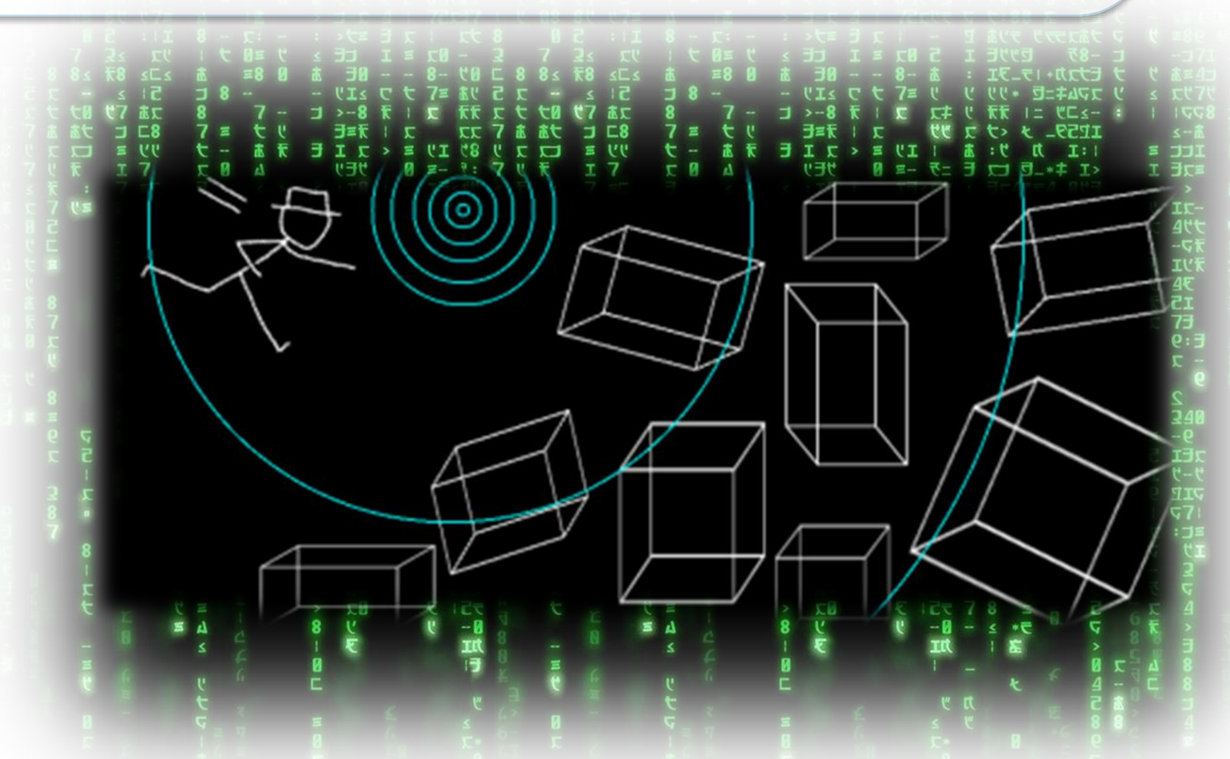
Acoustic Space Learning:
from Robots to Simulations

2009 - 2022







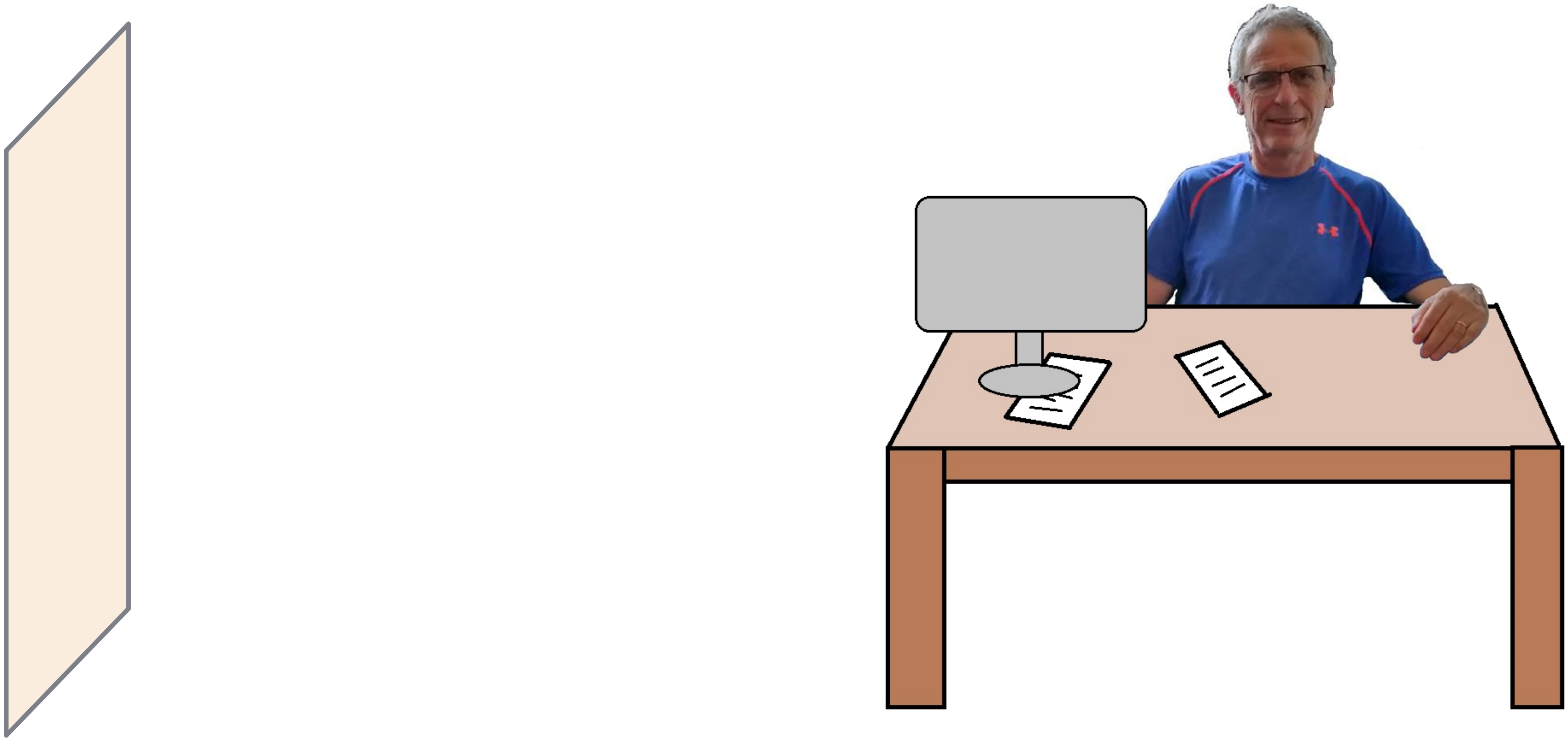
Antoine Deleforge, Inria Nancy – Grand Est, team MULTISPEECH

Antoine.Deleforge@inria.fr

How I met Radu

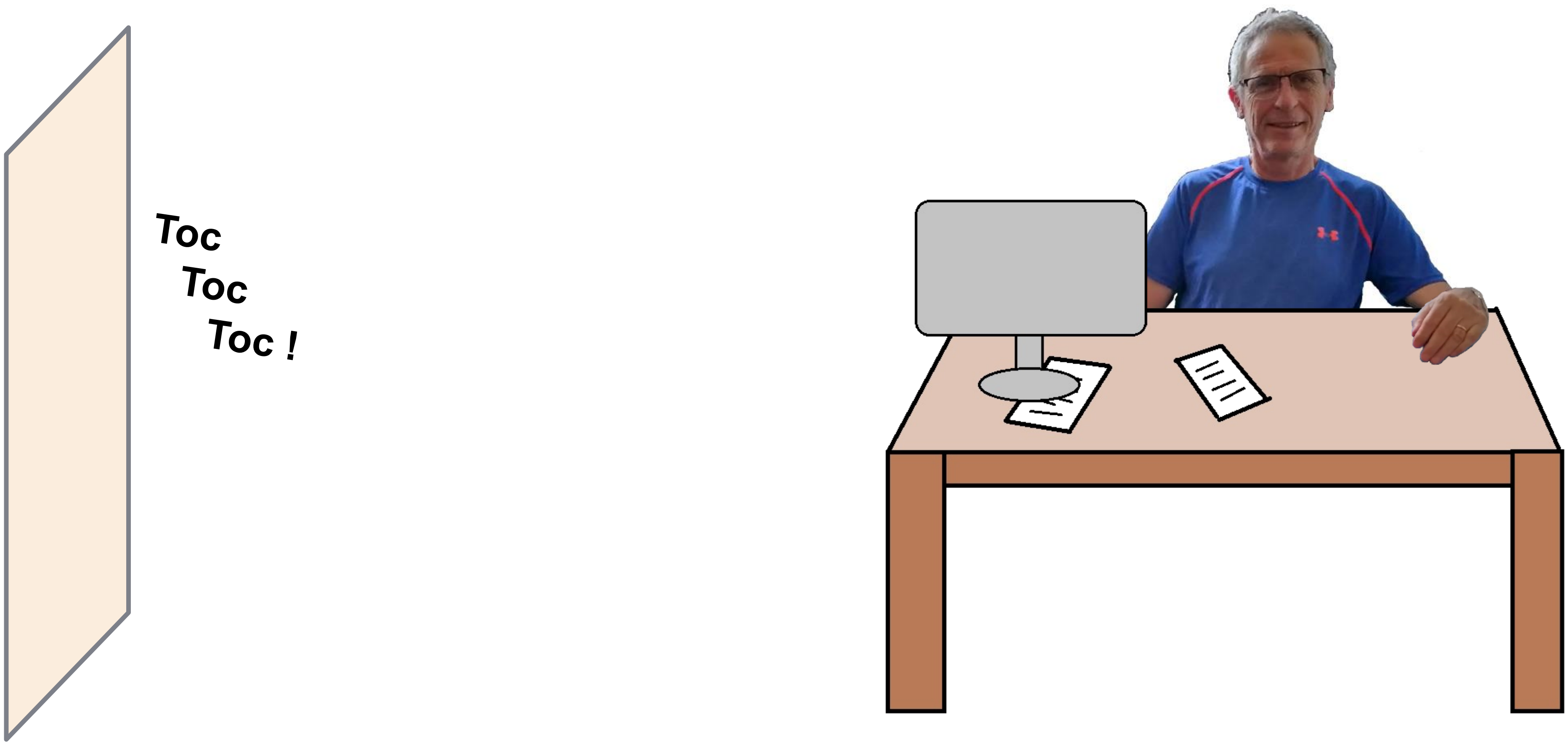
2009

March 2009, Radu's office, Inria Grenoble



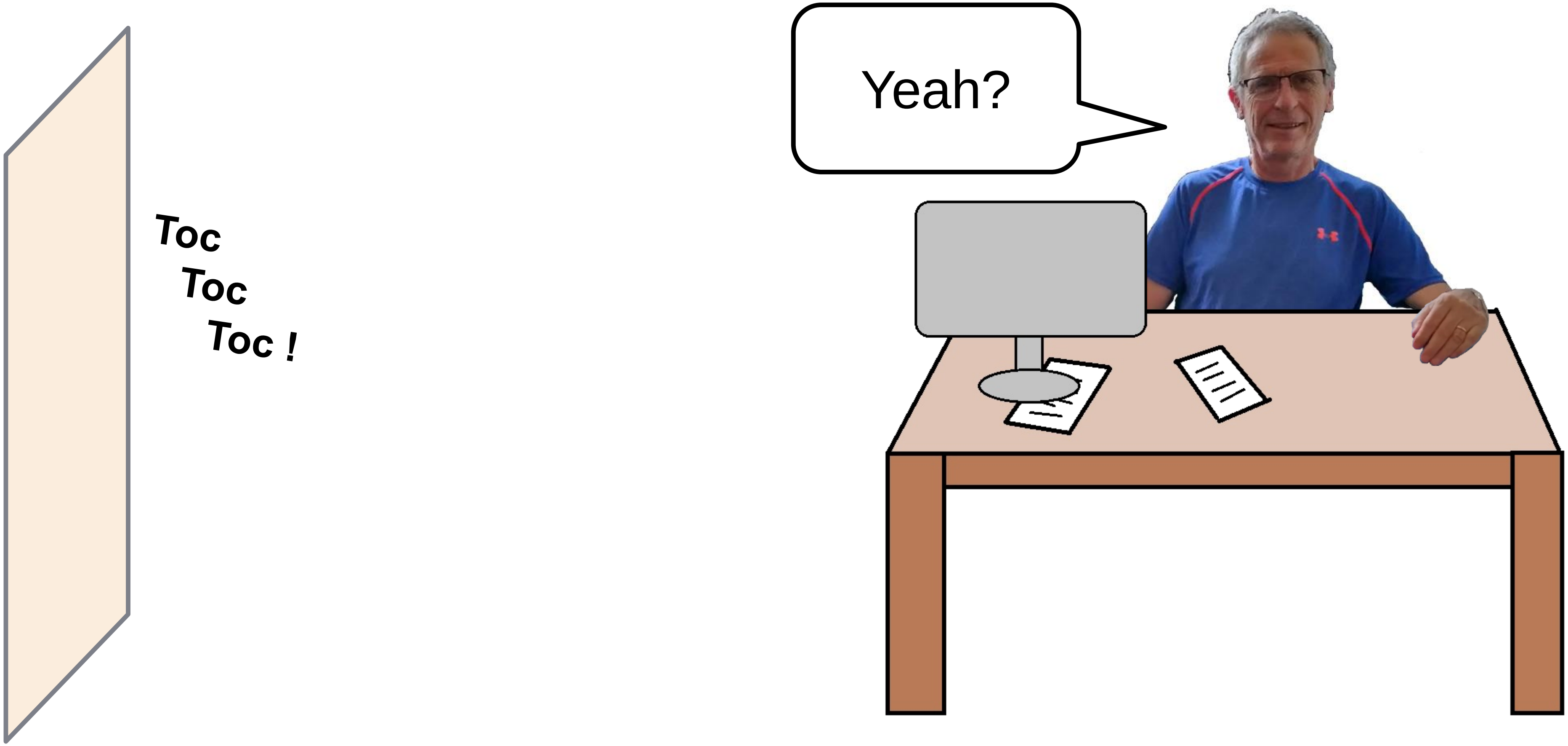
How I met Radu

March 2009, Radu’s office, Inria Grenoble



How I met Radu

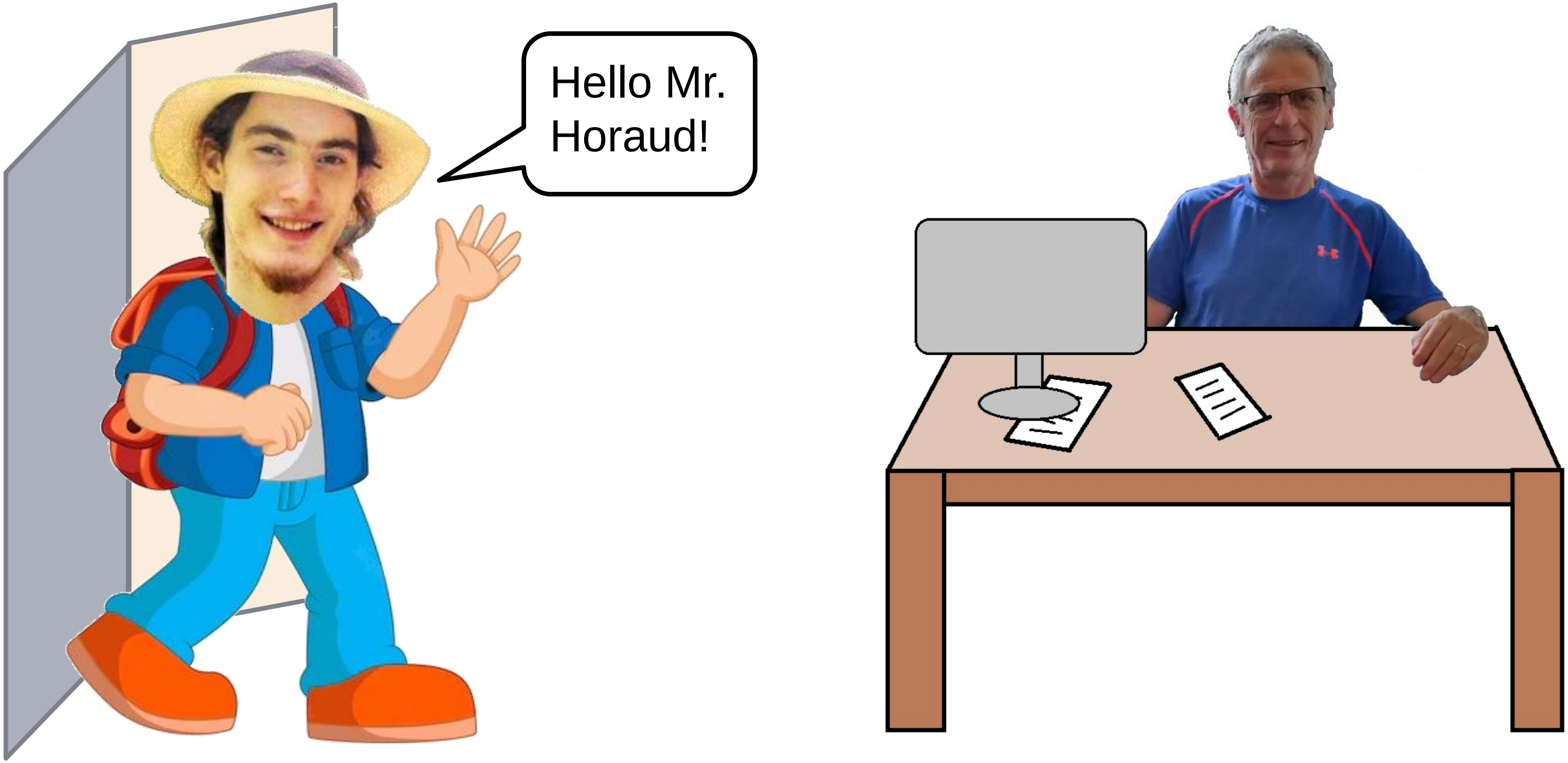
March 2009, Radu's office, Inria Grenoble



How I met Radu

2009

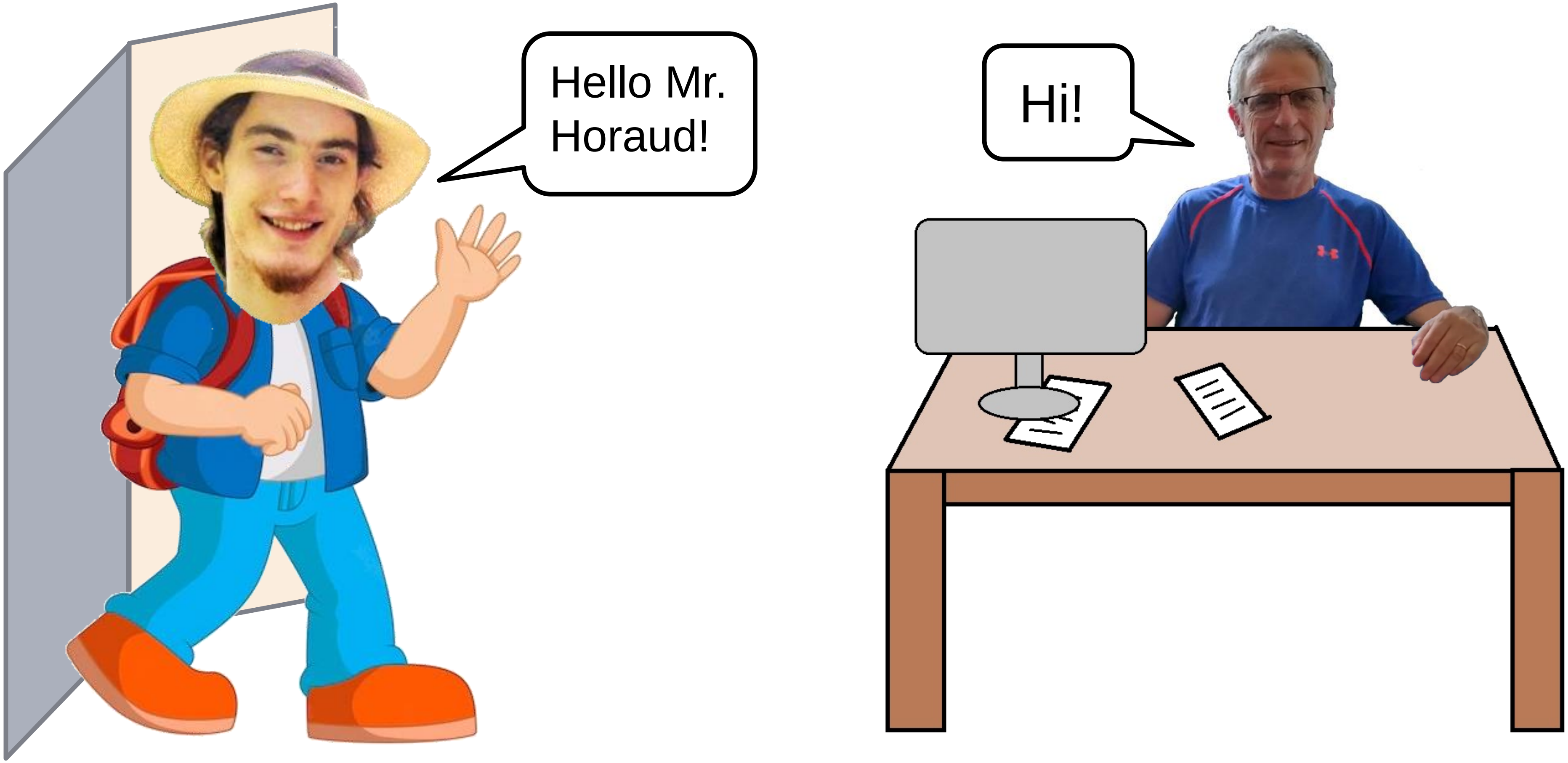
March 2009, Radu's office, Inria Grenoble



How I met Radu

2009

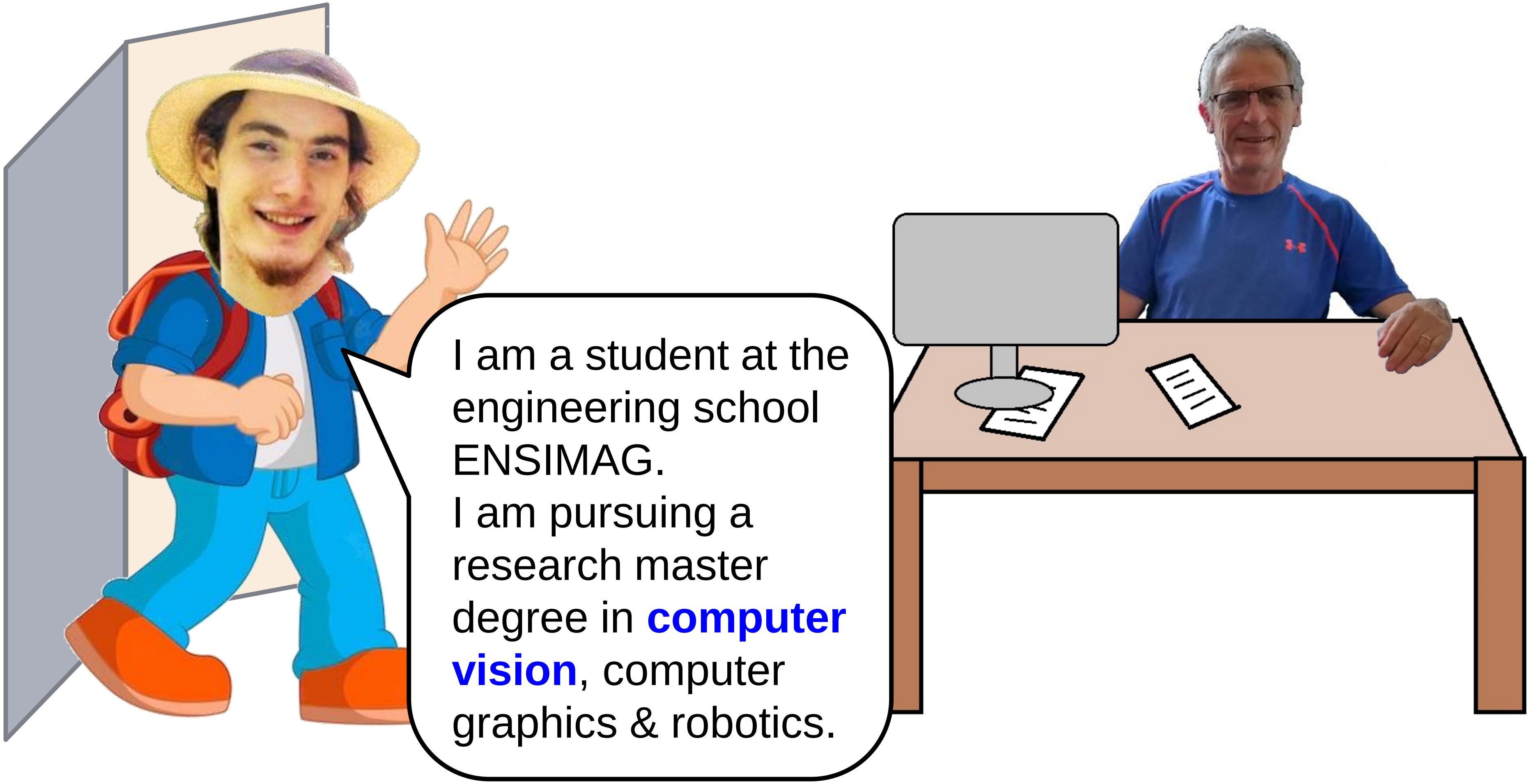
March 2009, Radu's office, Inria Grenoble



How I met Radu

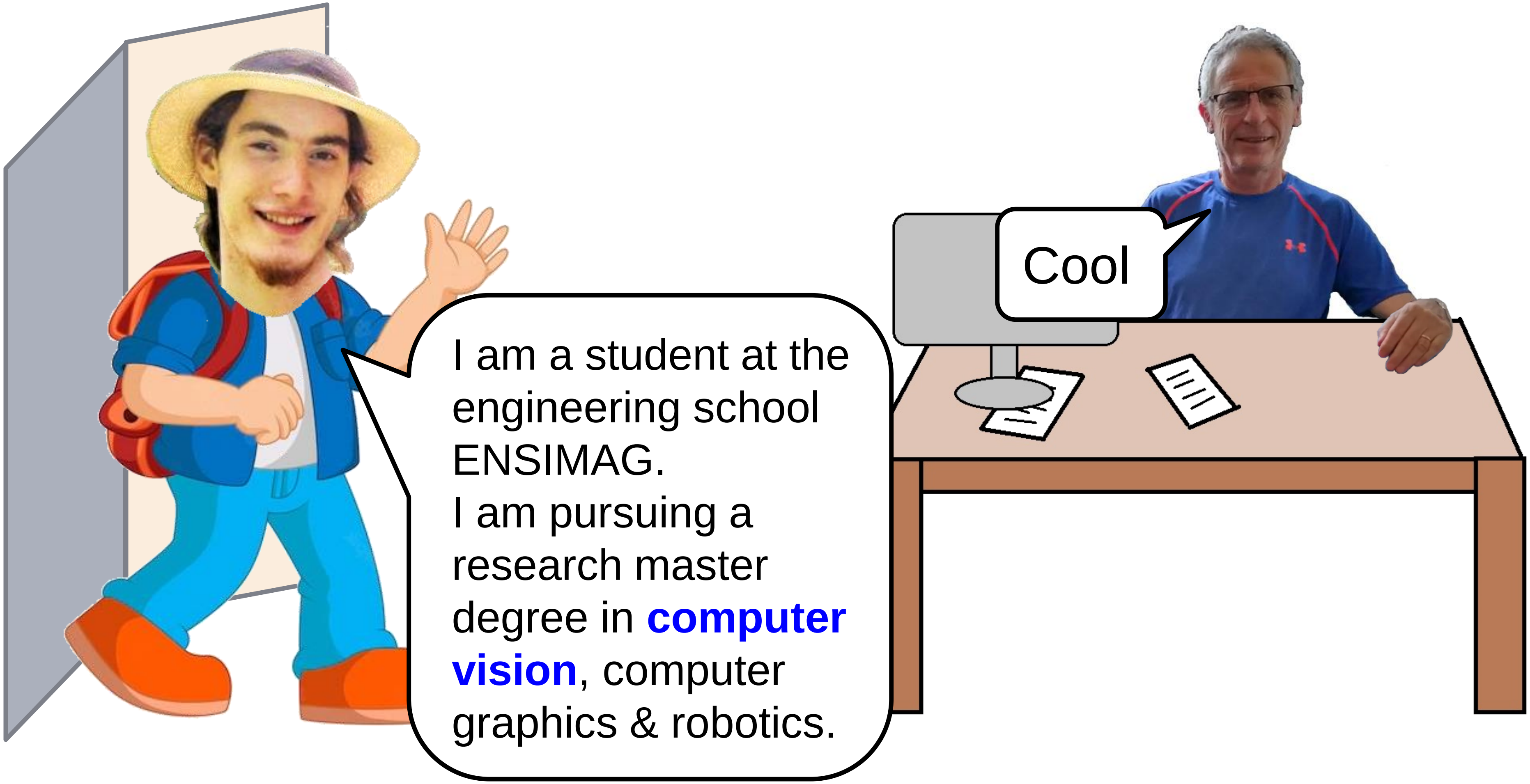
2009

March 2009, Radu’s office, Inria Grenoble



How I met Radu

March 2009, Radu's office, Inria Grenoble



How I met Radu

2009

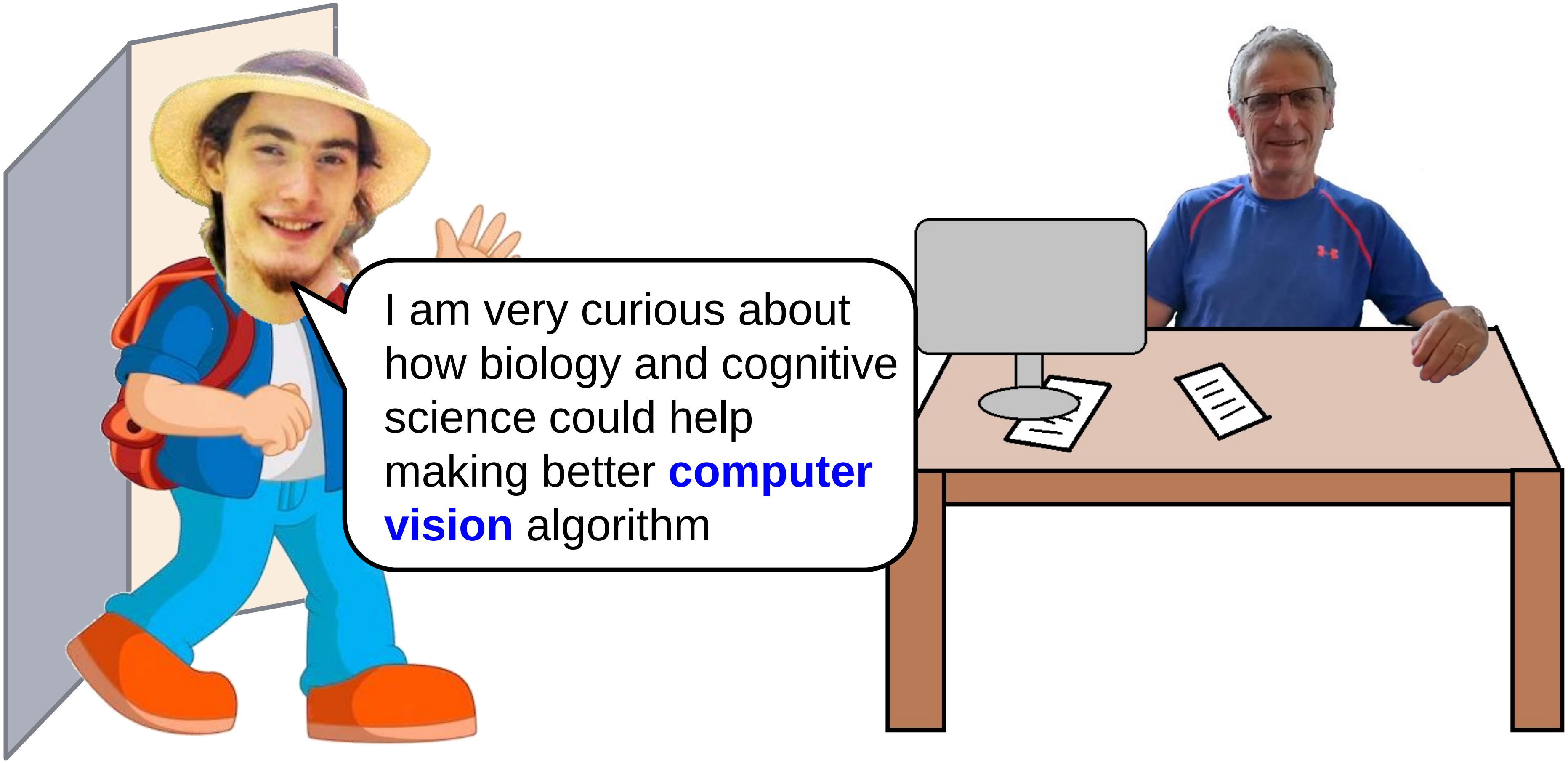
March 2009, Radu's office, Inria Grenoble

