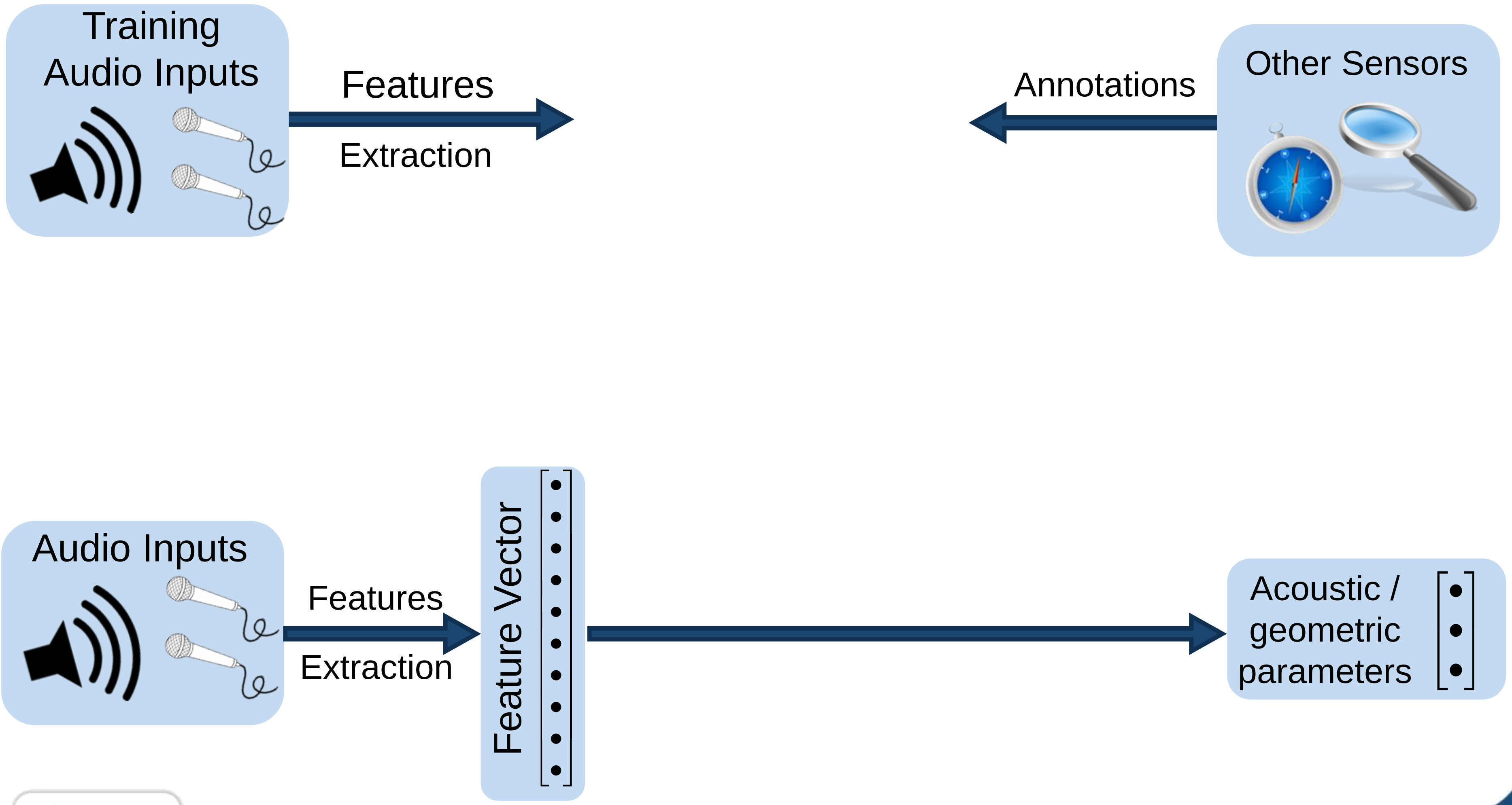
## b) Real-Data-Driven Approaches [1, 2, 3, 6]



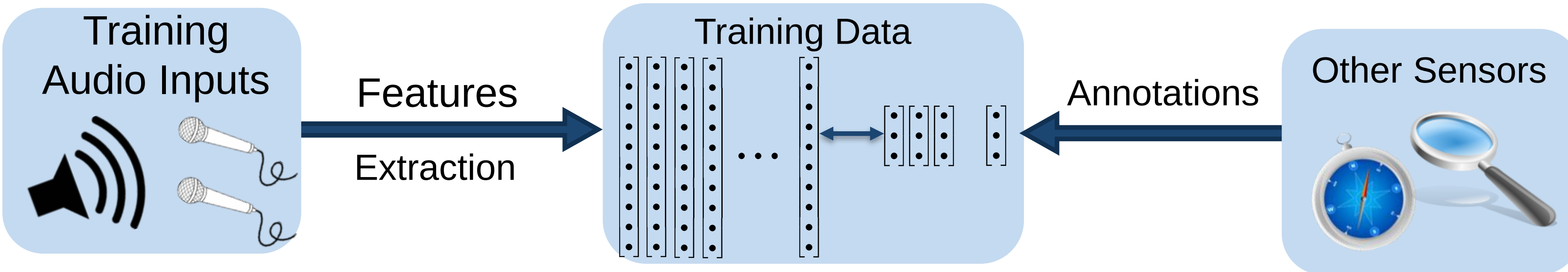
# b) Real-Data-Driven Approaches [1, 2, 3, 6]



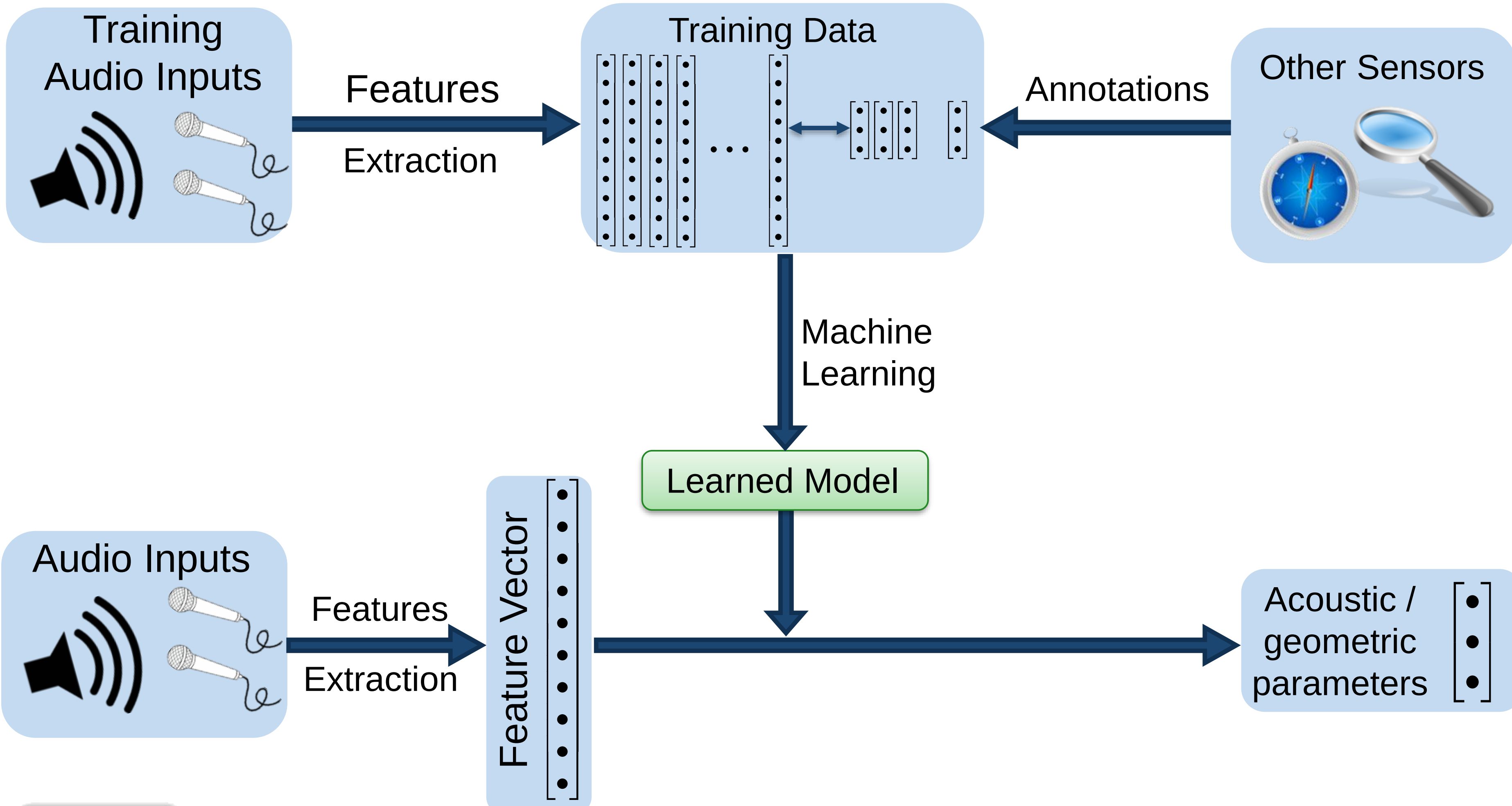

## b) Real-Data-Driven Approaches [1, 2, 3, 6]



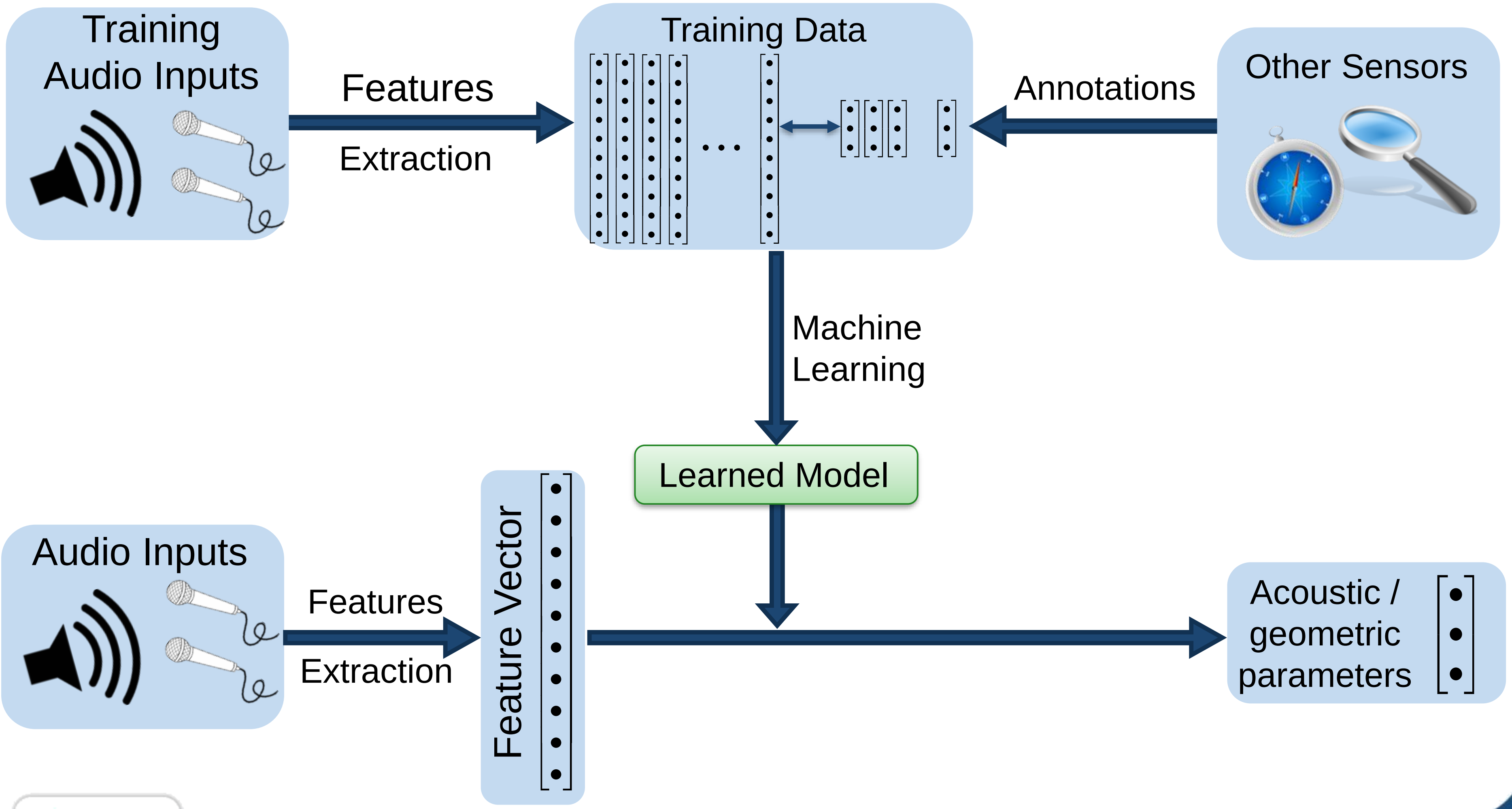
## b) Real-Data-Driven Approaches [1, 2, 3, 6]



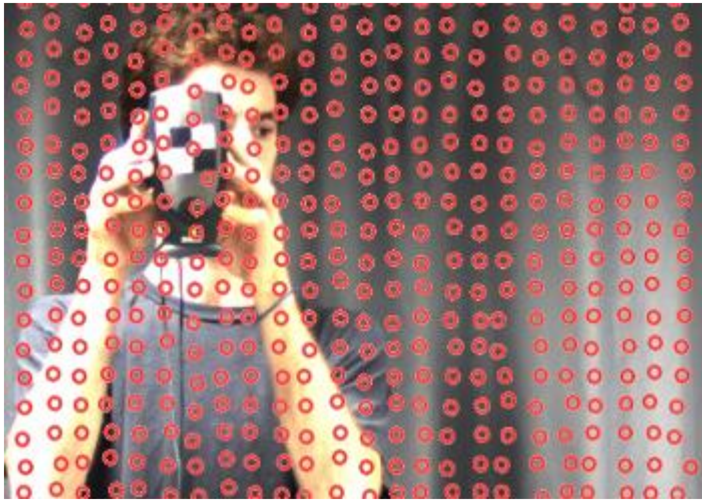

# b) Real-Data-Driven Approaches [1, 2, 3, 6]



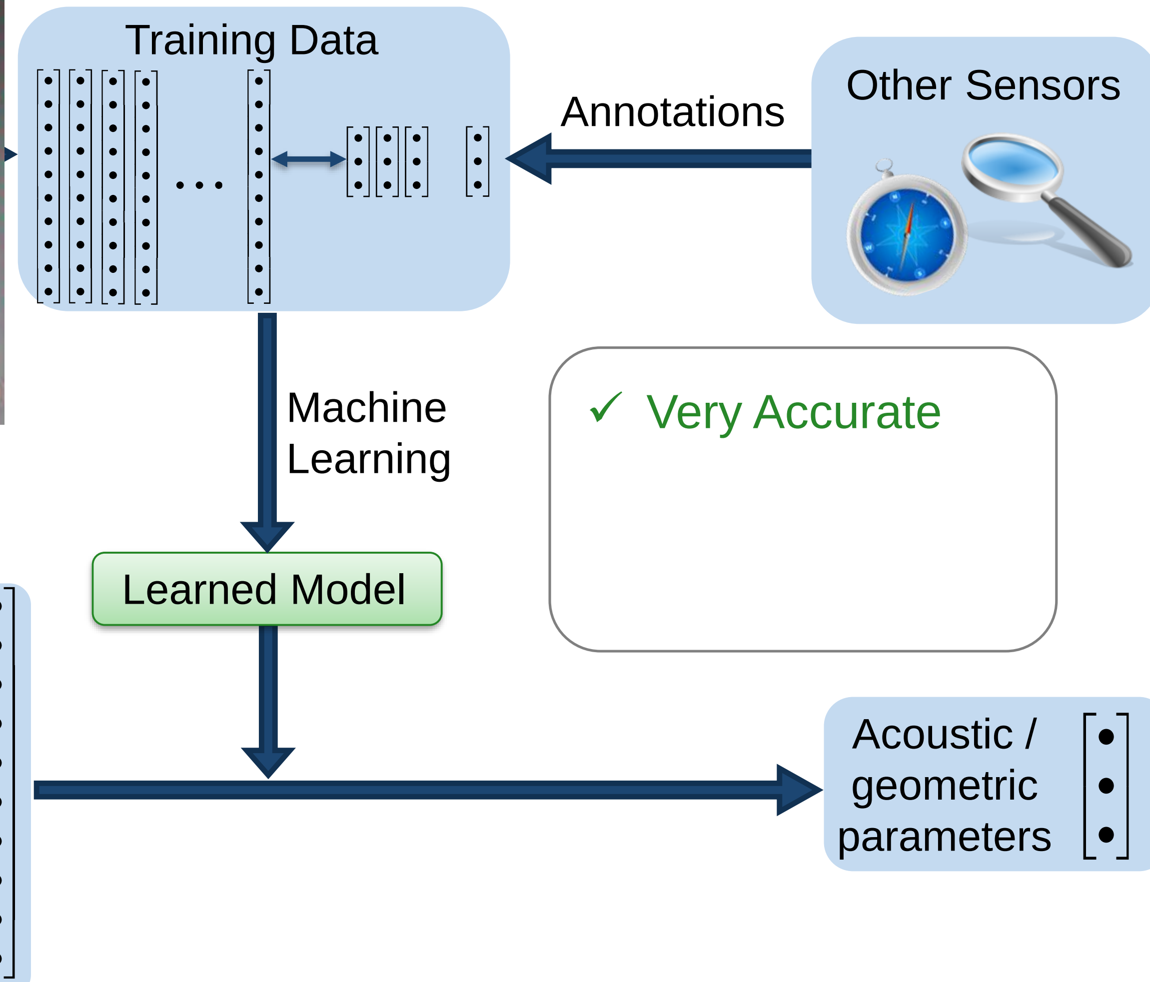
## b) Real-Data-Driven Approaches [1, 2, 3, 6]



## b) Real-Data-Driven Approaches [1, 2, 3, 6]

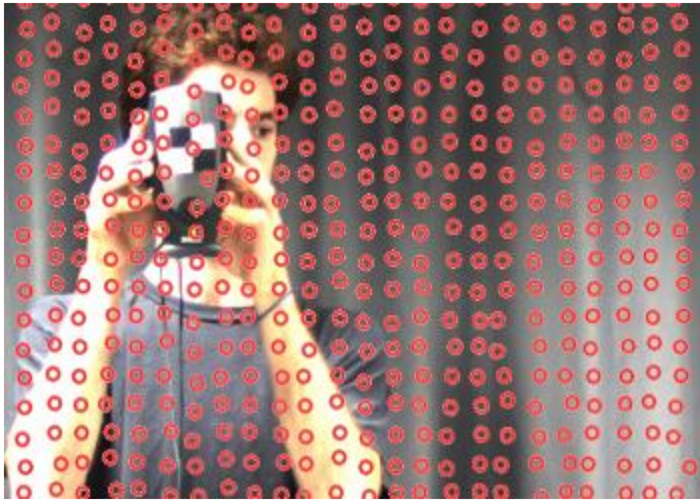


Check on Youtube:
https://youtu.be/mhOlcVpY7iA




✓ Very Accurate

# b) Real-Data-Driven Approaches [1, 2, 3, 6]



Check on Youtube:
https://youtu.be/mhOlcVpY7iA

Training Data

Annotations

Other Sensors

Machine Learning

Learned Model

- ✓ Very Accurate
- ✓ Noise-Robust

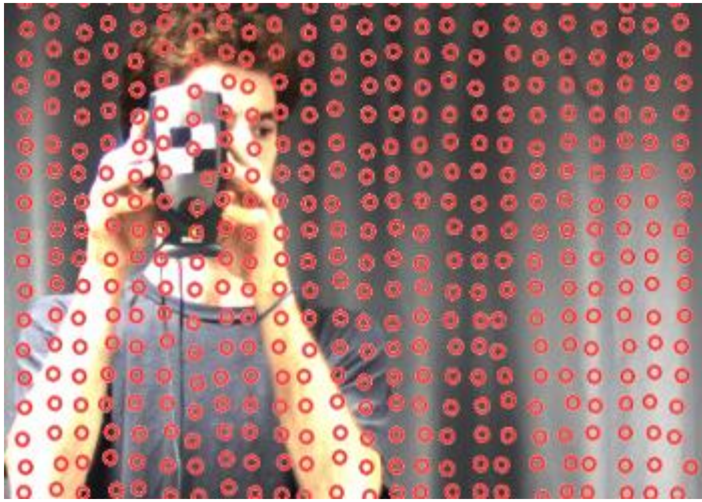
Audio Inputs

Features Extraction

Feature Vector

Acoustic / geometric parameters

# b) Real-Data-Driven Approaches [1, 2, 3, 6]



Check on Youtube:
https://youtu.be/mhOlcVpY7iA

Training Data

Annotations

Other Sensors

Machine Learning

Learned Model

- ✓ Very Accurate
- ✓ Noise-Robust
- ☒ Room-specific

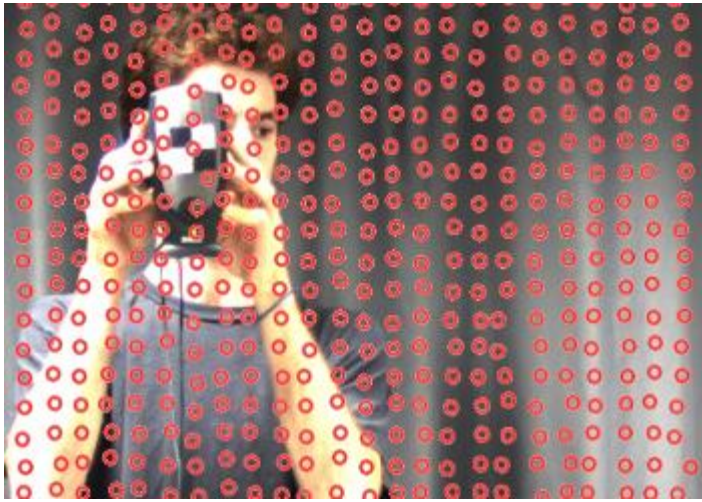
Audio Inputs

Features Extraction

Feature Vector

Acoustic / geometric parameters

## b) Real-Data-Driven Approaches [1, 2, 3, 6]



Check on Youtube:
https://youtu.be/mhOlcVpY7iA

Training Data

Annotations

Other Sensors

Machine Learning

Learned Model

- ✓ Very Accurate
- ✓ Noise-Robust
- ☒ Room-specific
- ☒ Costly to aquire

Audio Inputs

Features Extraction

Feature Vector

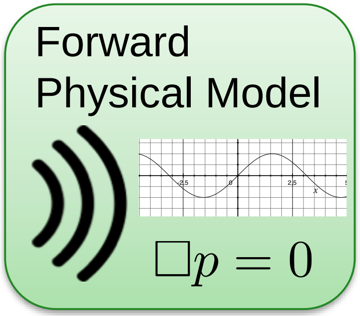
Acoustic / geometric parameters

## c) *Virtually-Supervised Learning*

[4, 5, 9, 16, 17]

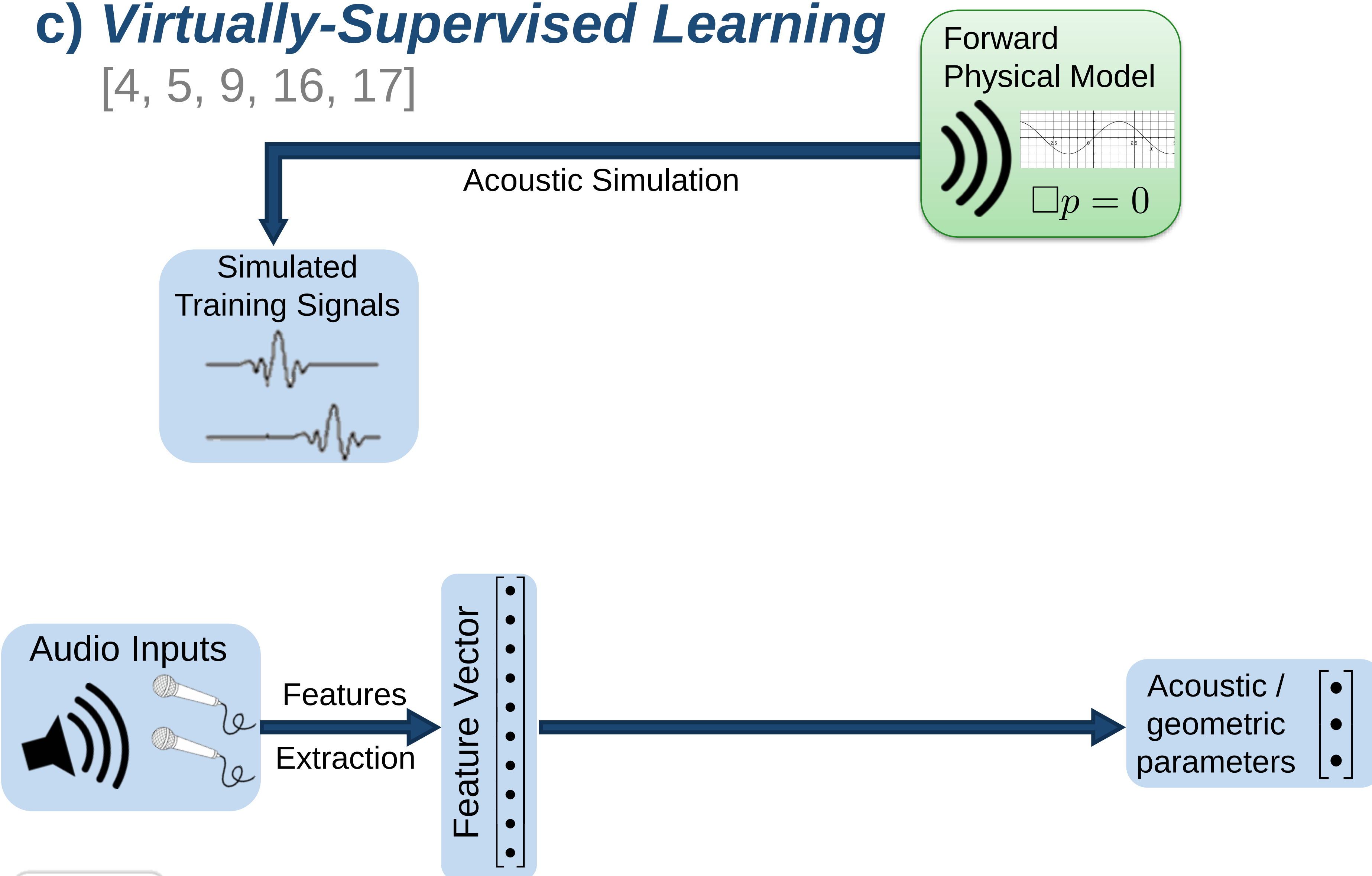
## c) *Virtually-Supervised Learning*

[4, 5, 9, 16, 17]



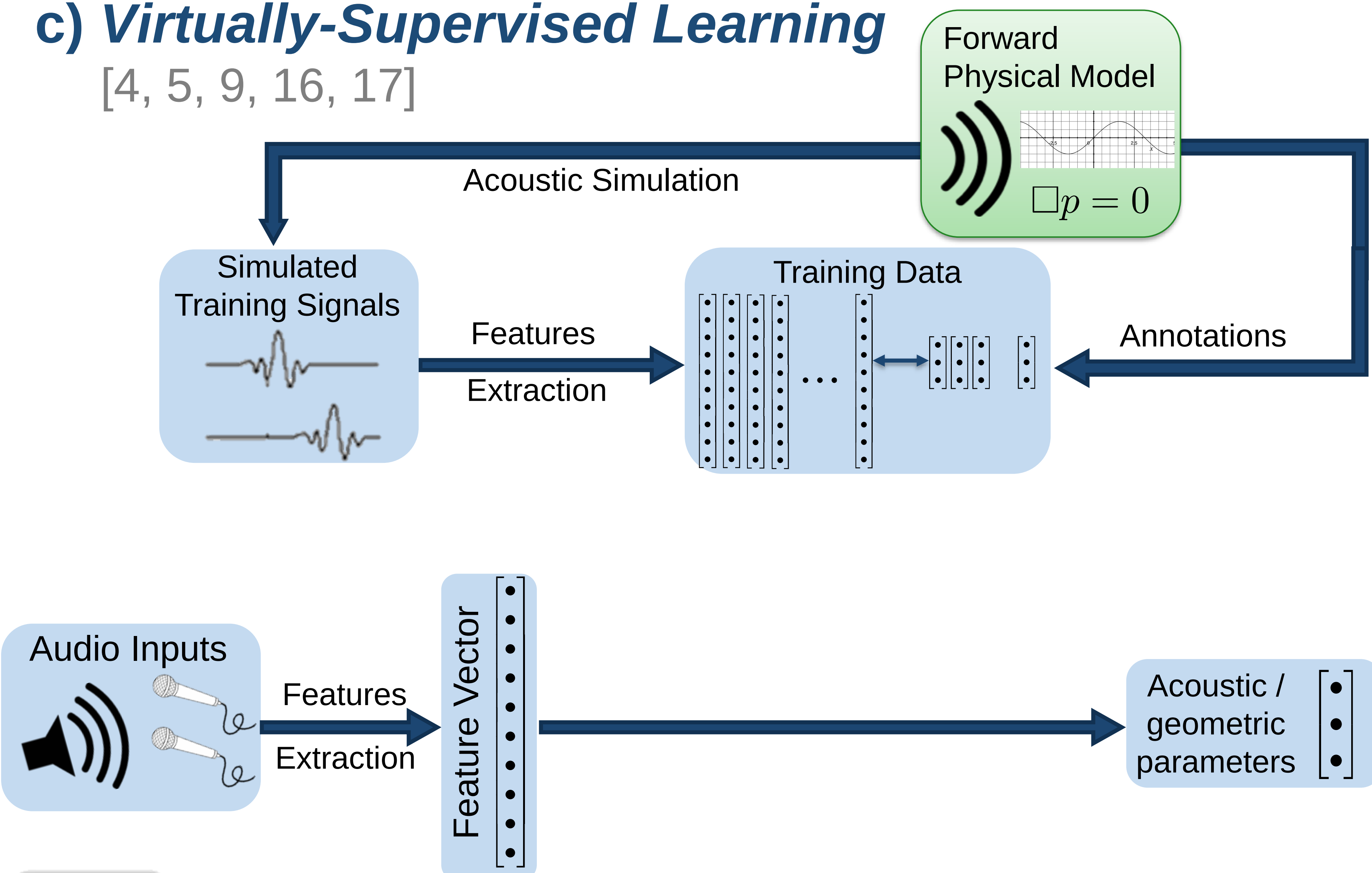

## c) *Virtually-Supervised Learning*

[4, 5, 9, 16, 17]



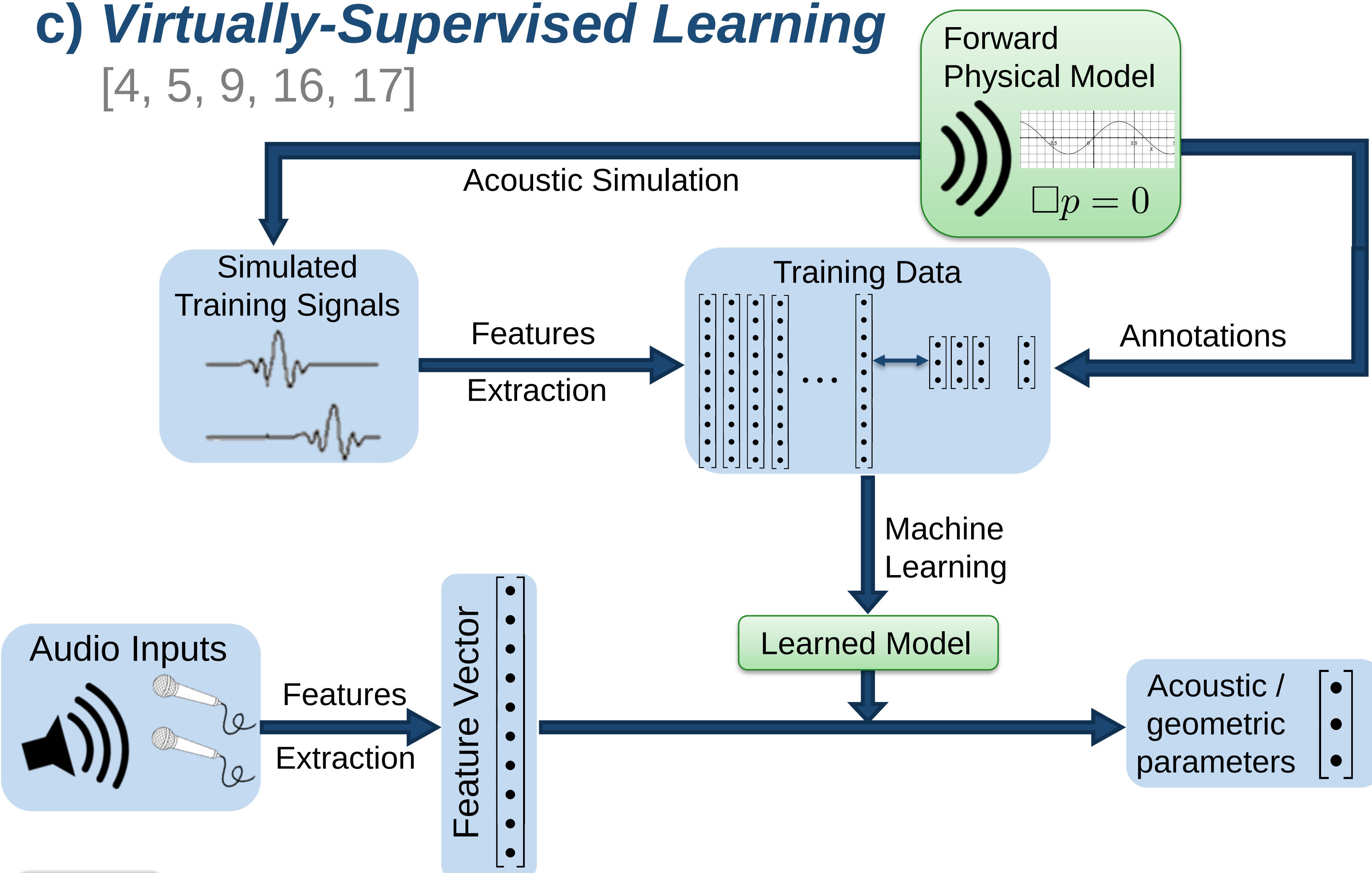
## c) *Virtually-Supervised Learning*

[4, 5, 9, 16, 17]



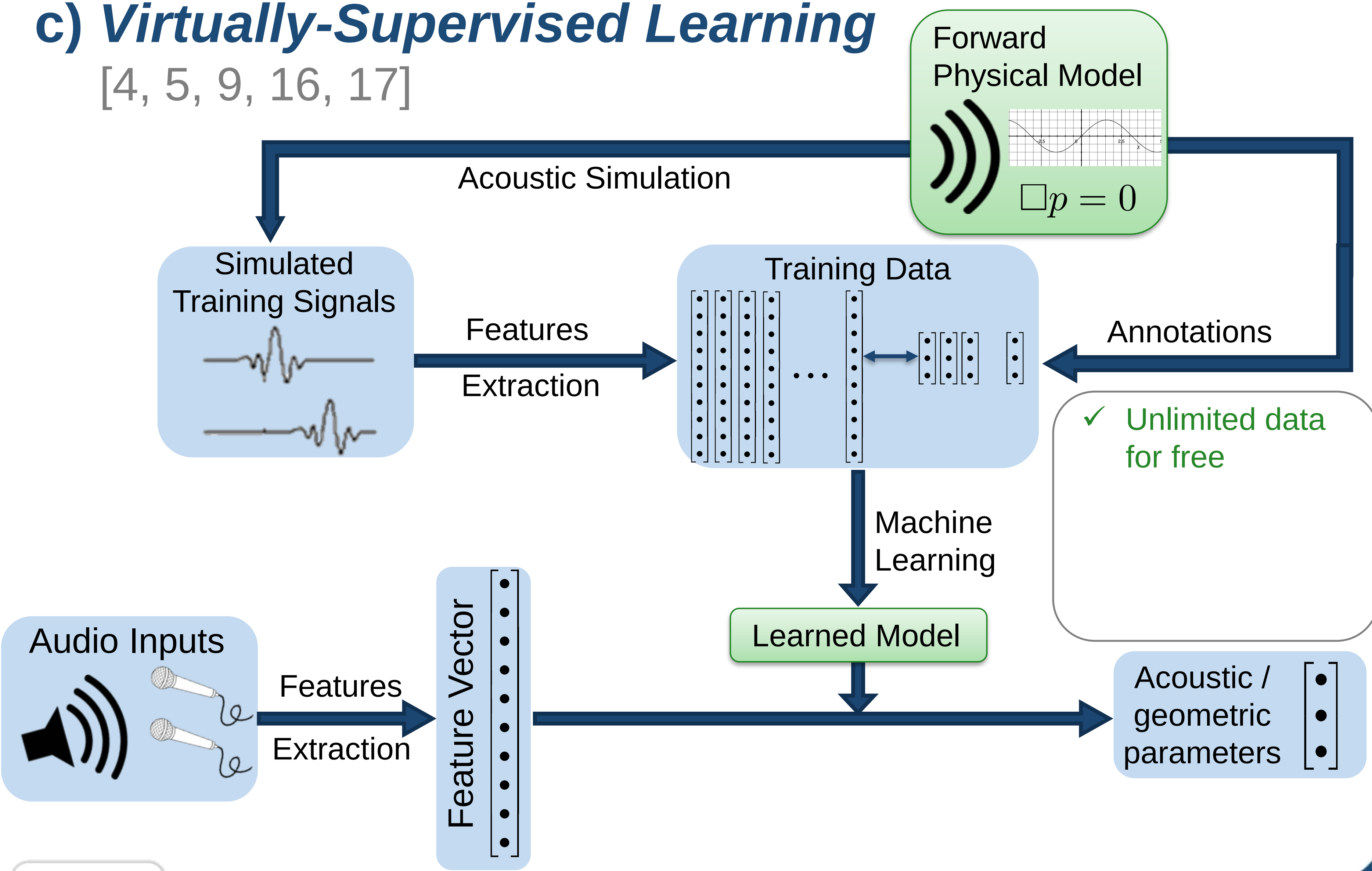
## c) *Virtually-Supervised Learning*

[4, 5, 9, 16, 17]



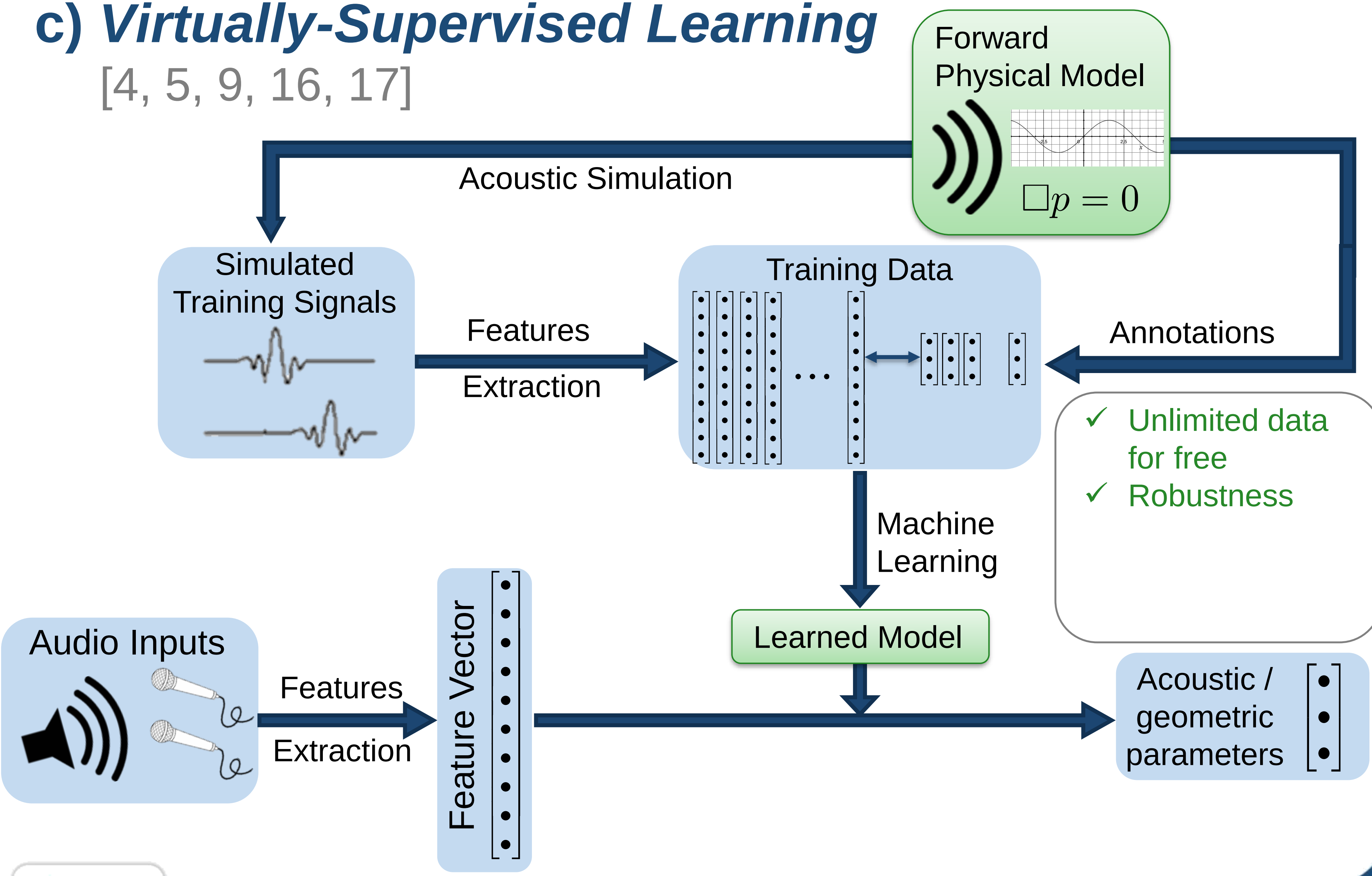
## c) *Virtually-Supervised Learning*

[4, 5, 9, 16, 17]



## c) *Virtually-Supervised Learning*

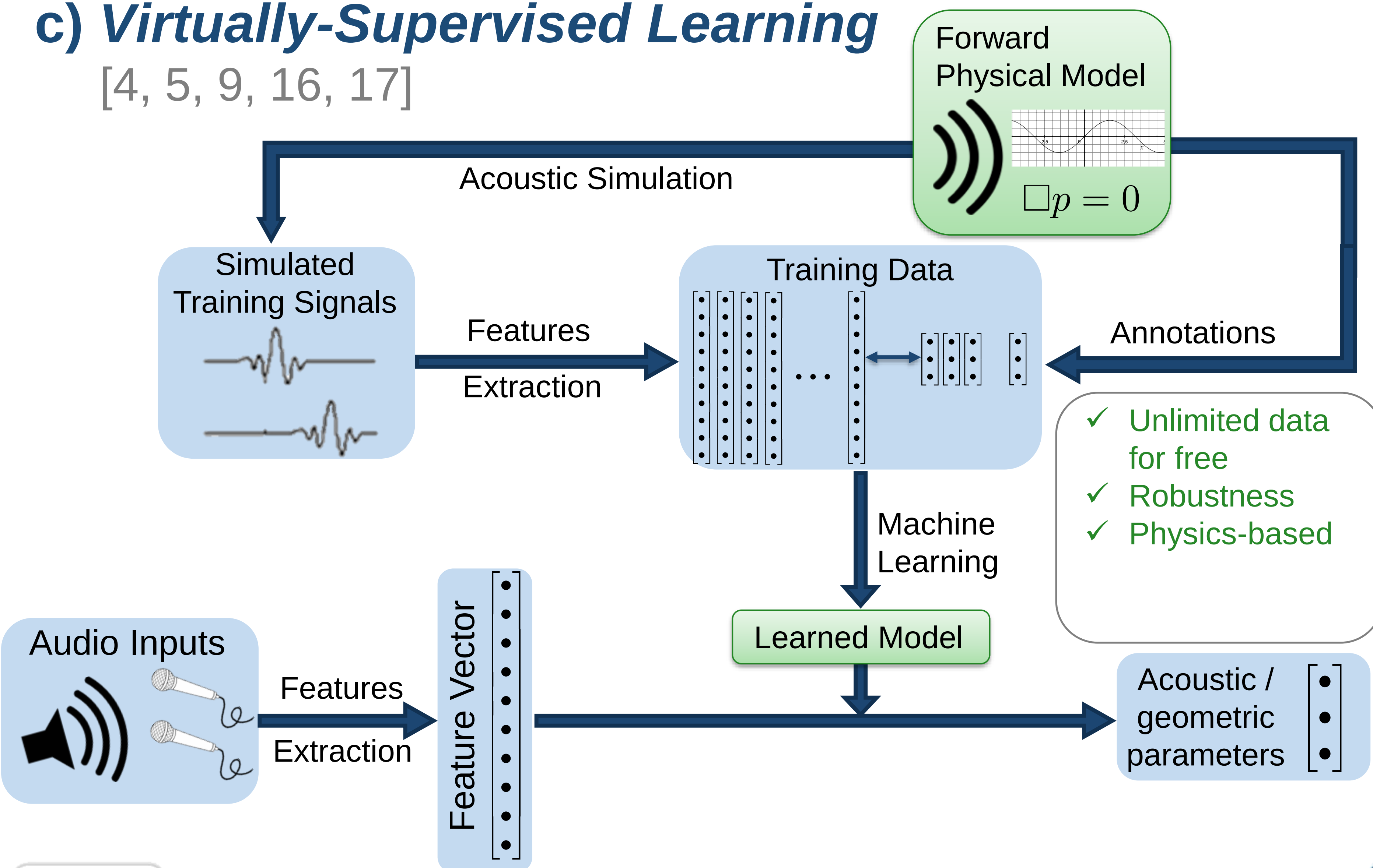
[4, 5, 9, 16, 17]