How I met Radu

March 2009, Radu's office, Inria Grenoble

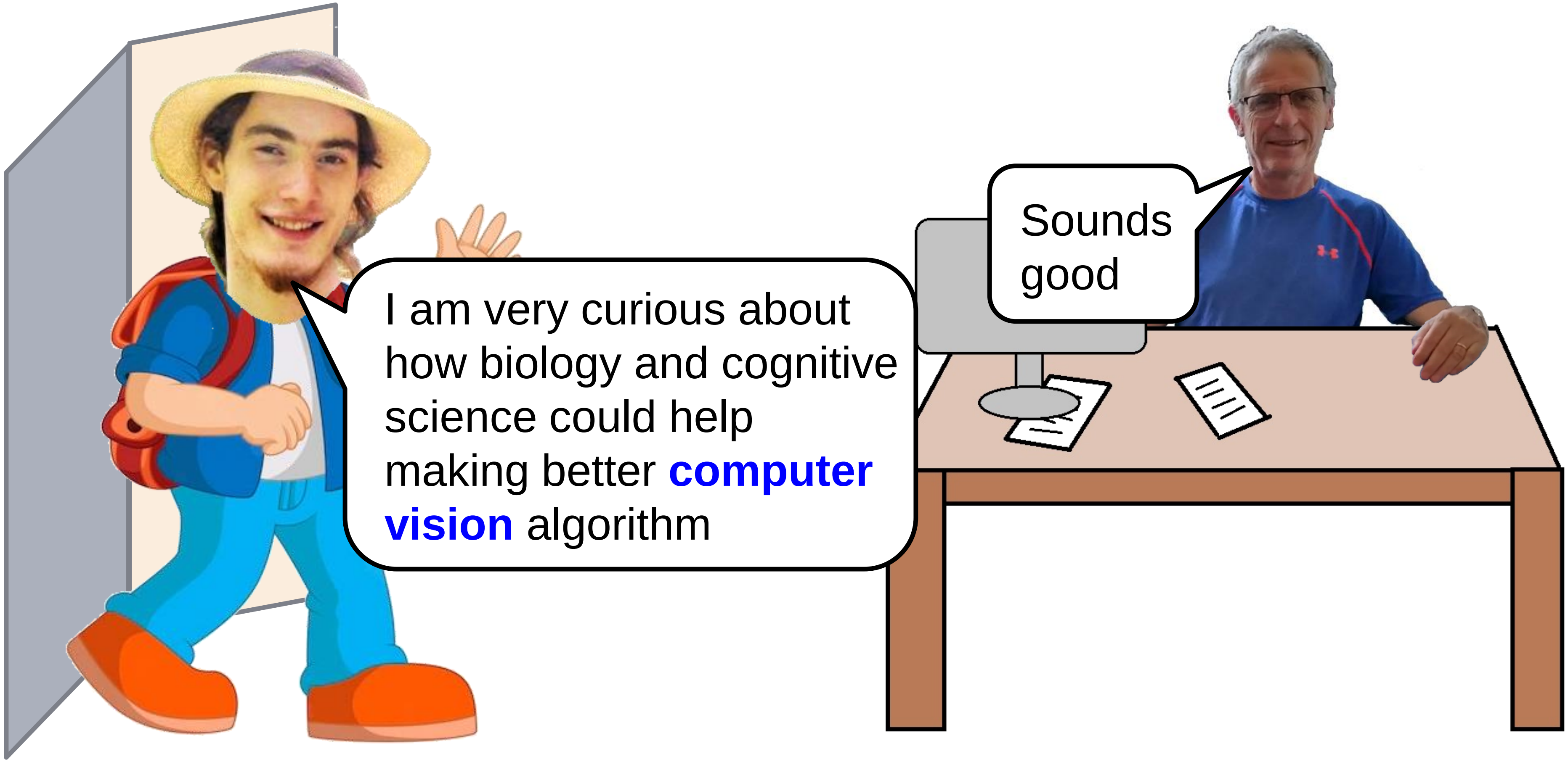
How I met Radu

March 2009, Radu's office, Inria Grenoble



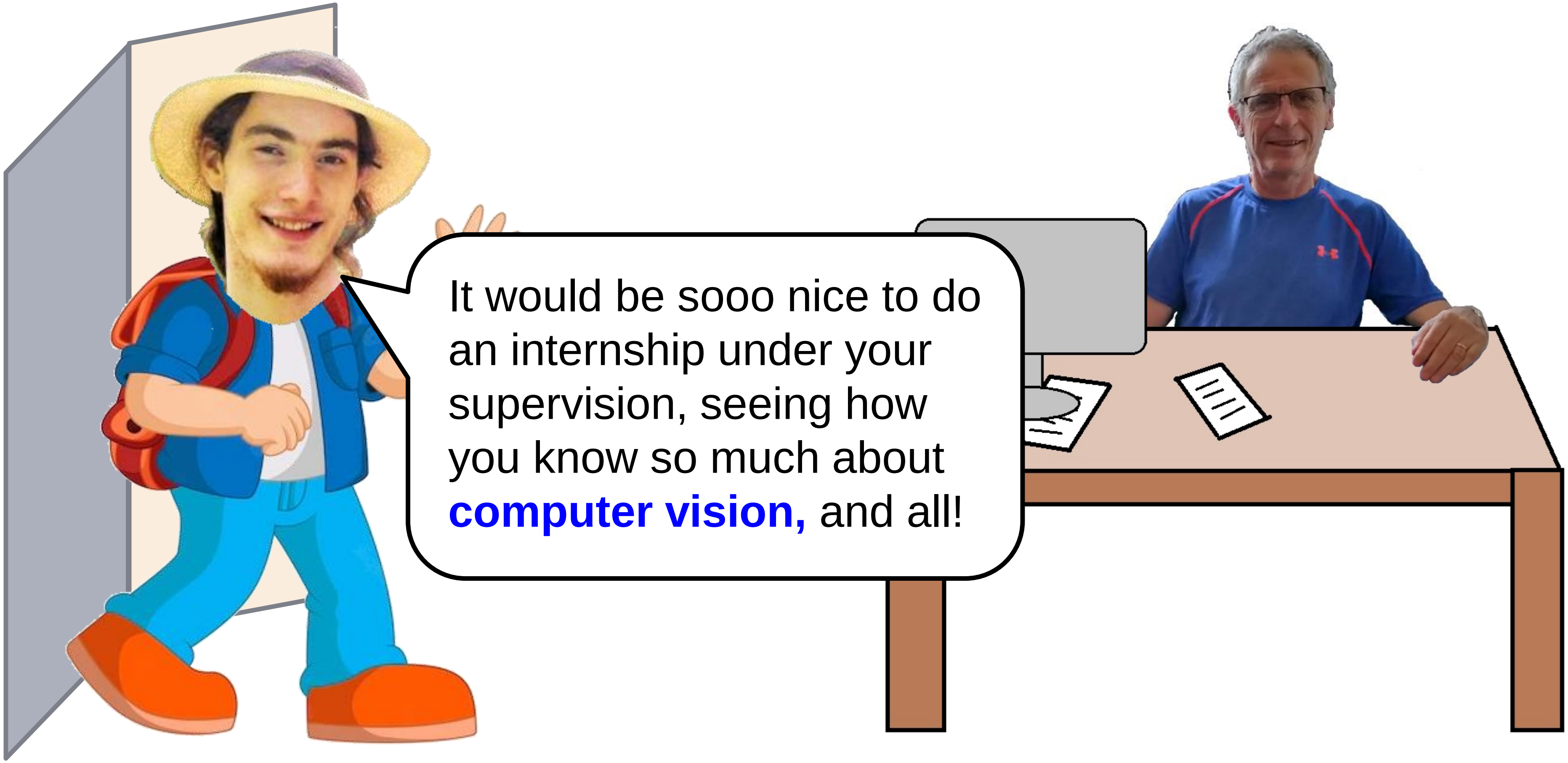
How I met Radu

March 2009, Radu's office, Inria Grenoble



How I met Radu

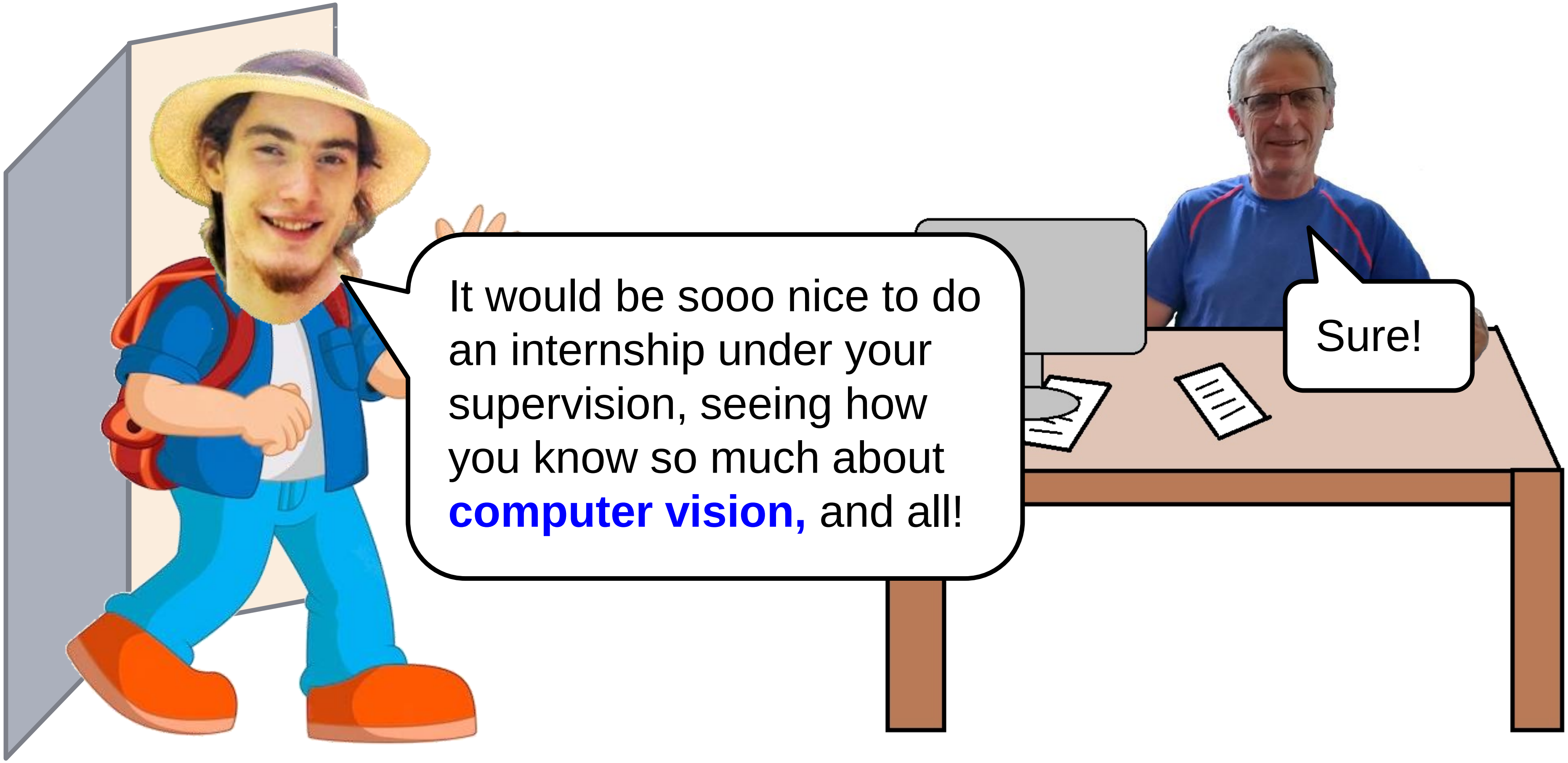
March 2009, Radu's office, Inria Grenoble



How I met Radu

2009

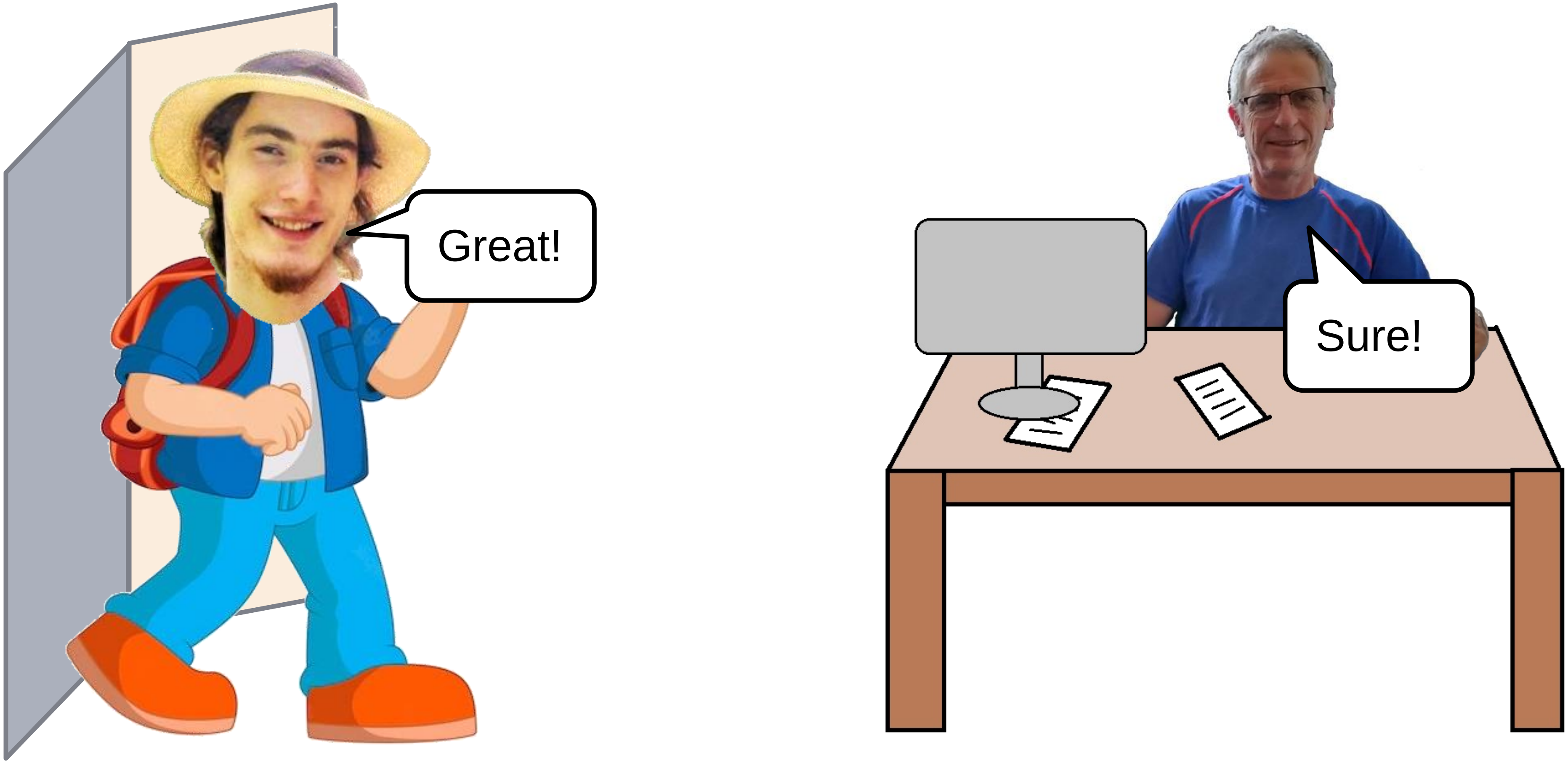
March 2009, Radu's office, Inria Grenoble



How I met Radu

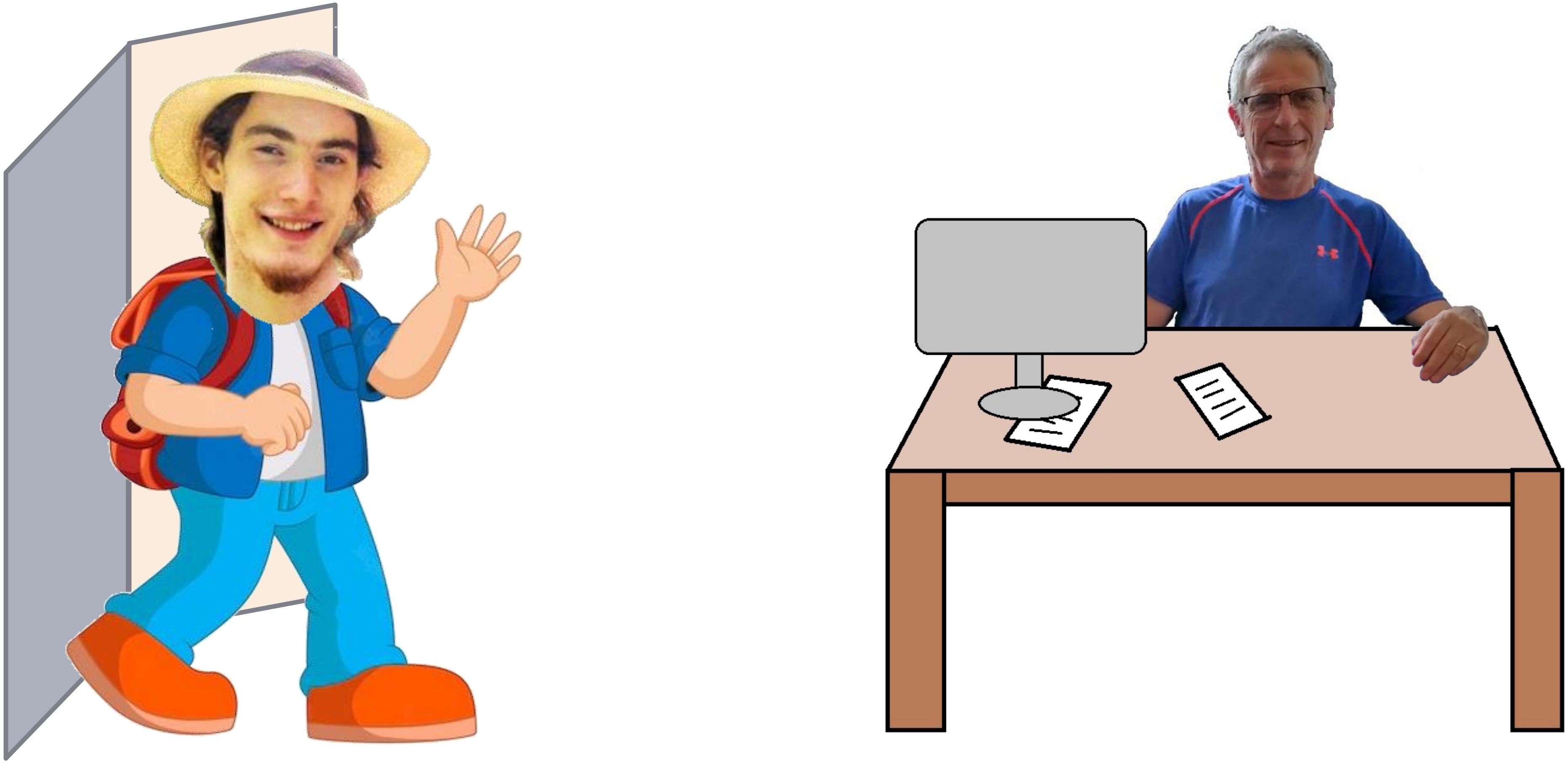
2009

March 2009, Radu's office, Inria Grenoble



How I met Radu

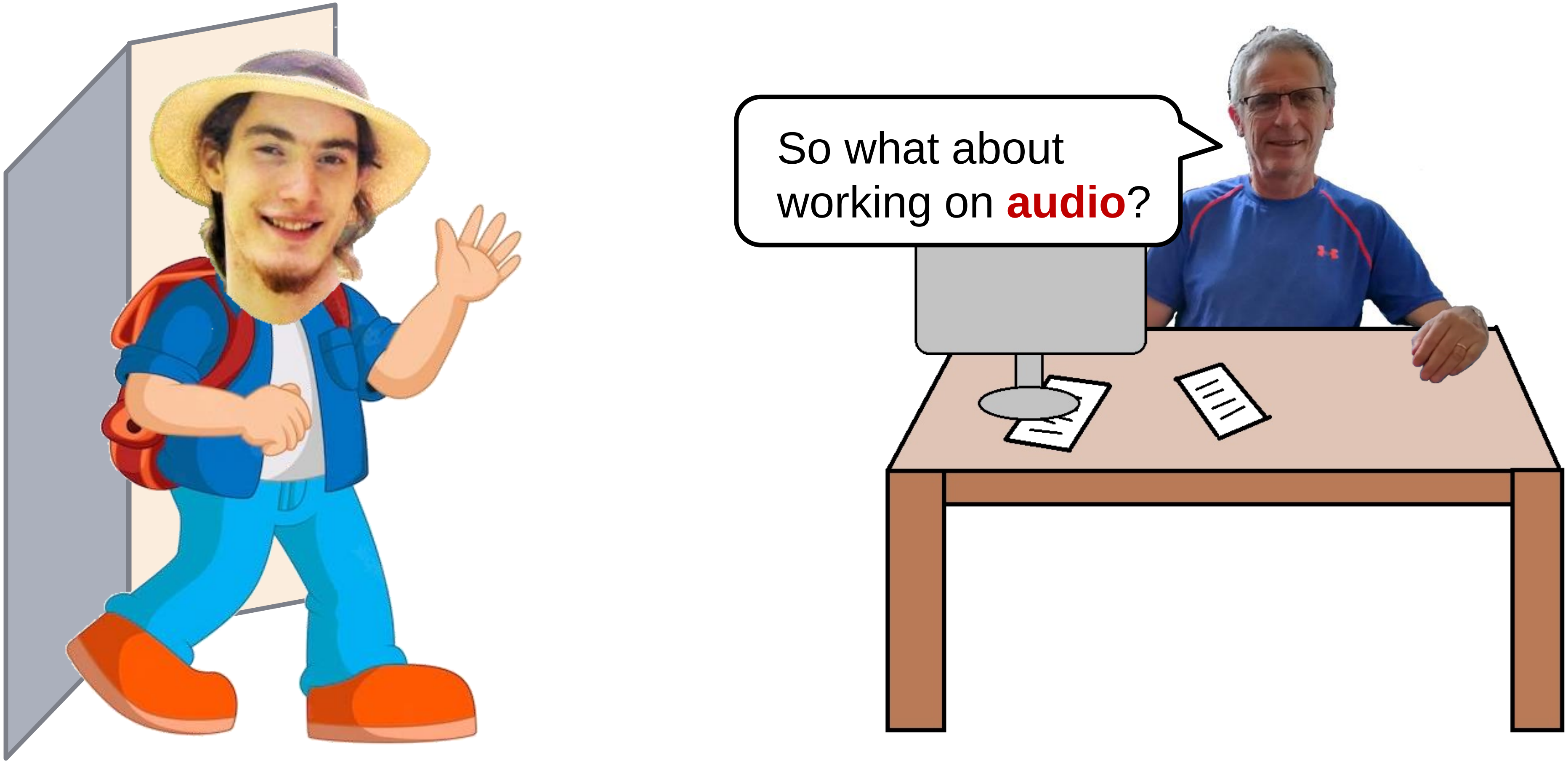
March 2009, Radu's office, Inria Grenoble



How I met Radu

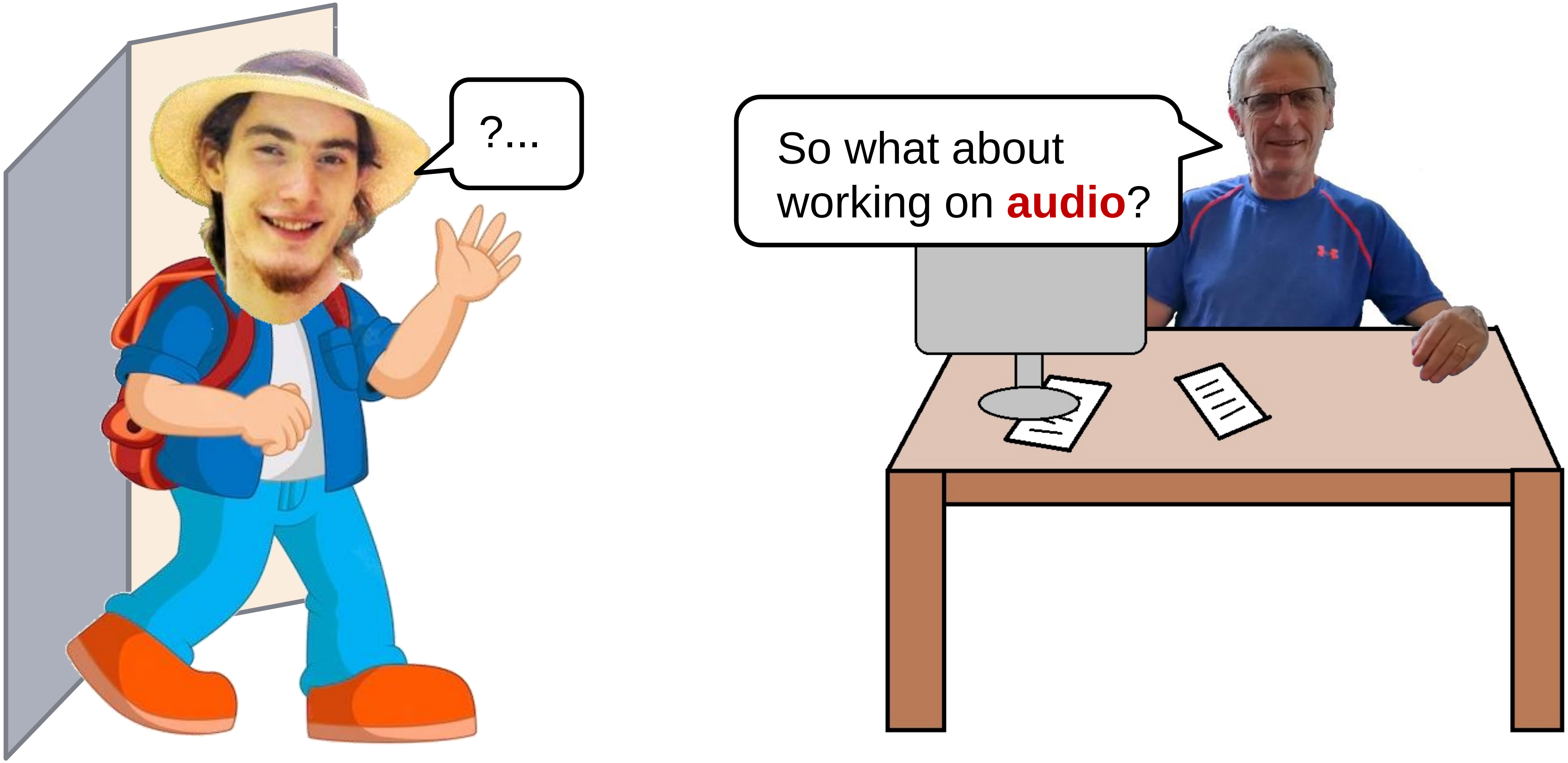
2009

March 2009, Radu's office, Inria Grenoble



How I met Radu

March 2009, Radu's office, Inria Grenoble



How I met Radu

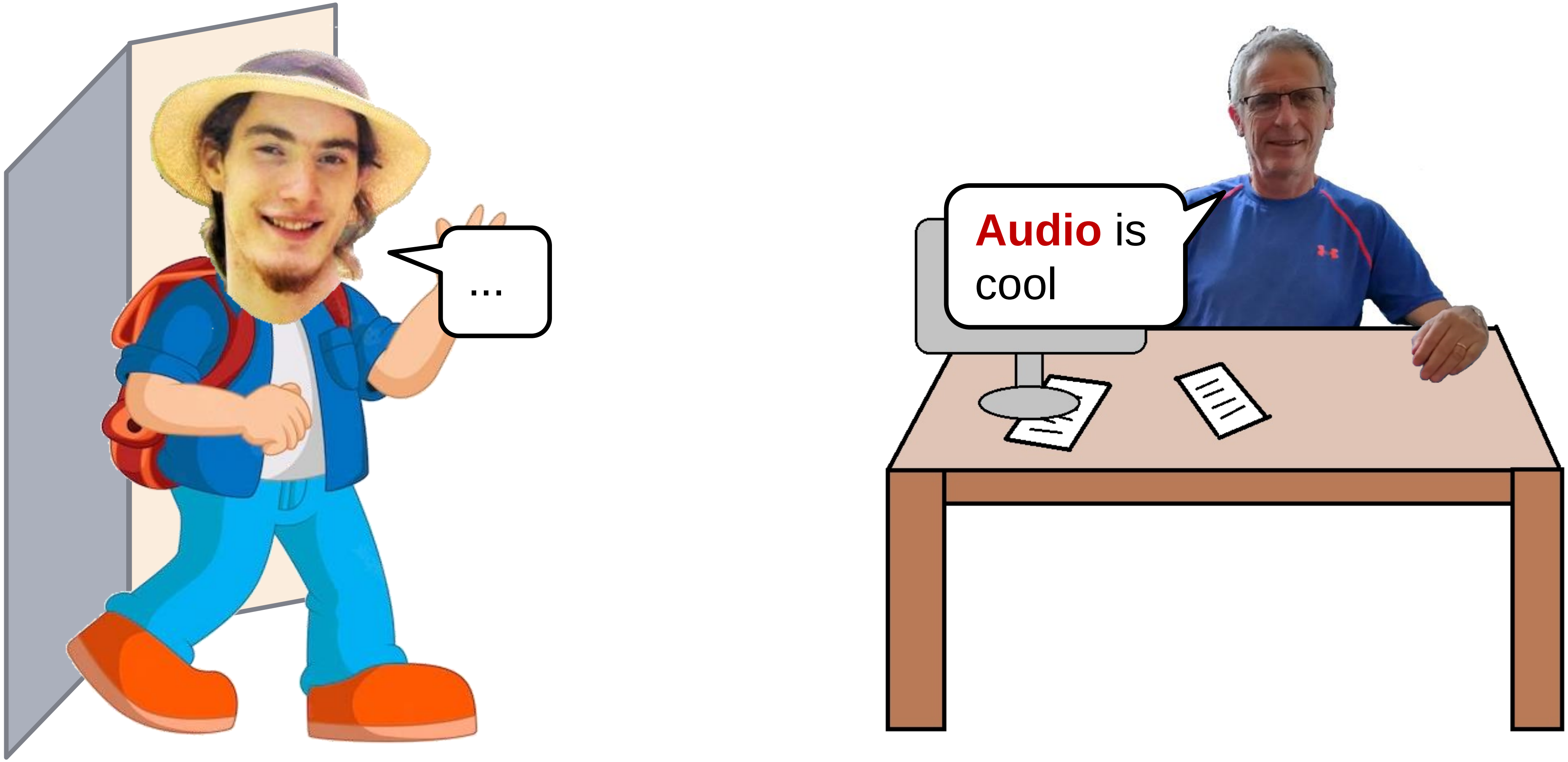
2009

March 2009, Radu's office, Inria Grenoble

How I met Radu

2009

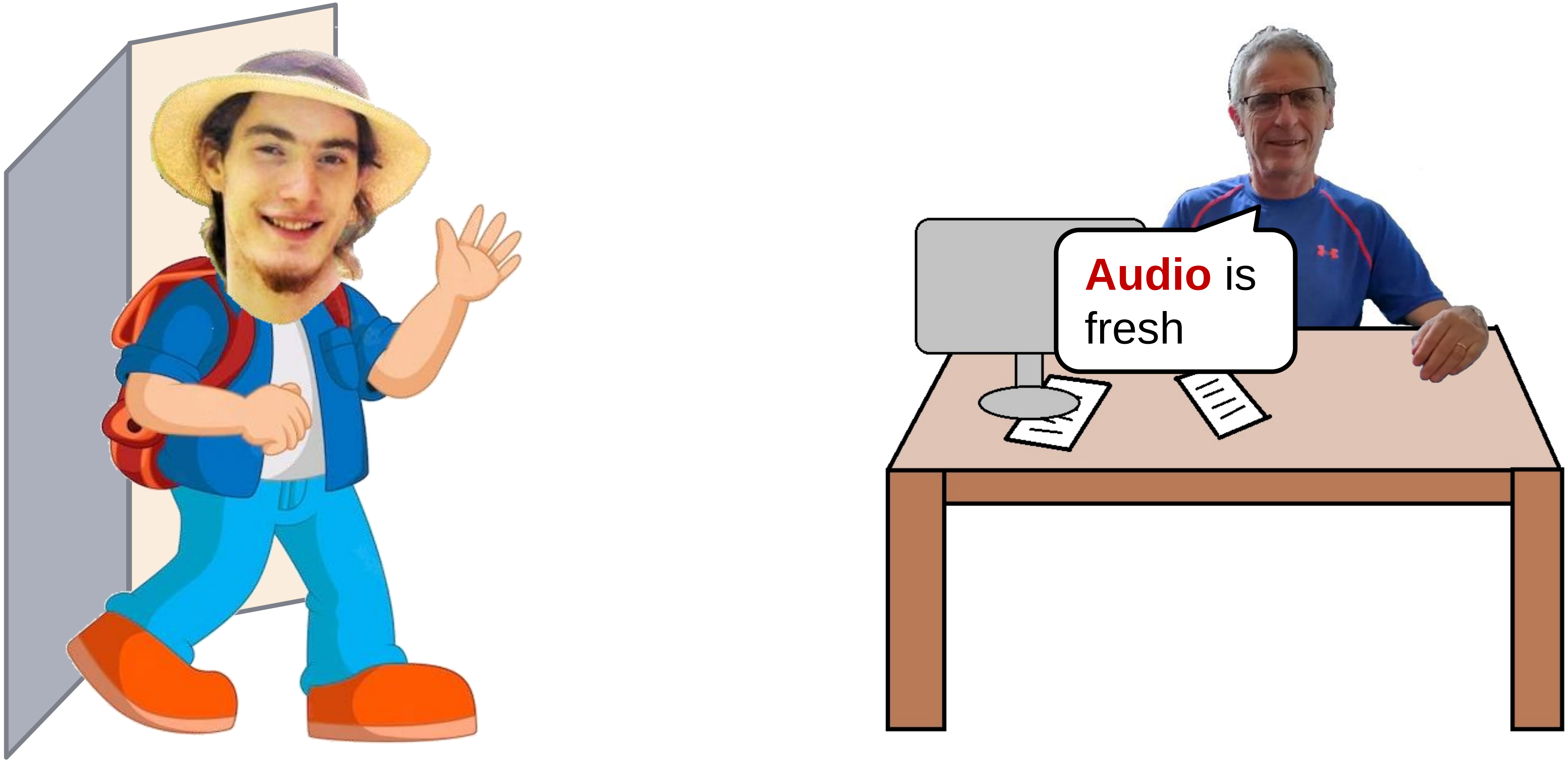
March 2009, Radu's office, Inria Grenoble



How I met Radu

2009

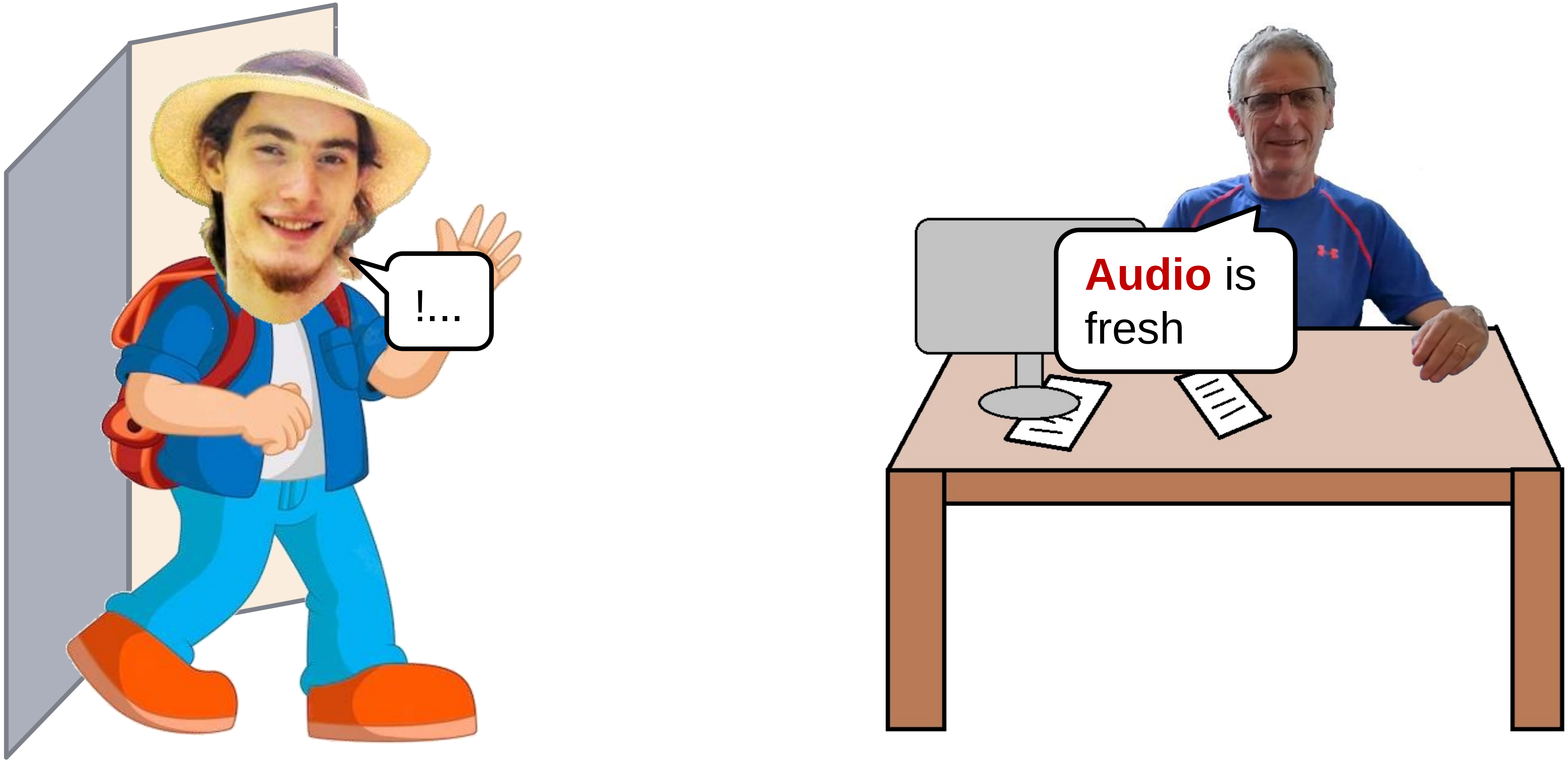
March 2009, Radu’s office, Inria Grenoble



How I met Radu

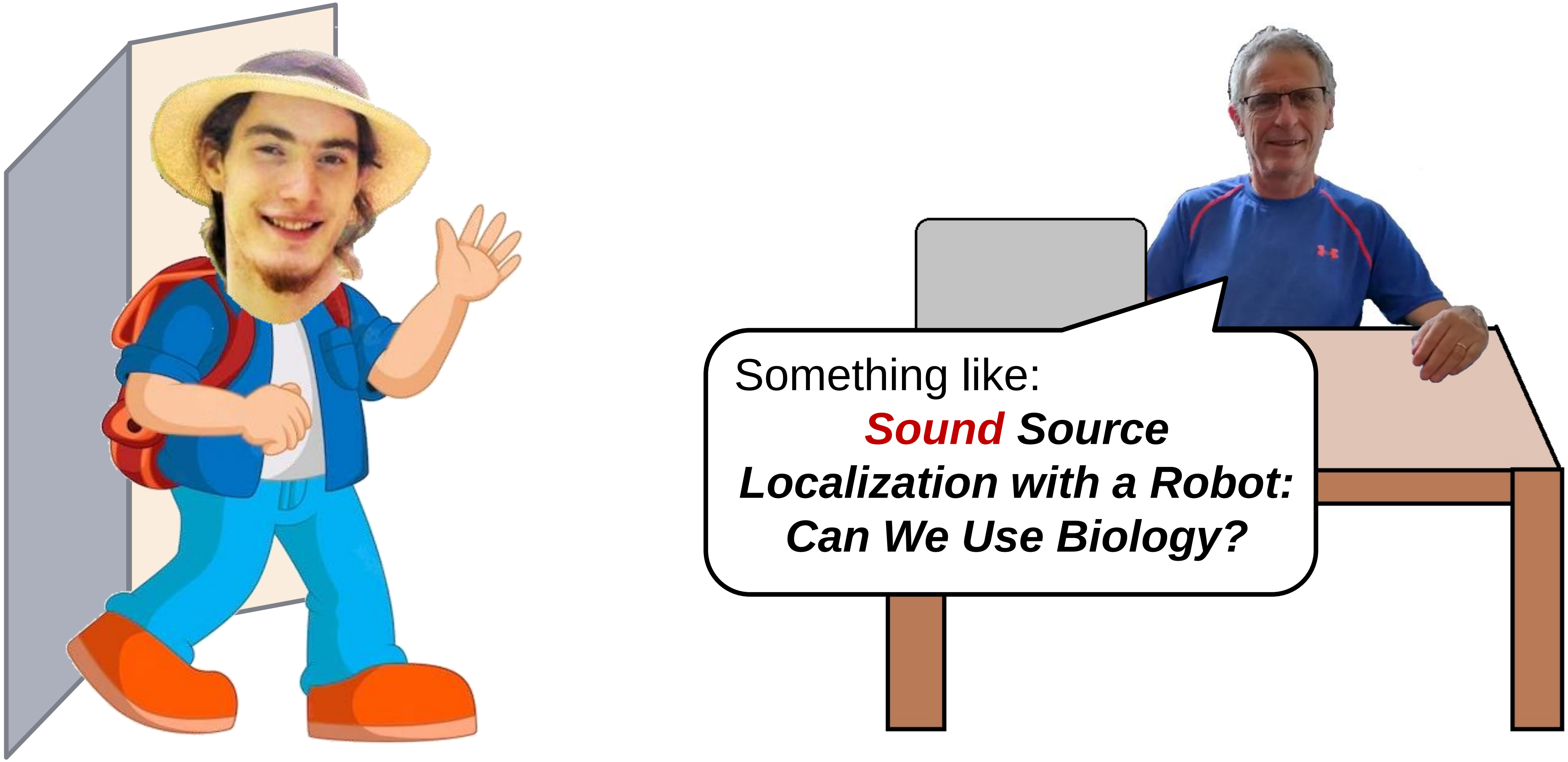
2009

March 2009, Radu’s office, Inria Grenoble



How I met Radu

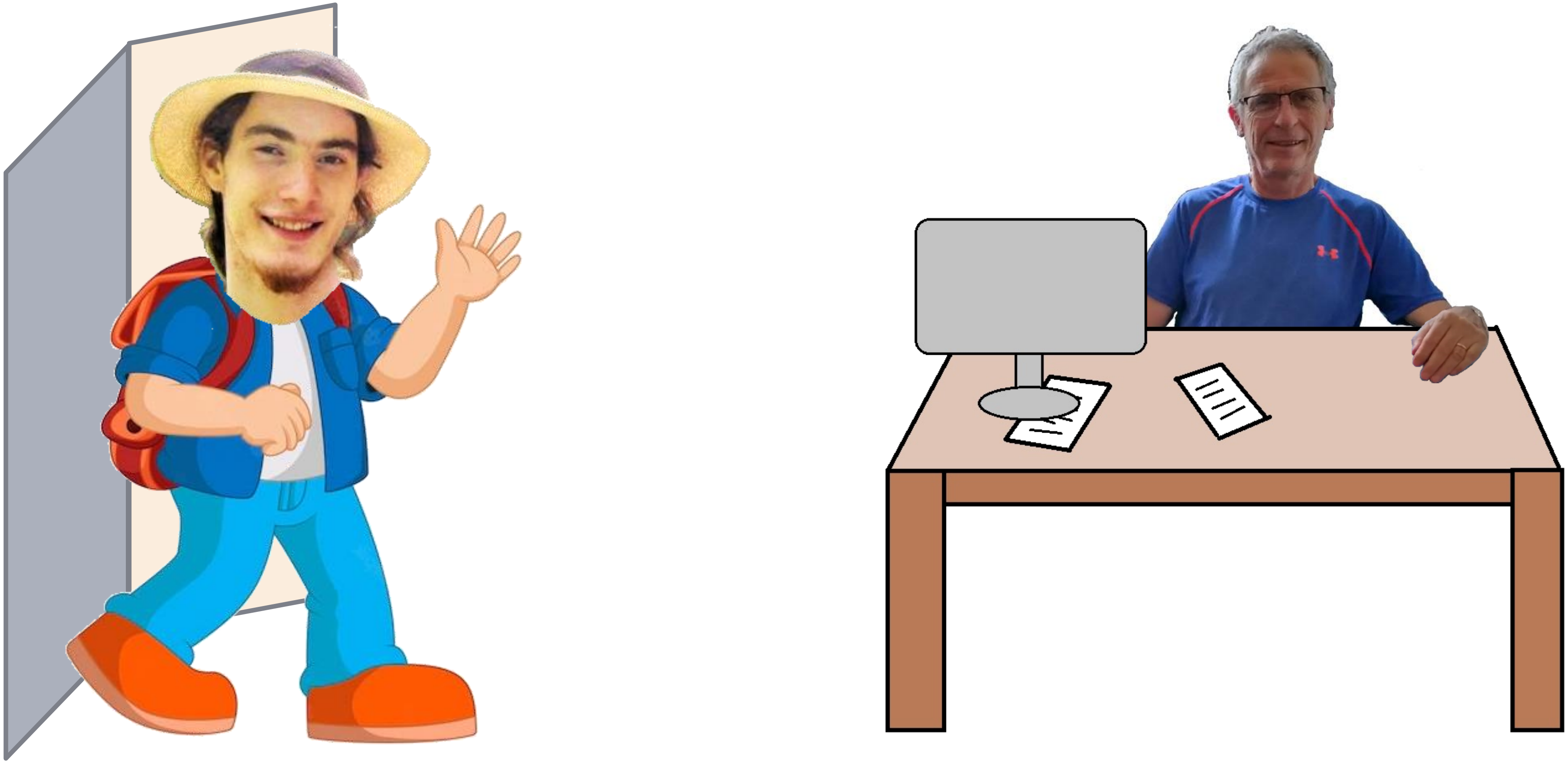
March 2009, Radu's office, Inria Grenoble



How I met Radu

2009

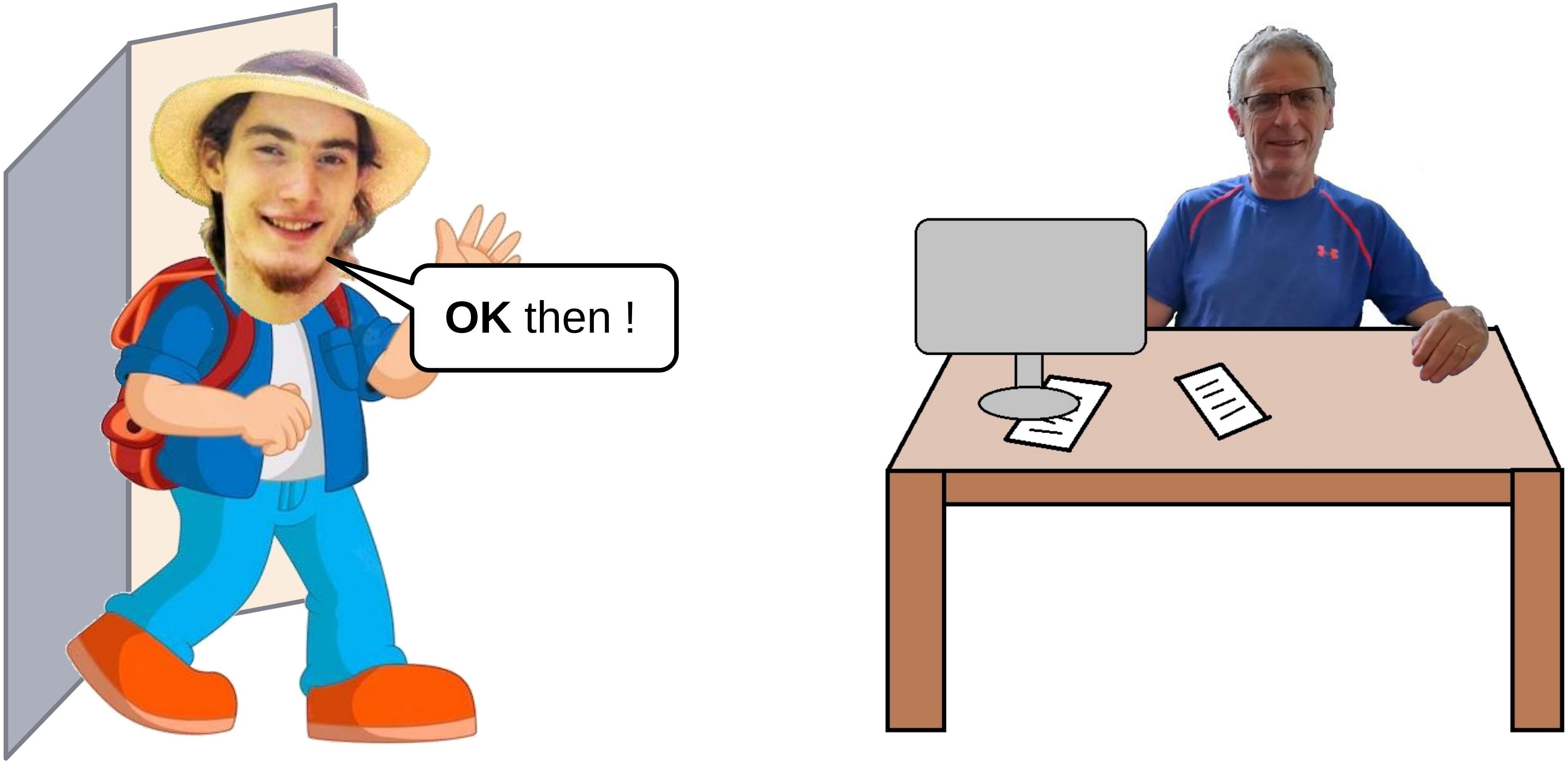
March 2009, Radu's office, Inria Grenoble



How I met Radu

2009

March 2009, Radu's office, Inria Grenoble

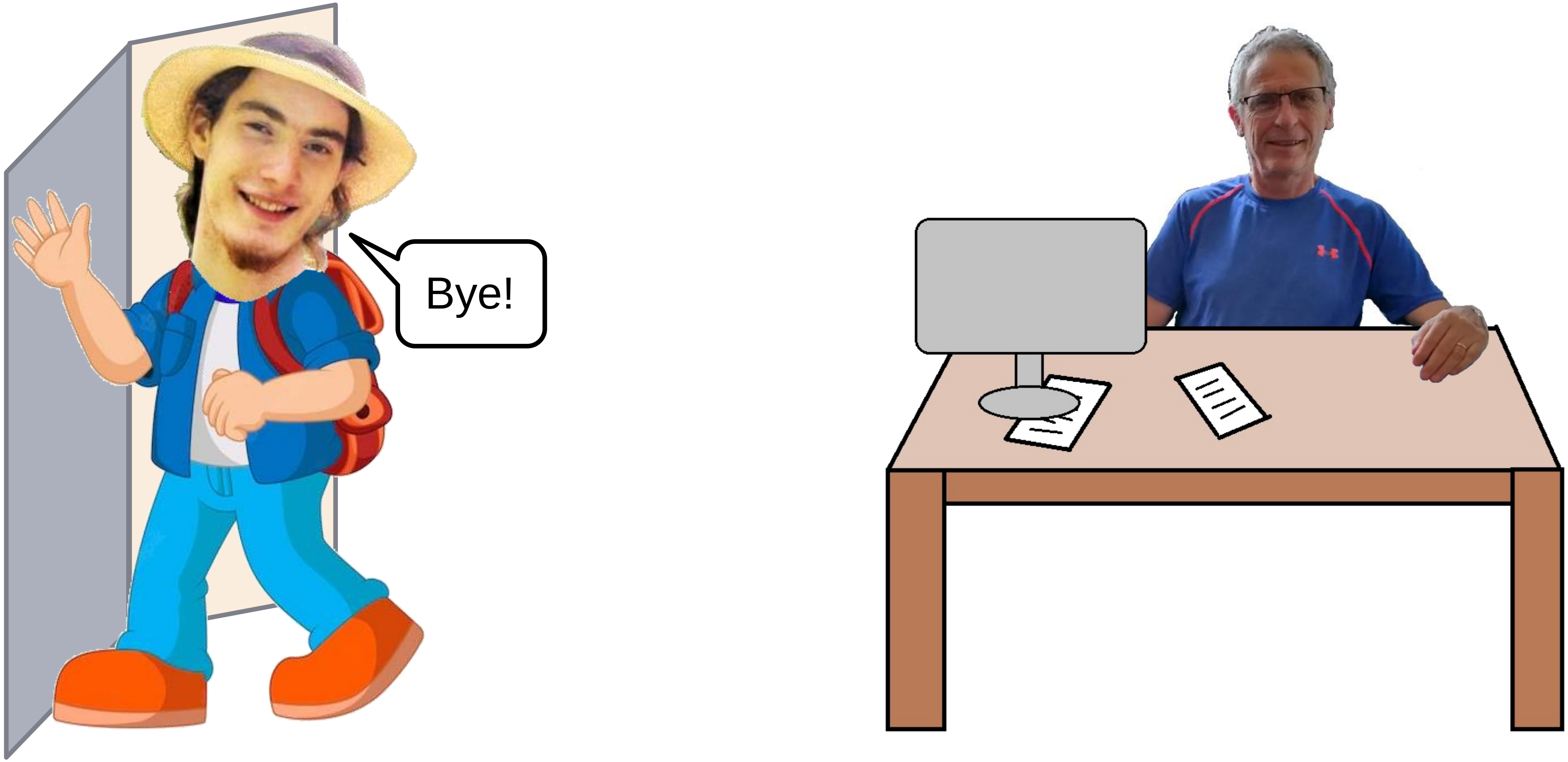


How I met Radu

March 2009, Radu's office, Inria Grenoble

How I met Radu

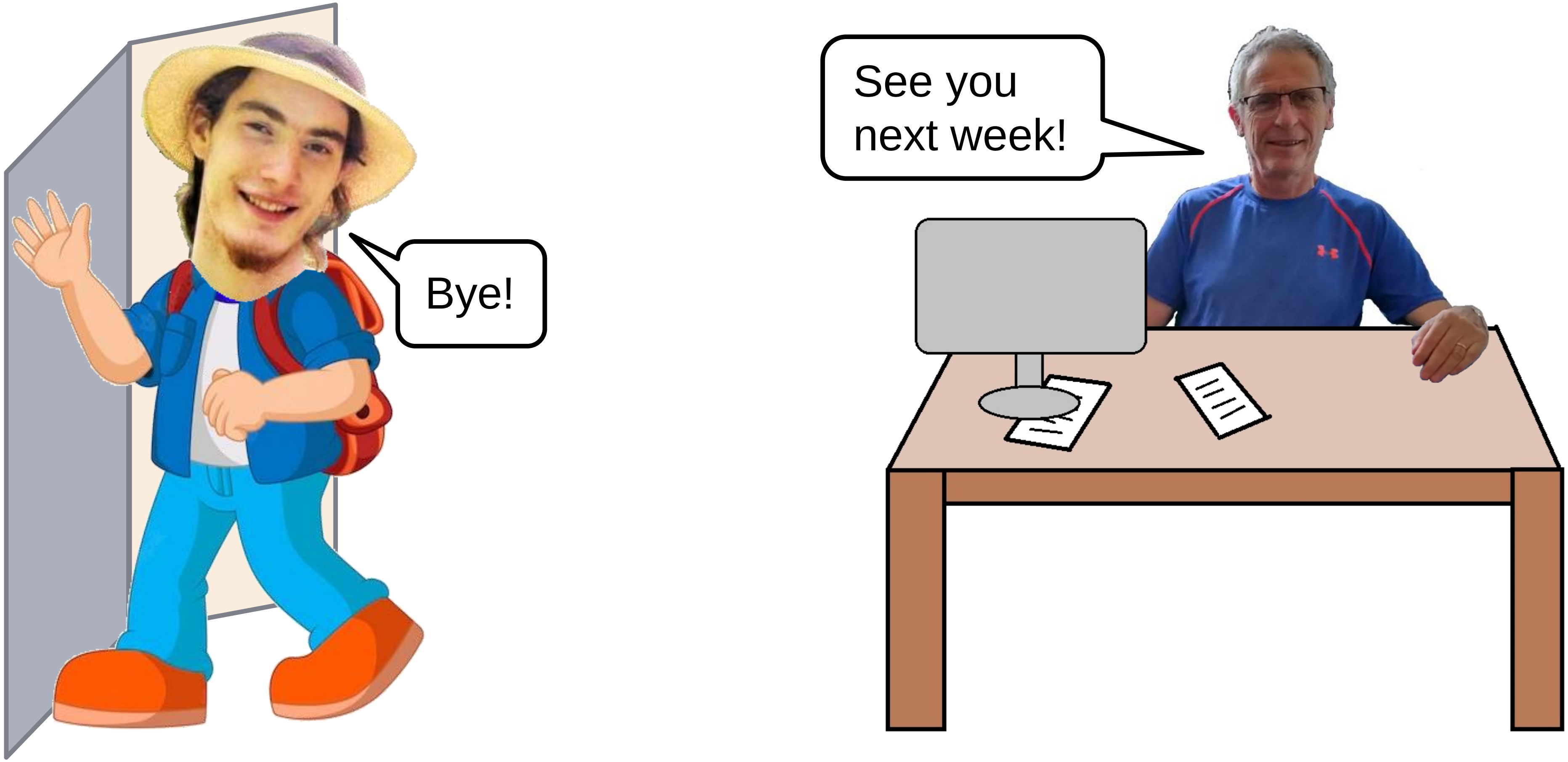
March 2009, Radu's office, Inria Grenoble



How I met Radu

2009

March 2009, Radu's office, Inria Grenoble



Master Internships

2009–2010

Master Internships

2009–2010

ARTICLE **Communicated by J. Kevin O'Regan**

A Sensorimotor Approach to Sound Localization

Murat Aytekin
aytekin@umd.edu
Neuroscience and Cognitive Science Program, University of Maryland, College Park, MD 20742, U.S.A.

Cynthia F. Moss
cmoss@psyc.umd.edu
Neuroscience and Cognitive Science Program, Department of Psychology and Institute of Systems Research, University of Maryland, College Park, MD 20742, U.S.A.

Jonathan Z. Simon
jzsimon@umd.edu
Neuroscience and Cognitive Science Program, Department of Electrical and Computer Engineering, Department of Biology, University of Maryland, College Park, MD 20742, U.S.A.

Master Internships

2009–2010

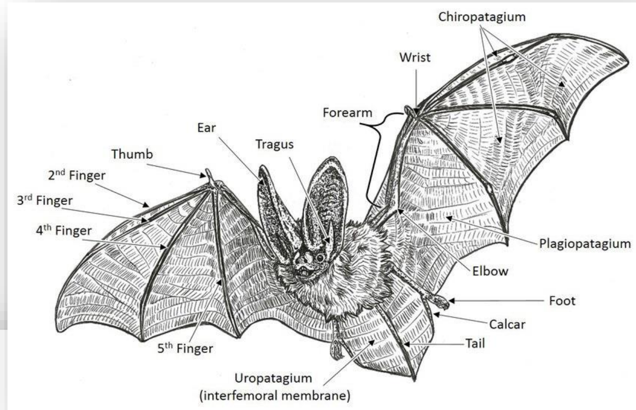
ARTICLE — Communicated by J. Kevin O'Regan

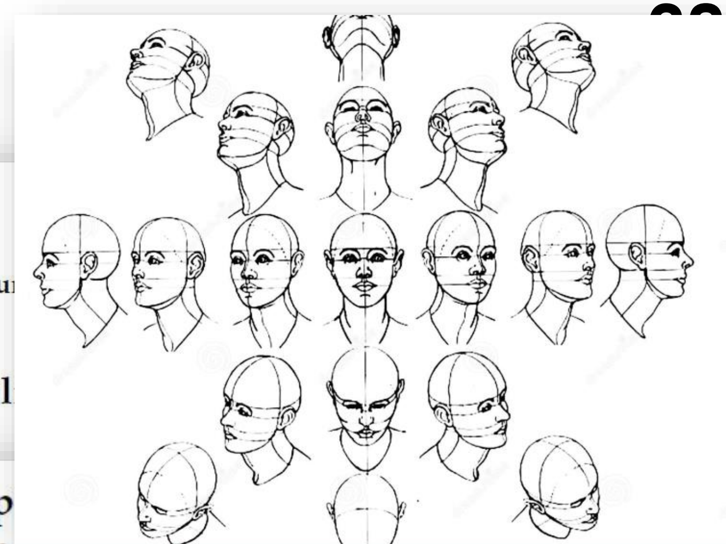
A Sensorimotor Approach to Sound Localization

Sound localization is known to be a complex phenomenon, combining multisensory information processing, experience-dependent plasticity, and movement. Here we present a sensorimotor model that addresses the question of how an organism could learn to localize sound sources without any a priori neural representation of its head-related transfer function or prior experience with auditory spatial information. We demonstrate quantitatively that the experience of the sensory consequences of its voluntary motor actions allows an organism to learn the spatial location of any sound source. Using examples from humans and echolocating bats, our model shows that a naive organism can learn the auditory space based solely on acoustic inputs and their relation to motor states.

Master Internships

09–2010





Commu

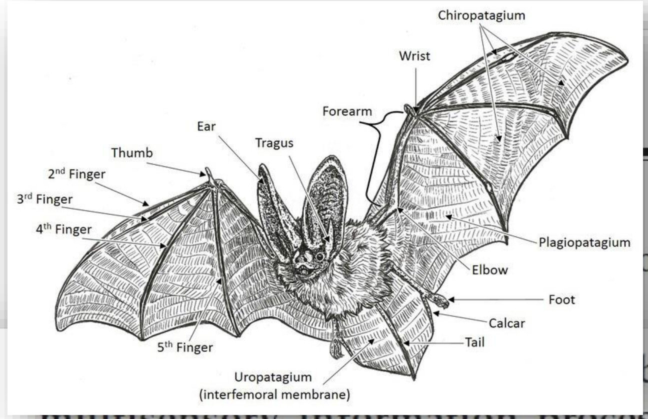
ound Local

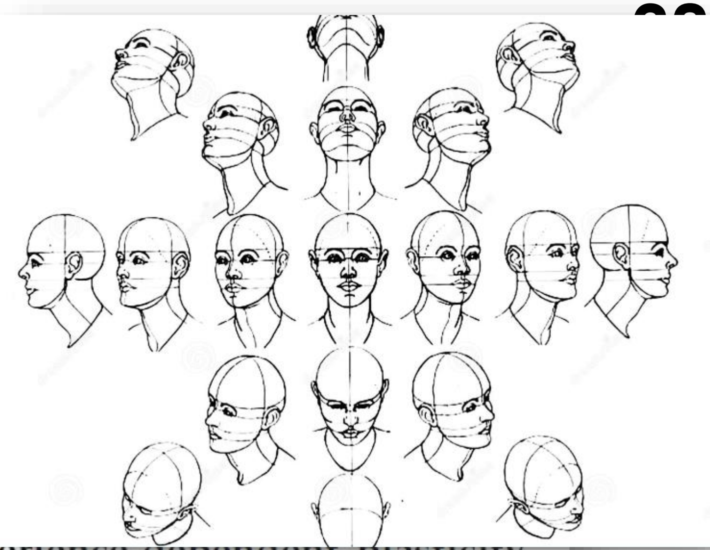
be a comp

multisensory information processing, experience-dependent plasticity, and movement. Here we present a sensorimotor model that addresses the question of how an organism could learn to localize sound sources without any a priori neural representation of its head-related transfer function or prior experience with auditory spatial information. We demonstrate quantitatively that the experience of the sensory consequences of its voluntary motor actions allows an organism to learn the spatial location of any sound source. Using examples from humans and echolocating bats, our model shows that a naive organism can learn the auditory space based solely on acoustic inputs and their relation to motor states.