✓ Unlimited data for free
✓ Robustness

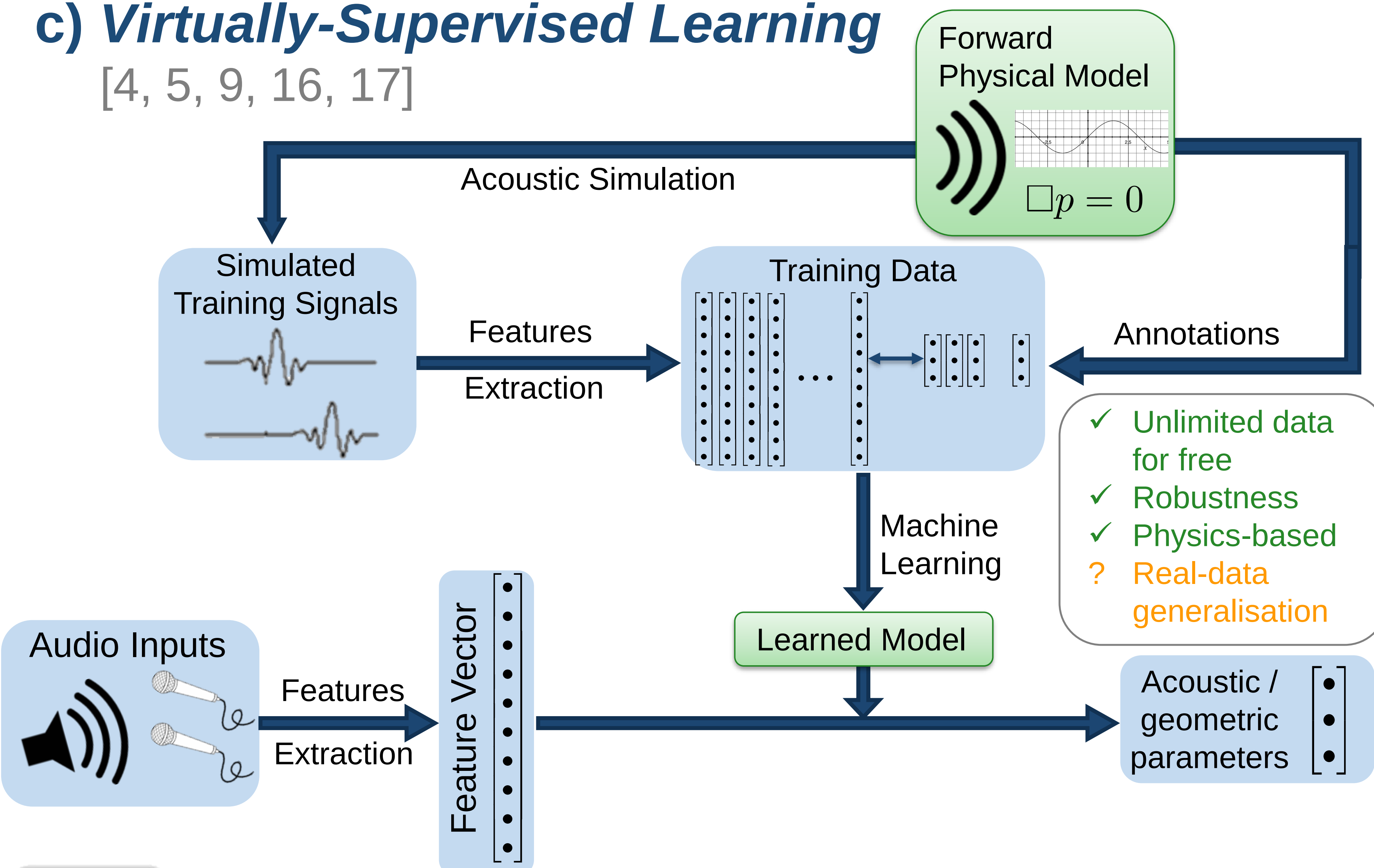
## c) *Virtually-Supervised Learning*

[4, 5, 9, 16, 17]



## c) *Virtually-Supervised Learning*

[4, 5, 9, 16, 17]




- ✓ Unlimited data for free
- ✓ Robustness
- ✓ Physics-based
- ? Real-data generalisation

# RIR Simulation Trade-offs

| Realism vs. Computational complexity | Diversity vs. Training set size |
| --- | --- |

# RIR Simulation Trade-offs

| Realism vs. Computational complexity | Diversity vs. Training set size |
| --- | --- |
| • Discretized wave equation solvers (e.g. FDTD) $$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$ | |

# RIR Simulation Trade-offs

| Realism vs. Computational complexity | Diversity vs. Training set size |
|---|---|
| • Discretized wave equation solvers (e.g. FDTD) $\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$ ✓ Solve everything ☒ Intractable above ~4 kHz | |

# RIR Simulation Trade-offs

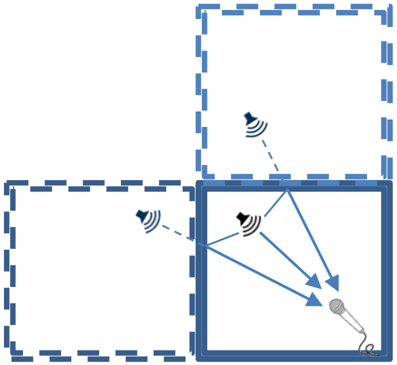
| **Realism vs. Computational complexity** | **Diversity vs. Training set size** |
|---|---|
| • Discretized wave equation solvers (e.g. FDTD) | |

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

✓ Solve everything

☒ Intractable above ~4 kHz

• Image source method [13]



# RIR Simulation Trade-offs

| Realism vs. Computational complexity | Diversity vs. Training set size |
|---|---|

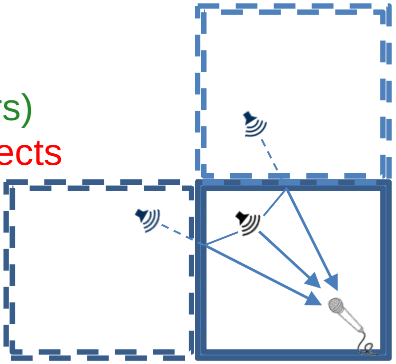
- Discretized wave equation solvers (e.g. FDTD)

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

✓ Solve everything
☒ Intractable above ~4 kHz

- Image source method [13]

✓ Fast (for low reflection orders)
☒ Doesn't capture low-freq effects
☒ Specular reflections only



# RIR Simulation Trade-offs

| Realism vs. Computational complexity | Diversity vs. Training set size |
|---|---|

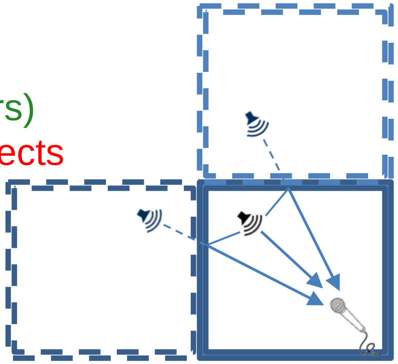
- Discretized wave equation solvers (e.g. FDTD)

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

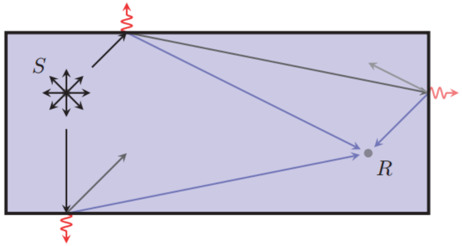
✓ Solve everything
☒ Intractable above ~4 kHz

- Image source method [13]

✓ Fast (for low reflection orders)
☒ Doesn't capture low-freq effects
☒ Specular reflections only



- Energy-based / Ray-based / Particle-based methods



# RIR Simulation Trade-offs

### Realism vs. Computational complexity

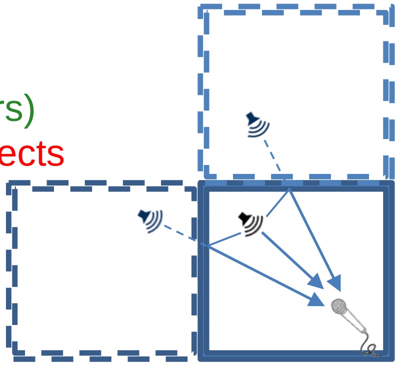
- Discretized wave equation solvers (e.g. FDTD)

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

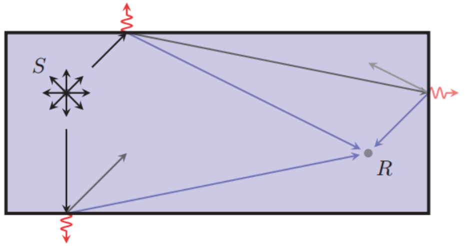
✓ Solve everything
☒ Intractable above ~4 kHz

- Image source method [13]

✓ Fast (for low reflection orders)
☒ Doesn’t capture low-freq effects
☒ Specular reflections only



- Energy-based / Ray-based / Particle-based methods




✓ Versatile
☒ Doesn’t capture low-freq effects
☒ Approx. TOAs

### Diversity vs. Training set size

# RIR Simulation Trade-offs

| Realism vs. Computational complexity | Diversity vs. Training set size |
|---|---|

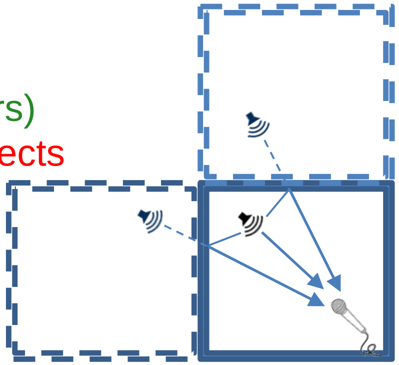
- Discretized wave equation solvers (e.g. FDTD)

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

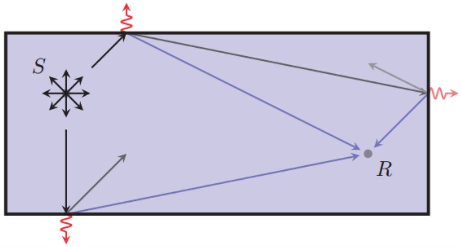
✓ Solve everything
☒ Intractable above ~4 kHz

- Image source method [13]

✓ Fast (for low reflection orders)
☒ Doesn't capture low-freq effects
☒ Specular reflections only



- Energy-based / Ray-based / Particle-based methods




✓ Versatile
☒ Doesn't capture low-freq effects
☒ Approx. TOAs

- **Simulators efficiently combining the last two:**
RoomSim [14], Pyroomacoustics [15]

# RIR Simulation Trade-offs

## Realism vs. Computational complexity

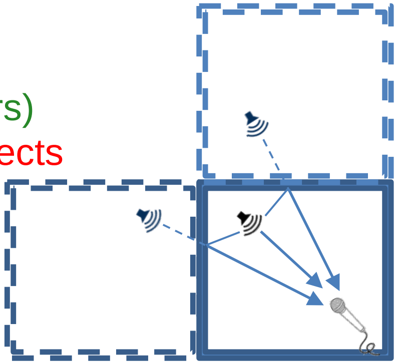
- Discretized wave equation solvers (e.g. FDTD)

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

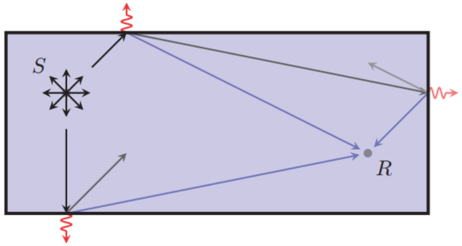
✓ Solve everything
☒ Intractable above ~4 kHz

- Image source method [13]

✓ Fast (for low reflection orders)
☒ Doesn't capture low-freq effects
☒ Specular reflections only



- Energy-based / Ray-based / Particle-based methods




✓ Versatile
☒ Doesn't capture low-freq effects
☒ Approx. TOAs

- **Simulators efficiently combining the last two:** RoomSim [14], Pyroomacoustics [15]

## Diversity vs. Training set size

- Room size? *Toilet, Office, Airport Hall*

# RIR Simulation Trade-offs

## Realism vs. Computational complexity

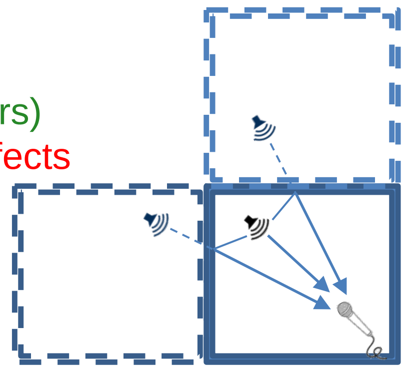
- Discretized wave equation solvers (e.g. FDTD)

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

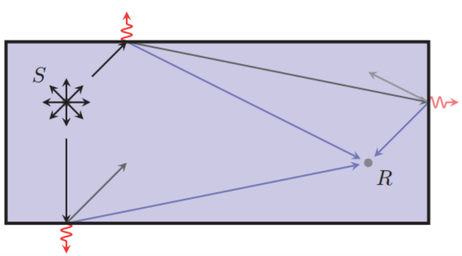
✓ Solve everything
☒ Intractable above ~4 kHz

- Image source method [13]

✓ Fast (for low reflection orders)
☒ Doesn't capture low-freq effects
☒ Specular reflections only



- Energy-based / Ray-based / Particle-based methods




✓ Versatile
☒ Doesn't capture low-freq effects
☒ Approx. TOAs

- **Simulators efficiently combining the last two:**
RoomSim [14], Pyroomacoustics [15]

## Diversity vs. Training set size

- Room size? *Toilet, Office, Airport Hall*
- Room shape? *Shoebox, Auditorium, Underground cave*

# RIR Simulation Trade-offs

## Realism vs. Computational complexity

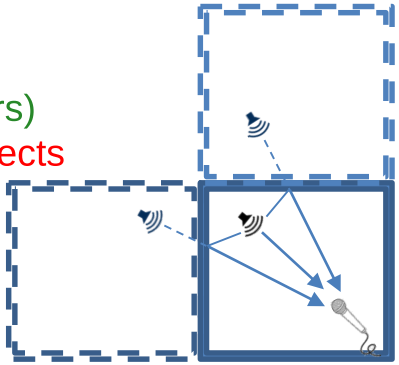
- Discretized wave equation solvers (e.g. FDTD)

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

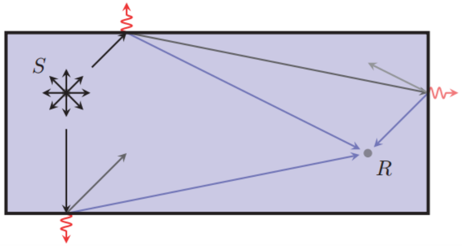
✓ Solve everything
☒ Intractable above ~4 kHz

- Image source method [13]

✓ Fast (for low reflection orders)
☒ Doesn't capture low-freq effects
☒ Specular reflections only



- Energy-based / Ray-based / Particle-based methods




✓ Versatile
☒ Doesn't capture low-freq effects
☒ Approx. TOAs

- **Simulators efficiently combining the last two:** RoomSim [14], Pyroomacoustics [15]

## Diversity vs. Training set size

- Room size? *Toilet, Office, Airport Hall*
- Room shape? *Shoebox, Auditorium, Underground cave*
- Room acoustics? *Abbey Road studio, Cathedral*

# RIR Simulation Trade-offs

## Realism vs. Computational complexity

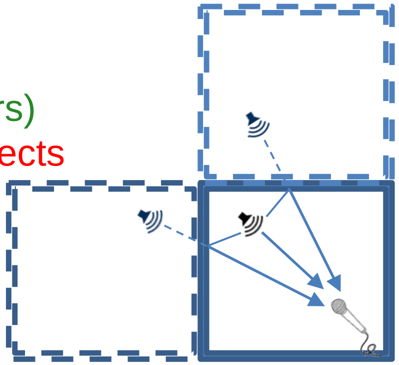
- Discretized wave equation solvers (e.g. FDTD)

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

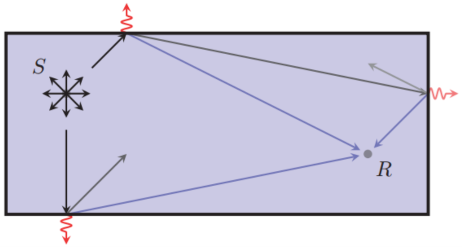
✓ Solve everything
☒ Intractable above ~4 kHz

- Image source method [13]

✓ Fast (for low reflection orders)
☒ Doesn't capture low-freq effects
☒ Specular reflections only



- Energy-based / Ray-based / Particle-based methods




✓ Versatile
☒ Doesn't capture low-freq effects
☒ Approx. TOAs

- **Simulators efficiently combining the last two:** RoomSim [14], Pyroomacoustics [15]

## Diversity vs. Training set size

- Room size? *Toilet, Office, Airport Hall*
- Room shape? *Shoebox, Auditorium, Underground cave*
- Room acoustics? *Abbey Road studio, Cathedral*
- Source/receiver types? *Omnidirectional, Cardiod, Human speaker, Hearing aids*

# RIR Simulation Trade-offs

## Realism vs. Computational complexity

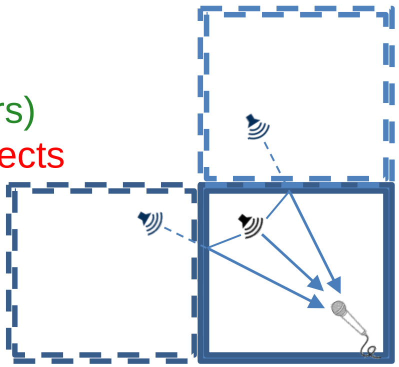
- Discretized wave equation solvers (e.g. FDTD)

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

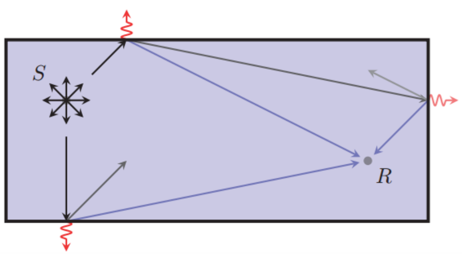
✓ Solve everything
☒ Intractable above ~4 kHz

- Image source method [13]

✓ Fast (for low reflection orders)
☒ Doesn't capture low-freq effects
☒ Specular reflections only



- Energy-based / Ray-based / Particle-based methods




✓ Versatile
☒ Doesn't capture low-freq effects
☒ Approx. TOAs

- **Simulators efficiently combining the last two:** RoomSim [14], Pyroomacoustics [15]

## Diversity vs. Training set size

- Room size? *Toilet, Office, Airport Hall*
- Room shape? *Shoebox, Auditorium, Underground cave*
- Room acoustics? *Abbey Road studio, Cathedral*
- Source/receiver types? *Omnidirectional, Cardiod, Human speaker, Hearing aids*

- Random shoebox rooms with:
  - length/width in [2m, 10m]
  - height in [2m, 6m]

# RIR Simulation Trade-offs

## Realism vs. Computational complexity

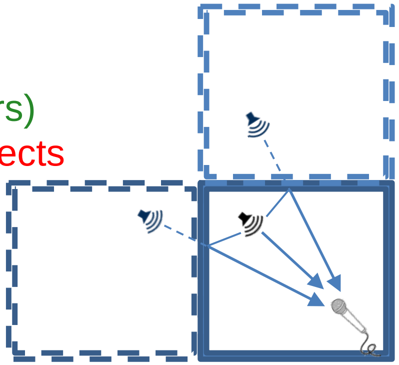
- Discretized wave equation solvers (e.g. FDTD)

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

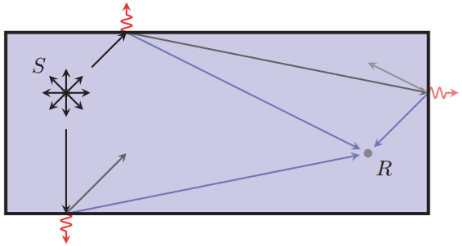
✓ Solve everything
☒ Intractable above ~4 kHz

- Image source method [13]

✓ Fast (for low reflection orders)
☒ Doesn't capture low-freq effects
☒ Specular reflections only



- Energy-based / Ray-based / Particle-based methods




✓ Versatile
☒ Doesn't capture low-freq effects
☒ Approx. TOAs

- **Simulators efficiently combining the last two:**
RoomSim [14], Pyroomacoustics [15]

## Diversity vs. Training set size

- Room size? *Toilet, Office, Airport Hall*
- Room shape? *Shoebox, Auditorium, Underground cave*
- Room acoustics? *Abbey Road studio, Cathedral*
- Source/receiver types? *Omnidirectional, Cardiod, Human speaker, Hearing aids*

- Random shoebox rooms with:
  - length/width in [2m, 10m]
  - height in [2m, 6m]
- Omnidirectional sources and receivers placed uniformly at random in the room with 1m « safe distance »

# RIR Simulation Trade-offs

## Realism vs. Computational complexity

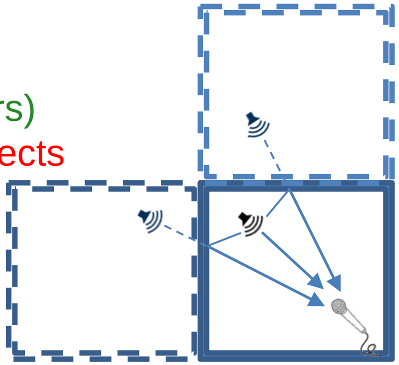
- Discretized wave equation solvers (e.g. FDTD)