Master Internships

2009–2010

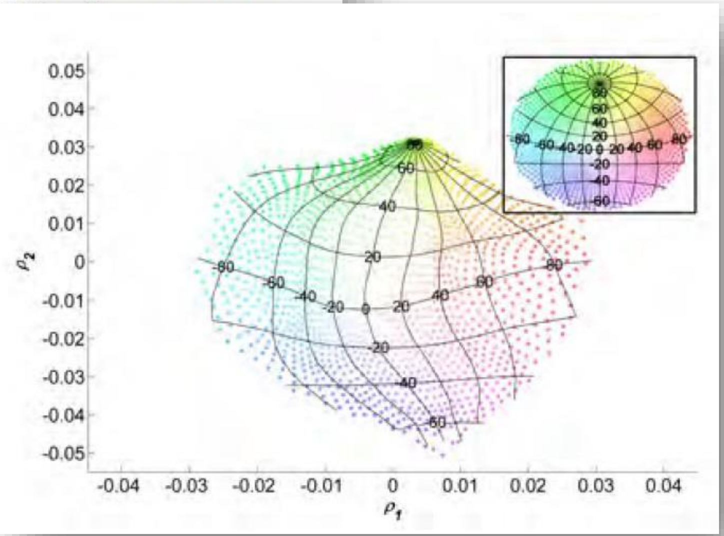




Commu

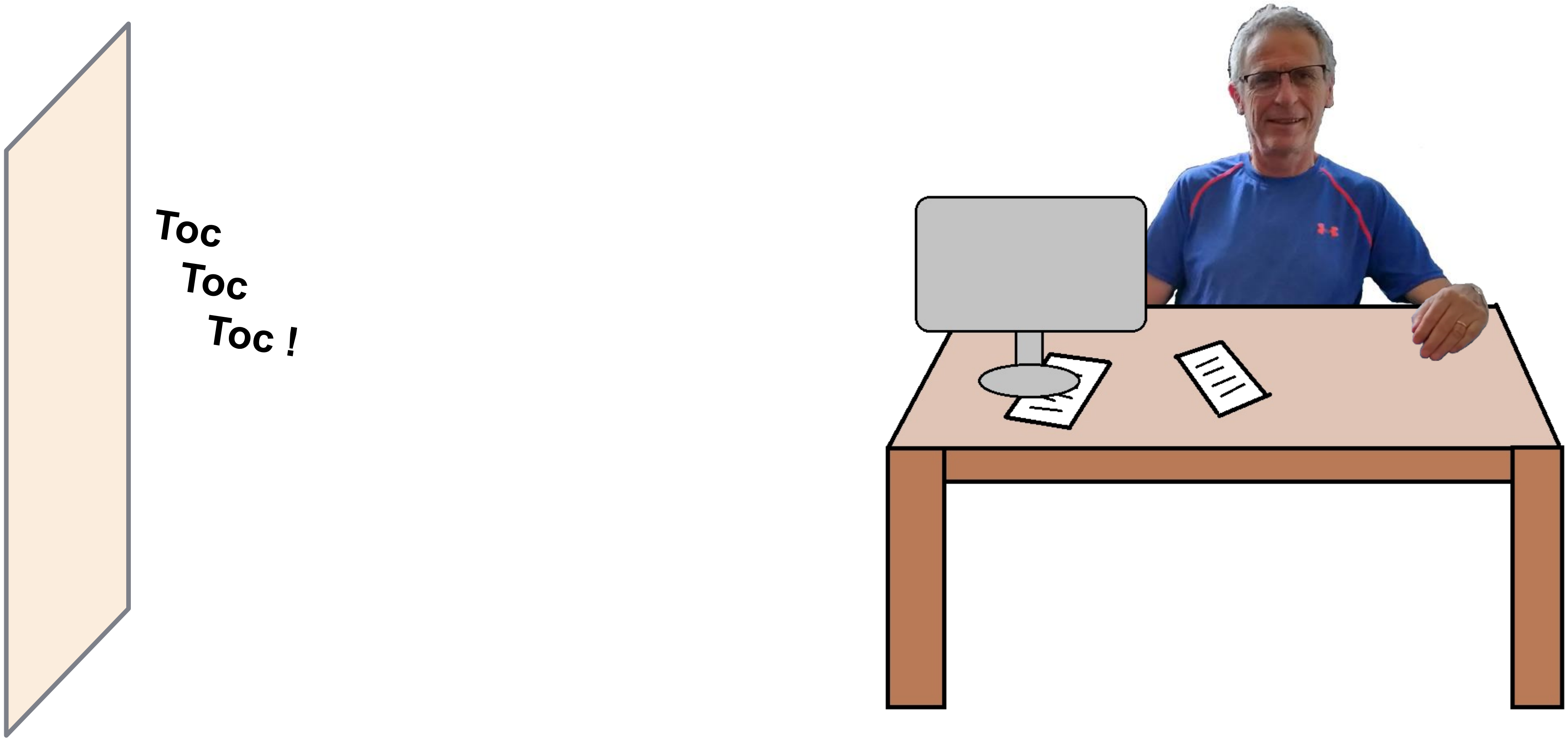
und Local

be a comp
multisensory information processing, experience-dependent plasticity, and movement. Here we present a sensorimotor model that addresses the question of how an organism could learn to localize sound sources without any a priori neural representation of its head-related transfer function or prior experience with auditory spatial information.
quantitatively that the experience of the sensory conseq
untary motor actions allows an organism to learn the sp
any sound source. Using examples from humans and e
our model shows that a naive organism can learn the audi
solely on acoustic inputs and their relation to motor stat



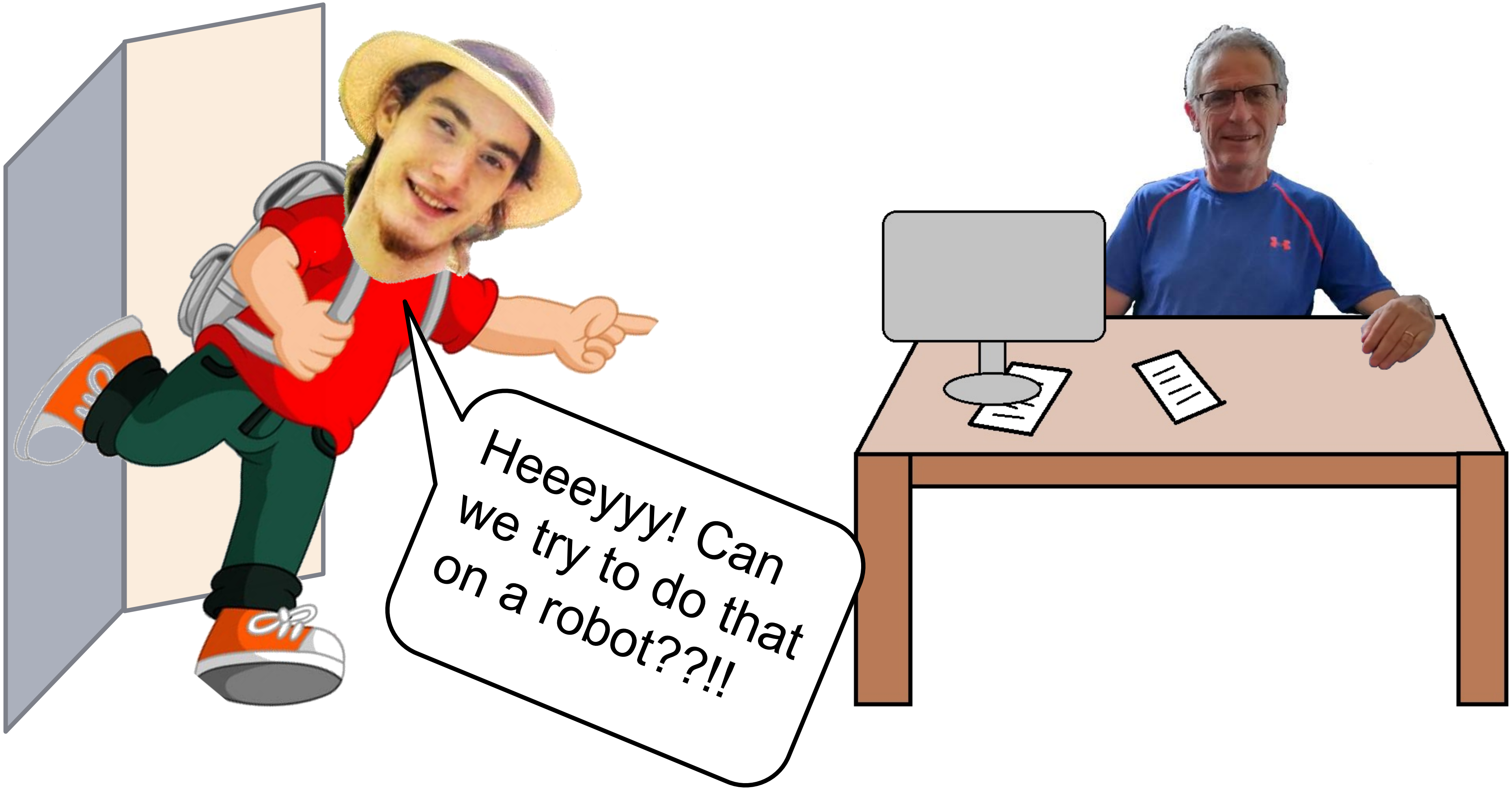
Master Internships

2009–2010



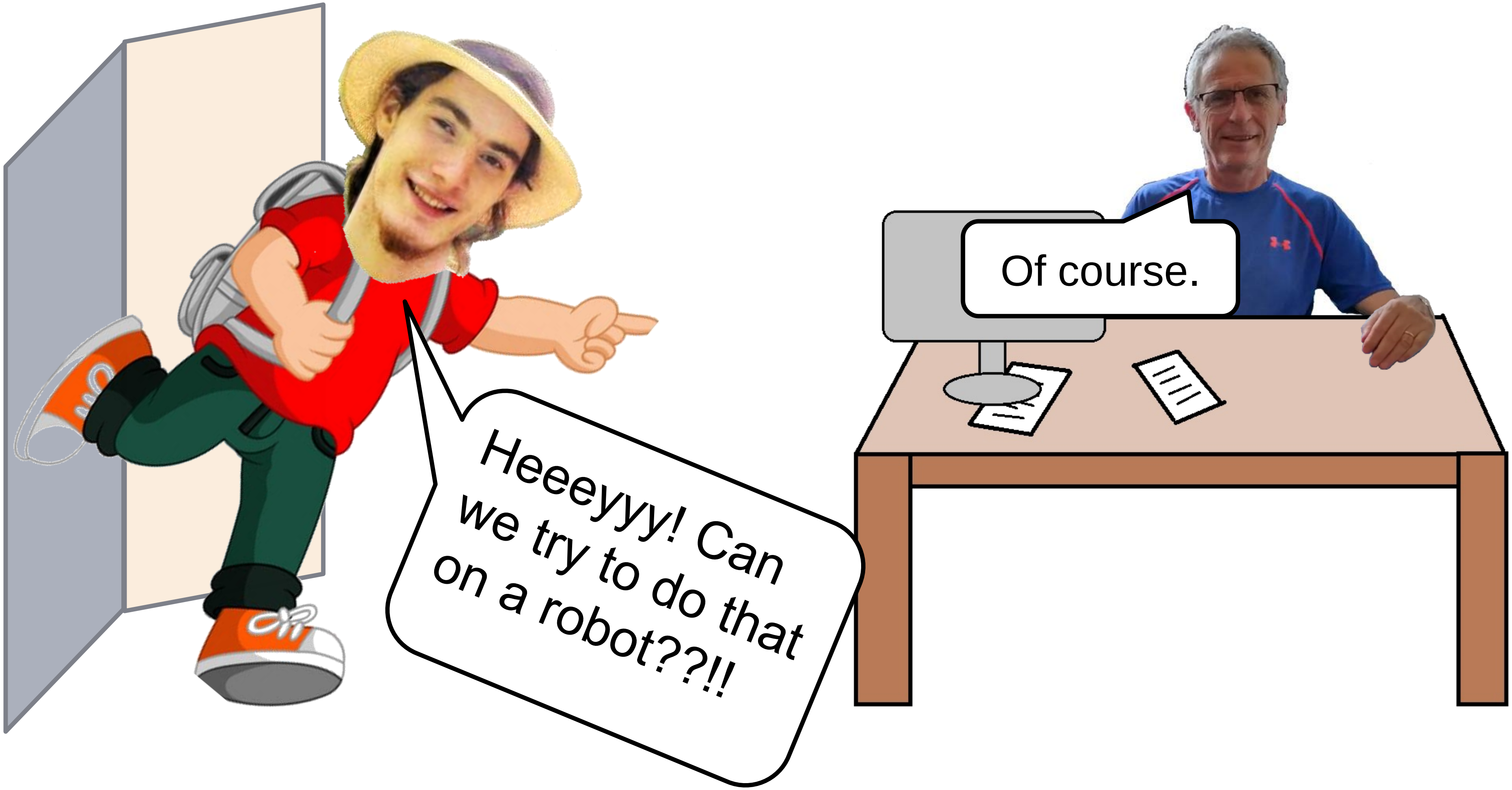
Master Internships

2009–2010



Master Internships

2009–2010



Master Internships

2009–2010

Master Internships

Master Internships

2009–2010

The POPEYE Robot

PhD thesis: Acoustic Space Sampling

2010–2013

- Use a **white noise emitter** to obtain interaural cues in all frequencies
- Two methods: the **listener** is moved or the **emitter** is moved

Audio-motor sampling

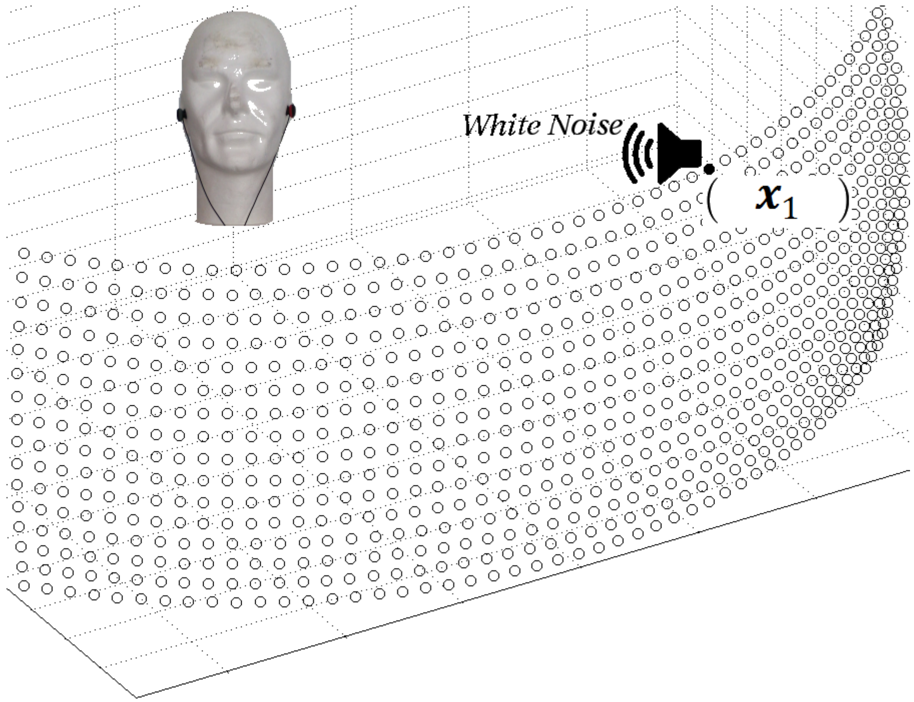
- 10,800 motor states (2 angles)
- 360° azimuth, 120° elevation

Audio-visual sampling

- 432 image positions (in pixels)
- Covering the camera field of view

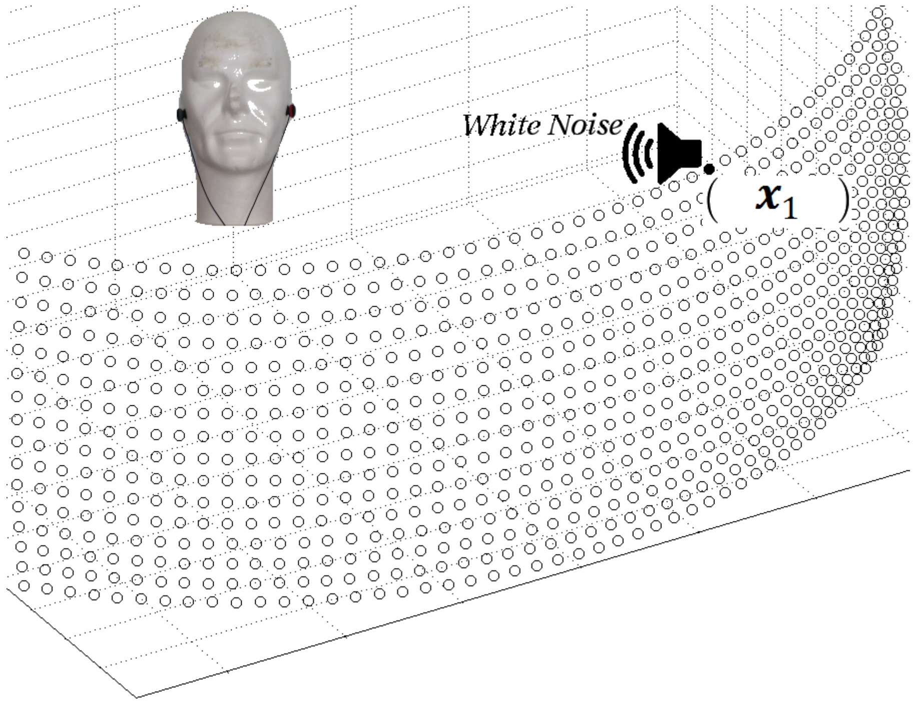
PhD thesis: Acoustic Space Sampling

2010–2013

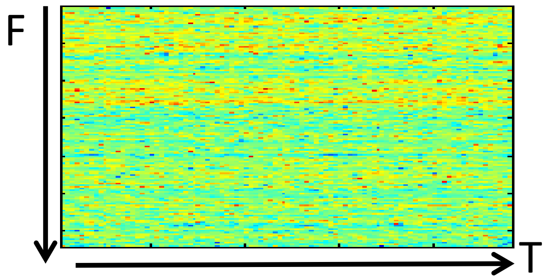


PhD thesis: Acoustic Space Sampling

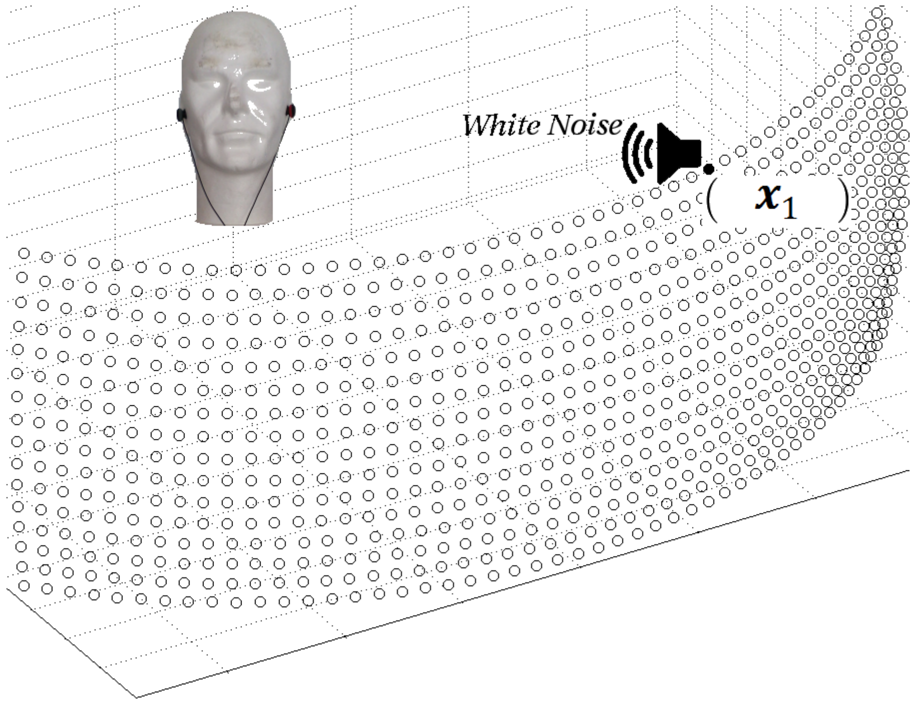
2010–2013



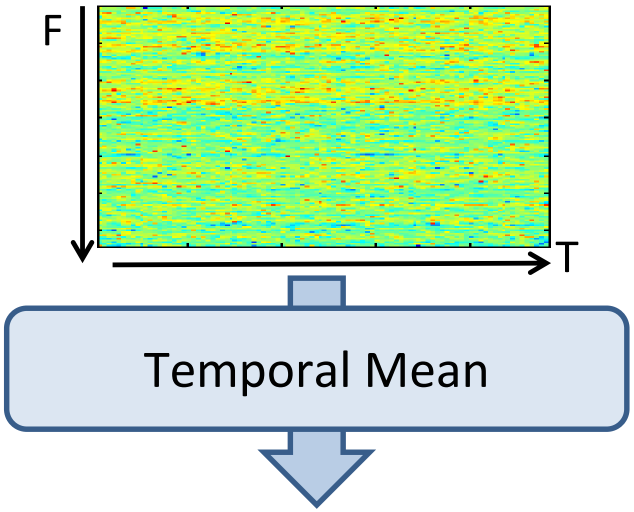
- Interaural Level Difference Spectrogram



PhD thesis: Acoustic Space Sampling

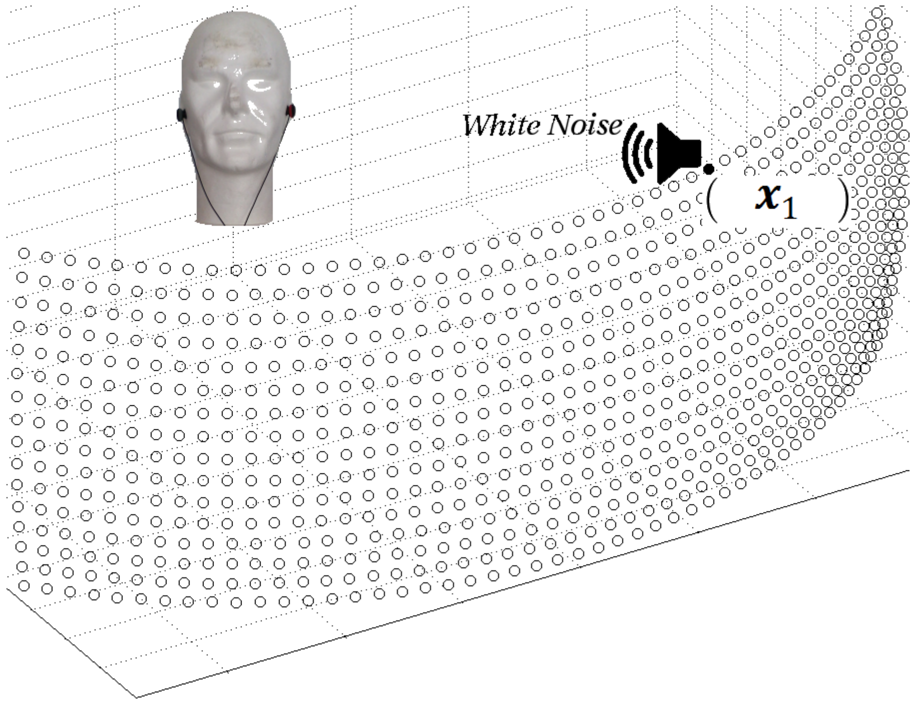


- Interaural Level Difference Spectrogram

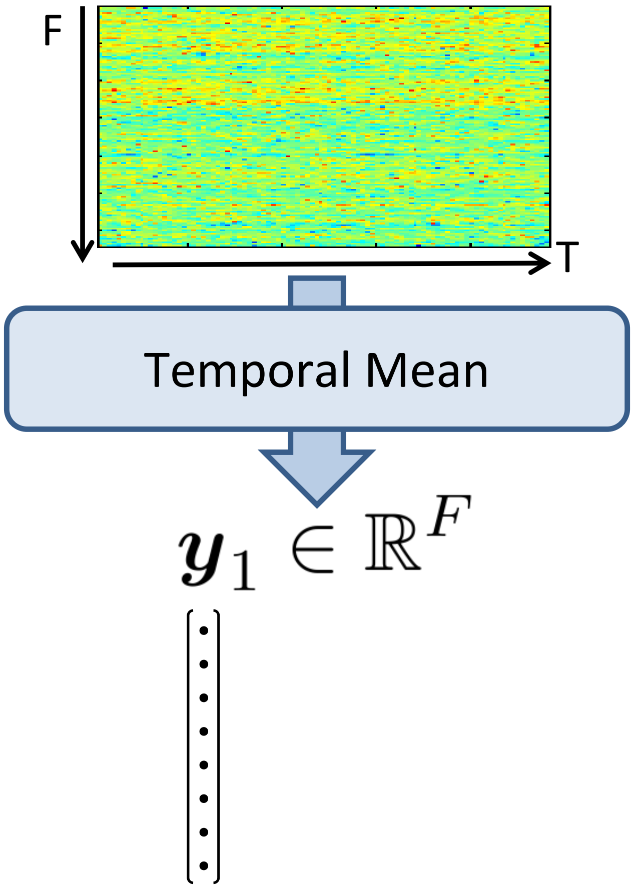


PhD thesis: Acoustic Space Sampling

2010–2013

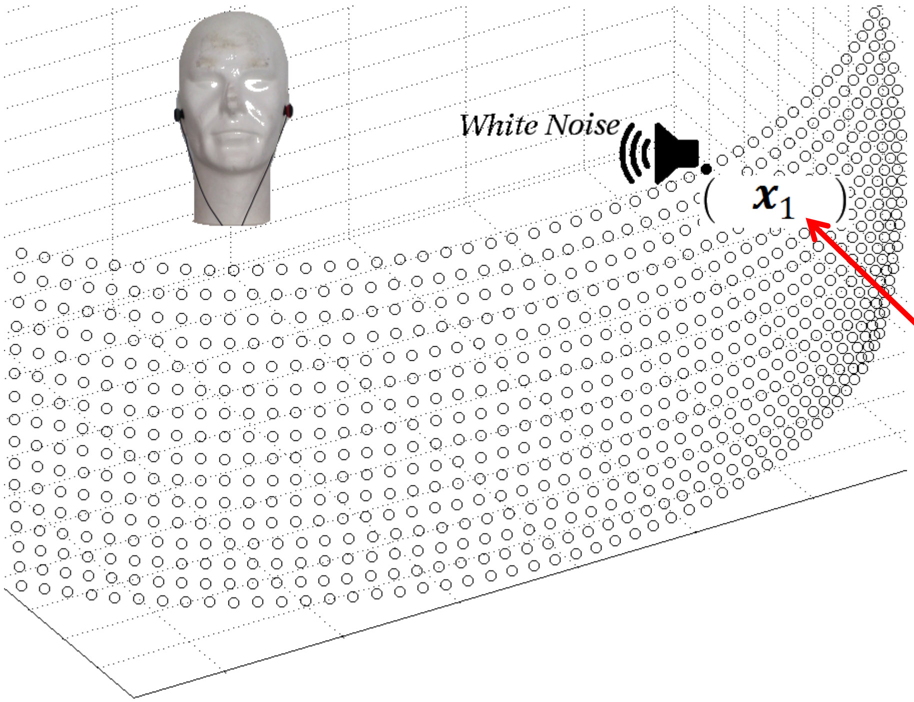


- Interaural Level Difference Spectrogram

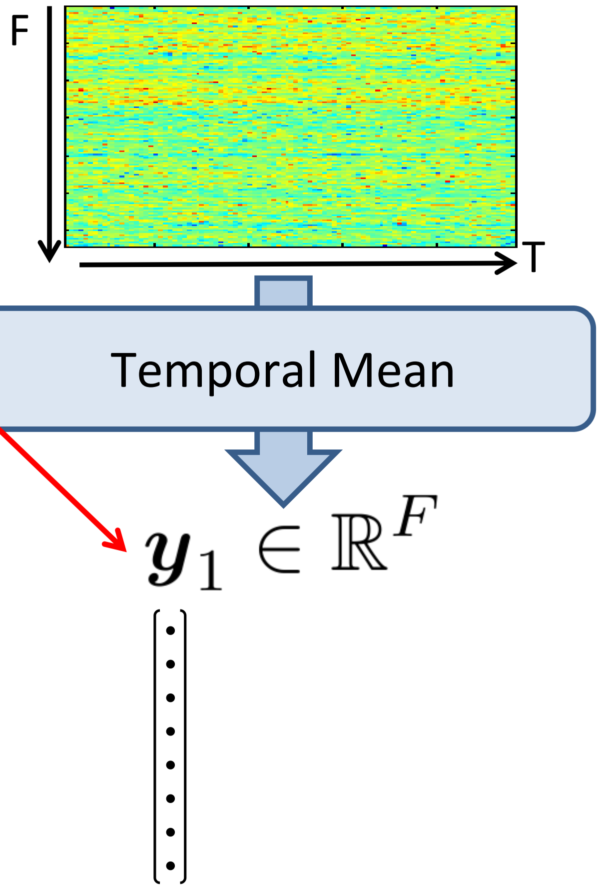


PhD thesis: Acoustic Space Sampling

2010–2013

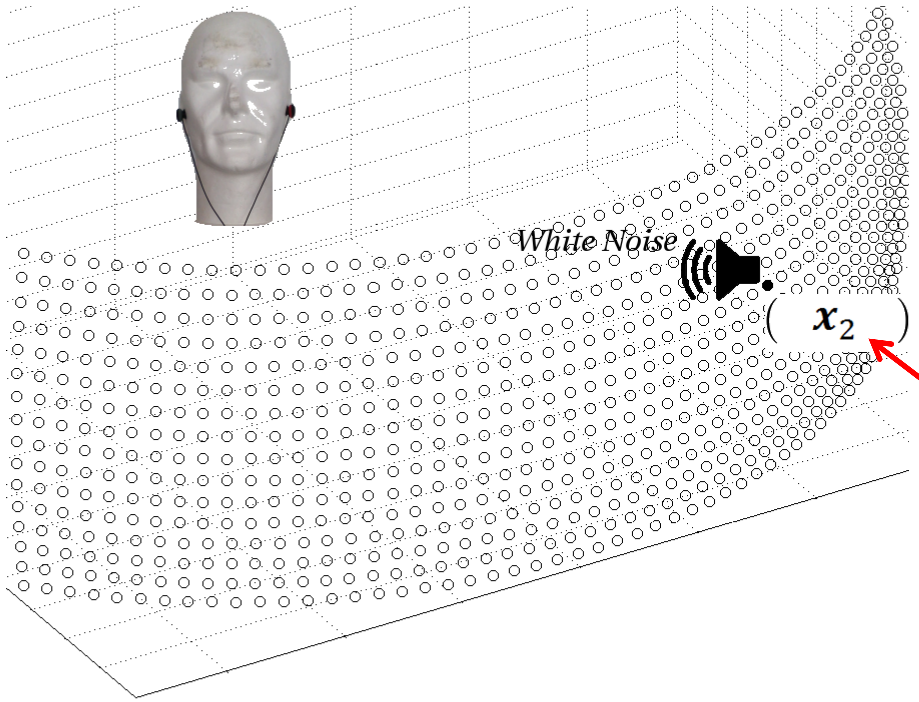


- Interaural Level Difference Spectrogram

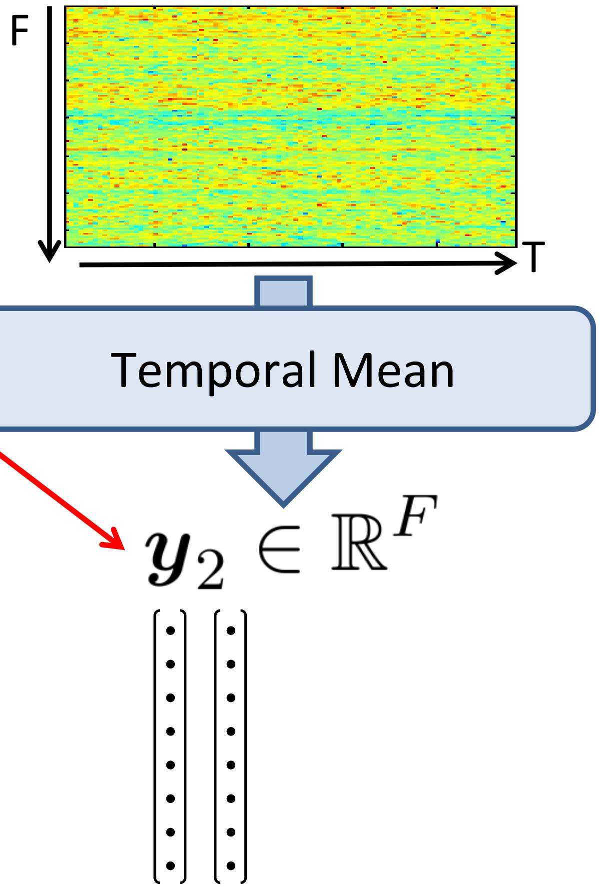


PhD thesis: Acoustic Space Sampling

2010–2013

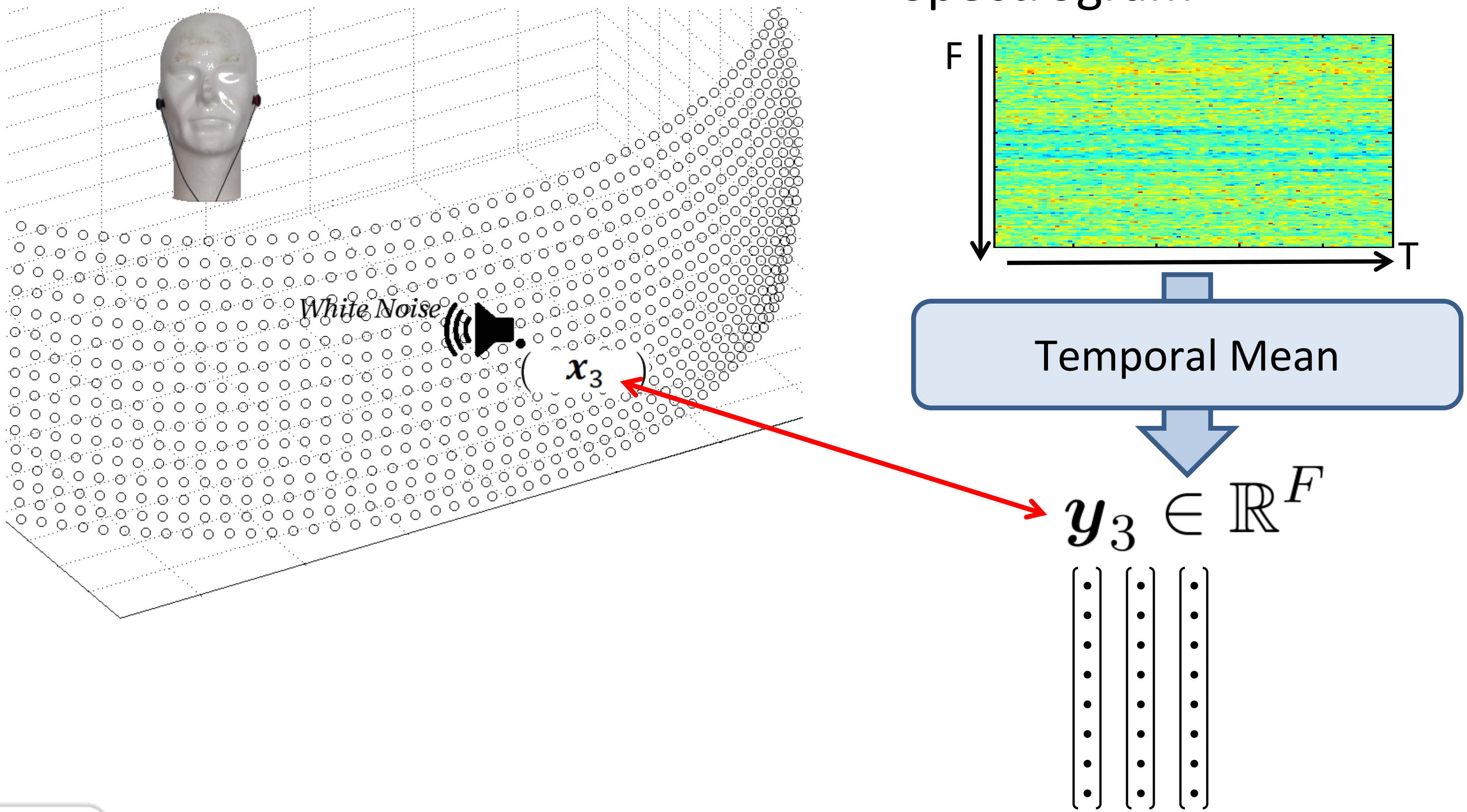


- Interaural Level Difference Spectrogram



PhD thesis: Acoustic Space Sampling

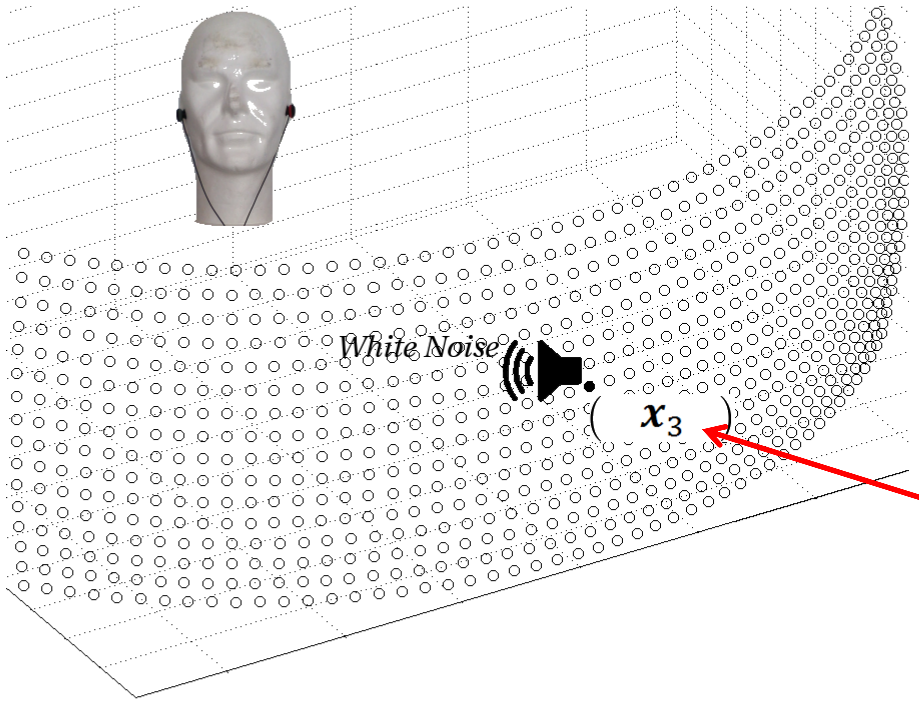
- Interaural Level Difference Spectrogram

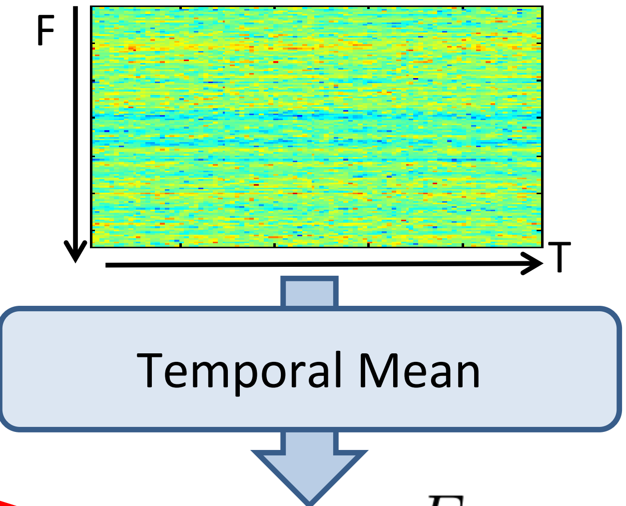


PhD thesis: Acoustic Space Sampling

2010–2013

- Interaural Level Difference Spectrogram





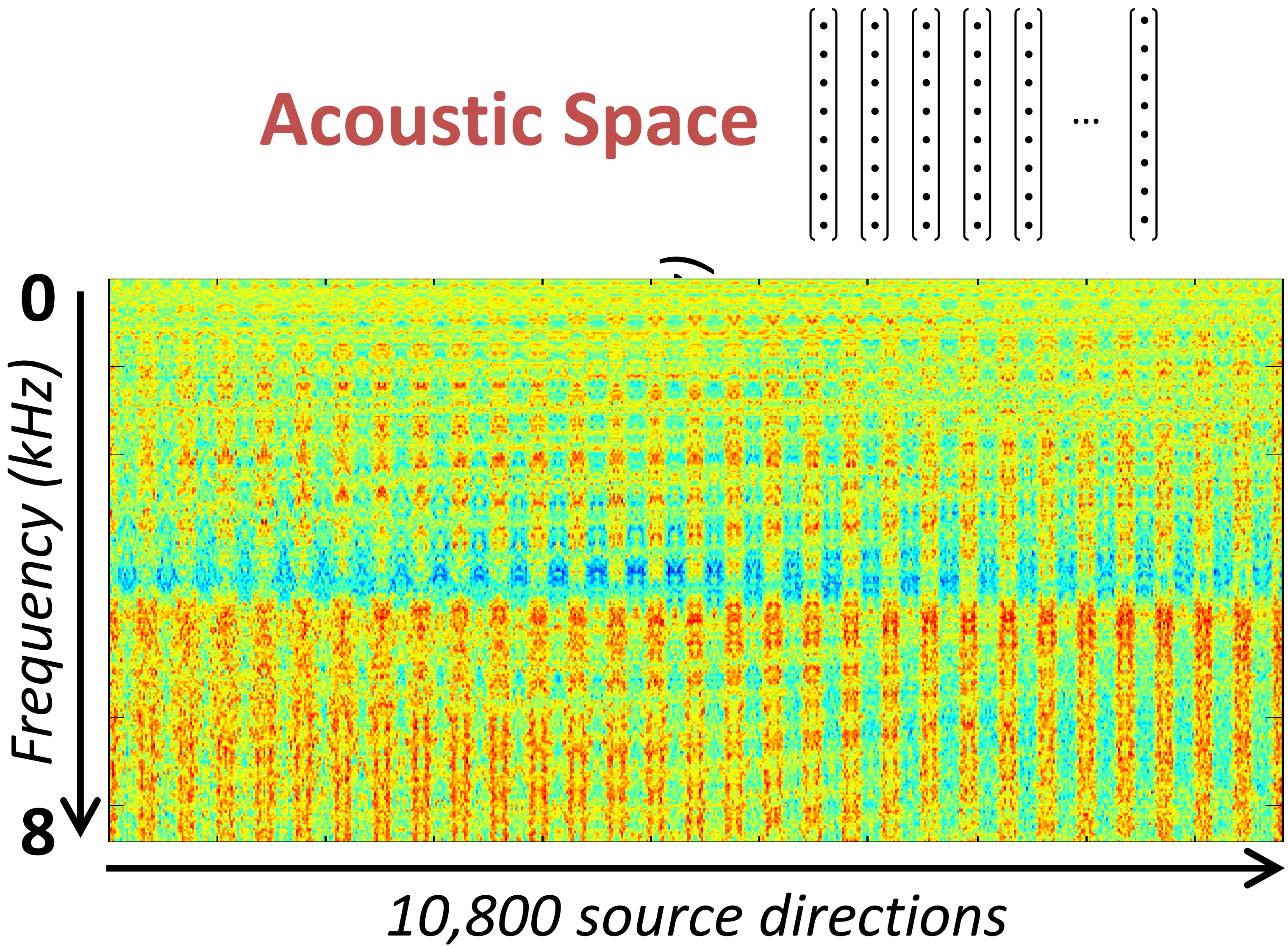
$$\boldsymbol{y}_3 \in \mathbb{R}^F$$

Acoustic Space

Antoine.Deleforge@inria.fr

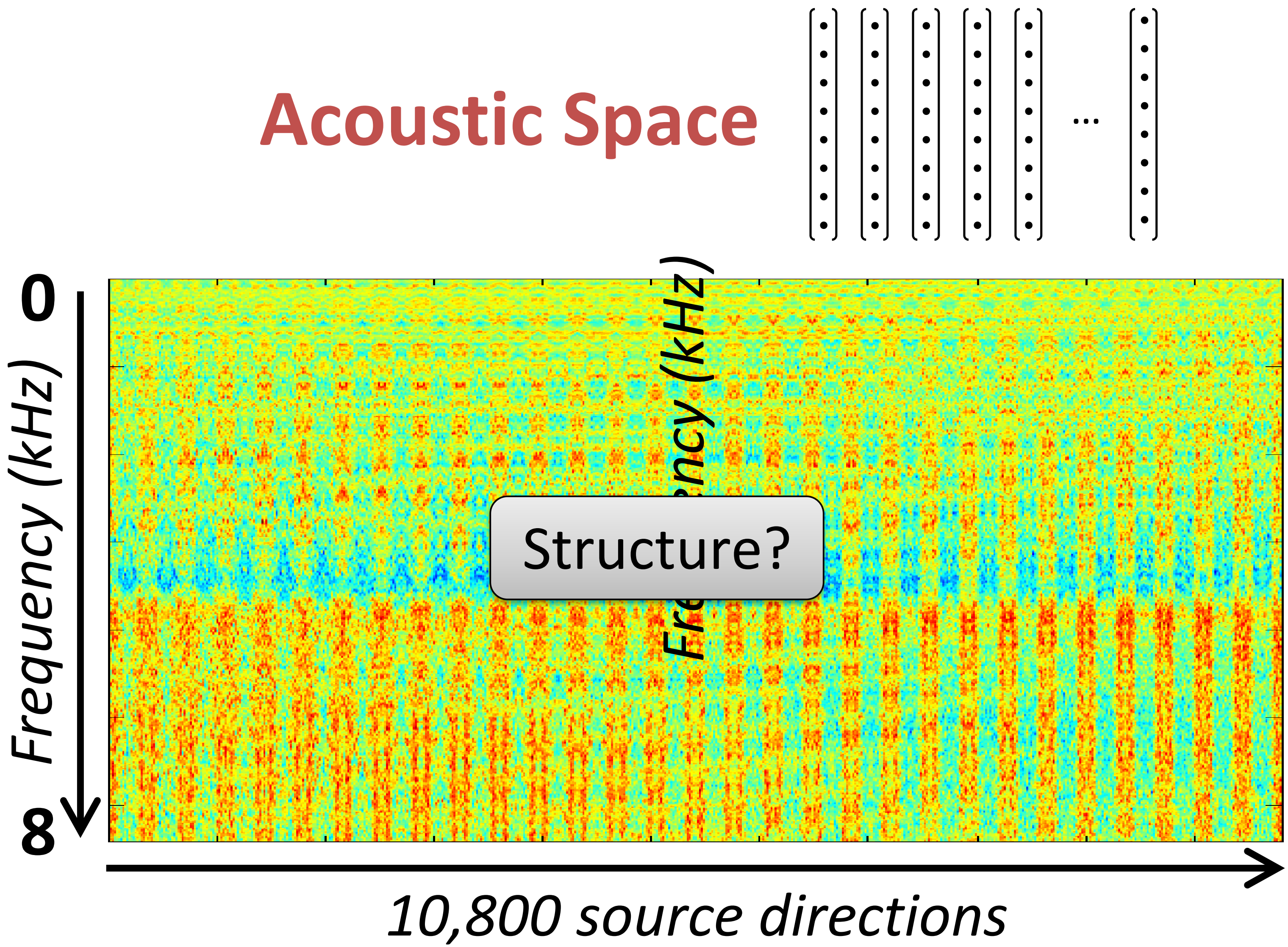
PhD thesis: Acoustic Space Sampling

2010–2013



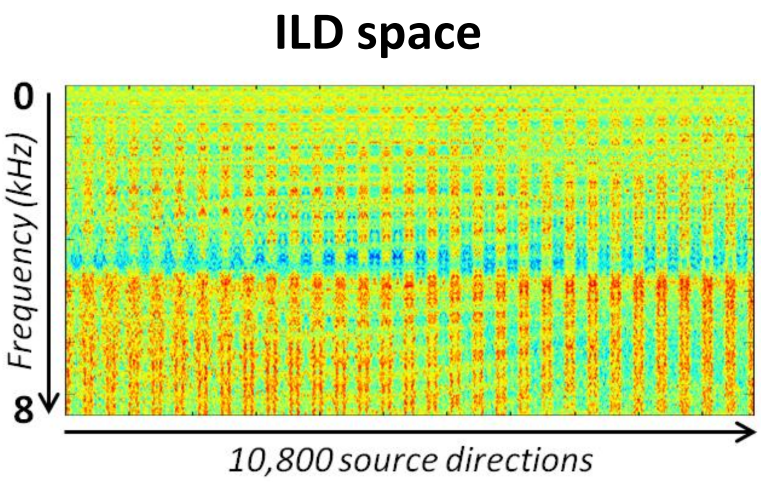
PhD thesis: Acoustic Space Sampling

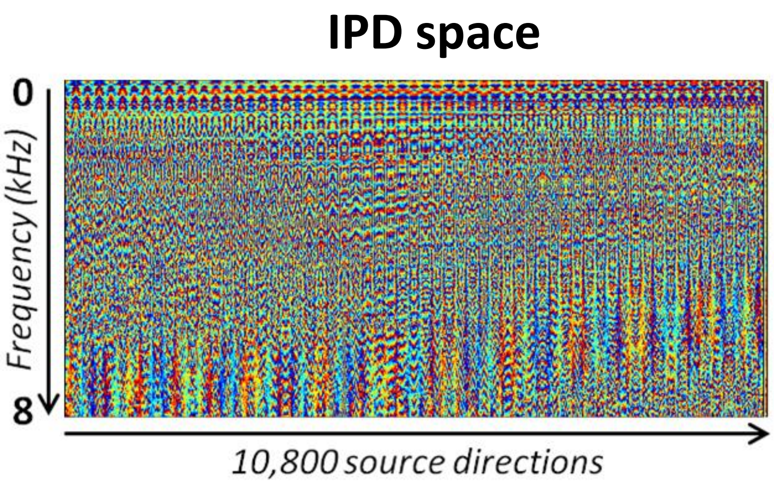
2010–2013



PhD thesis: Acoustic Space Sampling

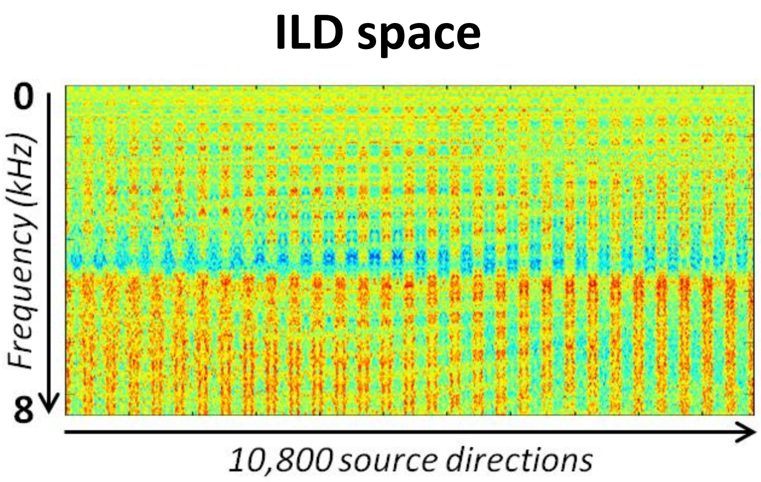
2010–2013

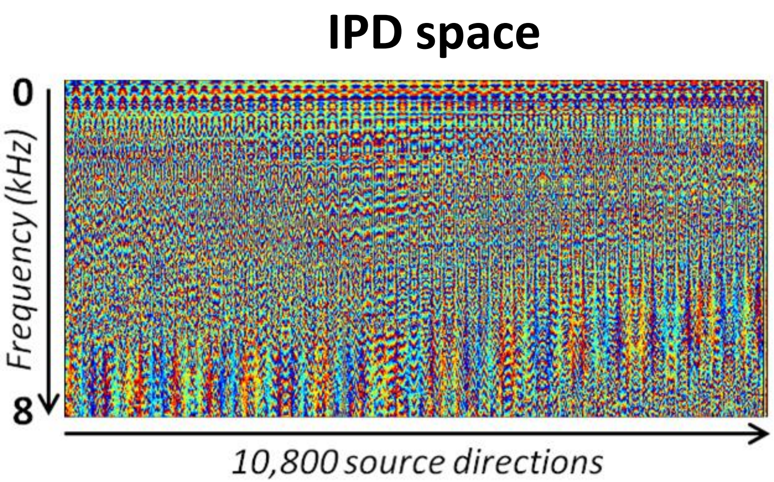




PhD thesis: Acoustic Space Sampling

2010–2013



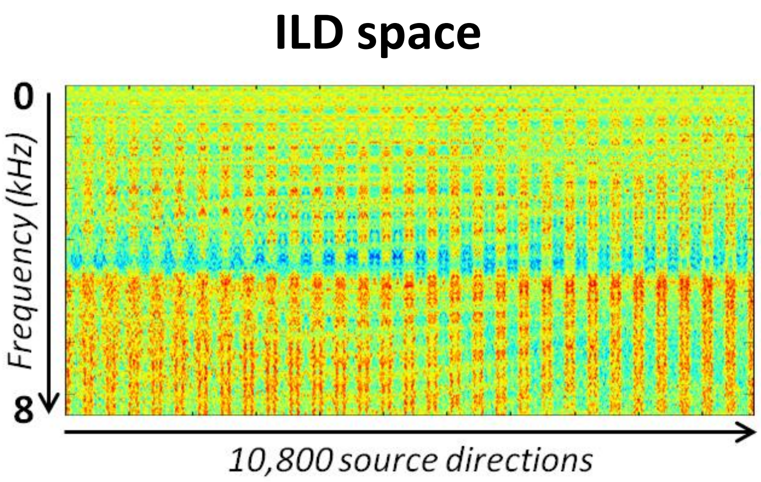


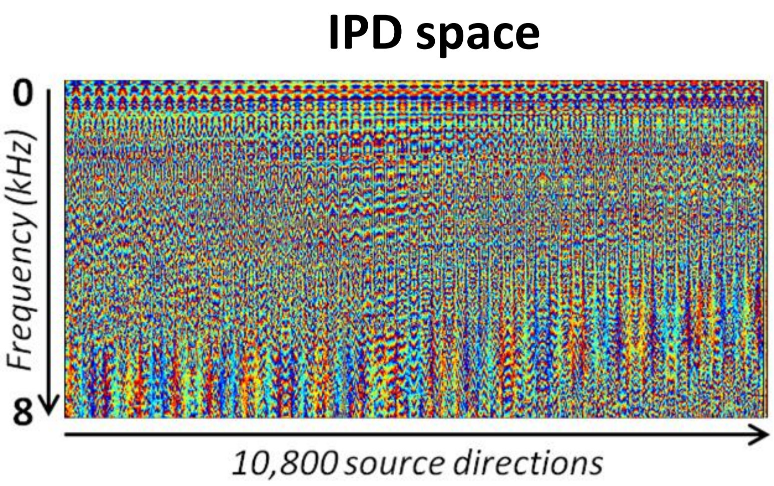
- These high-dimensional representations do not reveal the intrisic **structure** of acoustic spaces

Antoine.Deleforge@inria.fr

PhD thesis: Acoustic Space Sampling

2010–2013

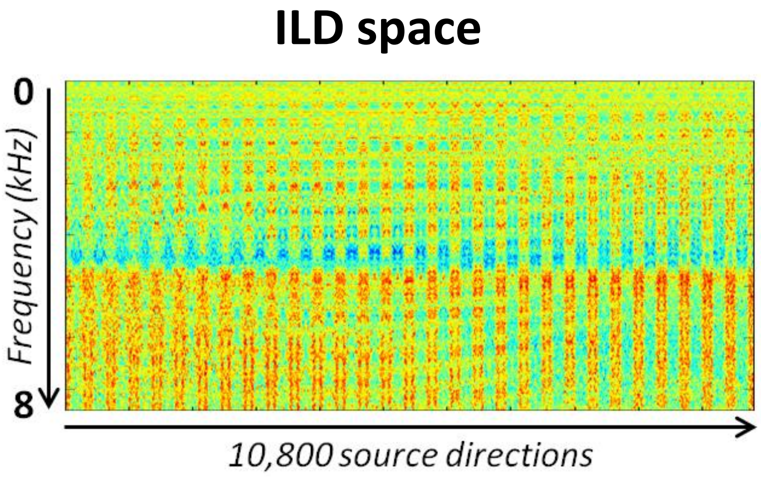


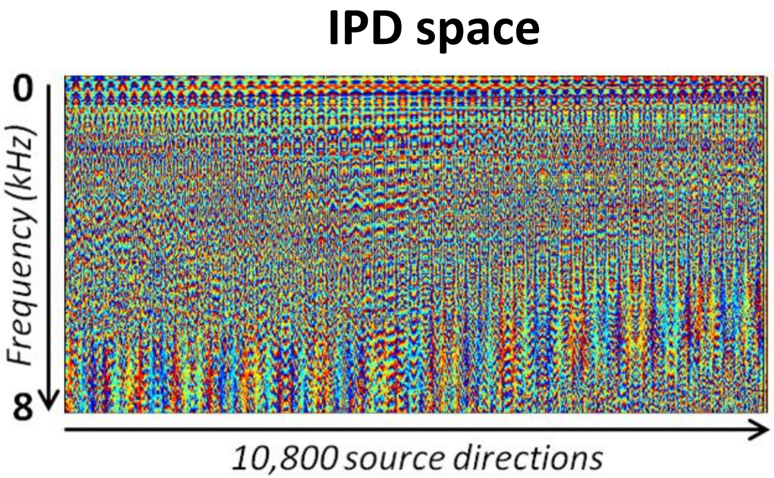


- These high-dimensional representations do not reveal the intrisic **structure** of acoustic spaces
- We seek a **low-dimensional** representation

PhD thesis: Acoustic Space Sampling

2010–2013

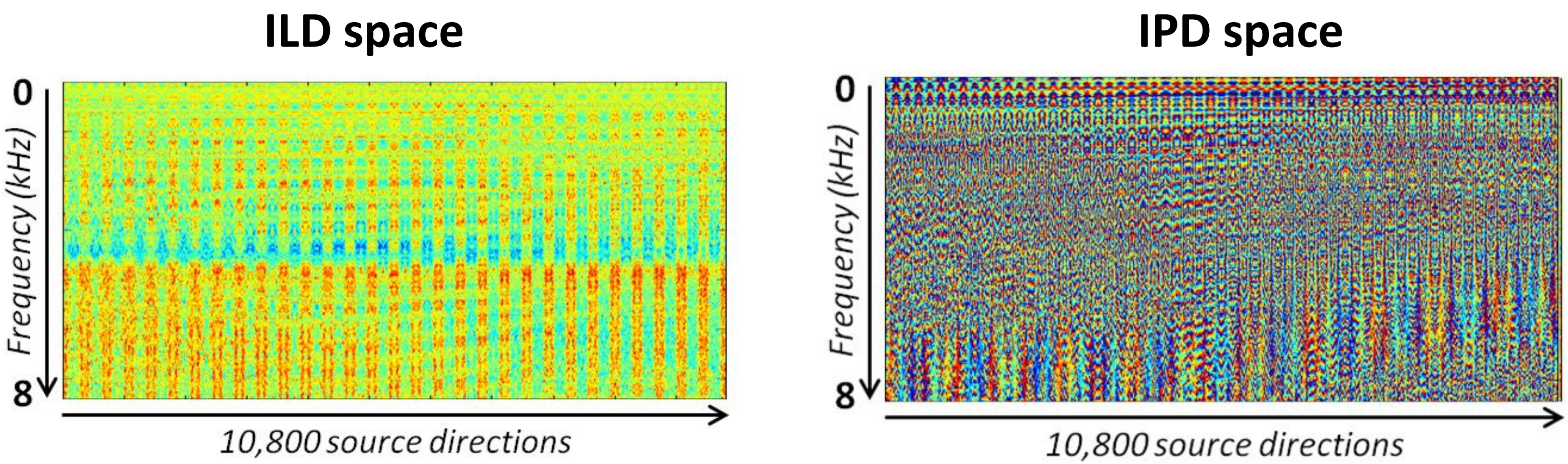




- These high-dimensional representations do not reveal the intrisic **structure** of acoustic spaces
- We seek a **low-dimensional** representation
- Can be obtained with dimensionality reduction techniques = ***unsupervised learning***

PhD thesis: Acoustic Space Sampling

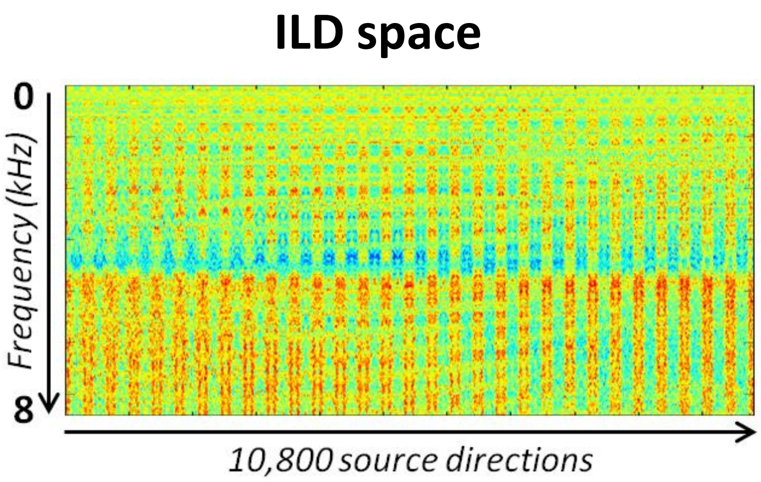
2010–2013

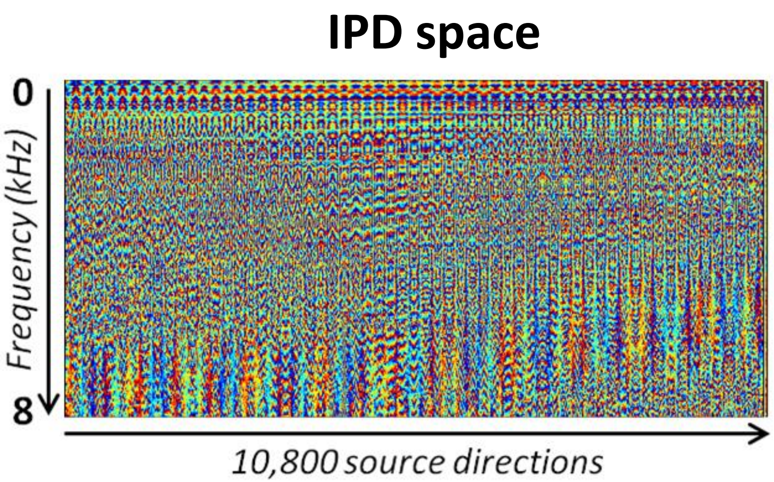


- These high-dimensional representations do not reveal the intrisic **structure** of acoustic spaces
- We seek a **low-dimensional** representation
- Can be obtained with dimensionality reduction techniques = ***unsupervised learning***
- **Map** data onto a lower dimensional space, using high dimensional data only

PhD thesis: Acoustic Space Sampling

2010–2013





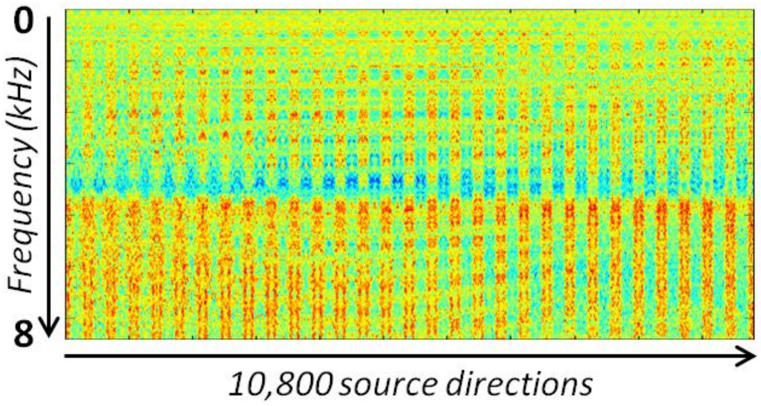
- These high-dimensional representations do not reveal the intrisic **structure** of acoustic spaces
- We seek a **low-dimensional** representation
- Can be obtained with dimensionality reduction techniques = ***unsupervised learning***
- **Map** data onto a lower dimensional space, using high dimensional data only
- **PCA**=linear, **manifold learning**=non-linear

Antoine.Deleforge@inria.fr

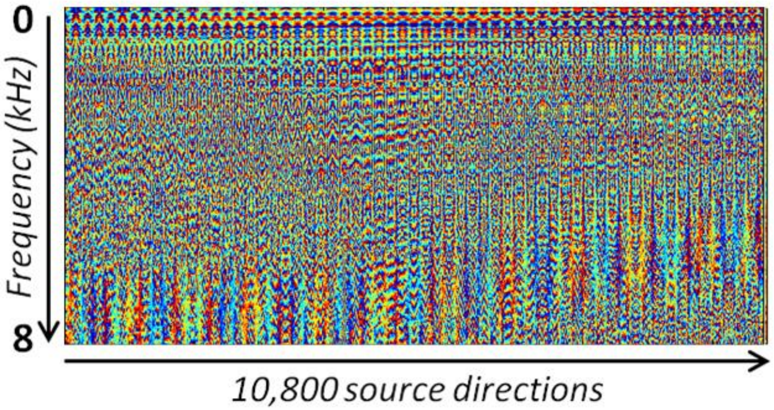
PhD thesis: Acoustic Space Sampling

2010–2013

ILD space



IPD space



Non-Linear dimensionality reduction (LTSA)

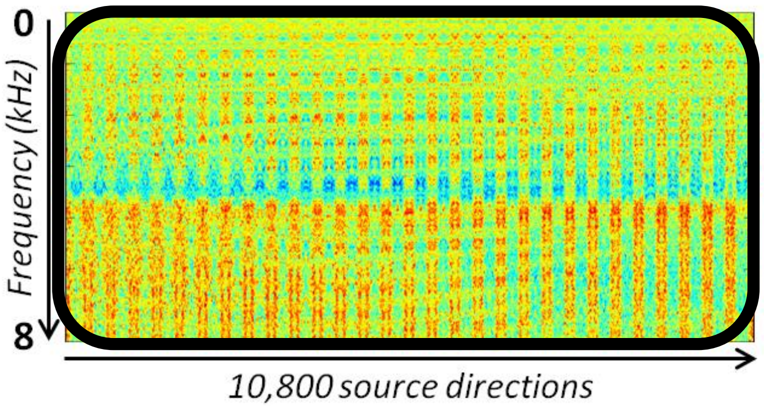
Zhang & Zha (2004)

- Apply PCA locally around each point
- Align globally each representation

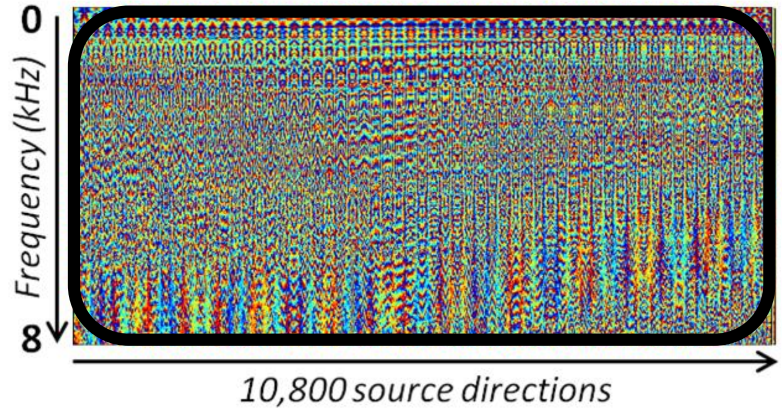
PhD thesis: Acoustic Space Sampling

2010–2013

ILD space



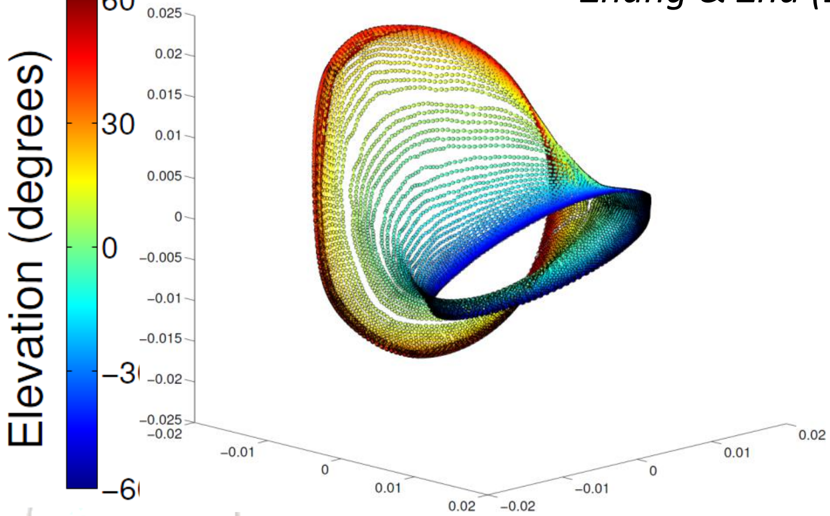
IPD space

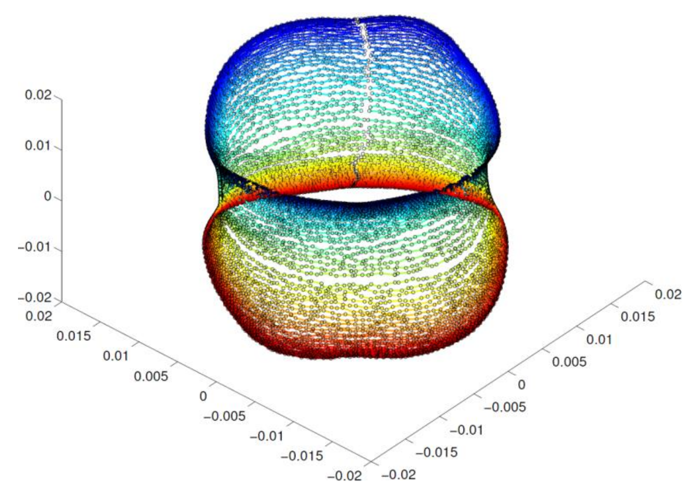


Non-Linear dimensionality reduction (LTSA)

Zhang & Zha (2004)