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0$$

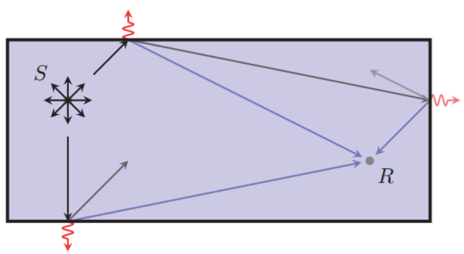
✓ Solve everything
☒ Intractable above ~4 kHz

- Image source method [13]

✓ Fast (for low reflection orders)
☒ Doesn't capture low-freq effects
☒ Specular reflections only



- Energy-based / Ray-based / Particle-based methods




✓ Versatile
☒ Doesn't capture low-freq effects
☒ Approx. TOAs

- **Simulators efficiently combining the last two:**
RoomSim [14], Pyroomacoustics [15]

## Diversity vs. Training set size

- Room size? *Toilet, Office, Airport Hall*
- Room shape? *Shoebox, Auditorium, Underground cave*
- Room acoustics? *Abbey Road studio, Cathedral*
- Source/receiver types? *Omnidirectional, Cardiod, Human speaker, Hearing aids*

- Random shoebox rooms with:
  - length/width in [2m, 10m]
  - height in [2m, 6m]
- Omnidirectional sources and receivers placed uniformly at random in the room with 1m « safe distance »
- 10k – 100k RIRs

# RIR Simulation Trade-offs

**What about the surface acoustic properties?**

# RIR Simulation Trade-offs

**What about the surface acoustic properties?**

- Diffusion coefficients:
  - Same random value for all surfaces
  - In [0, 0.3] < 500 Hz,   in [0.2, 1] > 500 Hz

# RIR Simulation Trade-offs

## What about the surface acoustic properties?

- Diffusion coefficients:
  - Same random value for all surfaces
  - In [0, 0.3] < 500 Hz,   in [0.2, 1] > 500 Hz

Meant to emulate typical surface diffusivity and room furnishing [X]

# RIR Simulation Trade-offs

**What about the surface acoustic properties?**

- Diffusion coefficients:
  - Same random value for all surfaces
  - In [0, 0.3] < 500 Hz,   in [0.2, 1] > 500 Hz

Meant to emulate typical surface diffusivity and room furnishing [X]

- Absorption coefficients:
  - Typically defined in octave bands $\left( b \in \{125, 250, 500, \ldots, 4000\} \text{ Hz} \right)$

# RIR Simulation Trade-offs

## What about the surface acoustic properties?

- Diffusion coefficients:
  - Same random value for all surfaces
  - In [0, 0.3] < 500 Hz, in [0.2, 1] > 500 Hz

Meant to emulate typical surface diffusivity and room furnishing [X]

- Absorption coefficients:
  - Typically defined in octave bands $\left( b \in \{125, 250, 500, \ldots, 4000\} \text{ Hz} \right)$

A **« reflectivity-biased »** acoustic sampling strategy [16]

For each surface type (wall, ceiling, floor) toss a coin:

- **On heads**: frequency-independent absorption coefficient in [0, 0.12] for **all** (*hard surfaces*)
- **On tails**: random absorption profile inside realistic ranges (*treated surface*)

# RIR Simulation Trade-offs

## What about the surface acoustic properties?

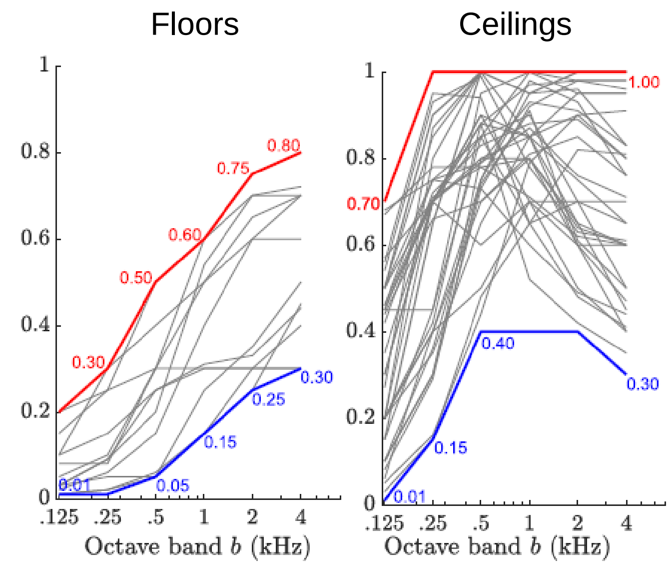
- Diffusion coefficients:
  - Same random value for all surfaces
  - In [0, 0.3] < 500 Hz,   in [0.2, 1] > 500 Hz

$\Big\}$ Meant to emulate typical surface diffusivity and room furnishing [X]

- Absorption coefficients:
  - Typically defined in octave bands $\left( b \in \{125, 250, 500, \ldots, 4000\} \text{ Hz} \right)$

A « **reflectivity-biased** » acoustic sampling strategy [16]

```
For each surface type (wall, ceiling, floor) toss
a coin:
  • On heads: frequency-independent absorption
  coefficient in [0, 0.12] for all (hard surfaces)
  • On tails: random absorption profile inside
  realistic ranges (treated surface)
```



# RIR Simulation Trade-offs

## What about the surface acoustic properties?

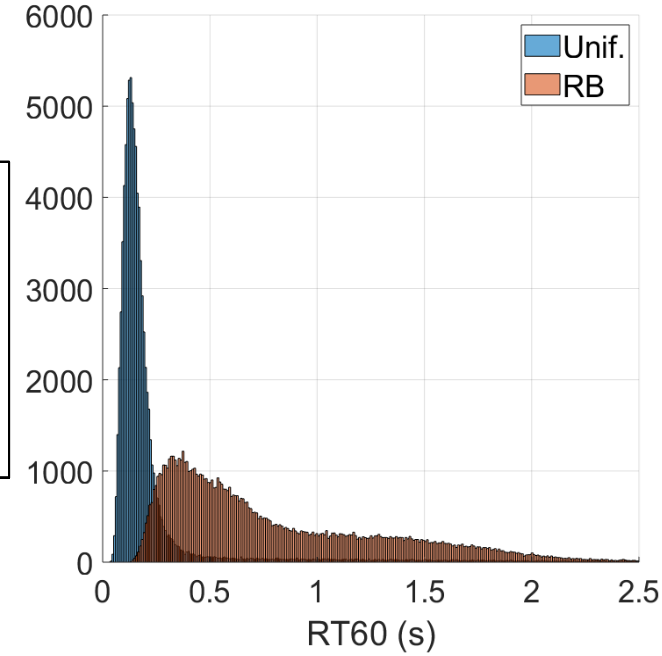
- Diffusion coefficients:
  - Same random value for all surfaces
  - In [0, 0.3] < 500 Hz, in [0.2, 1] > 500 Hz

Meant to emulate typical surface diffusivity and room furnishing [X]

- Absorption coefficients:
  - Typically defined in octave bands $\left( b \in \{125, 250, 500, \ldots, 4000\}\ \mathrm{Hz} \right)$

A « **reflectivity-biased** » acoustic sampling strategy [16]

```
For each surface type (wall, ceiling, floor) toss
a coin:
  • On heads: frequency-independent absorption
  coefficient in [0, 0.12] for all (hard surfaces)
  • On tails: random absorption profile inside
  realistic ranges (treated surface)
```



# OUTLINE

1) Intro & Background

**2) Virtually-Supervised Learning**

3) Examples and Results

4) Conclusions and Outlook

# OUTLINE

1) Intro & Background

2) Virtually-Supervised Learning

**3) Examples and Results**

4) Conclusions and Outlook

## Example 1: RIR -> Mean absorption profile of surfaces [16]

$$\bar{\alpha}(b) \overset{\text{def}}{=} \frac{1}{S_{\text{tot}}} \sum_{\text{surface } i} \alpha_i(b) S_i \qquad \left( b \in \{125, 250, 500, \ldots, 4000\} \text{ Hz} \right)$$

## Example 1: RIR -> Mean absorption profile of surfaces [16]

$$\bar{\alpha}(b) \overset{\text{def}}{=} \frac{1}{S_{\text{tot}}} \sum_{\text{surface } i} \alpha_i(b) S_i \qquad (\ b \in \{125, 250, 500, \ldots, 4000\} \text{ Hz } )$$

Absorption
coefficient in [0,1]

## Example 1: RIR -> Mean absorption profile of surfaces [16]

$$\bar{\alpha}(b) \overset{\text{def}}{=} \frac{1}{S_{\text{tot}}} \sum_{\text{surface } i} \alpha_i(b) S_i \qquad \left(\, b \in \{125, 250, 500, \ldots, 4000\} \text{ Hz} \,\right)$$

Area in m²

Antoine.Deleforge@inria.fr

## Example 1: RIR -> Mean absorption profile of surfaces [16]

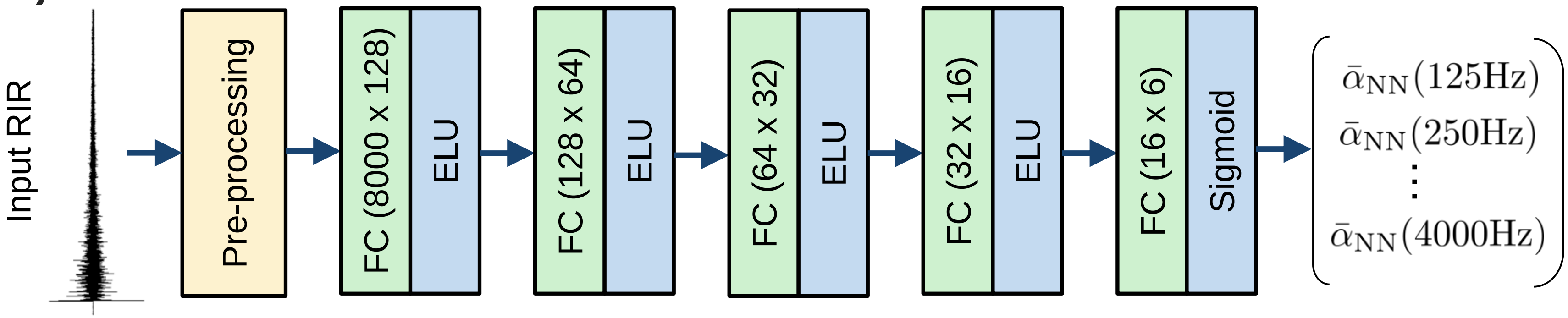
$$\bar{\alpha}(b) \overset{\text{def}}{=} \frac{1}{S_{\text{tot}}} \sum_{\text{surface } i} \alpha_i(b) S_i \qquad \left( b \in \{125, 250, 500, \ldots, 4000\} \text{ Hz} \right)$$

$\sum_i S_i$

## Example 1: RIR -> Mean absorption profile of surfaces [16]

$$\bar{\alpha}(b) \stackrel{\text{def}}{=} \frac{1}{S_{\text{tot}}} \sum_{\text{surface } i} \alpha_i(b) S_i \qquad \left( b \in \{125, 250, 500, \dots, 4000\} \text{ Hz} \right)$$

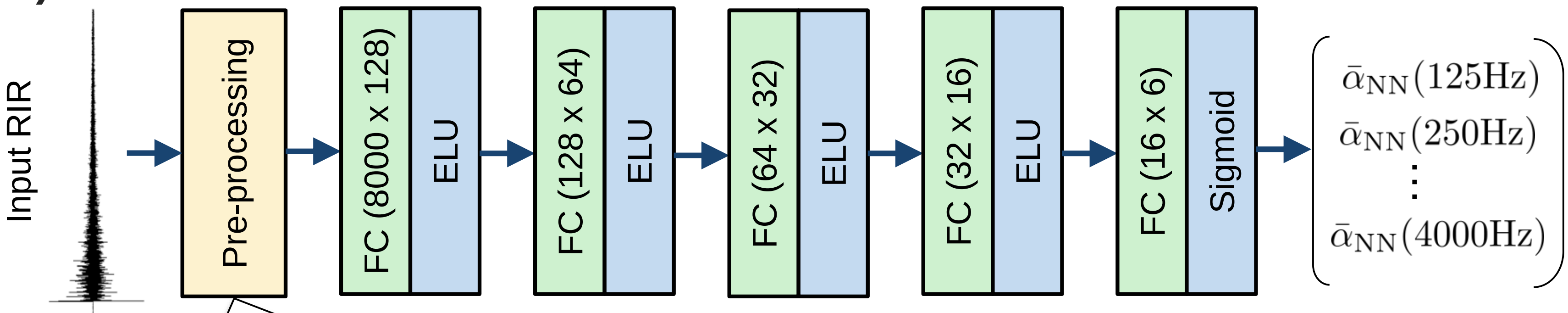
### 1) MLP



# Example 1: RIR -> Mean absorption profile of surfaces [16]

$$\bar{\alpha}(b) \overset{\text{def}}{=} \frac{1}{S_{\text{tot}}} \sum_{\text{surface } i} \alpha_i(b) S_i \qquad \left( b \in \{125, 250, 500, \ldots, 4000\} \text{ Hz} \right)$$

## 1) MLP



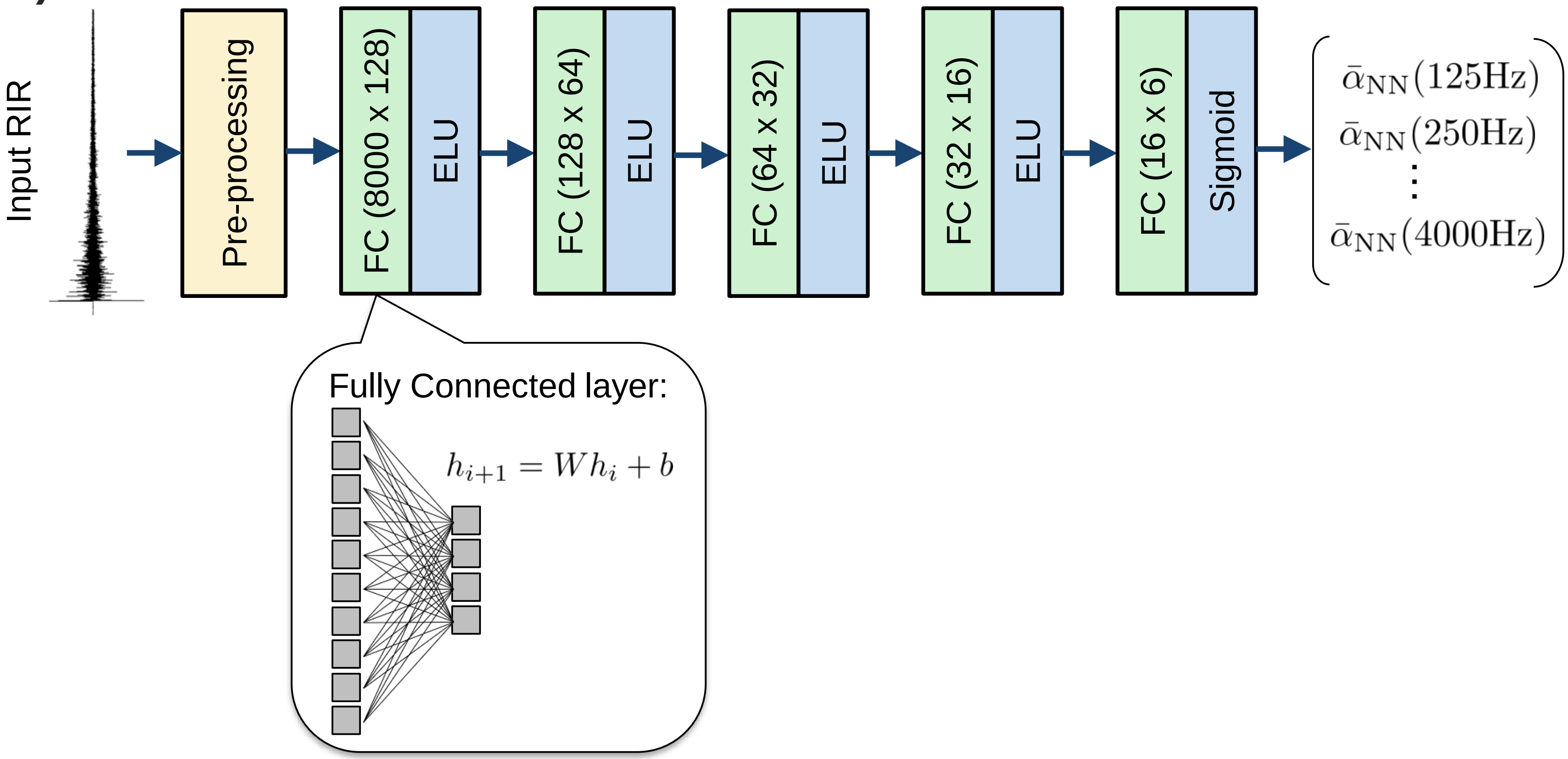

- Resample to 16 kHz
- Crop to 0.5 sec
- Normalize to max = 1
- Additive white Gaussian noise (SNR= 30 dB)

$\rightarrow$ Input vector in $\mathbb{R}^{8000}$

## Example 1: RIR -> Mean absorption profile of surfaces [16]

$$\bar{\alpha}(b) \stackrel{\text{def}}{=} \frac{1}{S_{\text{tot}}} \sum_{\text{surface } i} \alpha_i(b) S_i \qquad \left( b \in \{125, 250, 500, \ldots, 4000\} \text{ Hz} \right)$$

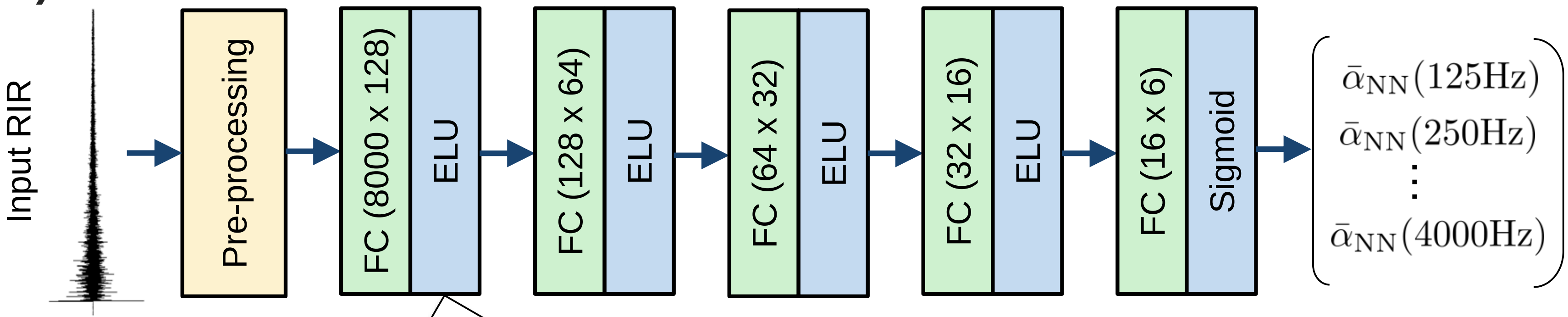
**1) MLP**



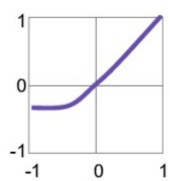
## Example 1: RIR -> Mean absorption profile of surfaces [16]

$$\bar{\alpha}(b) \stackrel{\text{def}}{=} \frac{1}{S_{\text{tot}}} \sum_{\text{surface } i} \alpha_i(b) S_i \qquad \left( b \in \{125, 250, 500, \ldots, 4000\} \text{ Hz} \right)$$

### 1) MLP




Exponential Linear Unit:

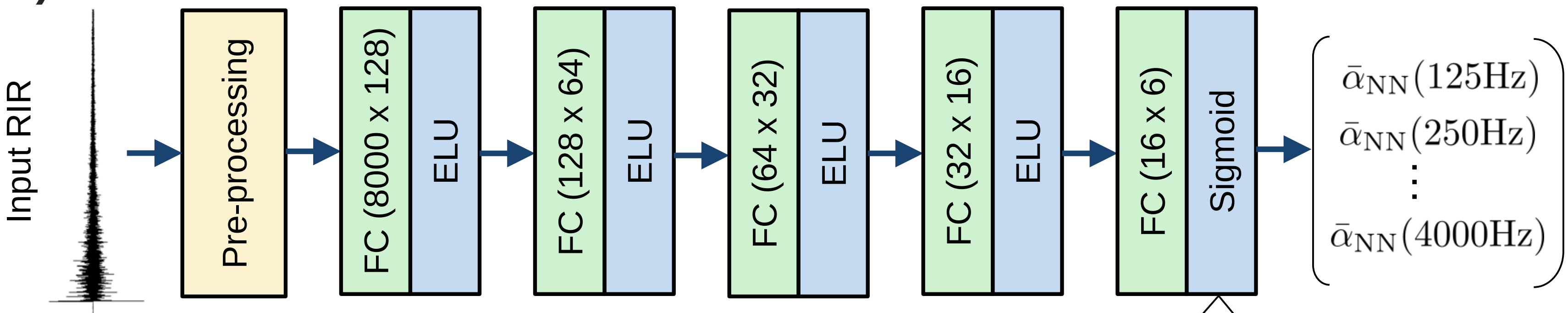



$$y = \begin{cases} x, & x \geq 0 \\ \alpha(e^x - 1), & x < 0 \end{cases}$$

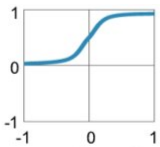
## Example 1: RIR -> Mean absorption profile of surfaces [16]

$$\bar{\alpha}(b) \stackrel{\text{def}}{=} \frac{1}{S_{\text{tot}}} \sum_{\text{surface } i} \alpha_i(b) S_i \qquad \left( b \in \{125, 250, 500, \ldots, 4000\} \text{ Hz} \right)$$

**1) MLP**




Sigmoid:

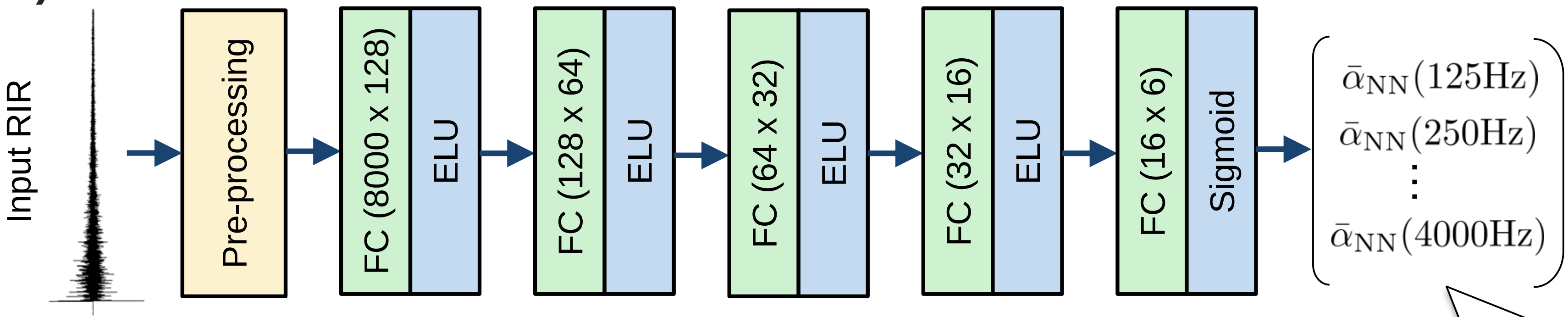



$y = 1/(1+e^{-x})$

# Example 1: RIR -> Mean absorption profile of surfaces [16]

$$\bar{\alpha}(b) \overset{\text{def}}{=} \frac{1}{S_{\text{tot}}} \sum_{\text{surface } i} \alpha_i(b) S_i \qquad \left( b \in \{125, 250, 500, \ldots, 4000\} \text{ Hz} \right)$$

## 1) MLP

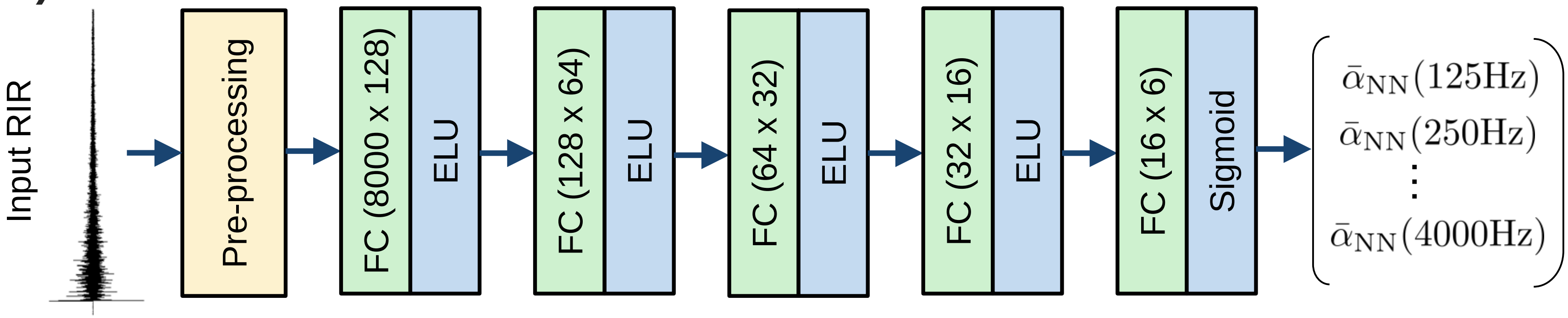



- Output vector in $[0, 1]^6$
- Loss Function = Mean Squared Error
- Optimal parameters on dev. set over 200 epochs

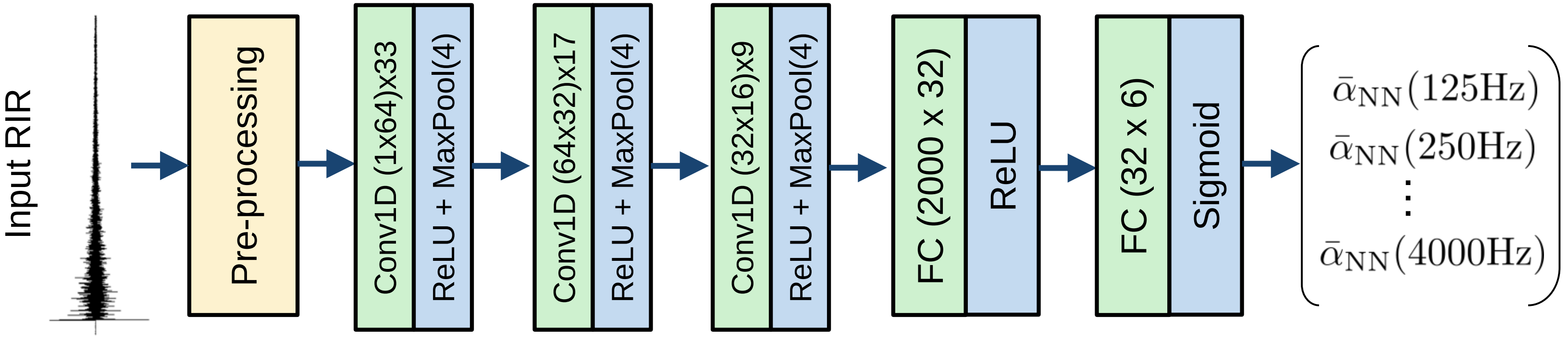
## Example 1: RIR -> Mean absorption profile of surfaces [16]

$$\bar{\alpha}(b) \overset{\text{def}}{=} \frac{1}{S_{\text{tot}}} \sum_{\text{surface } i} \alpha_i(b) S_i \qquad \left( b \in \{125, 250, 500, \ldots, 4000\} \text{ Hz} \right)$$

**1) MLP**



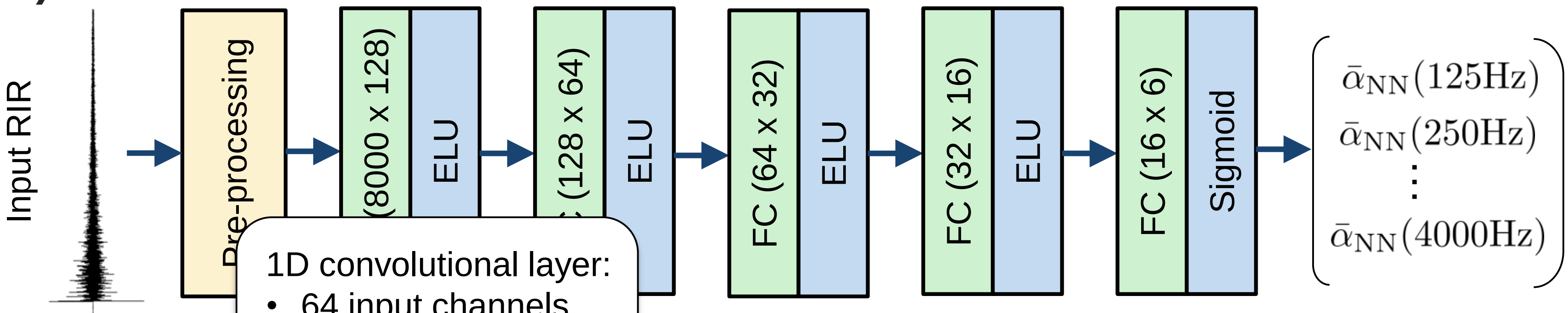

**2) CNN**



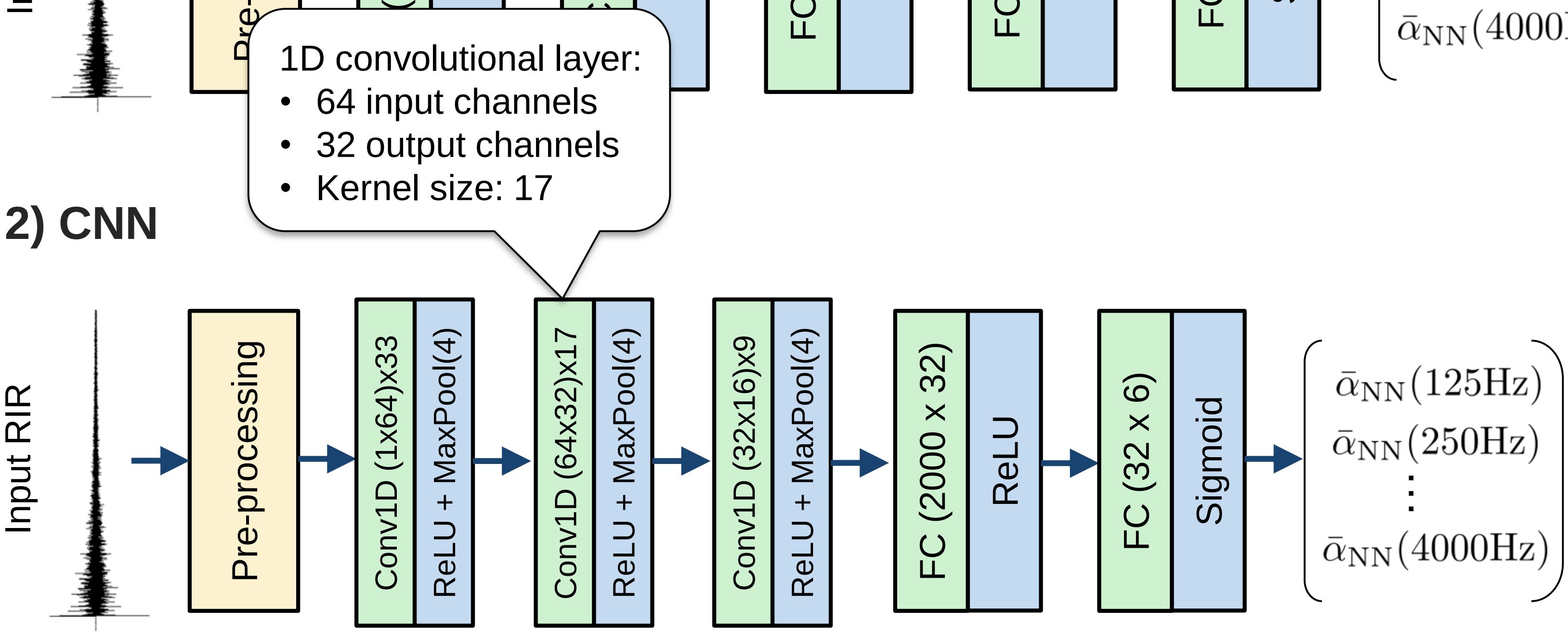
## Example 1: RIR -> Mean absorption profile of surfaces [16]

$$\bar{\alpha}(b) \overset{\text{def}}{=} \frac{1}{S_{\text{tot}}} \sum_{\text{surface } i} \alpha_i(b) S_i \qquad \left( b \in \{125, 250, 500, \ldots, 4000\} \text{ Hz} \right)$$

### 1) MLP

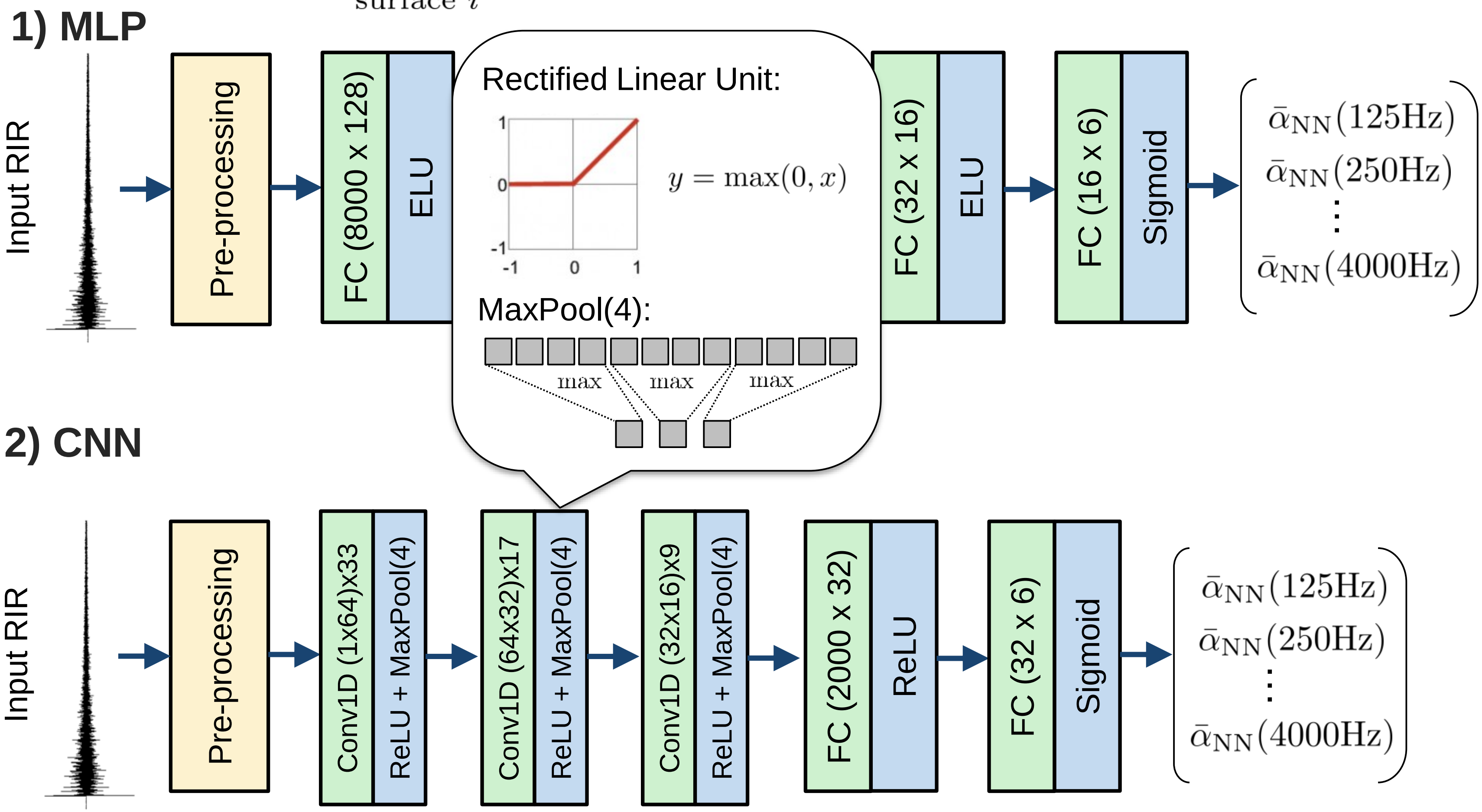



### 2) CNN



# Example 1: RIR -> Mean absorption profile of surfaces [16]

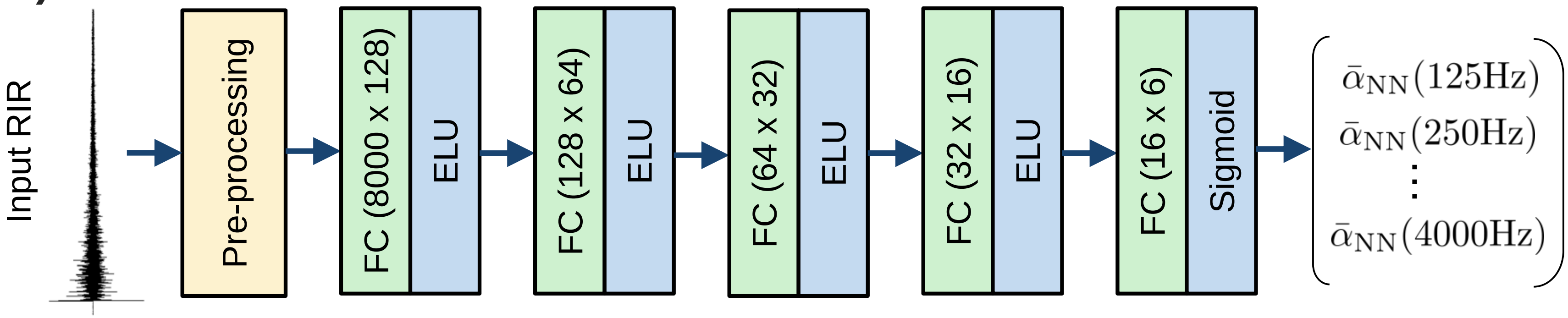
$$\bar{\alpha}(b) \stackrel{\text{def}}{=} \frac{1}{S_{\text{tot}}} \sum_{\text{surface } i} \alpha_i(b) S_i \qquad (\, b \in \{125, 250, 500, \ldots, 4000\} \text{ Hz } )$$



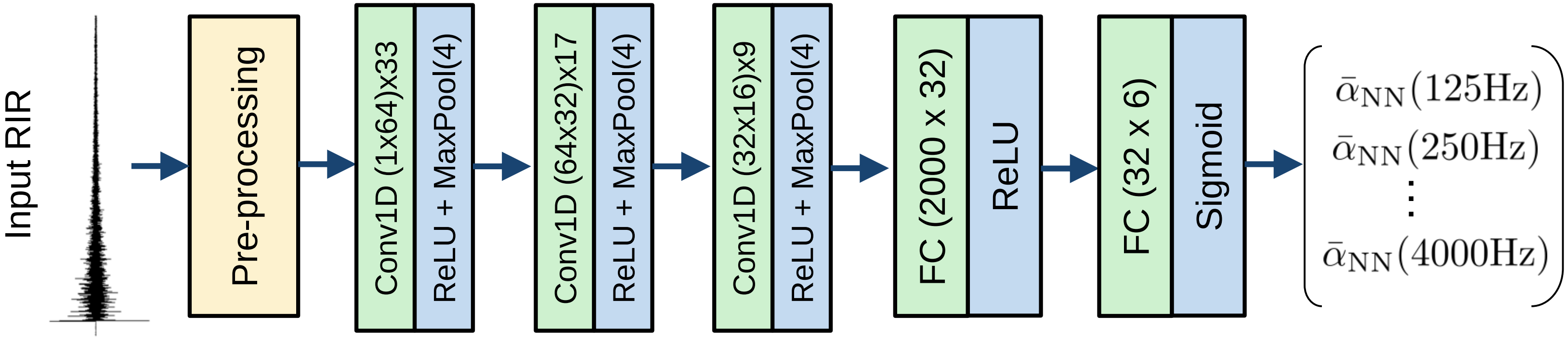
## Example 1: RIR -> Mean absorption profile of surfaces [16]

$$\bar{\alpha}(b) \stackrel{\text{def}}{=} \frac{1}{S_{\text{tot}}} \sum_{\text{surface } i} \alpha_i(b) S_i \qquad \left( b \in \{125, 250, 500, \dots, 4000\} \text{ Hz} \right)$$

**1) MLP**




**2) CNN**



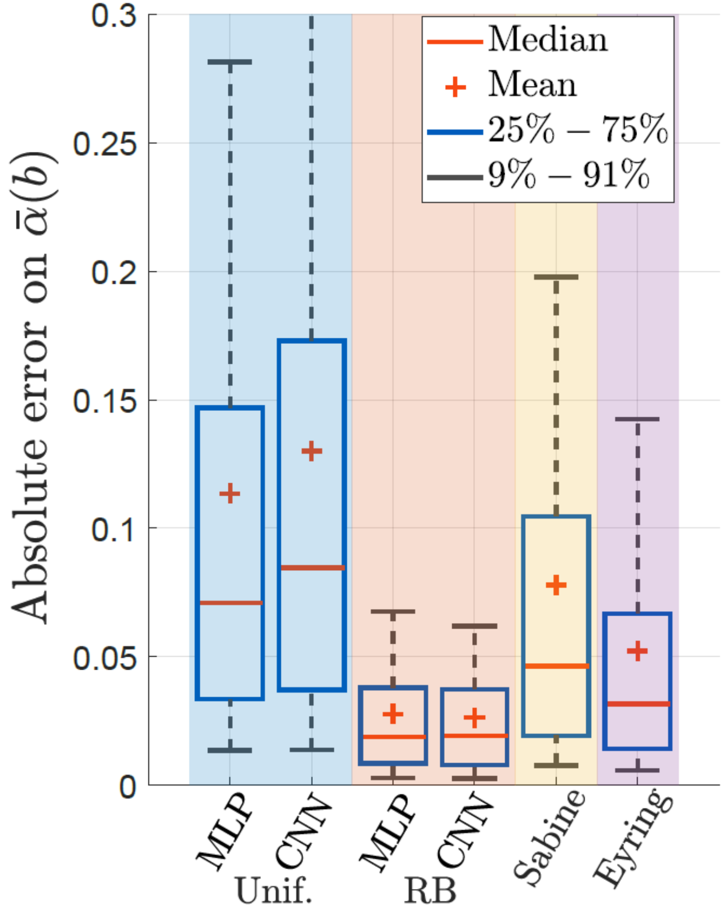
## Example 1: RIR -> Mean absorption profile of surfaces [16]

- **Simulated test results:** RoomSim, real absorption profiles, 5 room geometries, 500 RIRs
- Comparing two training sets (**Unif.**, **RB**) and the two neural networks (**MLP, CNN**) against **Sabine** and **Eyring**'s laws (given true $S_{tot}$ and $V$)

## Example 1: RIR -> Mean absorption profile of surfaces [16]

- **Simulated test results:** RoomSim, real absorption profiles, 5 room geometries, 500 RIRs

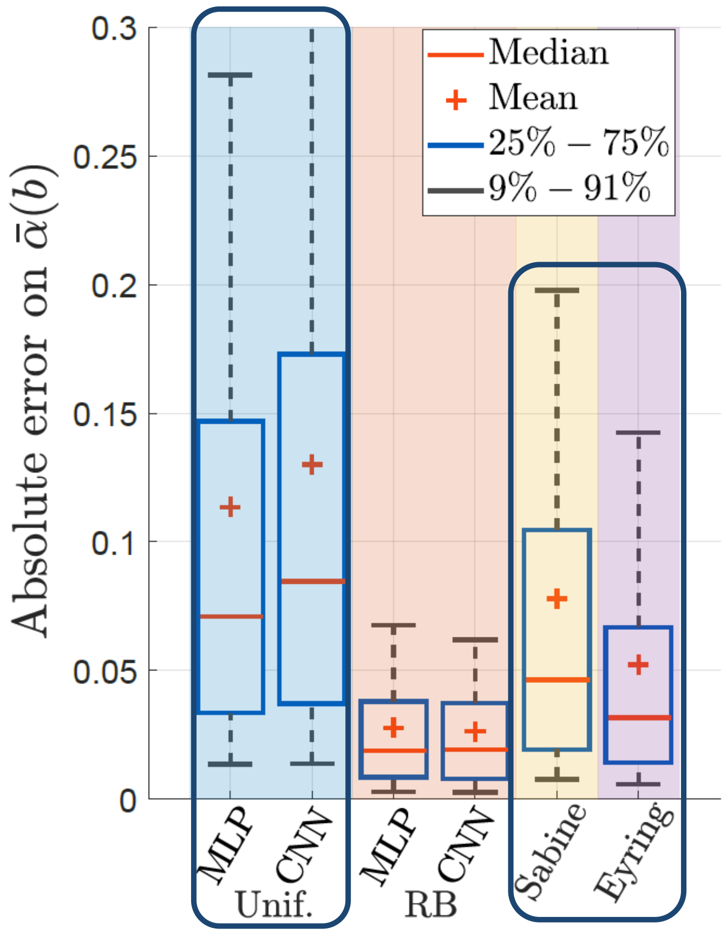
- Comparing two training sets (**Unif.**, **RB**) and the two neural networks (**MLP, CNN**) against **Sabine** and **Eyring**'s laws (given true $S_{tot}$ and $V$)