- Apply PCA locally around each point
- Align globally each representation

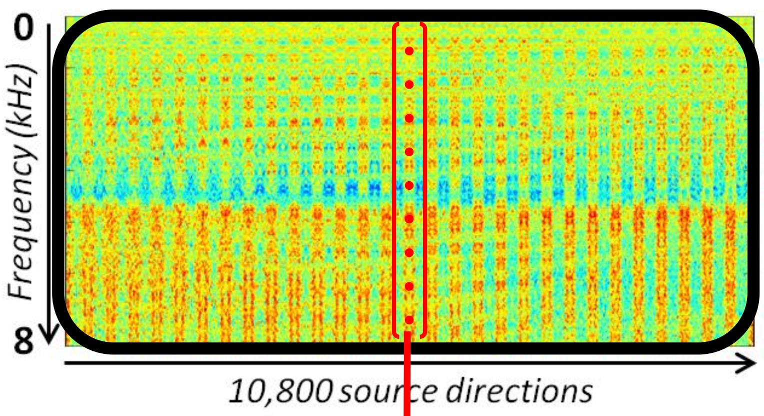




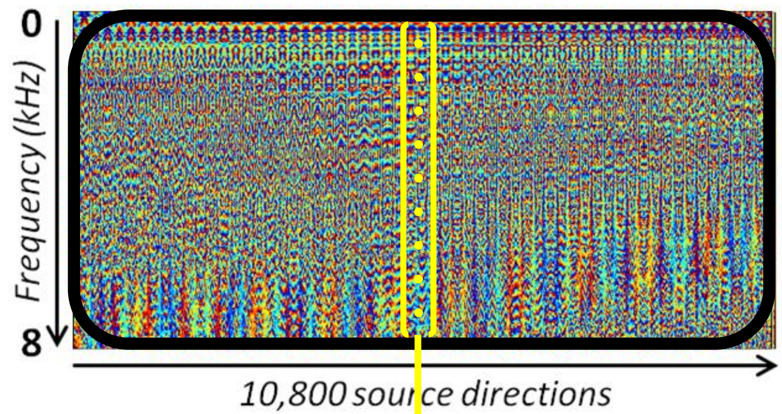
PhD thesis: Acoustic Space Sampling

2010–2013

ILD space



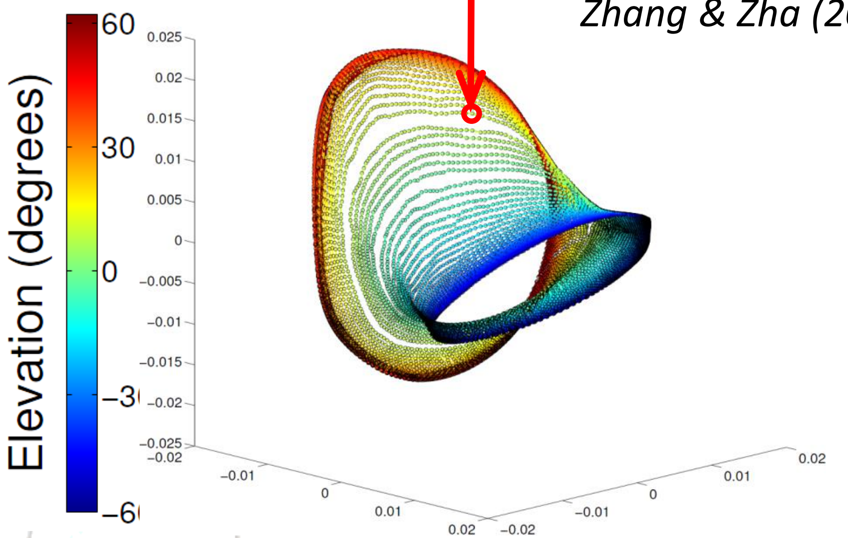
IPD space

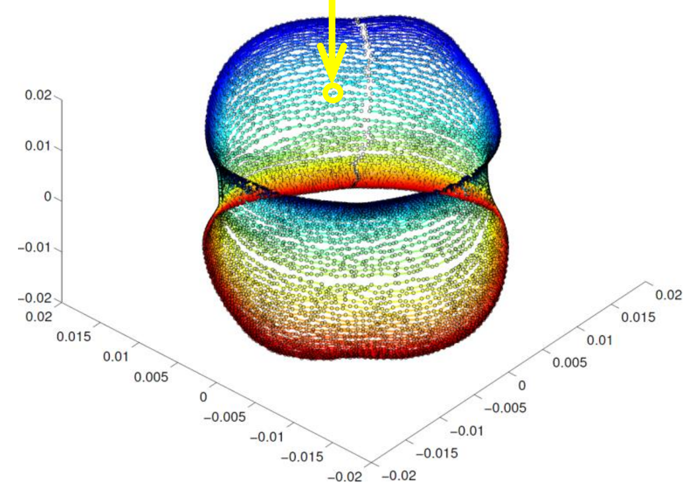


Non-Linear dimensionality reduction (LTSA)

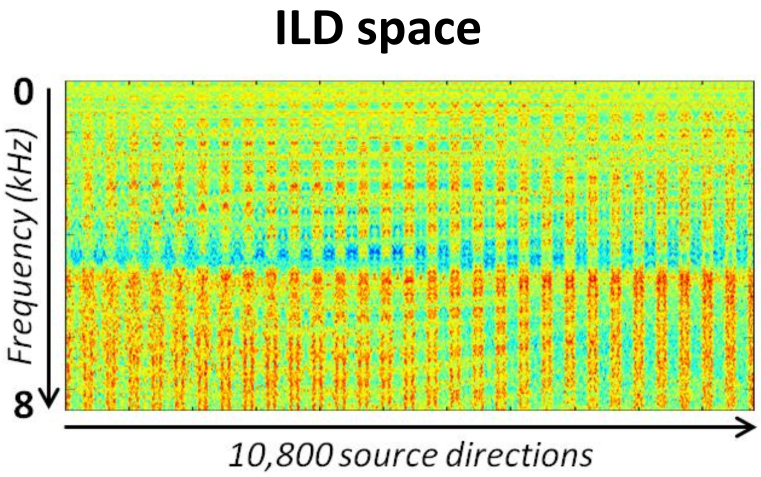
Zhang & Zha (2004)

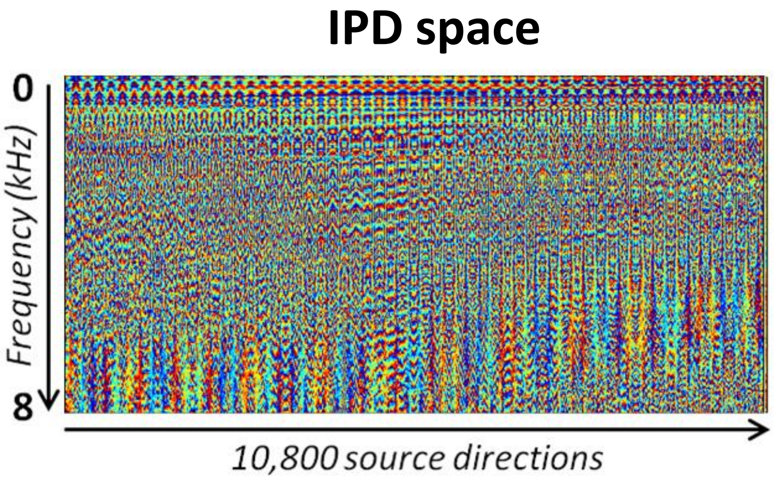
- Apply PCA locally around each point
- Align globally each representation





PhD thesis: Acoustic Space Sampling

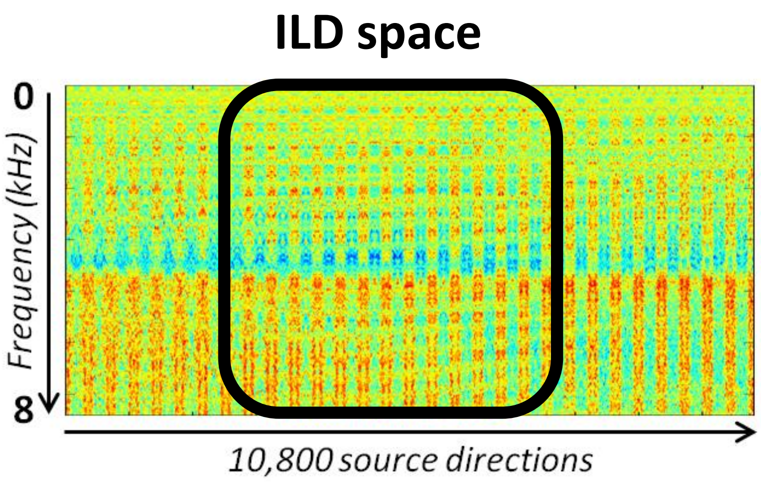


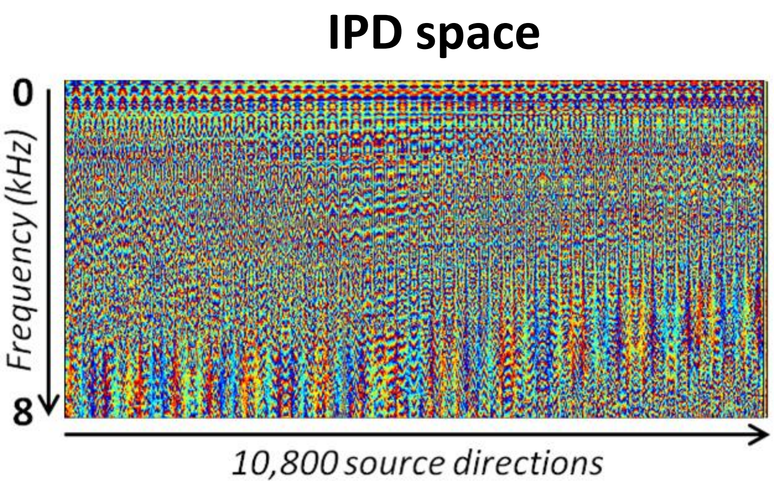


Linear dimensionality reduction (PCA)

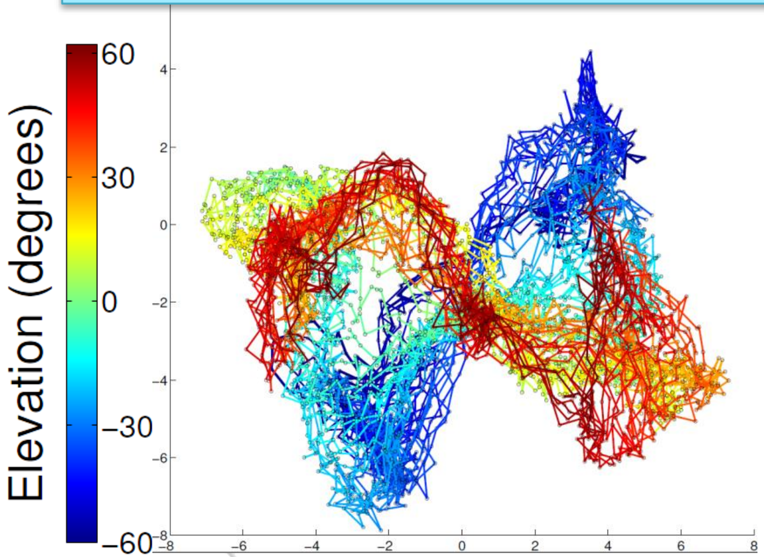
PhD thesis: Acoustic Space Sampling

2010–2013



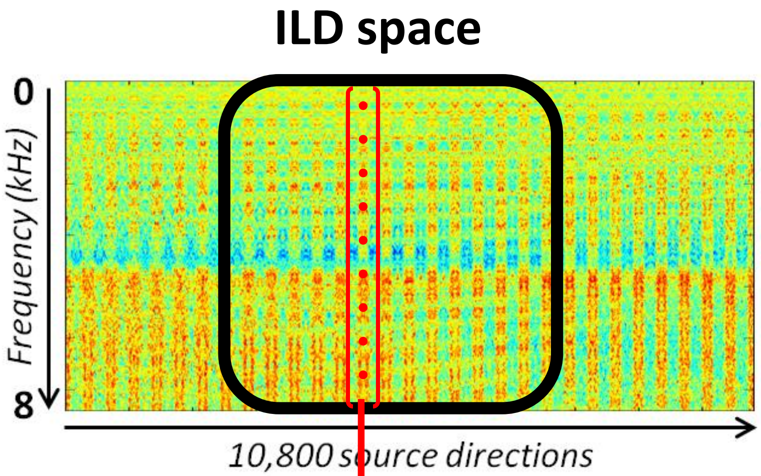


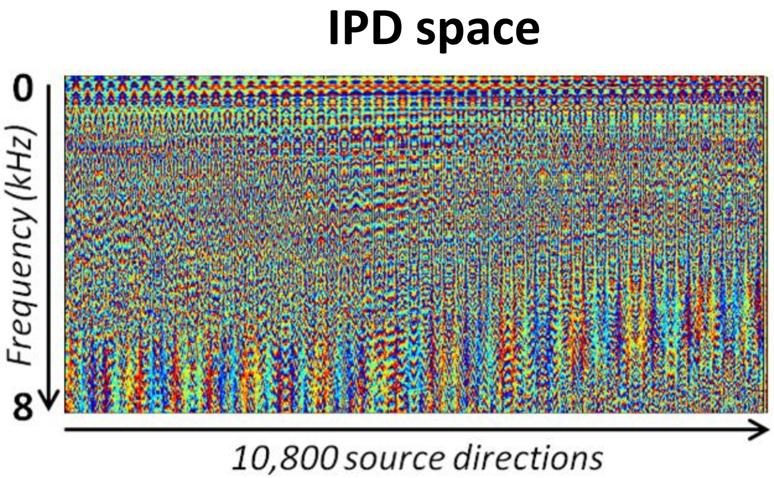
Linear dimensionality reduction (PCA)



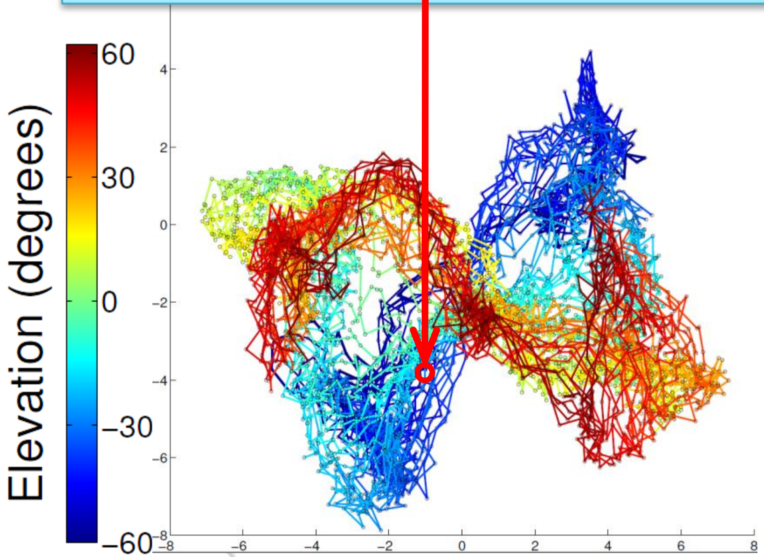
PhD thesis: Acoustic Space Sampling

2010–2013





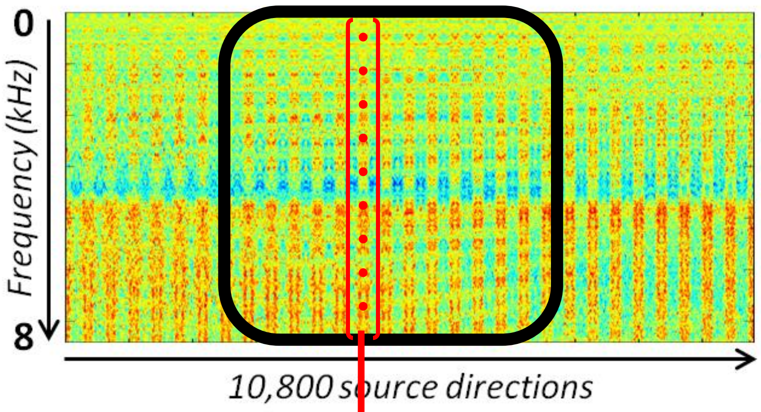
Linear dimensionality reduction (PCA)



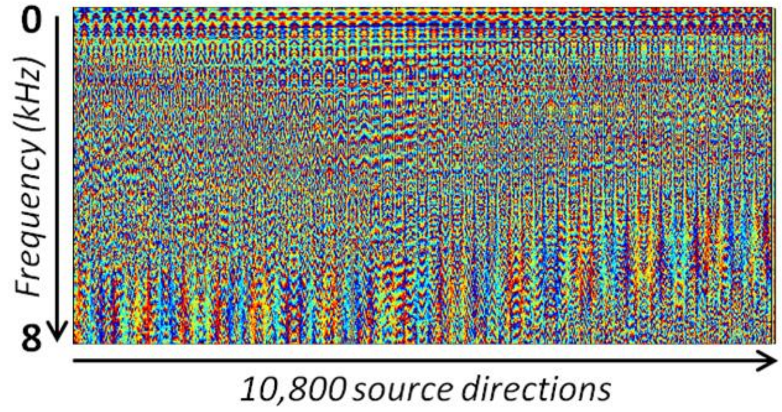
PhD thesis: Acoustic Space Sampling

2010–2013

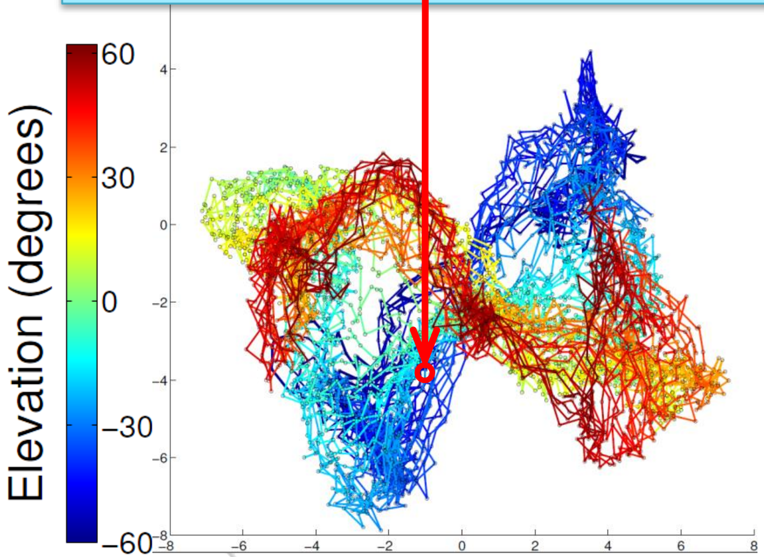
ILD space



IPD space



Linear dimensionality reduction (PCA)



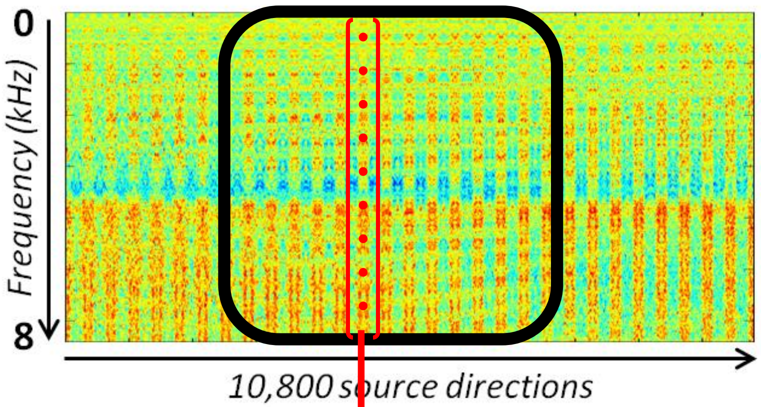
- Very distorted

=> Non-linear structure

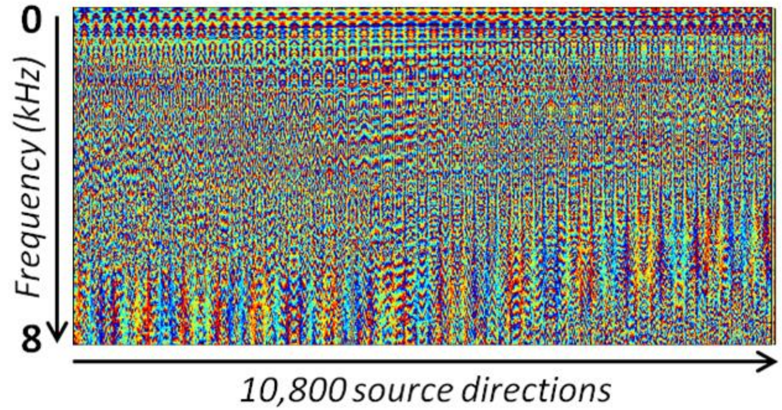
PhD thesis: Acoustic Space Sampling

2010–2013

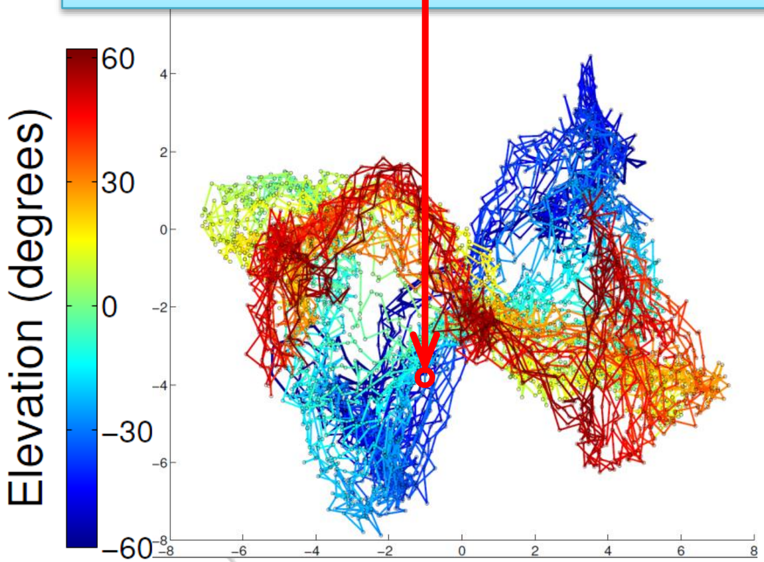
ILD space



IPD space



Linear dimensionality reduction (PCA)



• Very distorted
=> Non-linear structure

Conclusion on acoustic spaces

• Non-linear but locally-linear
• Lie on a low-dimensional manifold parameterized by source positions

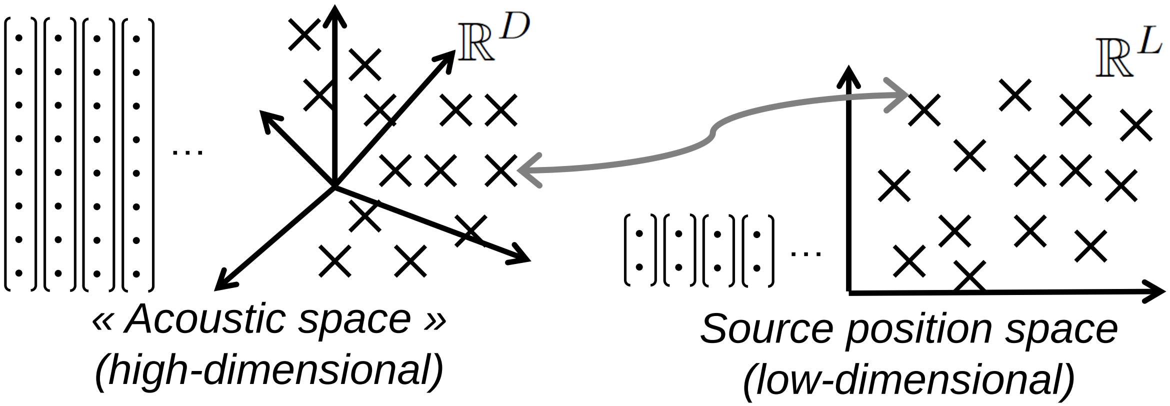
PhD thesis: Acoustic Space Learning

2010–2013

PhD thesis: Acoustic Space Learning

2010–2013

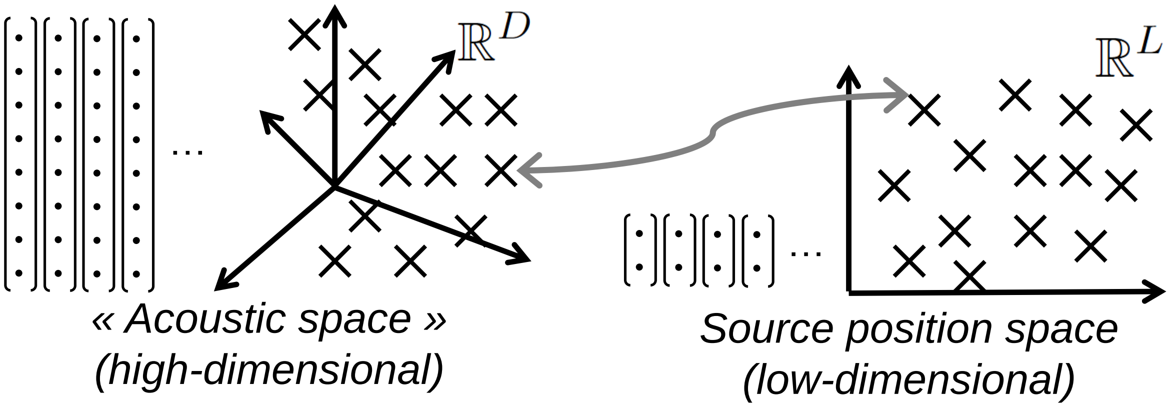
1. Use associated vectors as training data:



PhD thesis: Acoustic Space Learning

2010–2013

1. Use associated vectors as training data:

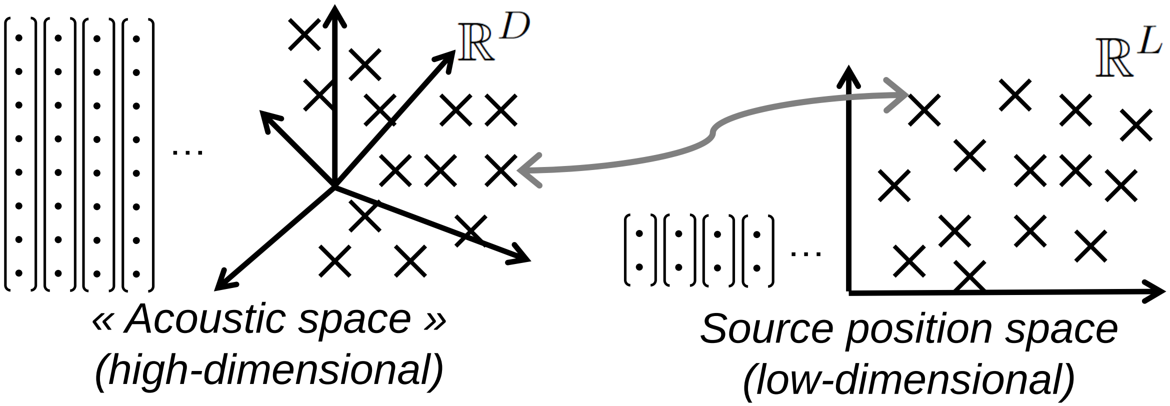


How to **collect** training data?

PhD thesis: Acoustic Space Learning

2010–2013

1. Use associated vectors as training data:

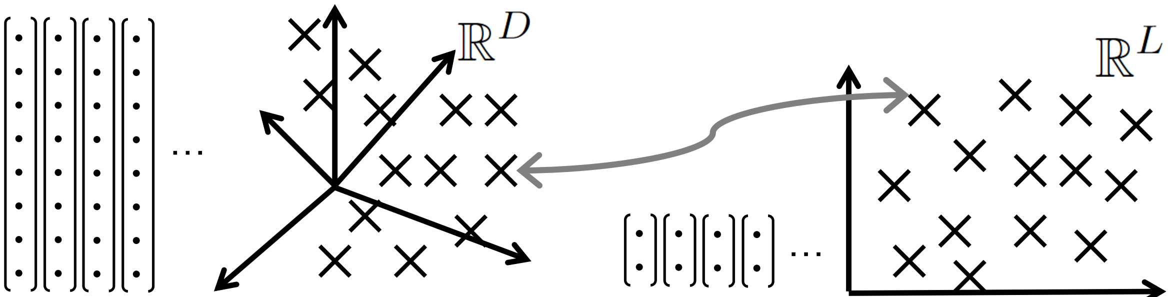


How to **collect** training data?

What is the **structure** of the acoustic space?

PhD thesis: Acoustic Space Learning

1. Use associated vectors as training data:



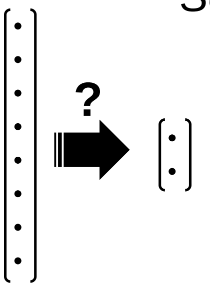
« Acoustic space »
(high-dimensional)

Source position space
(low-dimensional)

How to **collect** training data?

What is the **structure** of the acoustic space?

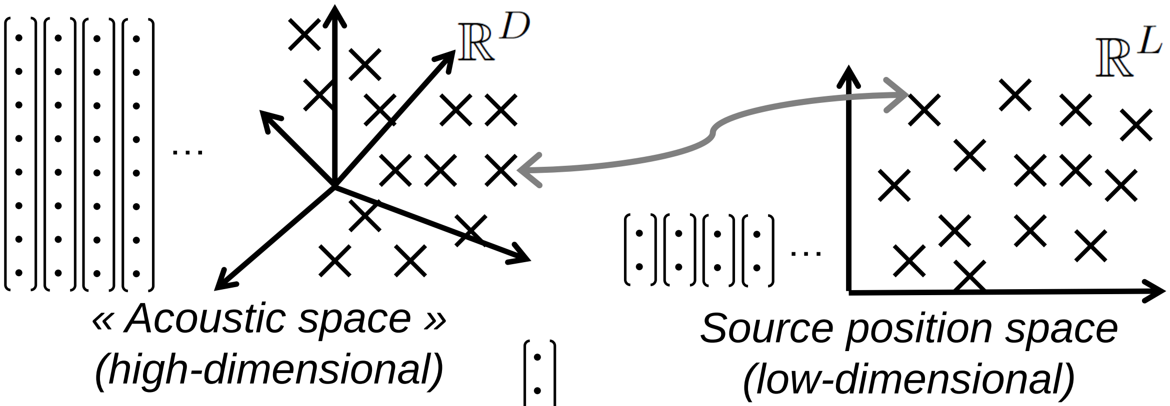
2. Learn the mapping:



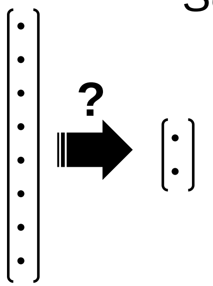
PhD thesis: Acoustic Space Learning

2010–2013

1. Use associated vectors as training data:



2. Learn the mapping:



How to **collect** training data?

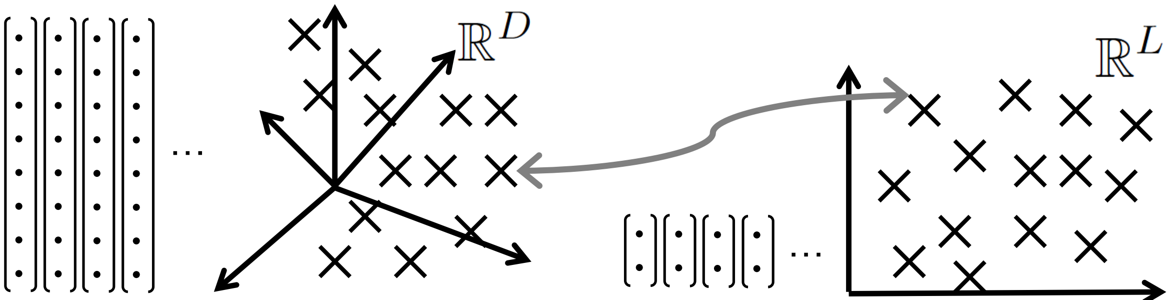
What is the **structure** of the acoustic space?

How to deal with **high-dimensional** input?

PhD thesis: Acoustic Space Learning

2010–2013

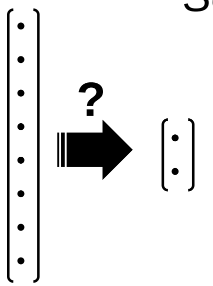
1. Use associated vectors as training data:



« Acoustic space »
(high-dimensional)

Source position space
(low-dimensional)

2. Learn the mapping:



How to **collect** training data?

What is the **structure** of the acoustic space?

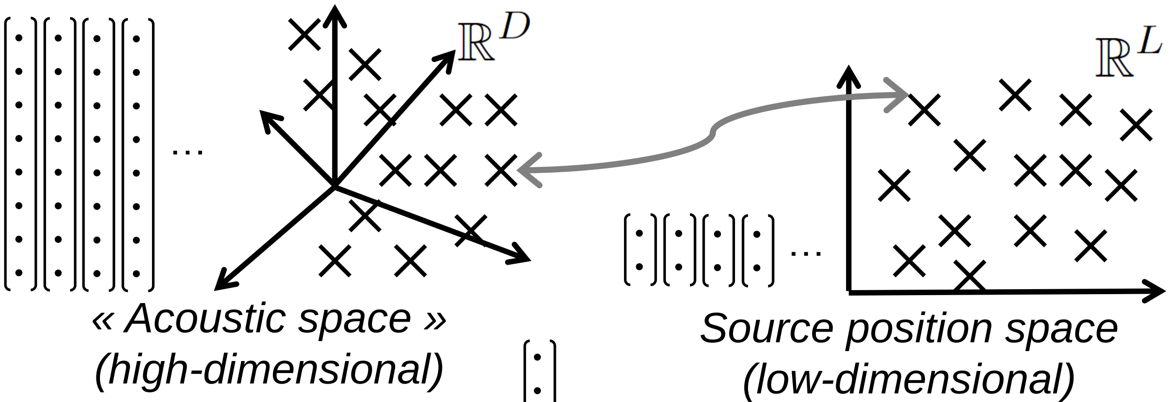
How to deal with **high-dimensional** input?

How to estimate a **non-linear** mapping?

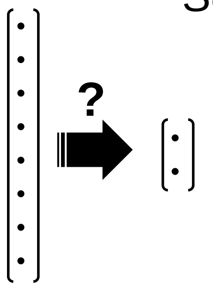
PhD thesis: Acoustic Space Learning

2010–2013

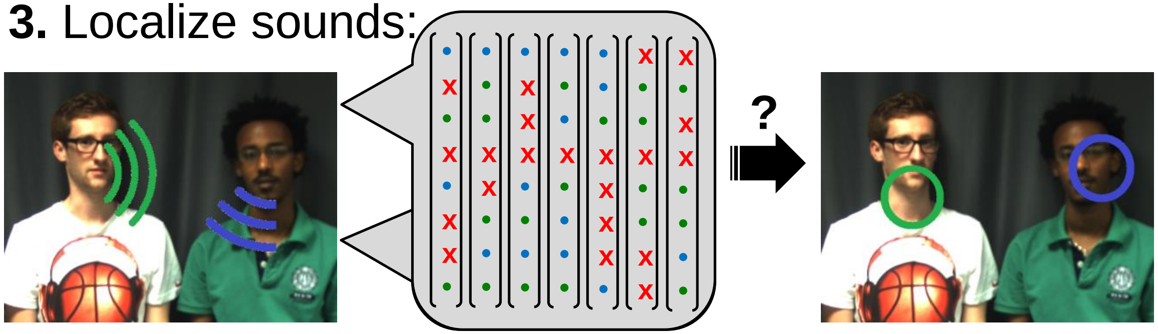
1. Use associated vectors as training data:



2. Learn the mapping:



3. Localize sounds:



How to **collect** training data?

What is the **structure** of the acoustic space?