# Example 1: RIR -> Mean absorption profile of surfaces [16]

- **Simulated test results:** RoomSim, real absorption profiles, 5 room geometries, 500 RIRs
- Comparing two training sets (**Unif.**, **RB**) and the two neural networks (**MLP, CNN**) against **Sabine** and **Eyring**'s laws (given true $S_{tot}$ and $V$)

➡ Training on uniformly sampled acoustics fails to outperform reverberation theory

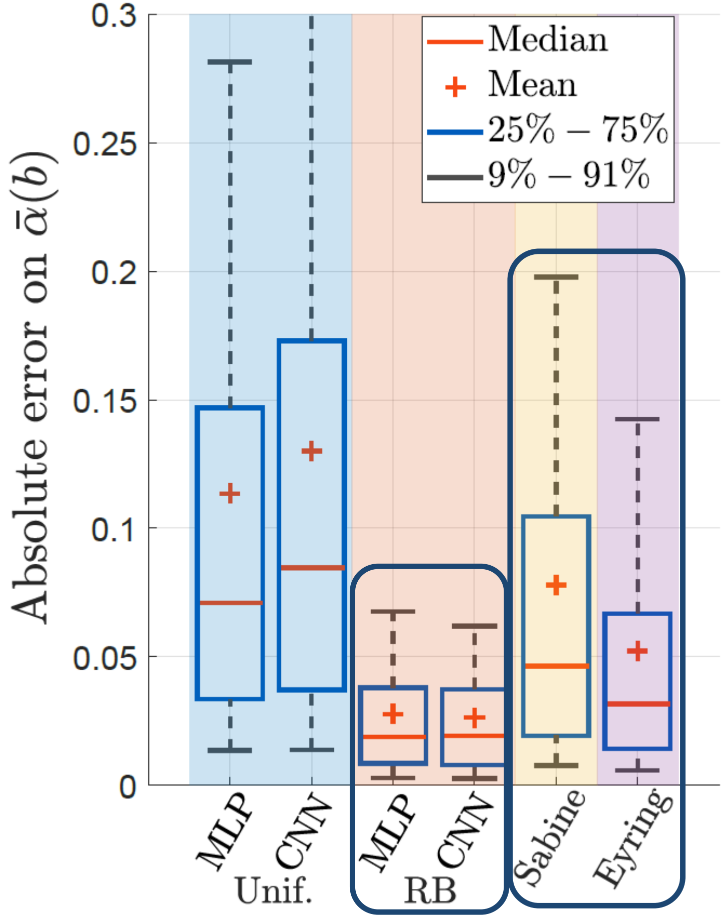


# Example 1: RIR -> Mean absorption profile of surfaces [16]

- **Simulated test results:** RoomSim, real absorption profiles, 5 room geometries, 500 RIRs
- Comparing two training sets (**Unif.**, **RB**) and the two neural networks (**MLP, CNN**) against **Sabine** and **Eyring**'s laws (given true $S_{tot}$ and $V$)

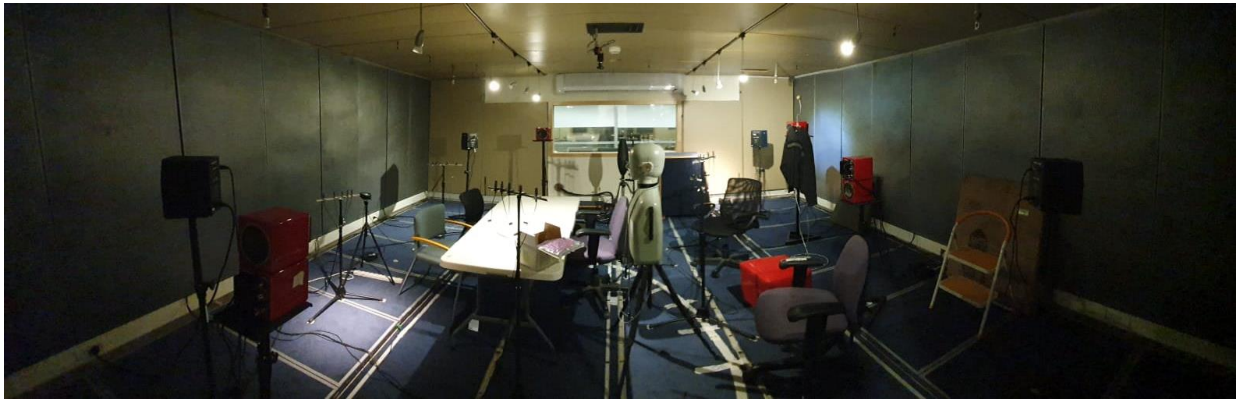
➡ Training on uniformly sampled acoustics fails to outperform reverberation theory

➡ Training on the reflectivity-biased set significantly outperforms both baselines



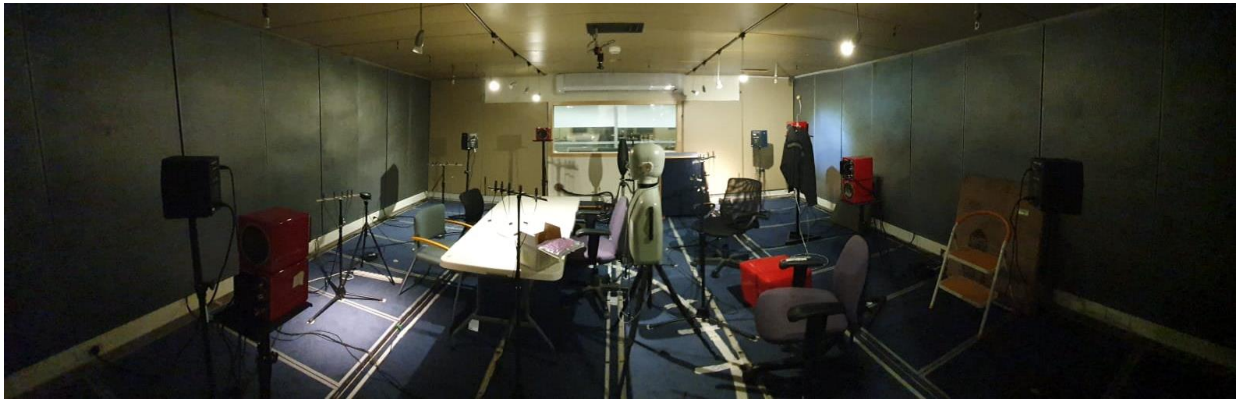
## Example 1: RIR -> Mean absorption profile of surfaces [16]

- Encouraging generalizability to **real data** (900 RIRs, 10 room configurations [12])



## Example 1: RIR -> Mean absorption profile of surfaces [16]

- Encouraging generalizability to **real data** (900 RIRs, 10 room configurations [12])

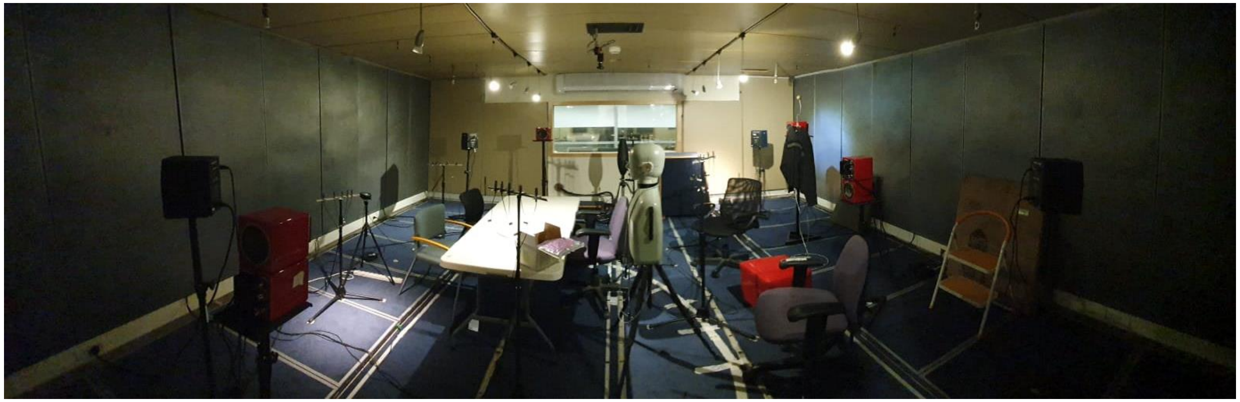


$\mathcal{A}$ : RIR featuring « *nice* » reverberation decay

$\mathcal{B}$ : RIR with « *unusual* » reveberation decay

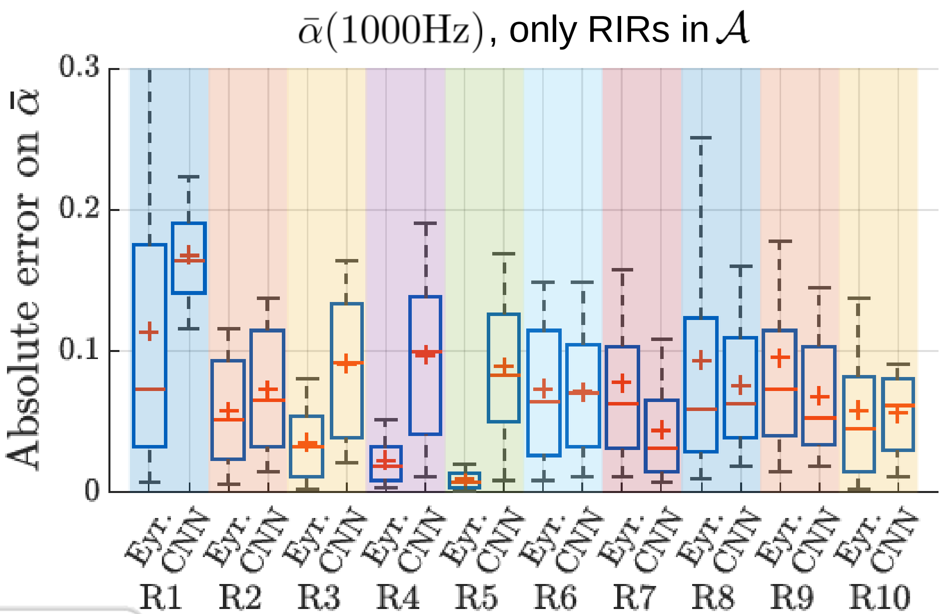
# Example 1: RIR -> Mean absorption profile of surfaces [16]

- Encouraging generalizability to **real data** (900 RIRs, 10 room configurations [12])



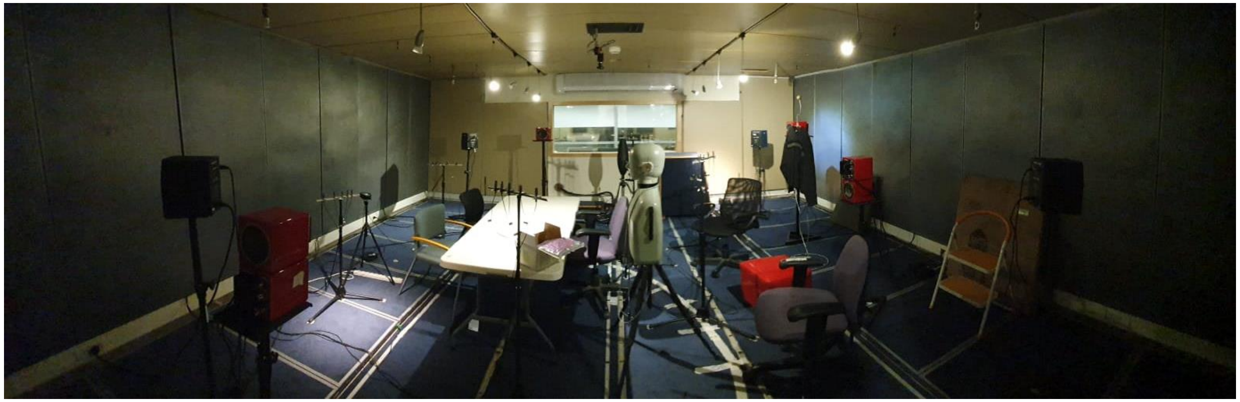
$\mathcal{A}$ : RIR featuring « *nice* » reverberation decay

$\mathcal{B}$ : RIR with « *unusual* » reveberation decay



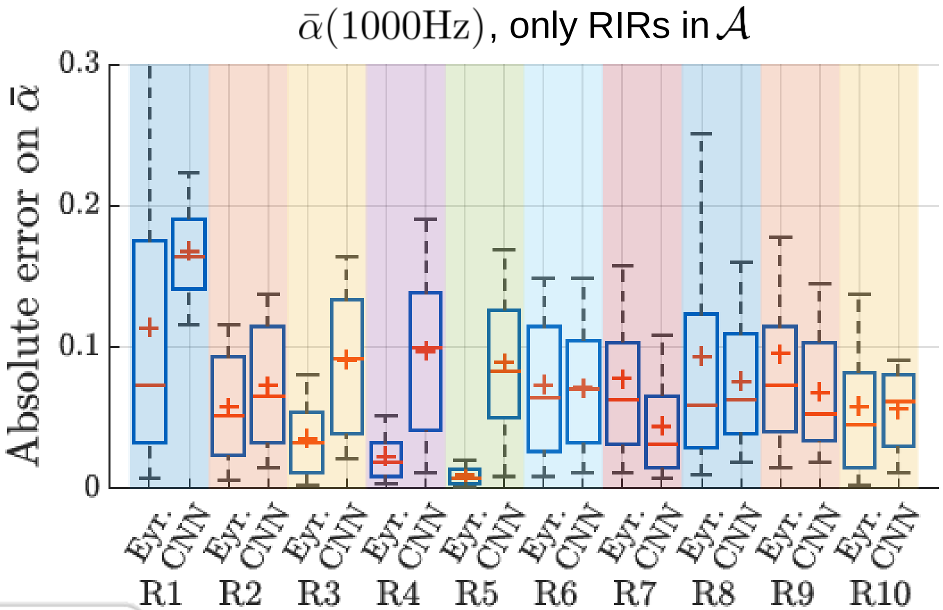
## Example 1: RIR -> Mean absorption profile of surfaces [16]

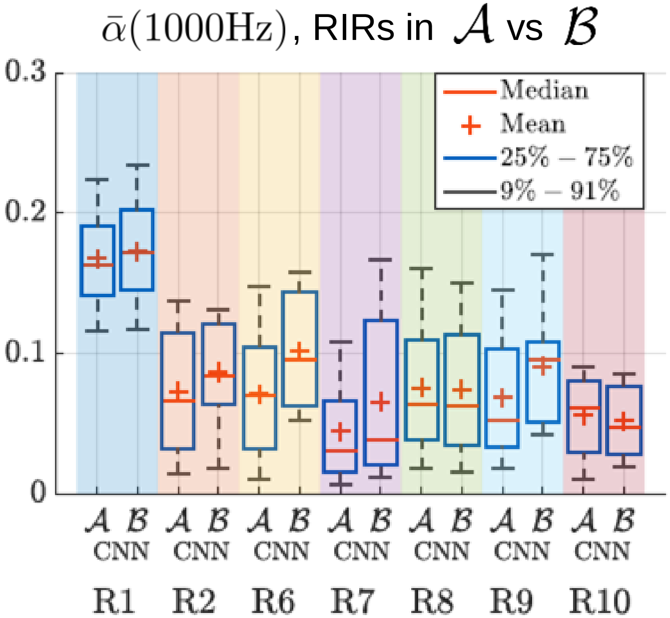
- Encouraging generalizability to **real data** (900 RIRs, 10 room configurations [12])



$\mathcal{A}$ : RIR featuring « *nice* » reverberation decay

$\mathcal{B}$ : RIR with « *unusual* » reveberation decay

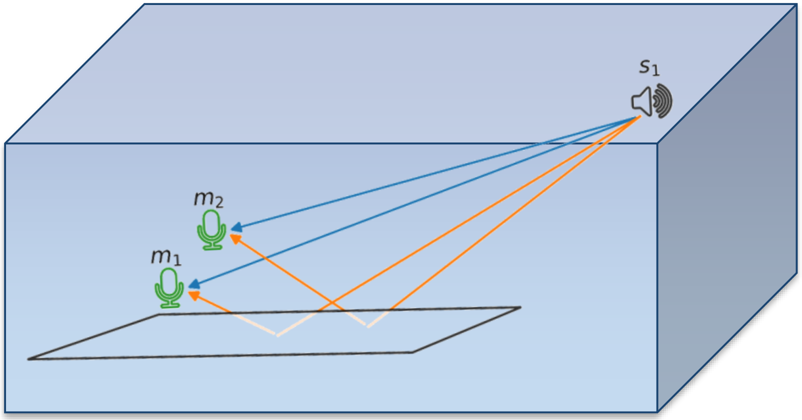


# Example 2: Blind echo estimation [4]

**A « pic-nic » dataset**

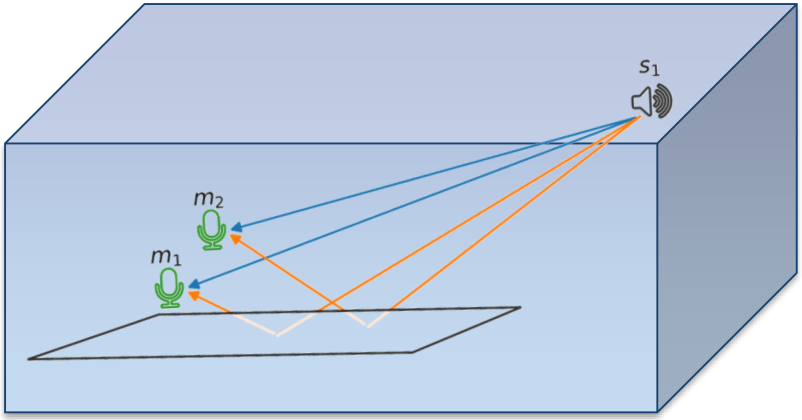
- One Source
- Two microphones
- Nearest surface is most reflective
- Random shoe-box rooms



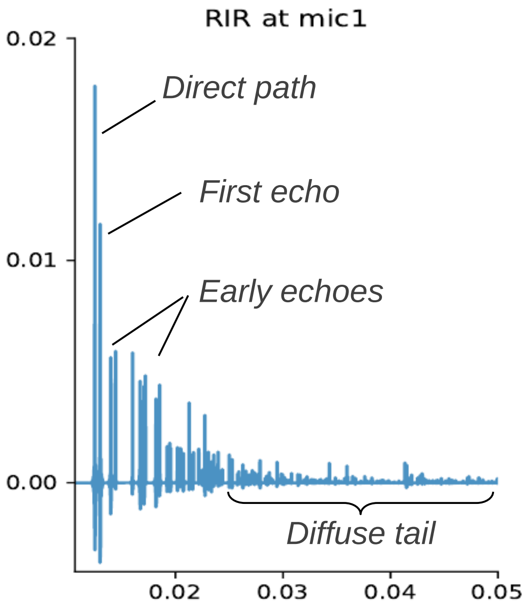
# Example 2: Blind echo estimation [4]

**A « pic-nic » dataset**

- One Source
- Two microphones
- Nearest surface is most reflective
- Random shoe-box rooms



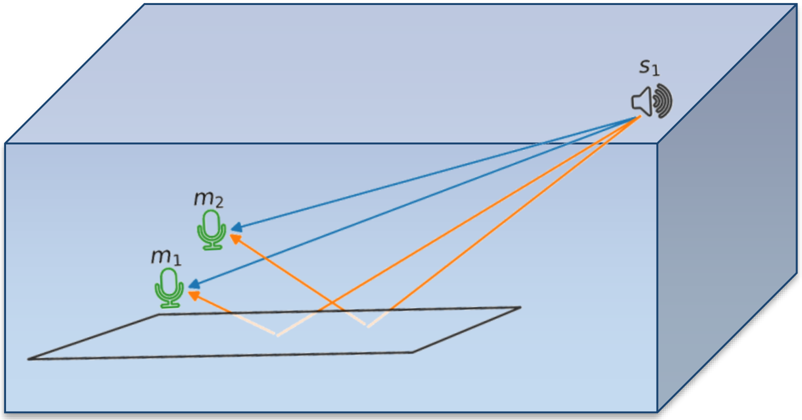

Room impulse responses look like this:



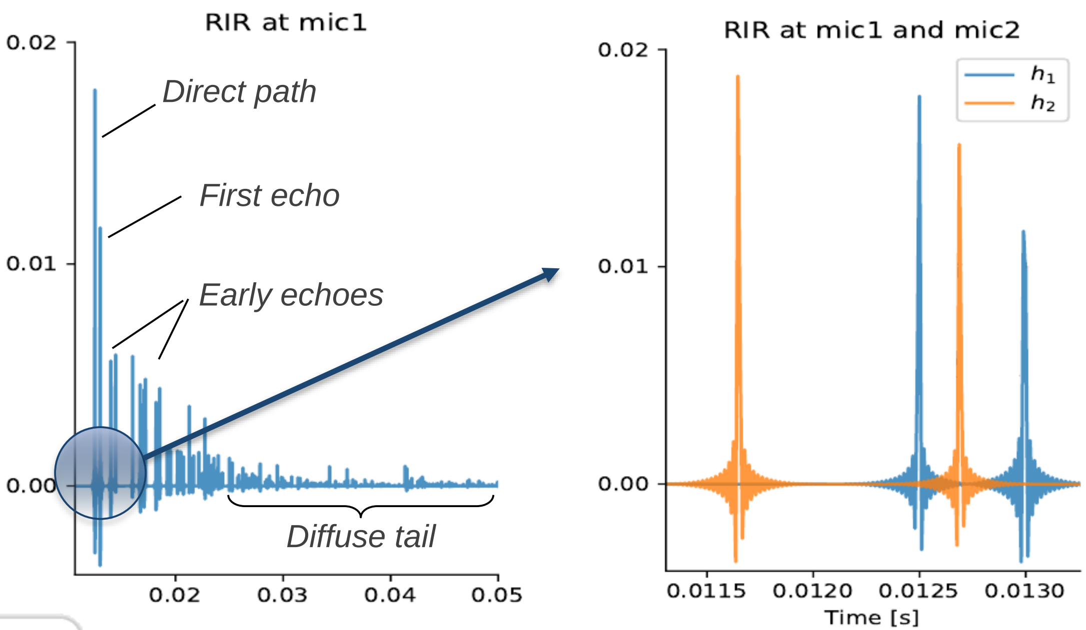
# Example 2: Blind echo estimation [4]

**A « pic-nic » dataset**

- One Source
- Two microphones
- Nearest surface is most reflective
- Random shoe-box rooms




➡ Room impulse responses look like this:

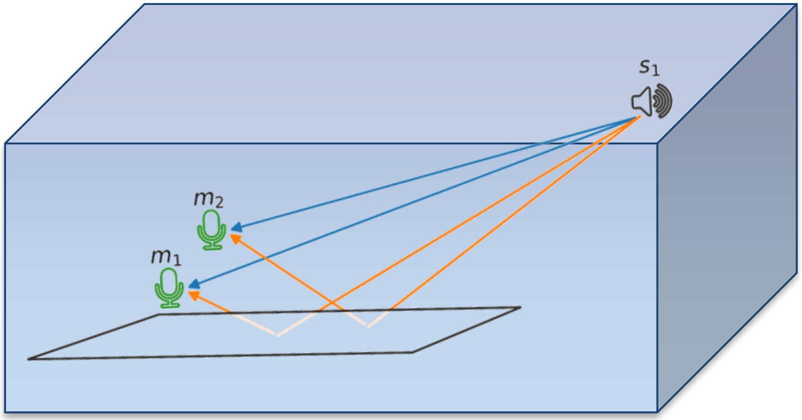



Antoine.Deleforge@inria.fr

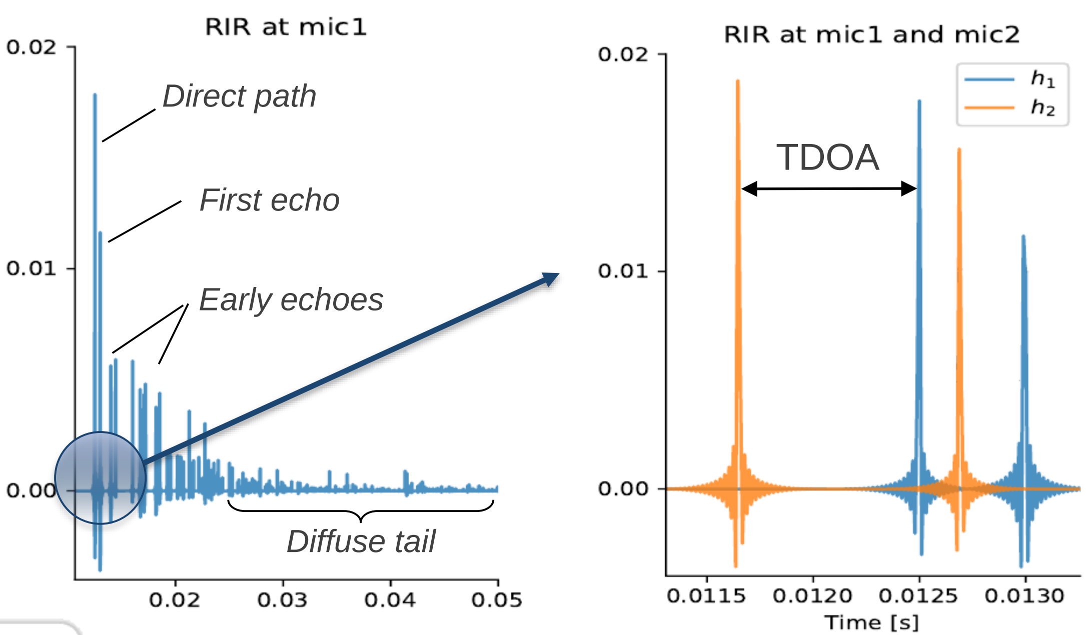
# Example 2: Blind echo estimation [4]

## A « pic-nic » dataset

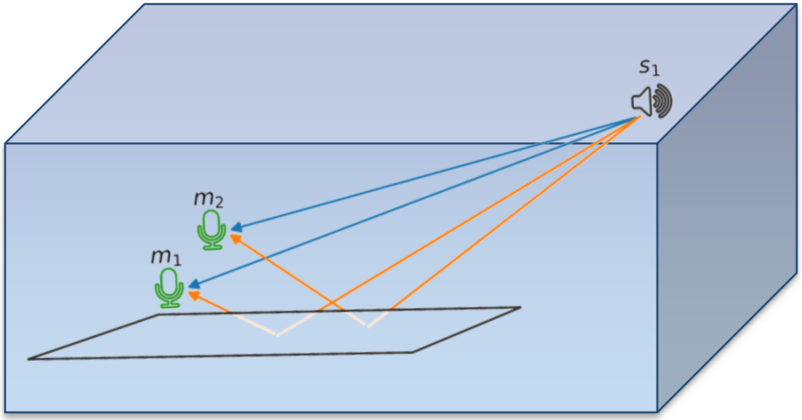
- One Source
- Two microphones
- Nearest surface is most reflective
- Random shoe-box rooms




➡ Room impulse responses look like this:



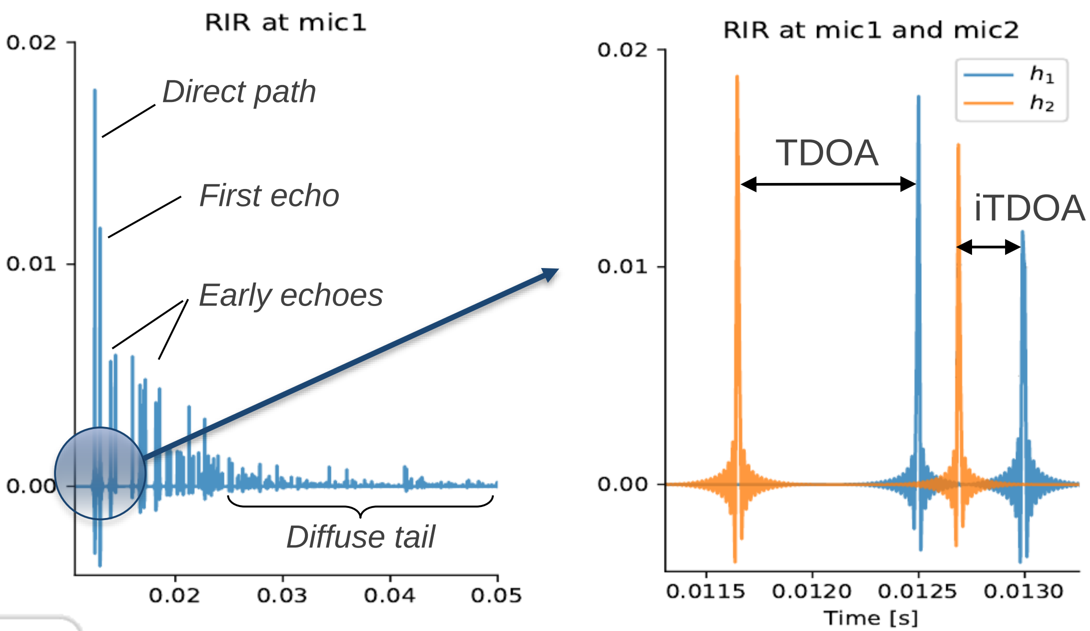
# Example 2: Blind echo estimation [4]



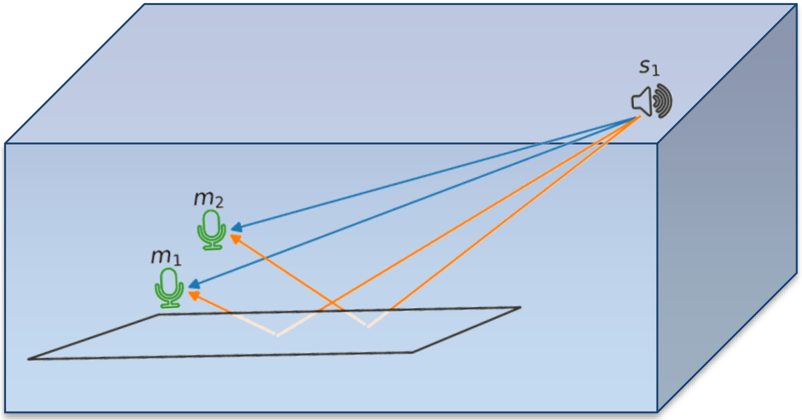

## A « pic-nic » dataset

- One Source
- Two microphones
- Nearest surface is most reflective
- Random shoe-box rooms

➡ Room impulse responses look like this:



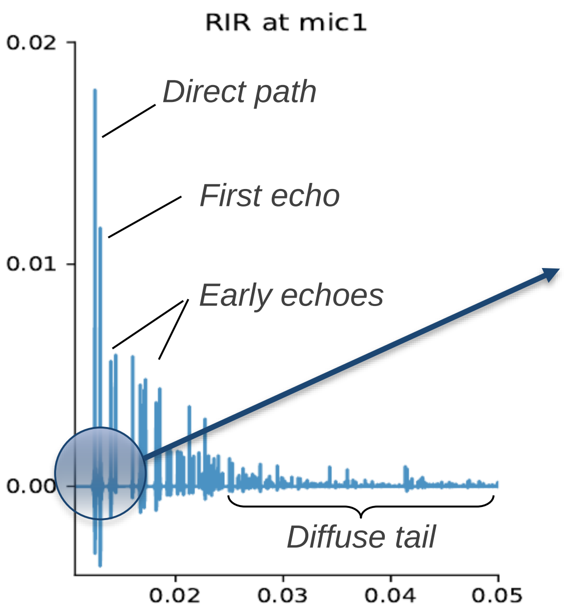
# Example 2: Blind echo estimation [4]

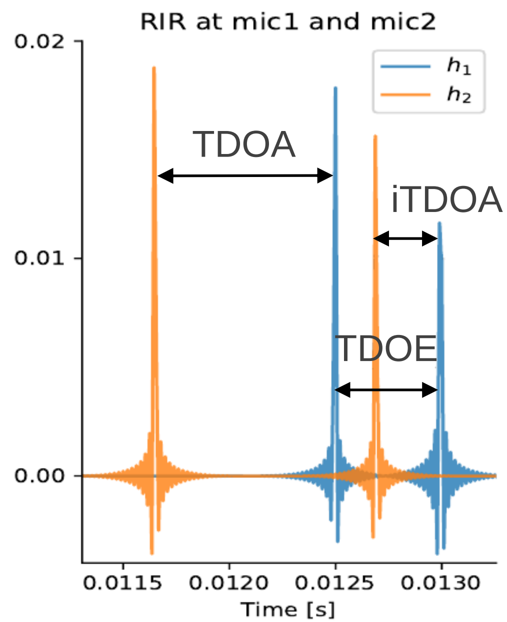



## A « pic-nic » dataset

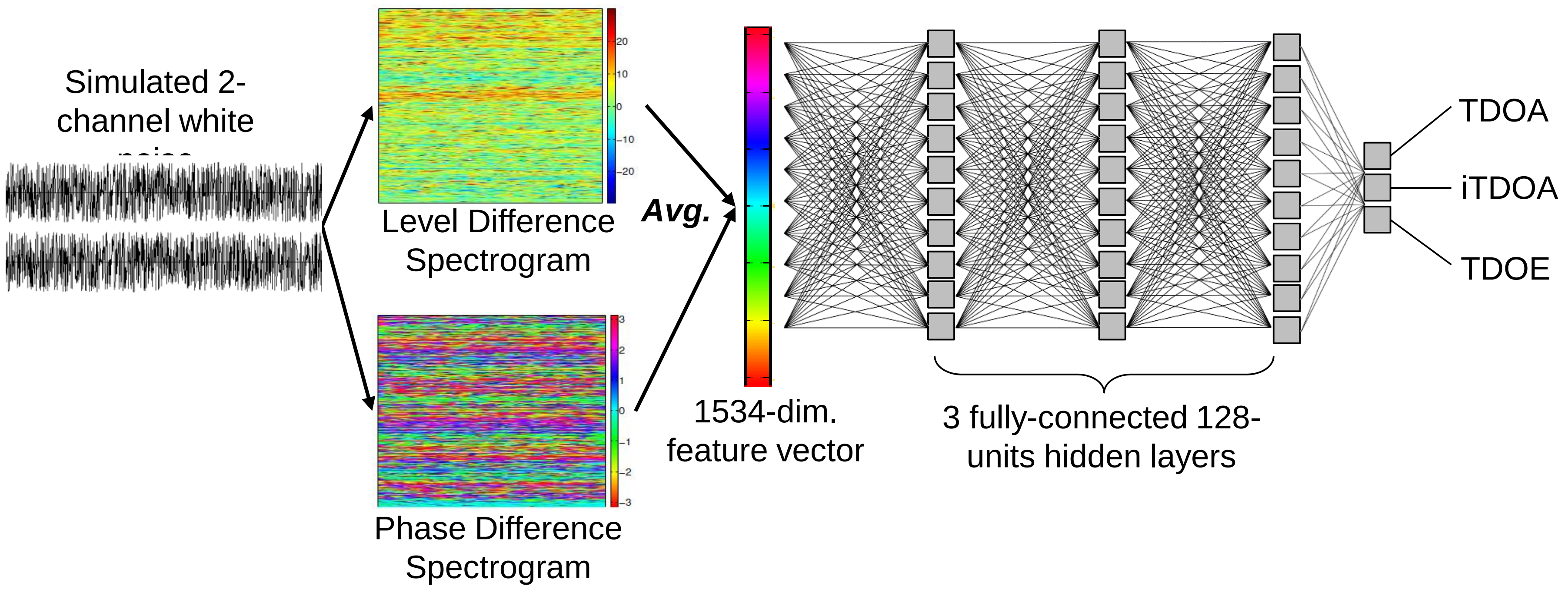
- One Source
- Two microphones
- Nearest surface is most reflective
- Random shoe-box rooms

➡ Room impulse responses look like this:

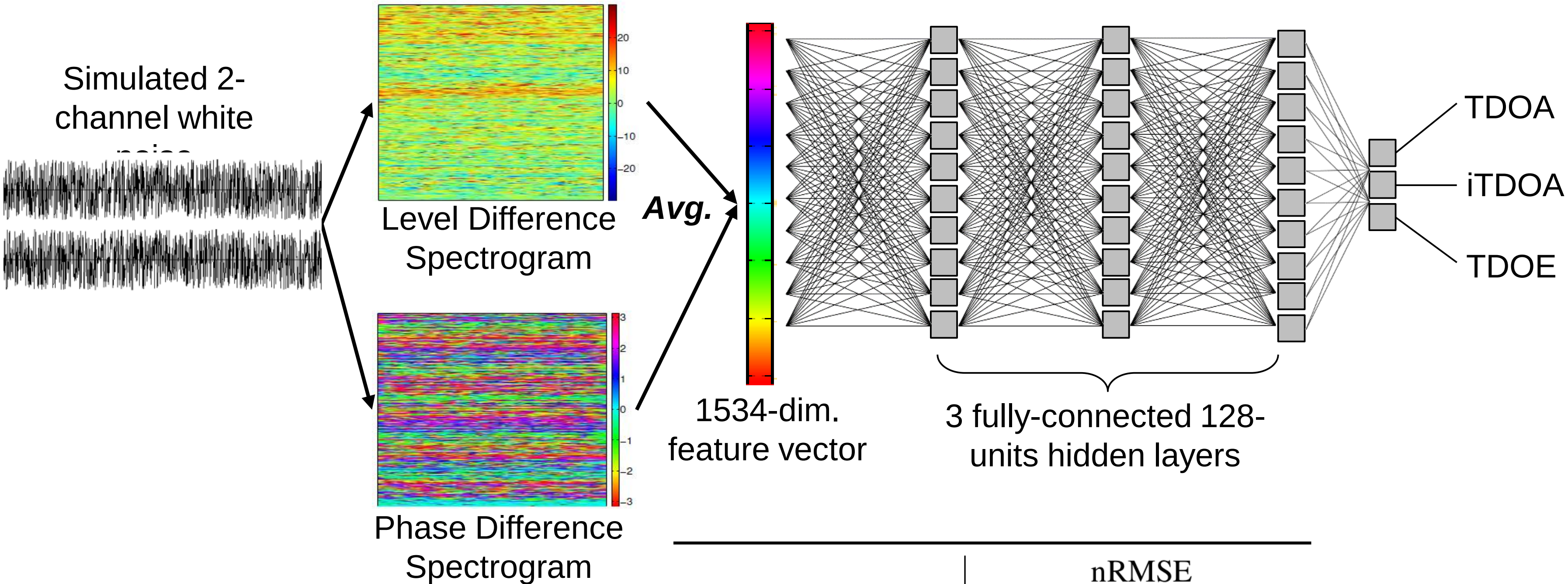


# Example 2: Blind echo estimation [4]



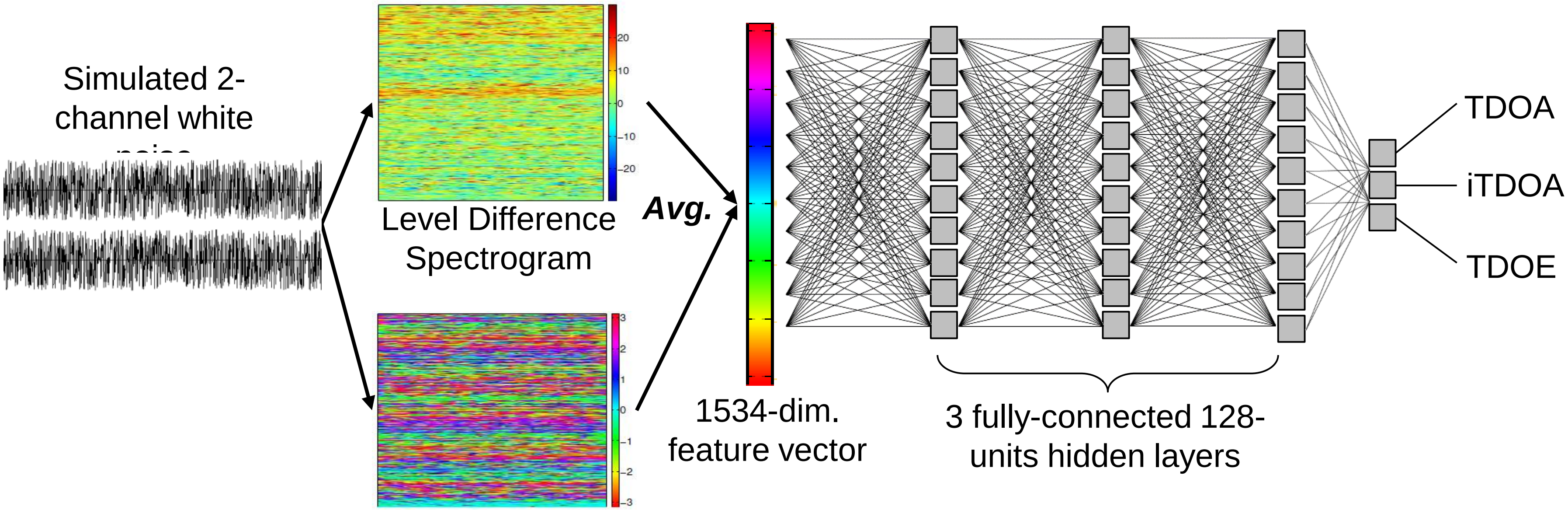
# Example 2: Blind echo estimation [4]




## Results on test set

| | Input | nRMSE TDOA | nRMSE iTDOA | nRMSE TDOE |
|---|---|---|---|---|
| MIRAGE | wn | 0.18 | 0.28 | 0.25 |
| MIRAGE | wn+n | 0.68 | 0.69 | 0.89 |
| MIRAGE | sp | 0.31 | 0.34 | 0.56 |
| MIRAGE | sp+n | 0.99 | 0.98 | 1.48 |
| GCC-PHAT | wn | 0.21 | - | - |
| GCC-PHAT | wn+n | 0.68 | - | - |
| GCC-PHAT | sp | 0.32 | - | - |
| GCC-PHAT | sp+n | 1.38 | - | - |

# Example 2: Blind echo estimation [4]

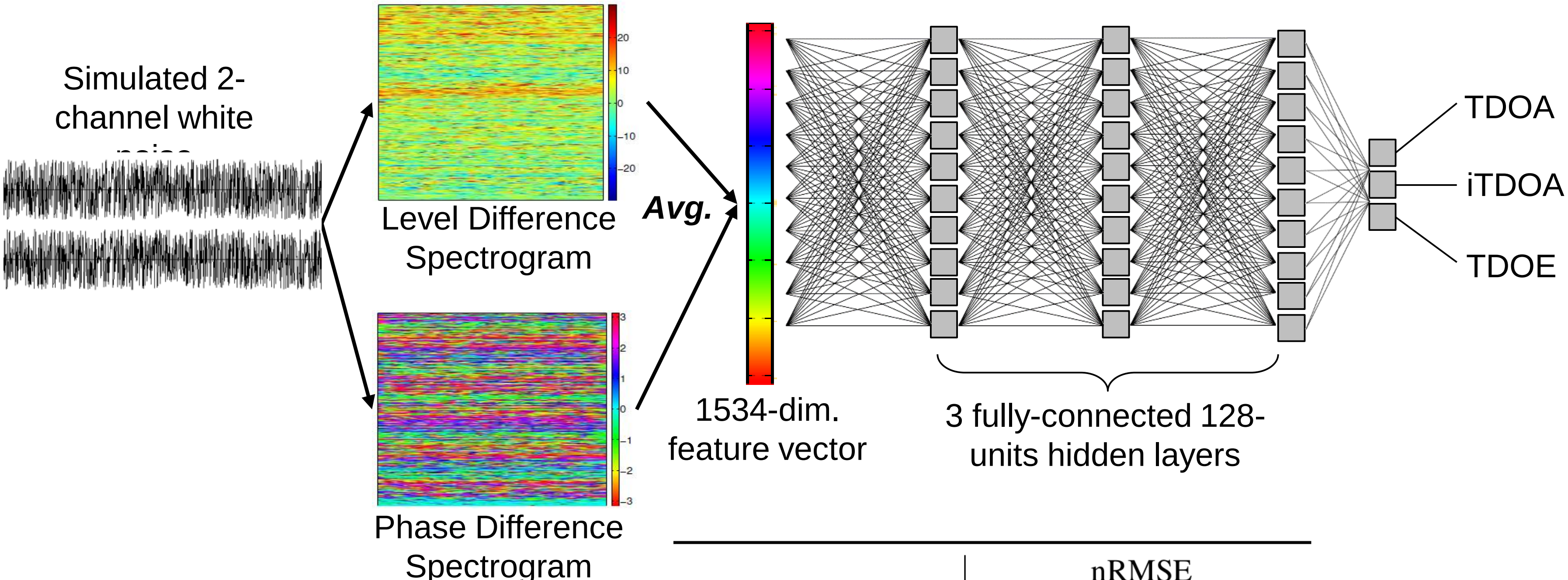



## Results on test set

✓ Good results with white noise

| | Input | nRMSE TDOA | nRMSE iTDOA | nRMSE TDOE |
|---|---|---|---|---|
| MIRAGE | wn | 0.18 | 0.28 | 0.25 |
| MIRAGE | wn+n | 0.68 | 0.69 | 0.89 |
| MIRAGE | sp | 0.31 | 0.34 | 0.56 |
| MIRAGE | sp+n | 0.99 | 0.98 | 1.48 |
| GCC-PHAT | wn | 0.21 | - | - |
| GCC-PHAT | wn+n | 0.68 | - | - |
| GCC-PHAT | sp | 0.32 | - | - |
| GCC-PHAT | sp+n | 1.38 | - | - |