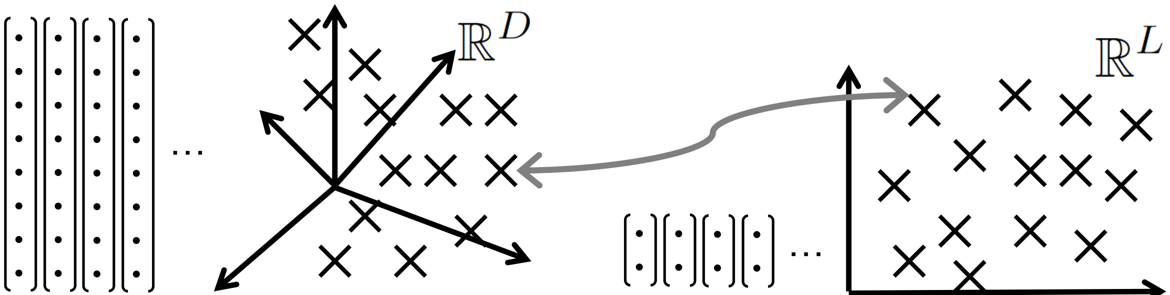
How to deal with **high-dimensional** input?

How to estimate a **non-linear** mapping?

PhD thesis: Acoustic Space Learning

2010–2013

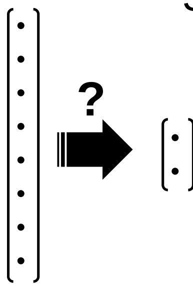
1. Use associated vectors as training data:



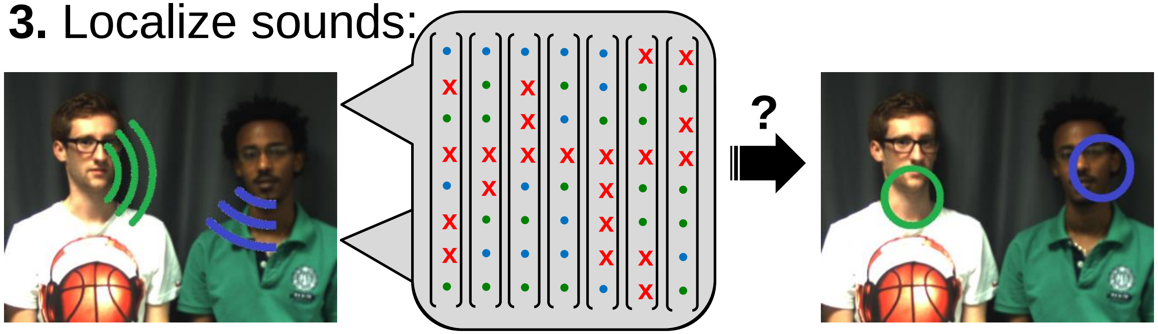
« Acoustic space » (high-dimensional)

Source position space (low-dimensional)

2. Learn the mapping:



3. Localize sounds:



How to **collect** training data?

What is the **structure** of the acoustic space?

How to deal with **high-dimensional** input?

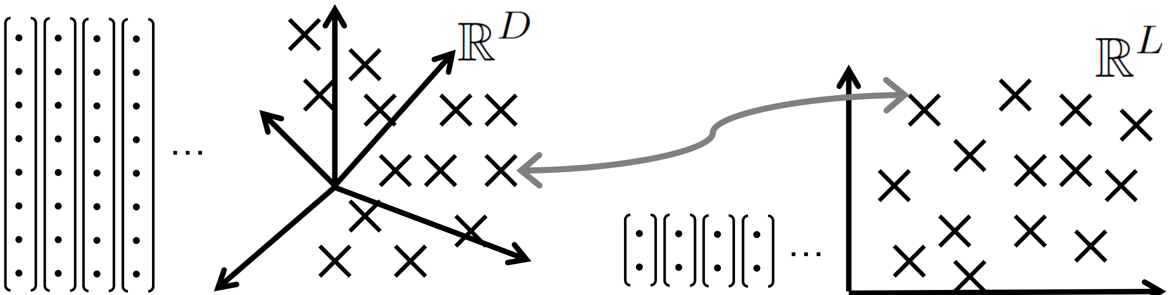
How to estimate a **non-linear** mapping?

How to handle **mixed series** with **missing values ?**

PhD thesis: Acoustic Space Learning

2010–2013

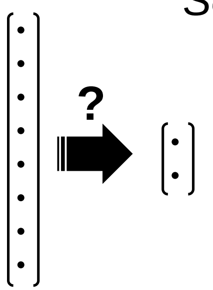
1. Use associated vectors as training data:



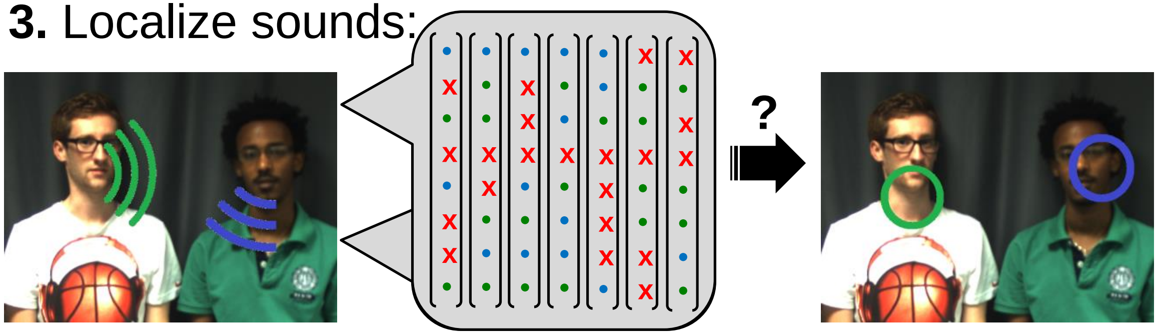
« Acoustic space »
(high-dimensional)

Source position space
(low-dimensional)

2. Learn the mapping:



3. Localize sounds:



How to **collect** training data?

What is the **structure** of the acoustic space?

How to deal with **high-dimensional** input?

How to estimate a **non-linear** mapping?

How to handle **mixed series** with **missing values ?**

How to **separate** sound sources?

PhD thesis: Some Results (1 source)

2010–2013

PhD thesis: Some Results (2 sources)

2010–2013

PhD thesis: Playing with Nao

2010–2013

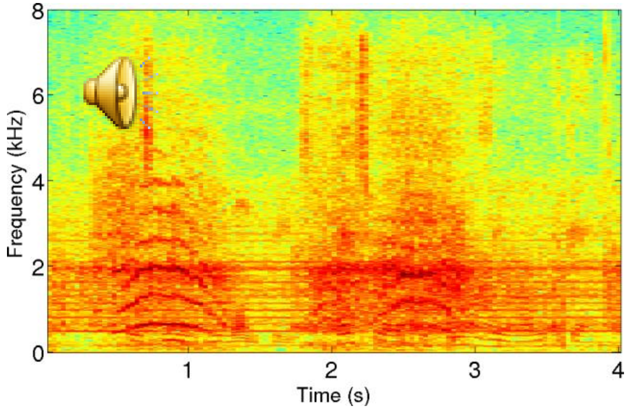
The Post-Doc

2014–2015

Or how I kept torturing robots with sounds for a living

Nao Waving

Nao Walking

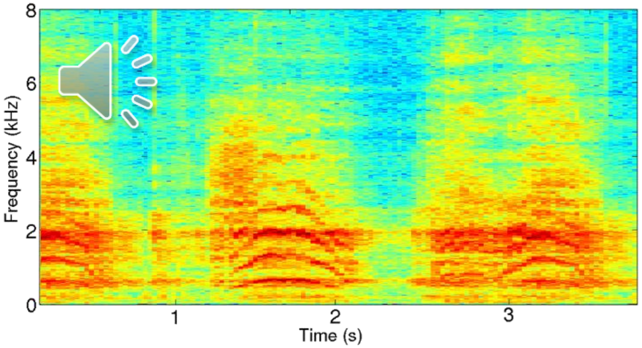


The Post-Doc

Dictionary-based egonoise reduction

Training Phase

≈ 30 seconds of noise only (fan removed with Wiener filtering)

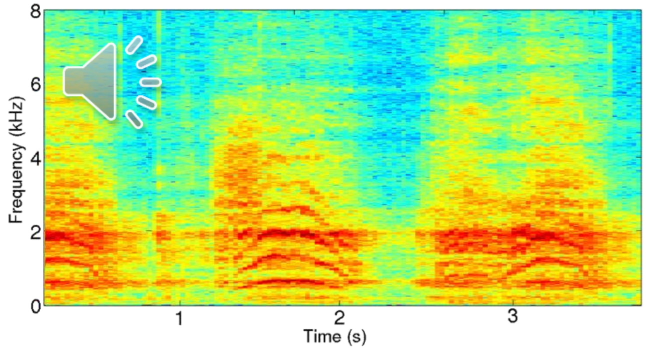


The Post-Doc

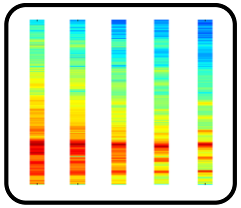
Dictionary-based egonoise reduction

Training Phase

≈ 30 seconds of noise only (fan removed with Wiener filtering)



DL algorithm
(e.g. NMF)



Dictionary
(*e.g.*: 30 ***atoms***)

Antoine.Deleforge@inria.fr

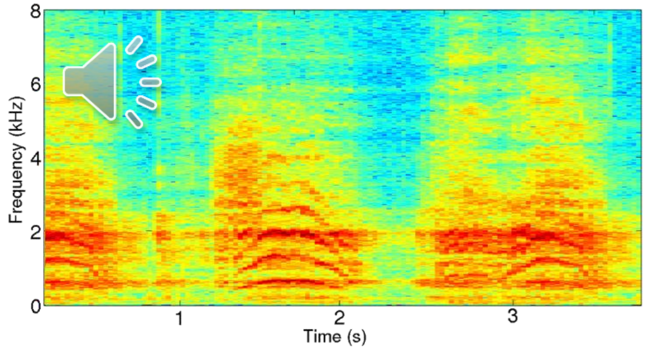
The Post-Doc

2014–2015

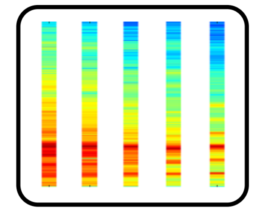
Dictionary-based egonoise reduction

Training Phase

≈ 30 seconds of noise only (fan removed with Wiener filtering)

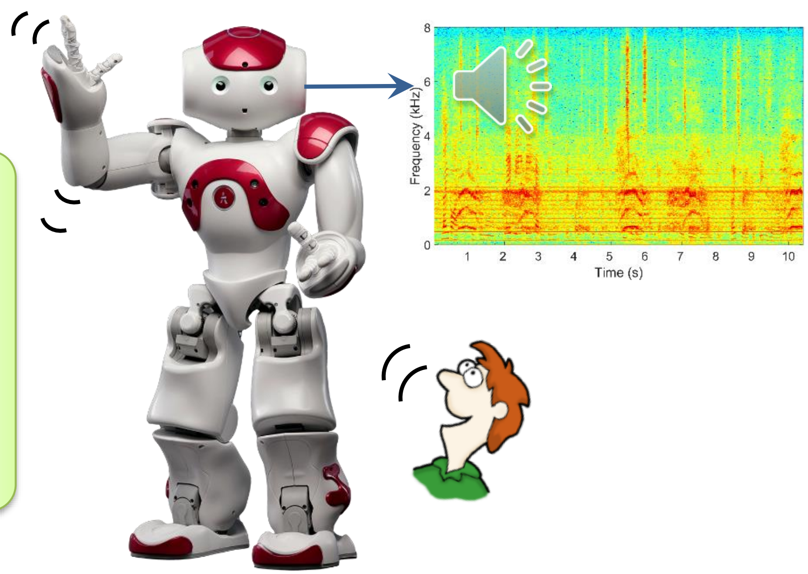


DL algorithm
(e.g. NMF)



Dictionary
(*e.g.*: 30 ***atoms***)

Testing Phase



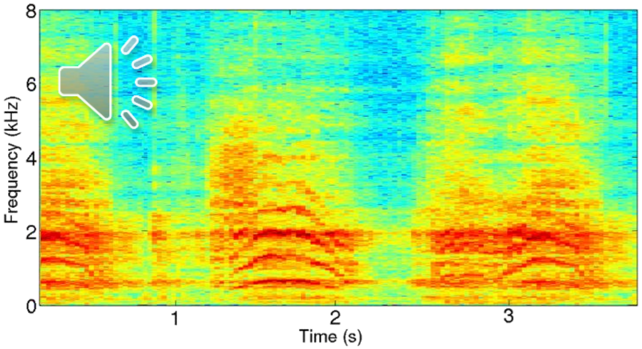
The Post-Doc

2014–2015

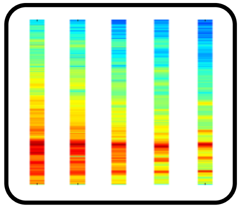
Dictionary-based egonoise reduction

Training Phase

≈ 30 seconds of noise only (fan removed with Wiener filtering)

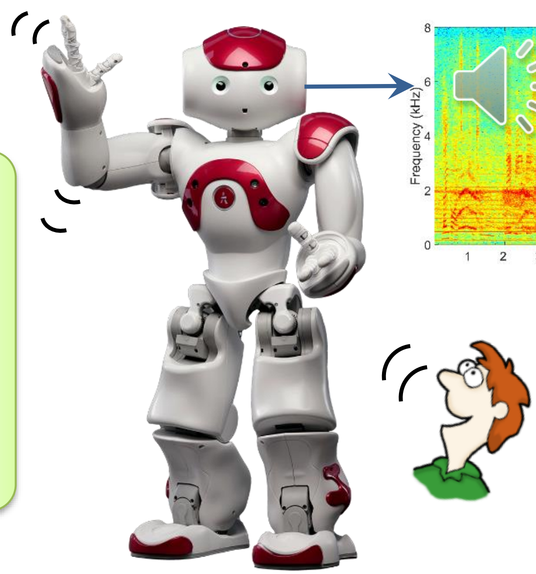


DL algorithm
(e.g. NMF)

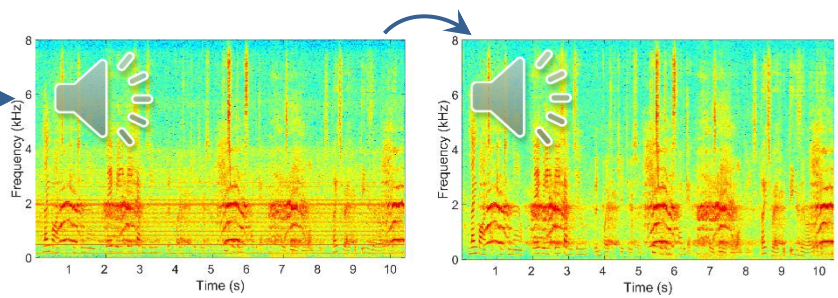


Dictionary
(*e.g.*: 30 ***atoms***)

Testing Phase



Wiener



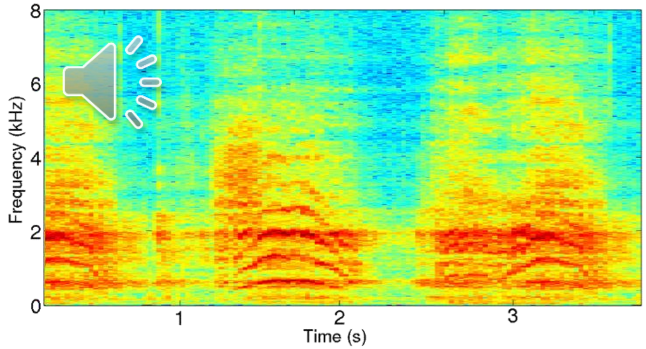
The Post-Doc

2014–2015

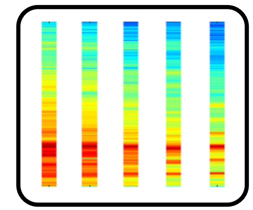
Dictionary-based egonoise reduction

≈ 30 seconds of noise only (fan removed with Wiener filtering)

Training Phase

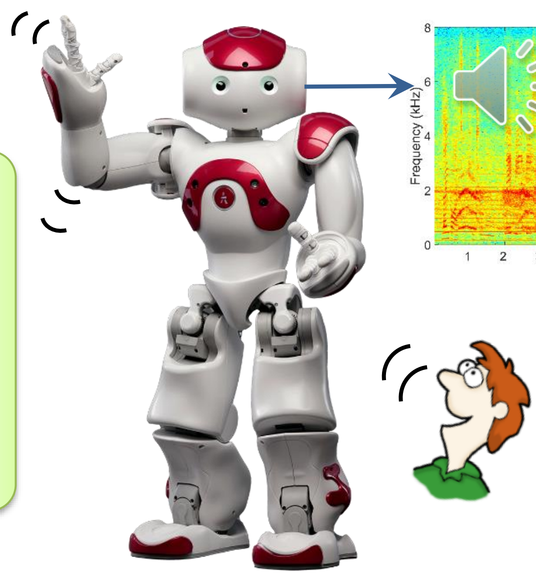


DL algorithm
(e.g. NMF)

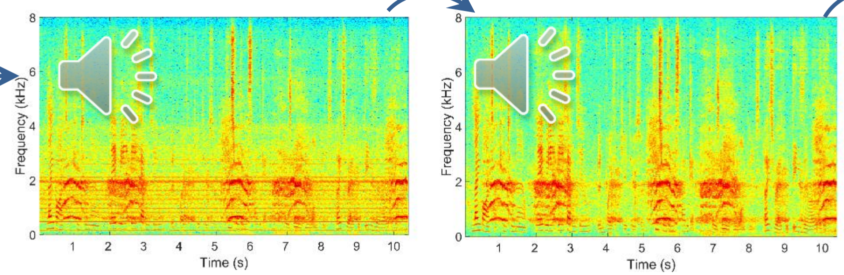


Dictionary
(*e.g.*: 30 ***atoms***)

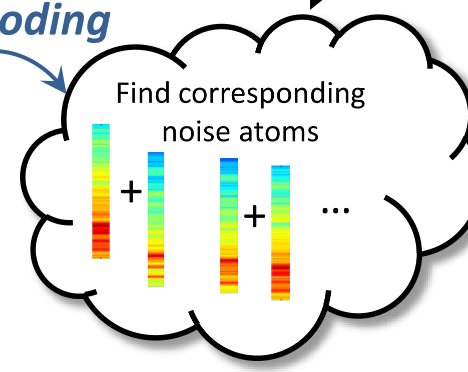
Testing Phase



Wiener



Sparse Coding

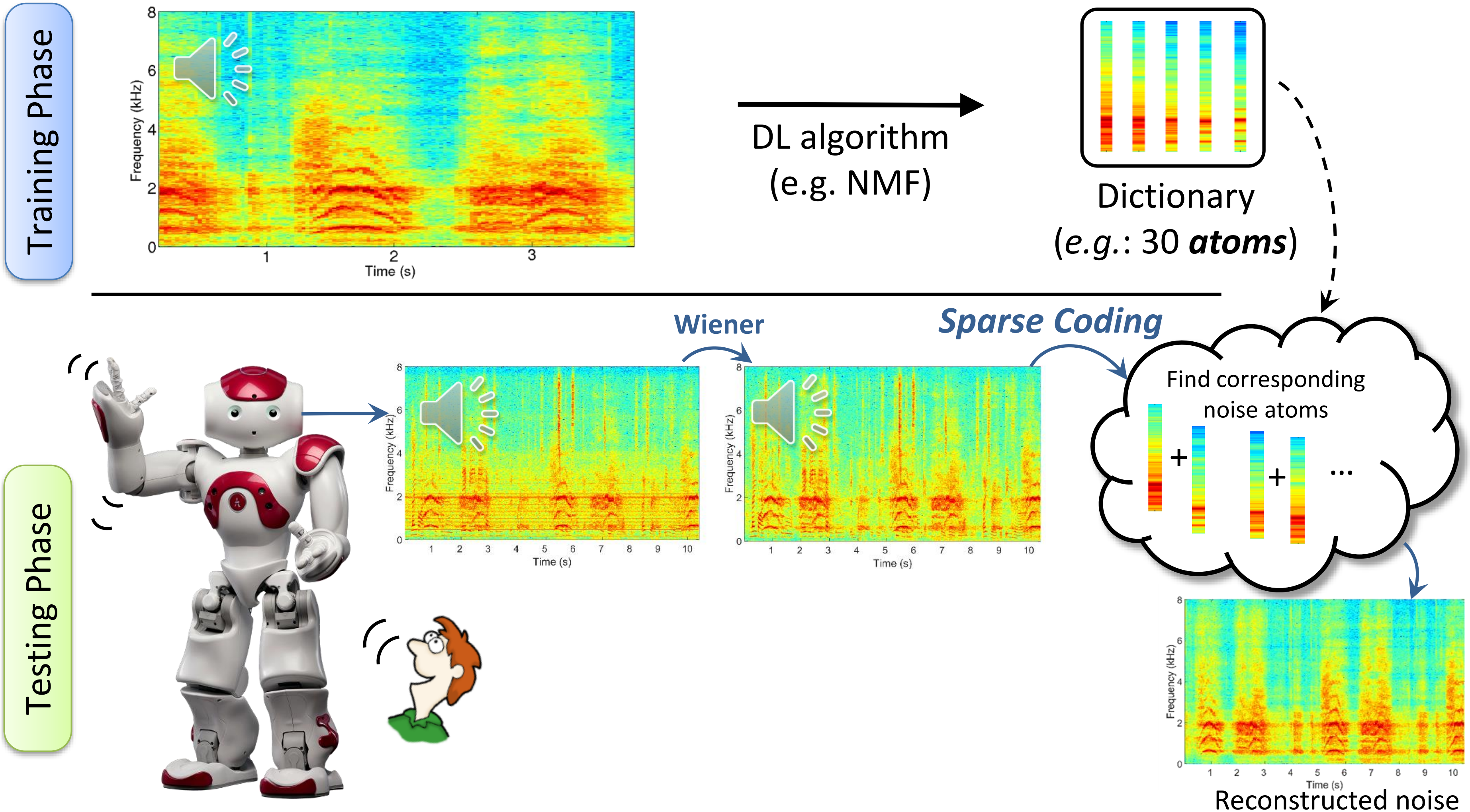


The Post-Doc

2014–2015

Dictionary-based egonoise reduction

≈ 30 seconds of noise only (fan removed with Wiener filtering)

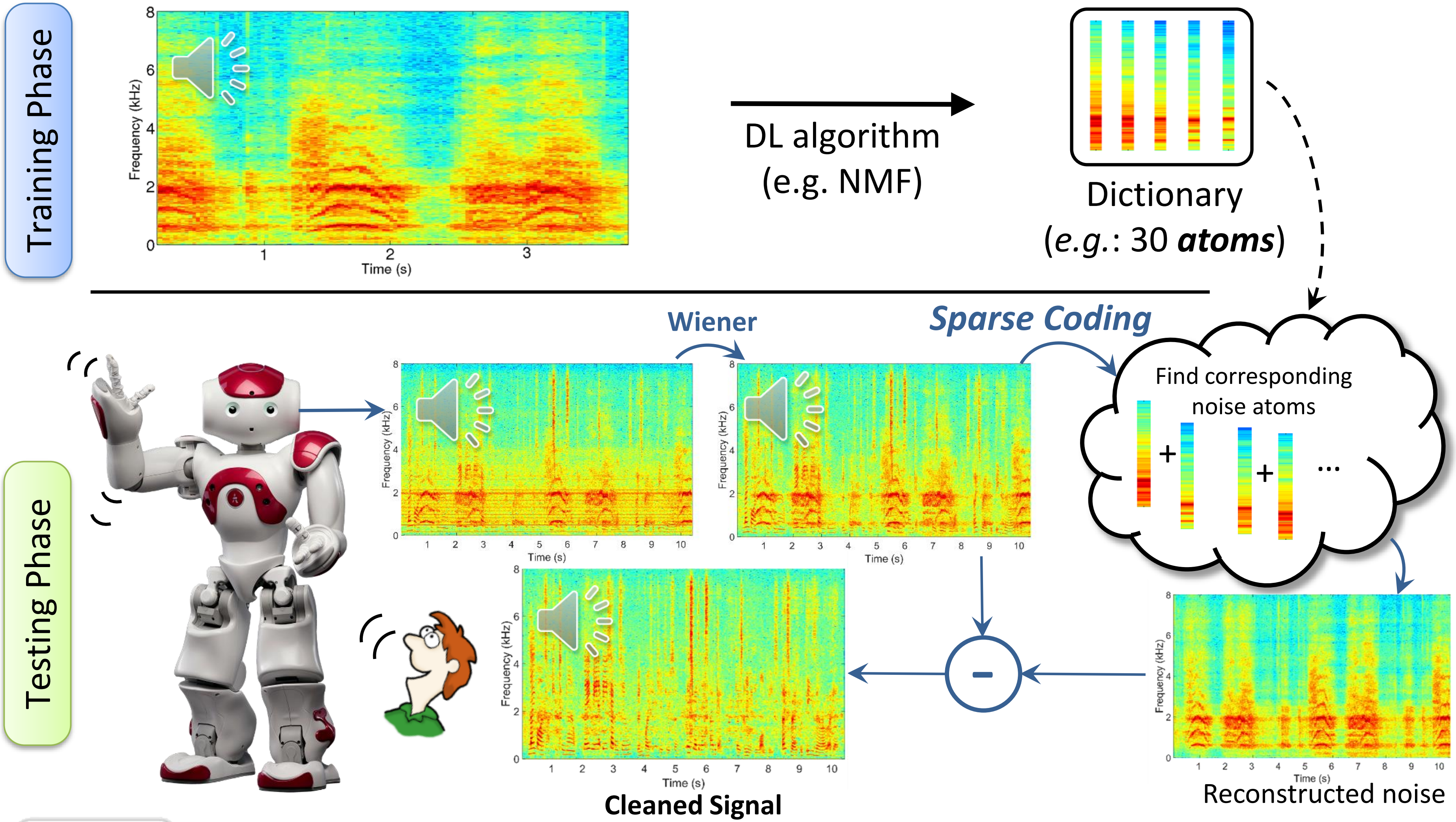


The Post-Doc

2014–2015

Dictionary-based egonoise reduction

≈ 30 seconds of noise only (fan removed with Wiener filtering)



Inria Rennes

2016–2018

Ok, let’s see on what robot I could release my psychopatic urges this time…

Inria Rennes

2016–2018

Ok, let's see on what robot I could release my psychopatic urges this time…

« Oh, hey, I heard you were working with ***Drones*** *? »*

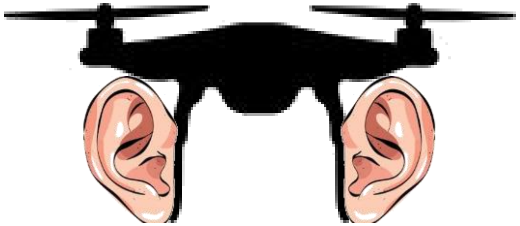
Inria Rennes

2016–2018

Ok, let's see on what robot I could release my psychopatic urges this time…

« Oh, hey, I heard you were working with ***Drones*** *? »*

« What do you think of a drone with ears? »



Inria Rennes

Ok, let's see on what robot I could release my psychopatic urges this time…

*« Oh, hey, I heard you were working with **Drones** ? »*

« What do you think of a drone with ears? »

Antoine.Deleforge@inria.fr

Inria Rennes

2016–2018

The DREGON dataset: Drone Audition for Search & Rescue

Present time, Inria Nancy

2018–2022

Or the sudden realization that gathering audio data to train robots is a massive pain in the… is very impractical.

Present time, Inria Nancy

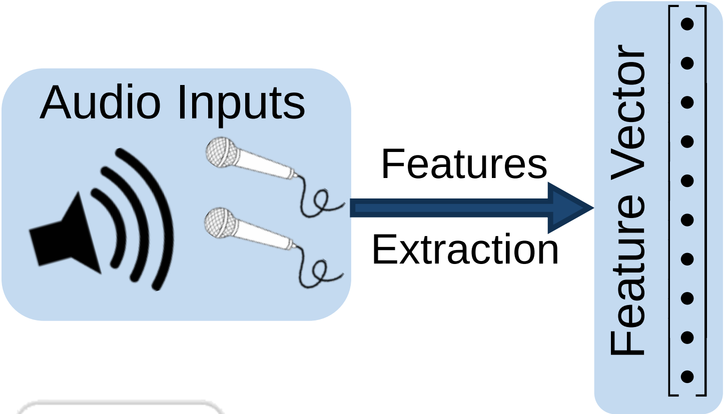
2018–2022

Or the sudden realization that gathering audio data to train robots is a massive pain in the… is very impractical.

Present time, Inria Nancy

2018–2022

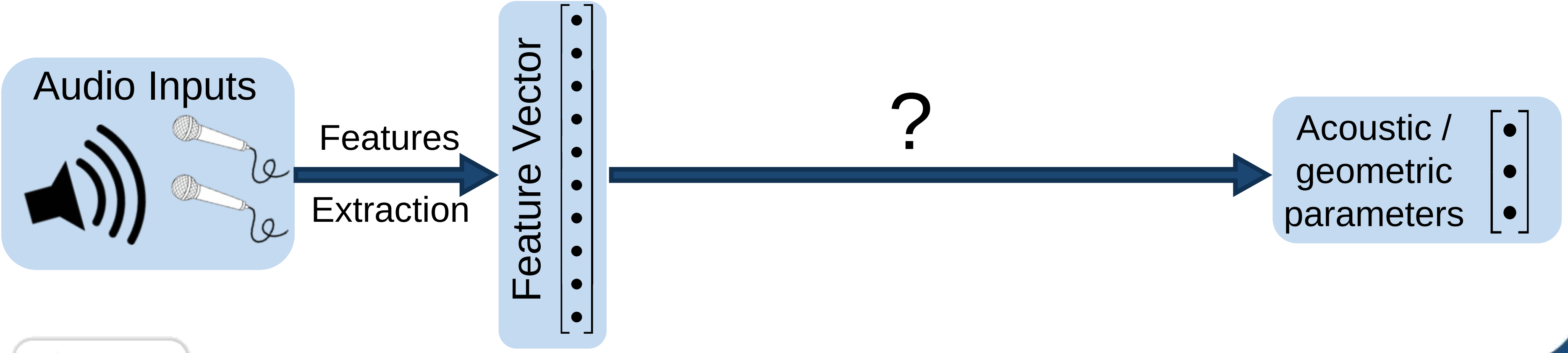
Or the sudden realization that gathering audio data to train robots is a massive pain in the… is very impractical.



Present time, Inria Nancy

2018–2022

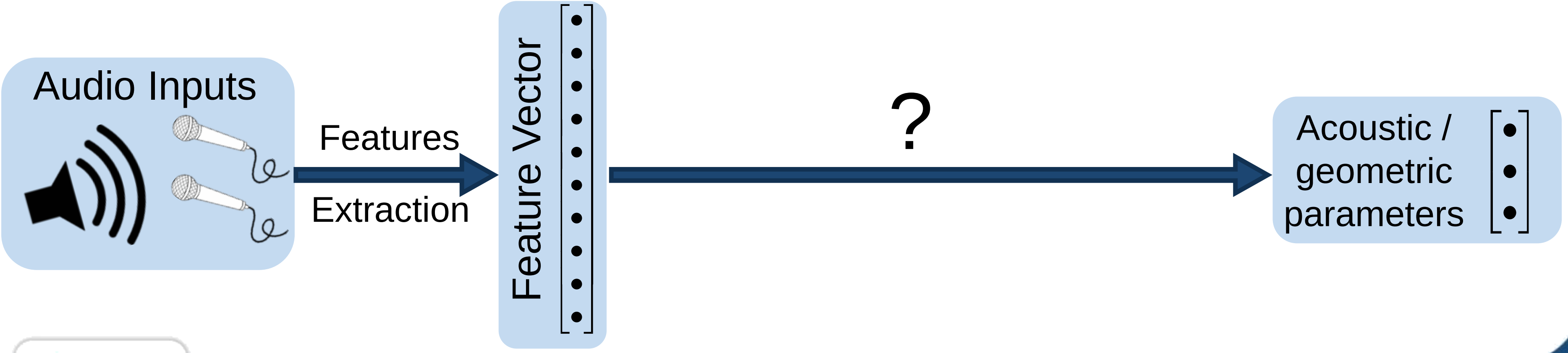
Or the sudden realization that gathering audio data to train robots is a massive pain in the... is very impractical.