# Example 2: Blind echo estimation [4]




# Results on test set

- ✓ Good results with white noise
- ☒ Poor generalization to noisy speech and real data

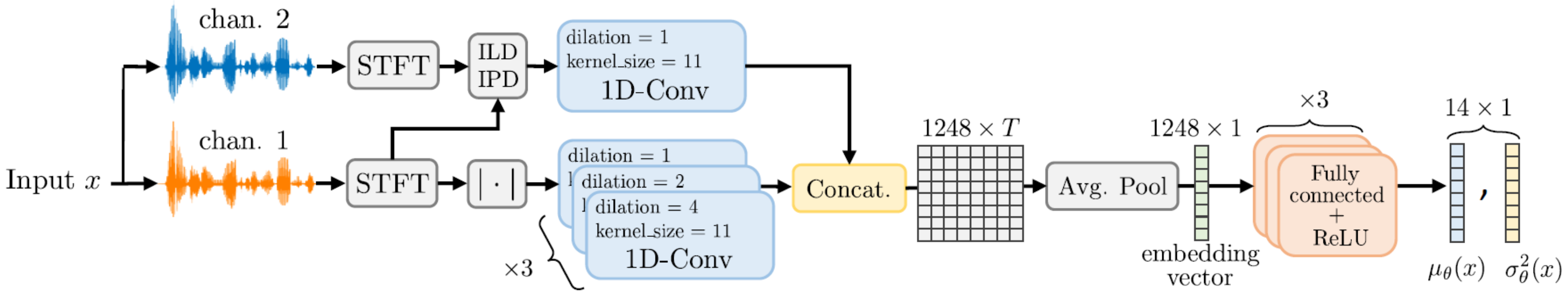
| | Input | nRMSE TDOA | nRMSE iTDOA | nRMSE TDOE |
|---|---|---|---|---|
| MIRAGE | wn | 0.18 | 0.28 | 0.25 |
| MIRAGE | wn+n | 0.68 | 0.69 | 0.89 |
| MIRAGE | sp | 0.31 | 0.34 | 0.56 |
| MIRAGE | sp+n | 0.99 | 0.98 | 1.48 |
| GCC-PHAT | wn | 0.21 | - | - |
| GCC-PHAT | wn+n | 0.68 | - | - |
| GCC-PHAT | sp | 0.32 | - | - |
| GCC-PHAT | sp+n | 1.38 | - | - |

# Example 3: Blind room parameter estimation [17]

- Joint estimation of **volume**, **total surface**, **RT**$_{60}$(b) and $\bar{\alpha}(b)$ from multiple, multichannel noisy speech recordings

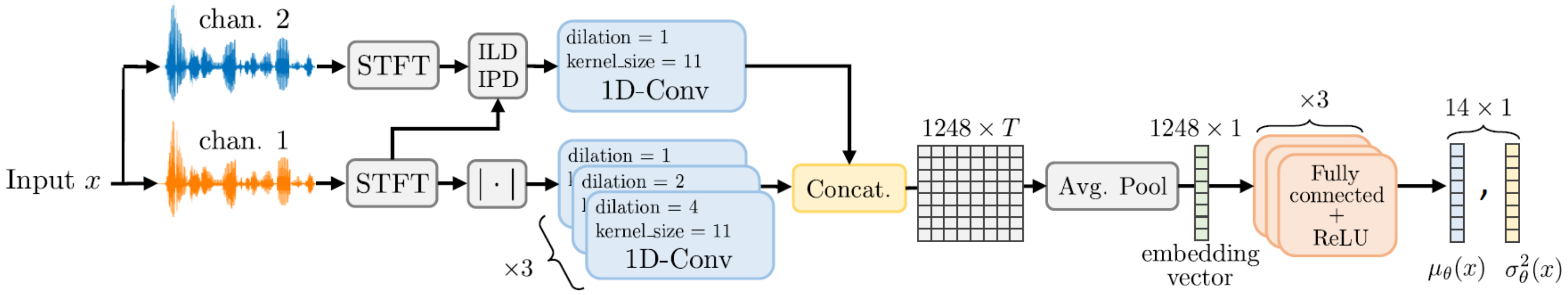
# Example 3: Blind room parameter estimation [17]

- Joint estimation of **volume**, **total surface**, **RT**$_{60}$(b) and $\bar{\alpha}(b)$ from multiple, multichannel noisy speech recordings



# Example 3: Blind room parameter estimation [17]

- Joint estimation of **volume**, **total surface**, **RT**$_{60}$(b) and $\bar{\alpha}(b)$ from multiple, multichannel noisy speech recordings

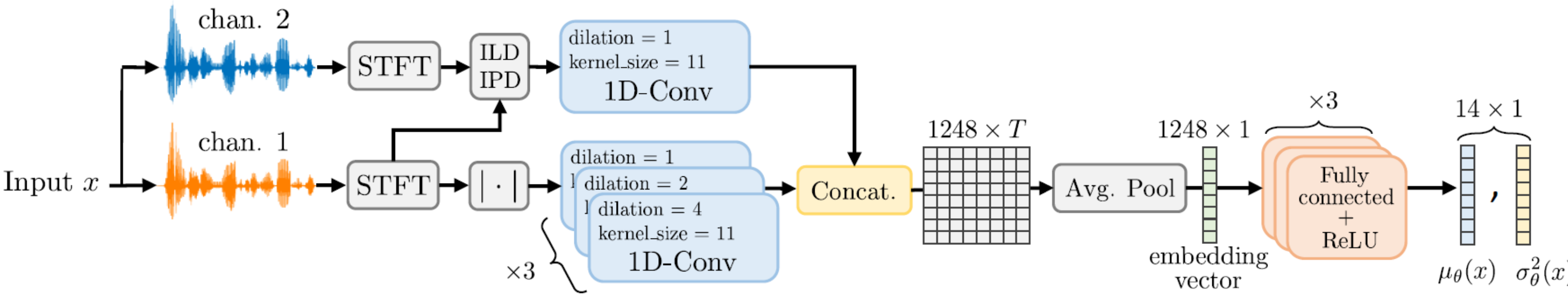



- A maximum-likelihood cost-function: $\mathcal{L}_\theta(\boldsymbol{x},\boldsymbol{y}) = -\log p_\theta(\boldsymbol{y}|\boldsymbol{x}) = -\log \mathcal{N}(\boldsymbol{y};\boldsymbol{\mu}_\theta(\boldsymbol{x}),\boldsymbol{\sigma}^2_\theta(\boldsymbol{x}))$

$$\stackrel{c}{=} \frac{1}{2}\sum_{d=1}^{D} \log \sigma^2_{d,\theta}(\boldsymbol{x}) + \frac{(y_d - \mu_{d,\theta}(\boldsymbol{x}))^2}{\sigma^2_{d,\theta}(\boldsymbol{x})}$$

# Example 3: Blind room parameter estimation [17]

- Joint estimation of **volume**, **total surface**, **RT**$_{60}$(b) and $\bar{\alpha}(b)$ from multiple, multichannel noisy speech recordings



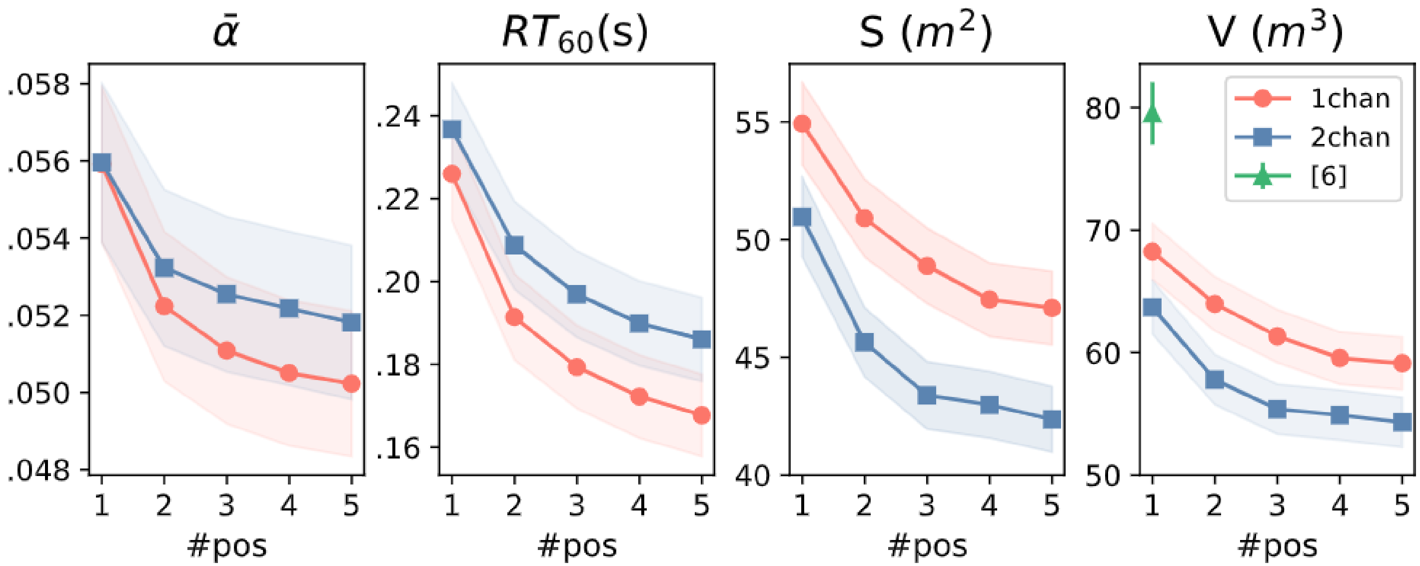

- A maximum-likelihood cost-function: $\mathcal{L}_\theta(\boldsymbol{x},\boldsymbol{y}) = -\log p_\theta(\boldsymbol{y}|\boldsymbol{x}) = -\log \mathcal{N}(\boldsymbol{y};\boldsymbol{\mu}_\theta(\boldsymbol{x}),\boldsymbol{\sigma}^2_\theta(\boldsymbol{x}))$

$$\stackrel{c}{=} \frac{1}{2}\sum_{d=1}^{D} \log \sigma^2_{d,\theta}(\boldsymbol{x}) + \frac{(y_d - \mu_{d,\theta}(\boldsymbol{x}))^2}{\sigma^2_{d,\theta}(\boldsymbol{x})}$$

- Allows aggreting multiple source-receiver recordings via Bayes' theorem:

$$\boxed{p_\theta(y_d|\bar{\boldsymbol{x}} = [\boldsymbol{x}_1,\dots,\boldsymbol{x}_J]) = \mathcal{N}(y_d;\bar{\mu}_{d,\theta}(\bar{\boldsymbol{x}}), 1/\bar{\gamma}^2_{d,\theta}(\bar{\boldsymbol{x}}))} \quad \bar{\mu}_{d,\theta}(\bar{\boldsymbol{x}}) = \sum_{j=1}^{J} \frac{\gamma^2_{d,\theta}(\boldsymbol{x}_j)}{\bar{\gamma}^2_{d,\theta}(\bar{\boldsymbol{x}})}\mu_{d,\theta}(\boldsymbol{x}_j),\ \bar{\gamma}^2_{d,\theta}(\bar{\boldsymbol{x}}) = \sum_{j=1}^{J}\gamma^2_{d,\theta}(\boldsymbol{x}_j)$$

# Example 3: Blind room parameter estimation [17]




| Method | Features | # pos | $\bar{\alpha}$ | $RT_{60}$ | $S$ | $V$ |
|---|---|---|---|---|---|---|
| [6] | SC | 1 | - | - | - | 137.8 |
| Ours | SC | 1 | **0.061** | 0.134 | 129.6 | 154.5 |
| Ours | SC | 5 | **0.060** | 0.097 | 125.8 | 149.1 |
| Ours | SC+IC | 1 | 0.084 | 0.101 | 89.4 | 107.6 |
| Ours | SC+IC | 5 | 0.094 | **0.062** | **50.2** | **68.8** |

Mean results on 3 real rooms [12] (30 rec. per room)

- ✓ 2-channels help V and S
- ✓ Multiple observati ons help
- ☒ Poor results for mean absorption below 1000 Hz

# OUTLINE

1) Intro & Background
2) Virtually-Supervised Learning
**3) Examples and Results**
4) Conclusions and Outlook

# OUTLINE

1) Intro & Background
2) Virtually-Supervised Learning
3) Examples and Results
**4) Conclusions and Outlook**

# 4) Conclusions & Outlook

- Recovering the geometry and acoustic profiles of a room from a clap recording is …. far from solved! In particular at low-freq (<1000 Hz)

# 4) Conclusions & Outlook

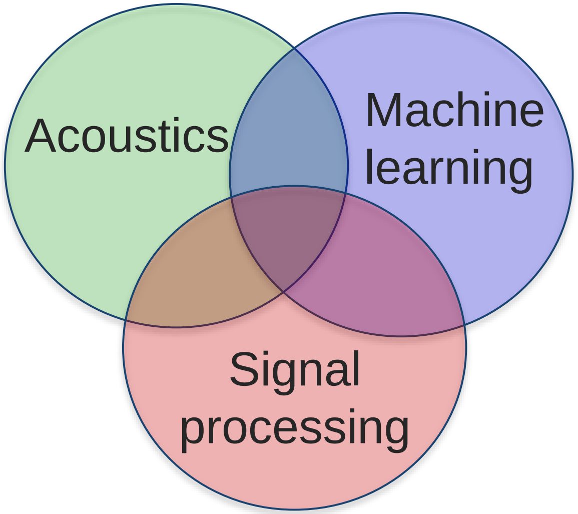
- Recovering the geometry and acoustic profiles of a room from a clap recording is …. far from solved! In particular at low-freq (<1000 Hz)
- Next challenge: estimation of individual surfaces/objects in the room?

# 4) Conclusions & Outlook

- Recovering the geometry and acoustic profiles of a room from a clap recording is …. far from solved! In particular at low-freq (<1000 Hz)
- Next challenge: estimation of individual surfaces/objects in the room?
- A variety of open and useful inverse problems in acoustics

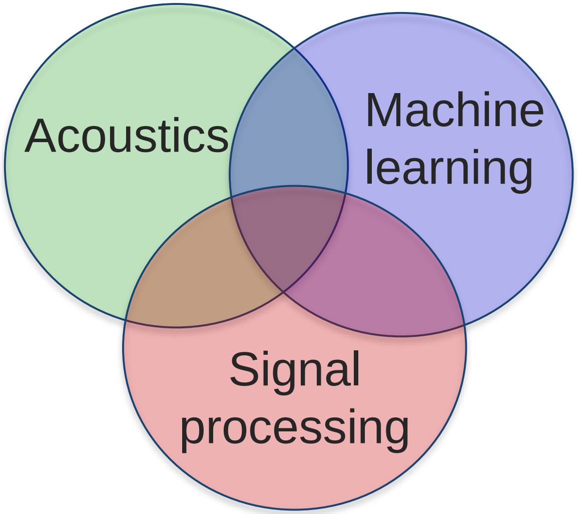
# 4) Conclusions & Outlook

- Recovering the geometry and acoustic profiles of a room from a clap recording is …. far from solved! In particular at low-freq (<1000 Hz)
- Next challenge: estimation of individual surfaces/objects in the room?
- A variety of open and useful inverse problems in acoustics
- *How to develop « hybrid » models that are jointly driven by data, signal processing and physics?*



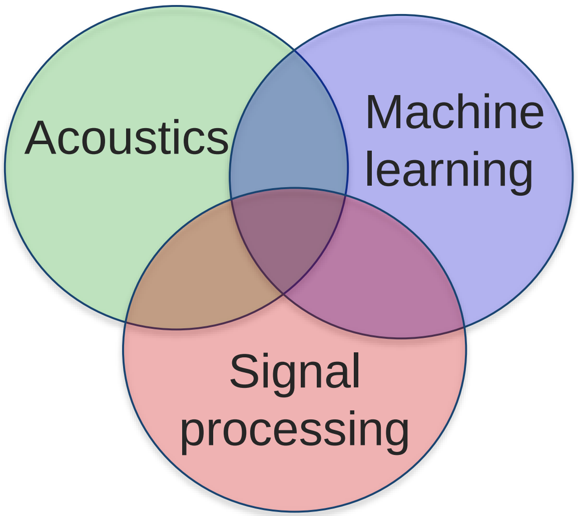
# 4) Conclusions & Outlook

- Recovering the geometry and acoustic profiles of a room from a clap recording is …. far from solved! In particular at low-freq (<1000 Hz)
- Next challenge: estimation of individual surfaces/objects in the room?
- A variety of open and useful inverse problems in acoustics
- *How to develop « hybrid » models that are jointly driven by data, signal processing and physics?*
- Promising directions: domain adaptation, transfer learning, generative models, adversarial networks, neural RIR generation, …



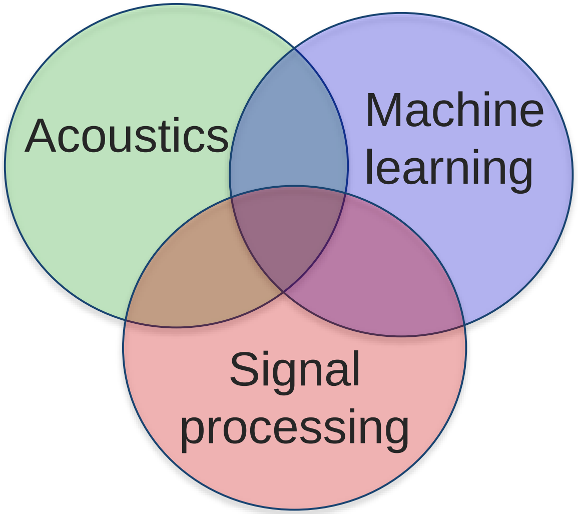
# 4) Conclusions & Outlook

- Recovering the geometry and acoustic profiles of a room from a clap recording is …. far from solved! In particular at low-freq (<1000 Hz)
- Next challenge: estimation of individual surfaces/objects in the room?
- A variety of open and useful inverse problems in acoustics
- *How to develop « hybrid » models that are jointly driven by data, signal processing and physics?*
- Promising directions: domain adaptation, transfer learning, generative models, adversarial networks, neural RIR generation, …
- *Coming soon:* extension of Pyroomacoustics that loads measured source and receiver directivity profiles



# 4) Conclusions & Outlook

- Recovering the geometry and acoustic profiles of a room from a clap recording is …. far from solved! In particular at low-freq (<1000 Hz)
- Next challenge: estimation of individual surfaces/objects in the room?
- A variety of open and useful inverse problems in acoustics
- *How to develop « hybrid » models that are jointly driven by data, signal processing and physics?*
- Promising directions: domain adaptation, transfer learning, generative models, adversarial networks, neural RIR generation, …
- *Coming soon:* extension of Pyroomacoustics that loads measured source and receiver directivity profiles




***Thank You ! Questions?***

# Bibliography

[1] A Deleforge, F Forbes, R Horaud (2015), "Acoustic space learning for sound-source separation and localization on binaural manifolds", International journal of neural systems 25(01):1440003

[2] A Deleforge, R Horaud, YY Schechner, L Girin (2015), "Co-localization of audio sources in images using binaural features and locally-linear regression", IEEE/ACM Transactions on Audio, Speech, and Language Processing 23(4):718-731

[3] A Deleforge, R Horaud (2012), "2D sound-source localization on the binaural manifold", IEEE International Workshop on Machine Learning for Signal Processing, 1-6

[4] D Di Carlo, A Deleforge, N Bertin (2019), "Mirage: 2d source localization using microphone pair augmentation with echoes", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

[5] S Kataria, C Gaultier, A Deleforge (2017), "Hearing in a shoe-box: binaural source position and wall absorption estimation using virtually supervised learning", IEEE International Conference on Acoustics, Speech and Signal and Signal Processing (ICASSP)

[6] A F Genovese, H Gamper, V Pulkki, N Raghuvanshi, I J Tashev (2019), "Blind room volume estimation from single-channel noisy speech.", IEEE International Conference on Acoustics, Speech and Signal and Signal Processing (ICASSP)

[7] R Scheibler, D Di Carlos, A Deleforge, I Dokmanic (2018), "Separake: Source separation with a little help from echoes", IEEE International Conference on Acoustics, Speech and Signal and Signal Processing (ICASSP)

[8] I Dokmanić, R Scheibler, M Vetterli (2015), "Raking the cocktail party", IEEE journal of selected topics in signal processing 9.5:825-836.

[9] C Gaultier, S Kataria, A Deleforge (2017), "VAST: The virtual acoustic space traveler dataset", International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)

# Bibliography


[10] HP Tukuljac, A Deleforge, R Gribonval (2018), "Mulan: A blind and off-grid method for multichannel echo retrieval", Neural Information Processing Conference (NeurIPS)

[11] D Di Carlo, C Elvira, A Deleforge, N Bertin, R Gribonval (2020), "Blaster: An off-grid method for blind and regularized acoustic echoes retrieval", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

[12] D Di Carlo, P Tandeitnik, C Foy, A Deleforge, N Bertin, S Gannot, "dEchorate: a Calibrated Room Impulse Response Database for Echo-aware Signal Processing (2021)", EURASIP Journal on Audio, Speech, and Music Processing

[13] B Allen, D A Berkley (1979), "Image method for efficiently simulating small-room acoustics.", The Journal of the Acoustical Society of America 65.4:943-950.

[14] Schimmel, Steven M., Martin F. Muller, and Norbert Dillier (2009), "A fast and accurate shoebox room acoustics simulator", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

[15] R Scheibler, E Bezzam, and I Dokmanić (2018), "Pyroomacoustics: A python package for audio room simulation and array processing algorithms", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

[16] C Foy, A Deleforge, D Di Carlo, "Mean absorption estimation from room impulse responses using virtually supervised learning", The Journal of the Acoustical Society of America 150 (2), 1286-1299

[17] P Srivastava, A Deleforge, E Vincent (2021), "Blind Room Parameter Estimation Using Multiple-Multichannel Speech Recordings", IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)

[18] I Dokmanić, R Parhizkar, A Walther, Y M Lu, M Vetterli (2013), "Acoustic echoes reveal room shape", Proceedings of the National Academy of Sciences 110(30):12186-12191.