Present time, Inria Nancy

2018–2022

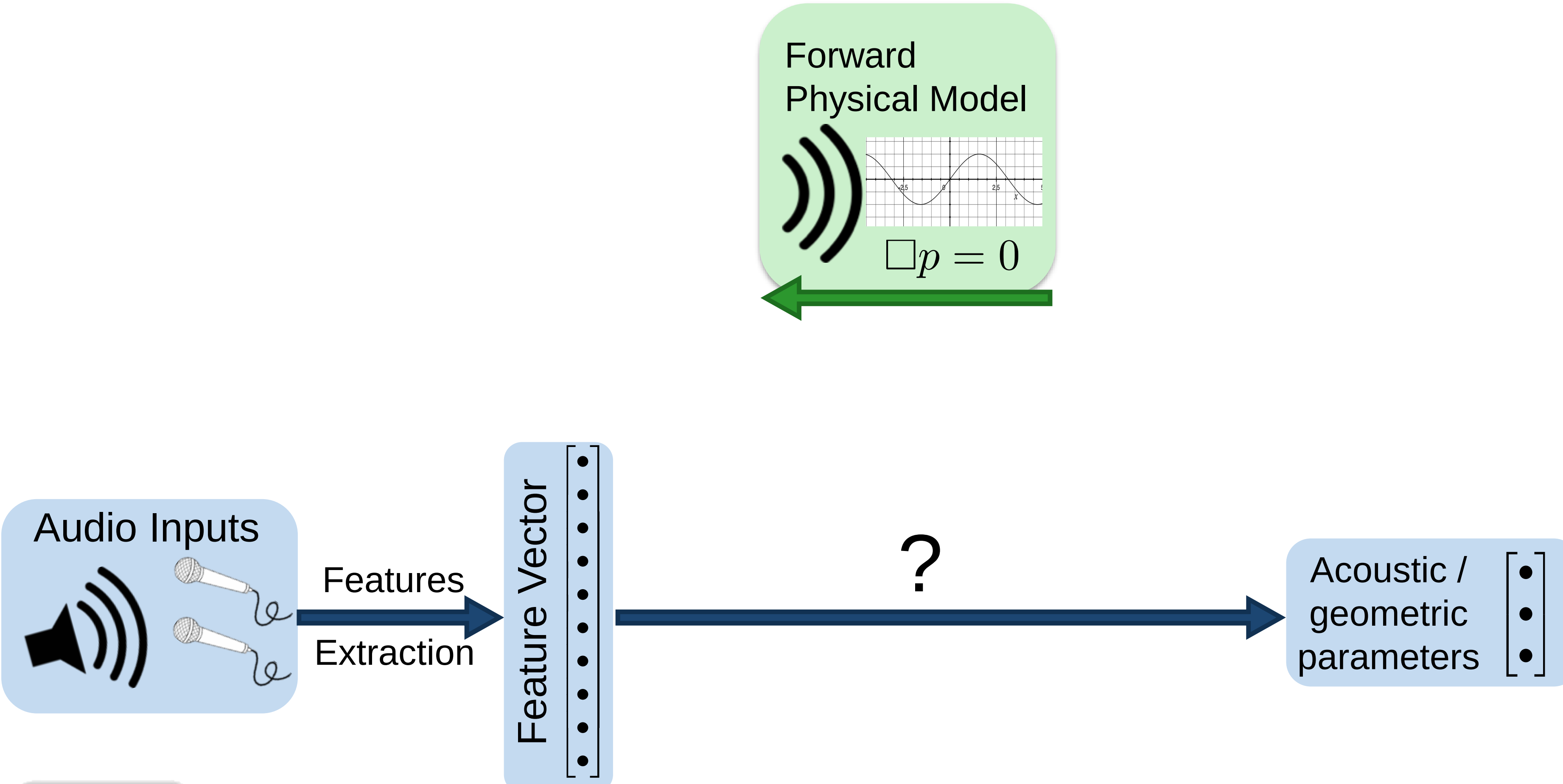
a) Physics-Driven / Traditional Approaches



Present time, Inria Nancy

2018–2022

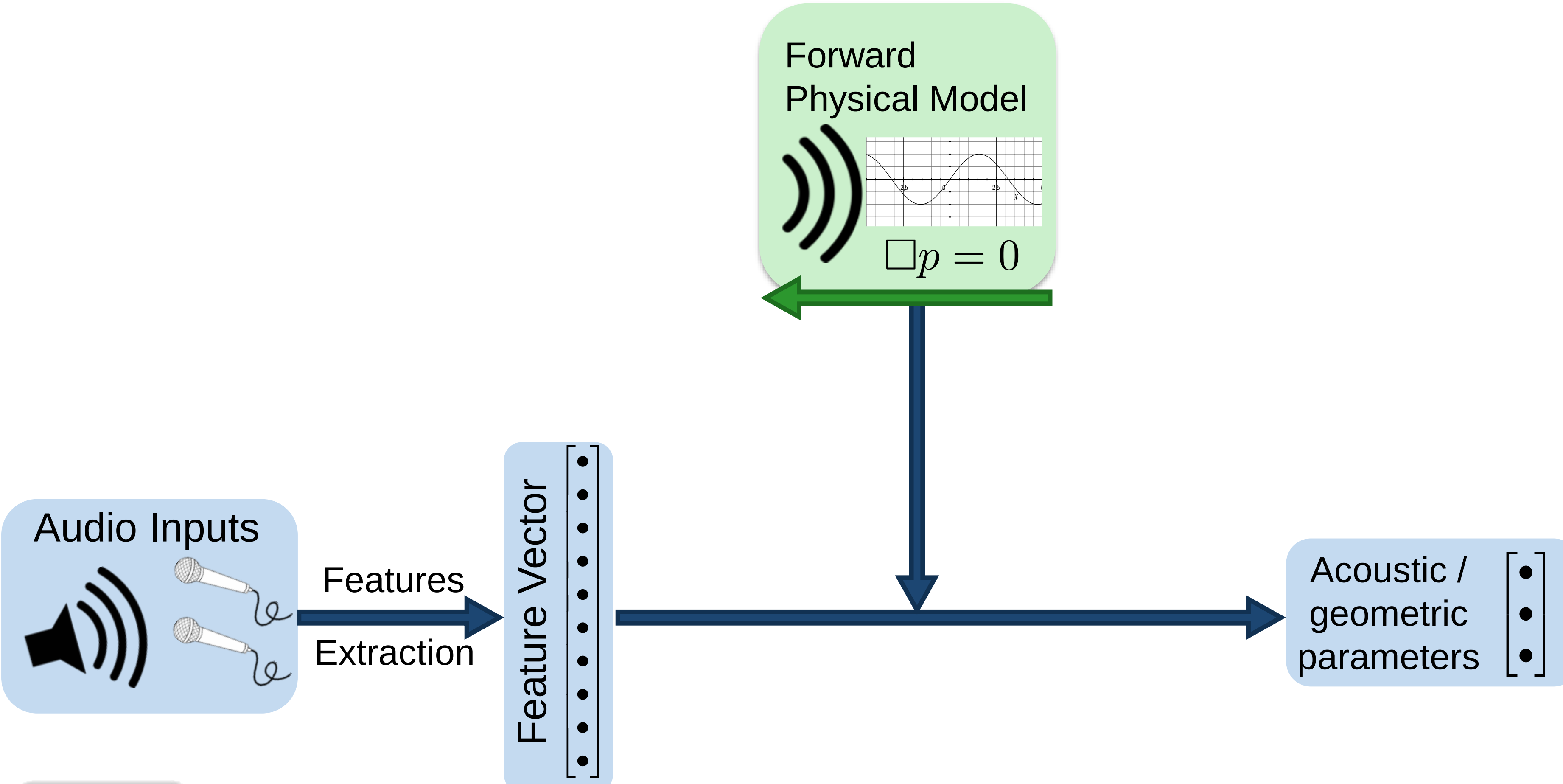
a) Physics-Driven / Traditional Approaches



Present time, Inria Nancy

2018–2022

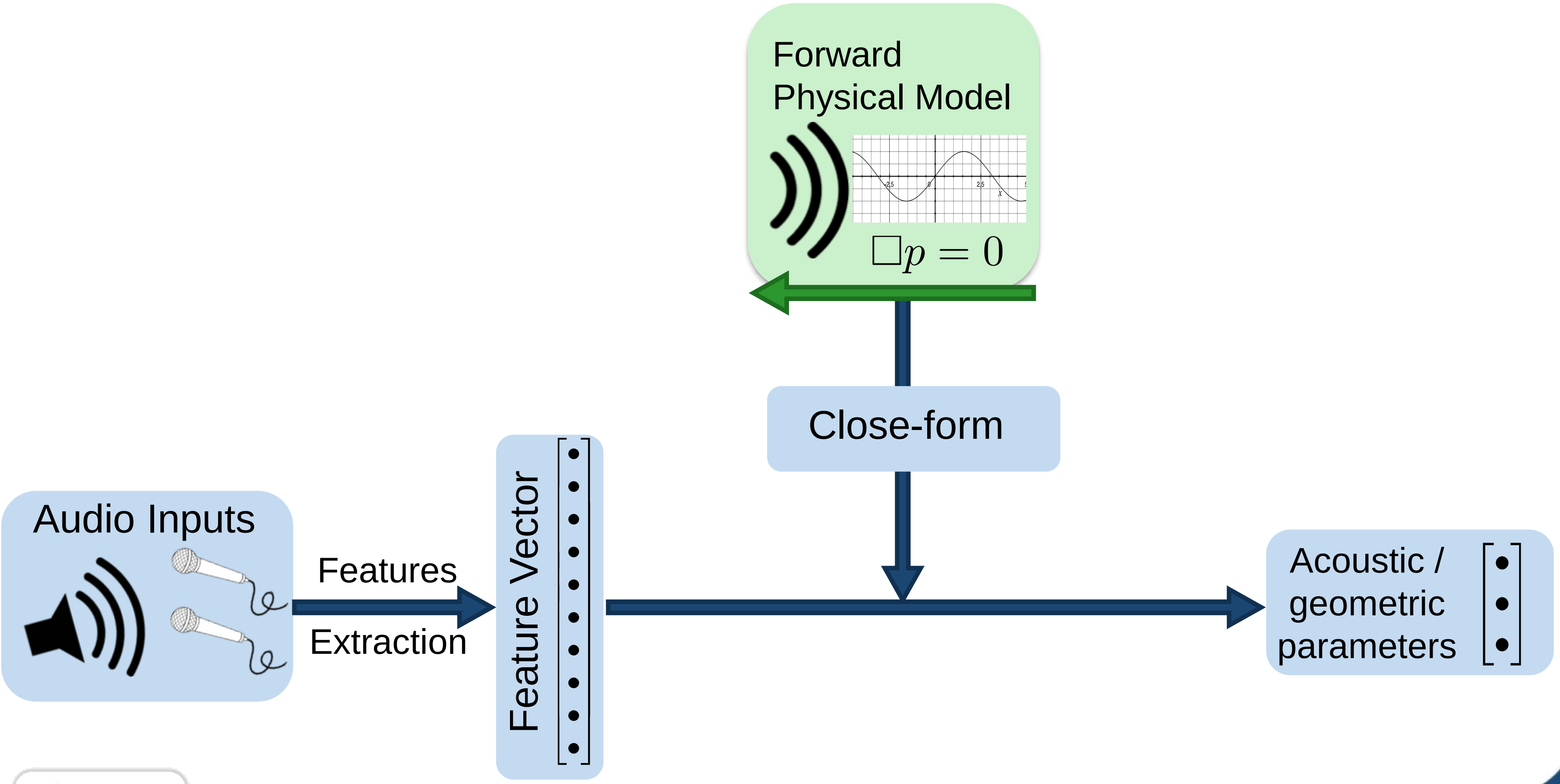
a) Physics-Driven / Traditional Approaches



Present time, Inria Nancy

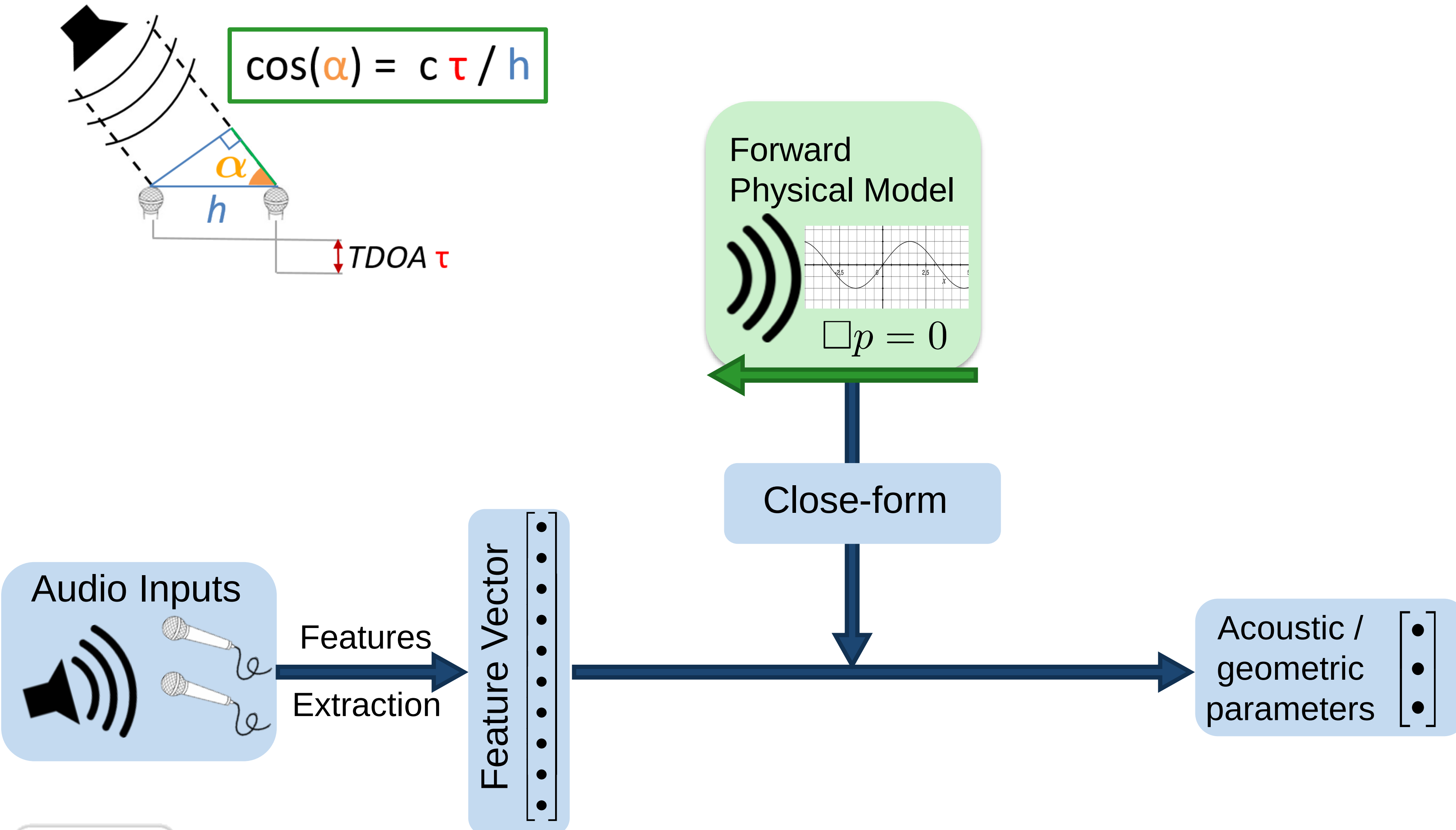
2018–2022

a) Physics-Driven / Traditional Approaches



Present time, Inria Nancy

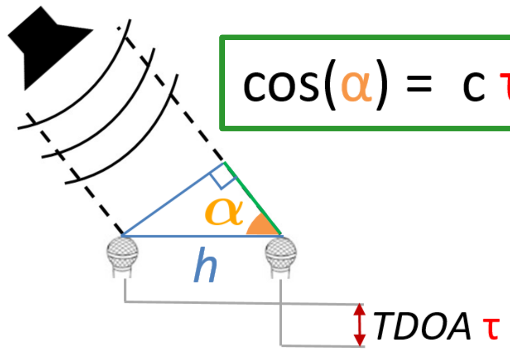
a) Physics-Driven / Traditional Approaches



Present time, Inria Nancy

2018–2022

a) Physics-Driven / Traditional Approaches



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

Sabine's law:

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

Forward Physical Model

$\Box p = 0$

Close-form

Audio Inputs

Features Extraction

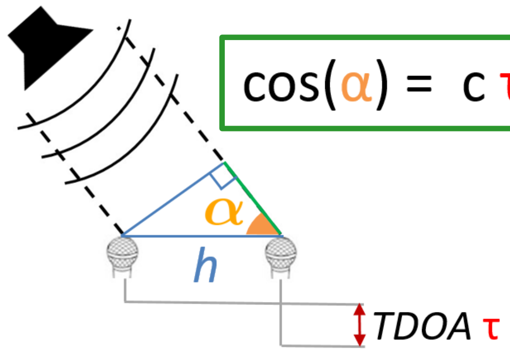
Feature Vector

Acoustic / geometric parameters

Present time, Inria Nancy

2018–2022

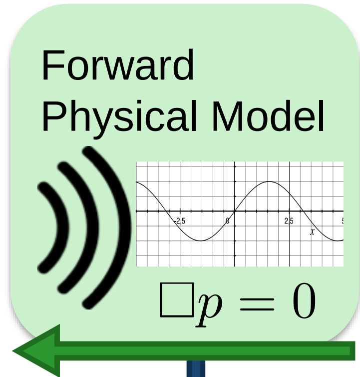
a) Physics-Driven / Traditional Approaches

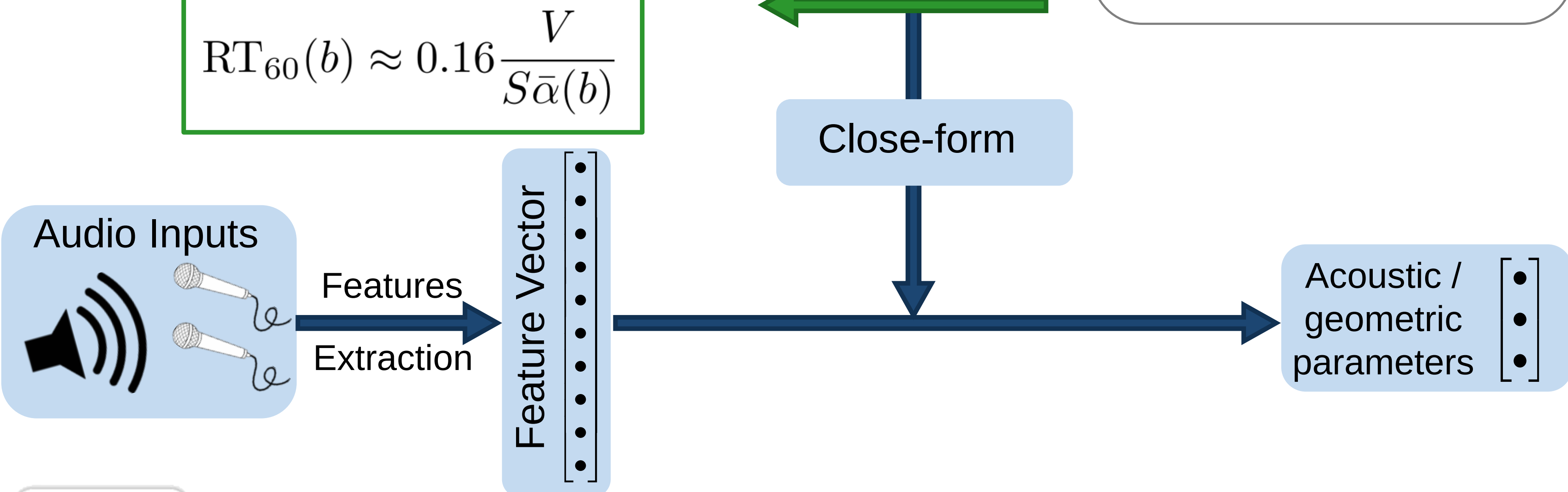


$$\boxed{\cos(\alpha) = c\,\tau / h}$$

Sabine's law:

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



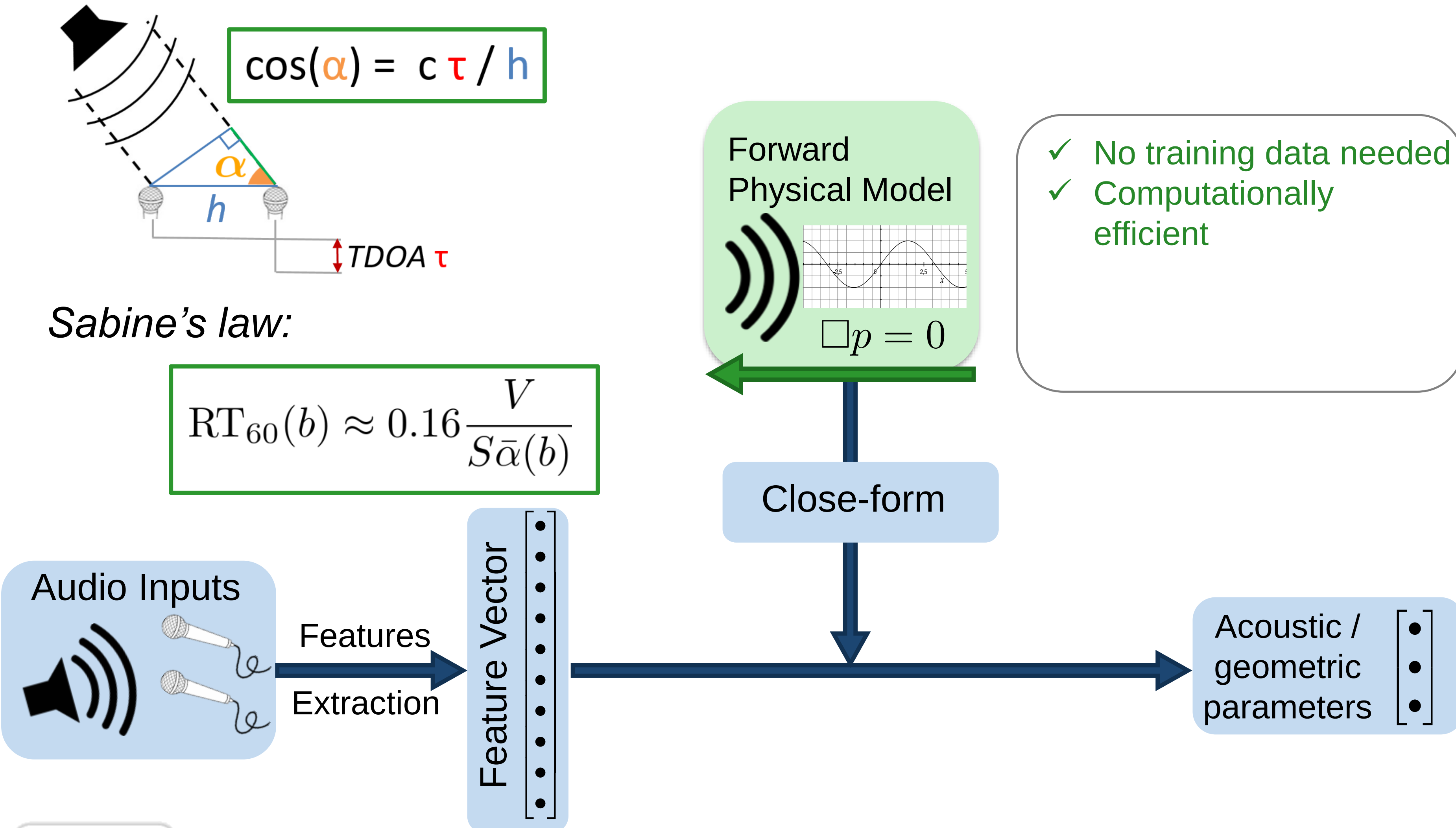


✓ No training data needed

Present time, Inria Nancy

2018–2022

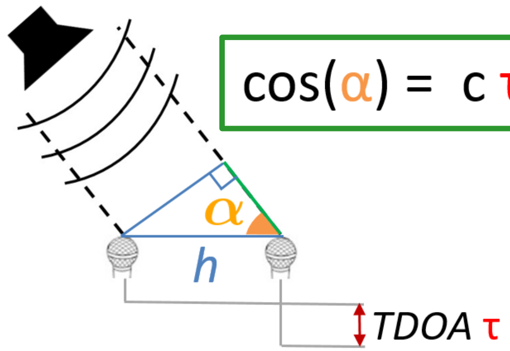
a) Physics-Driven / Traditional Approaches



Present time, Inria Nancy

2018–2022

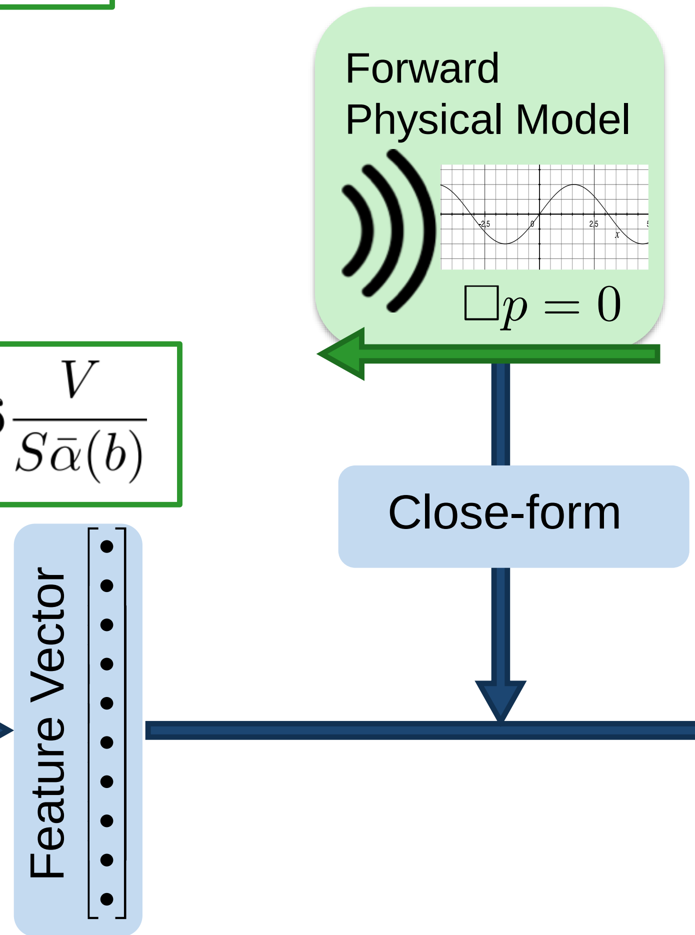
a) Physics-Driven / Traditional Approaches



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

Sabine's law:

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

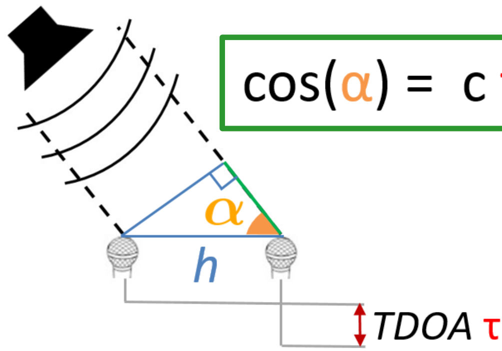


- ✓ No training data needed
- ✓ Computationally efficient
- ☒ Suffers in complex conditions

Present time, Inria Nancy

2018–2022

a) Physics-Driven / Traditional Approaches

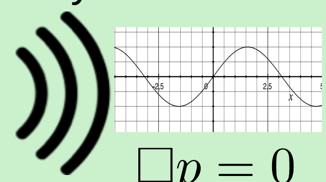


$\cos(\alpha) = c\,\tau / h$

Sabine's law:

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

Forward Physical Model



$\square p = 0$

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

Audio Inputs

Features Extraction

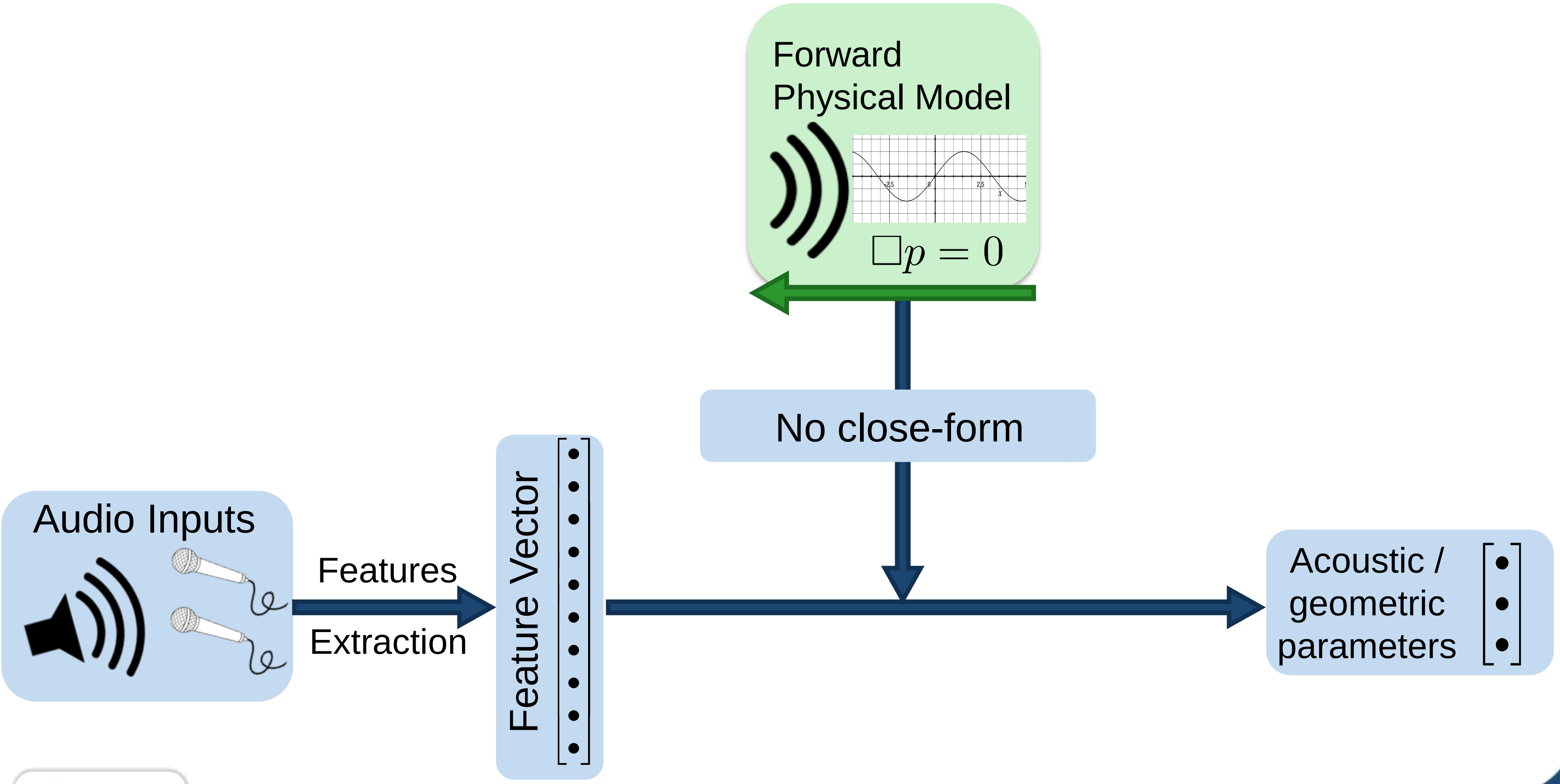
Feature Vector

Close-form

Acoustic / geometric parameters

Present time, Inria Nancy

a) Physics-Driven / Traditional Approaches

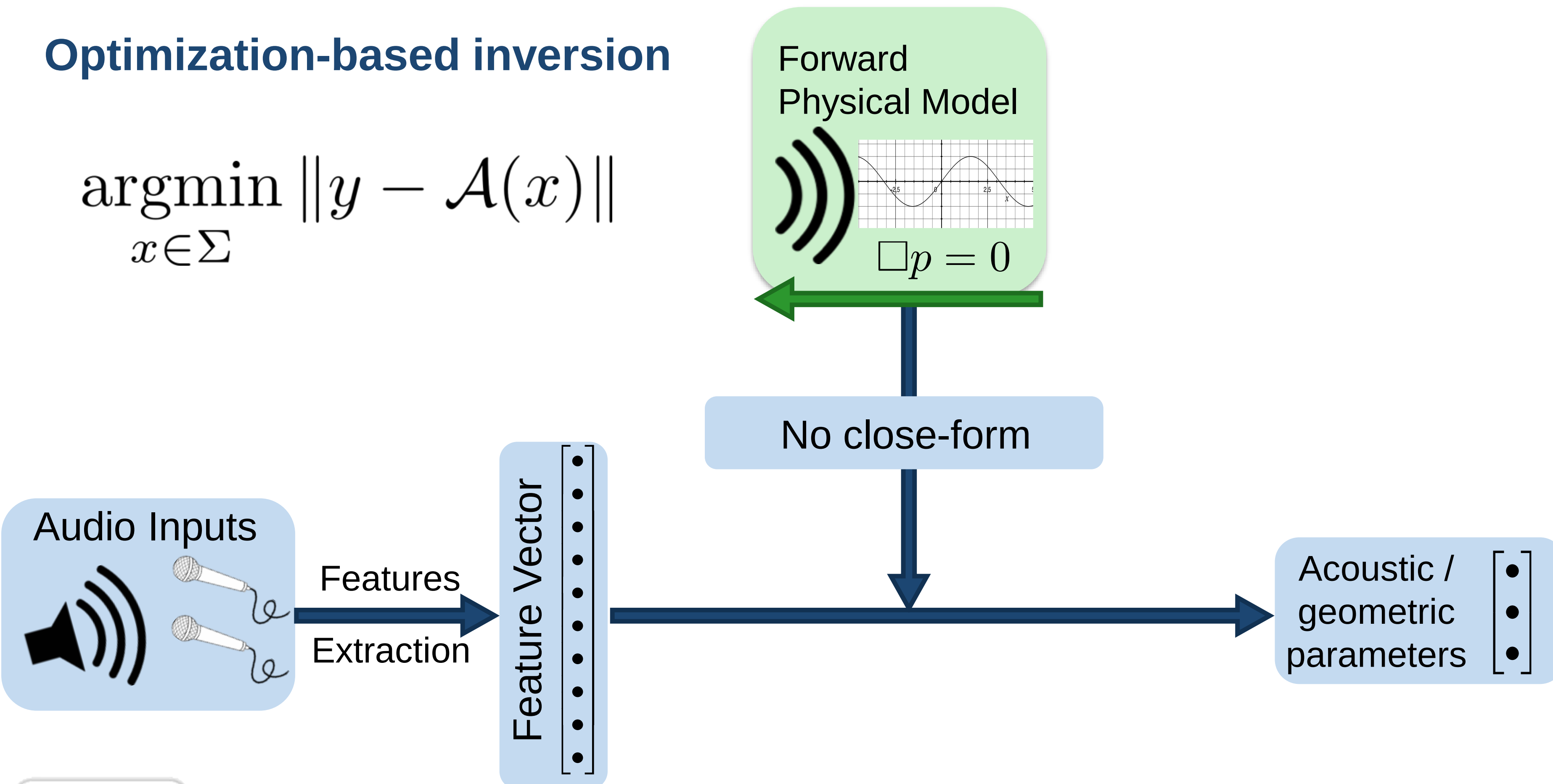


Present time, Inria Nancy

a) Physics-Driven / Traditional Approaches

Optimization-based inversion

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



Present time, Inria Nancy

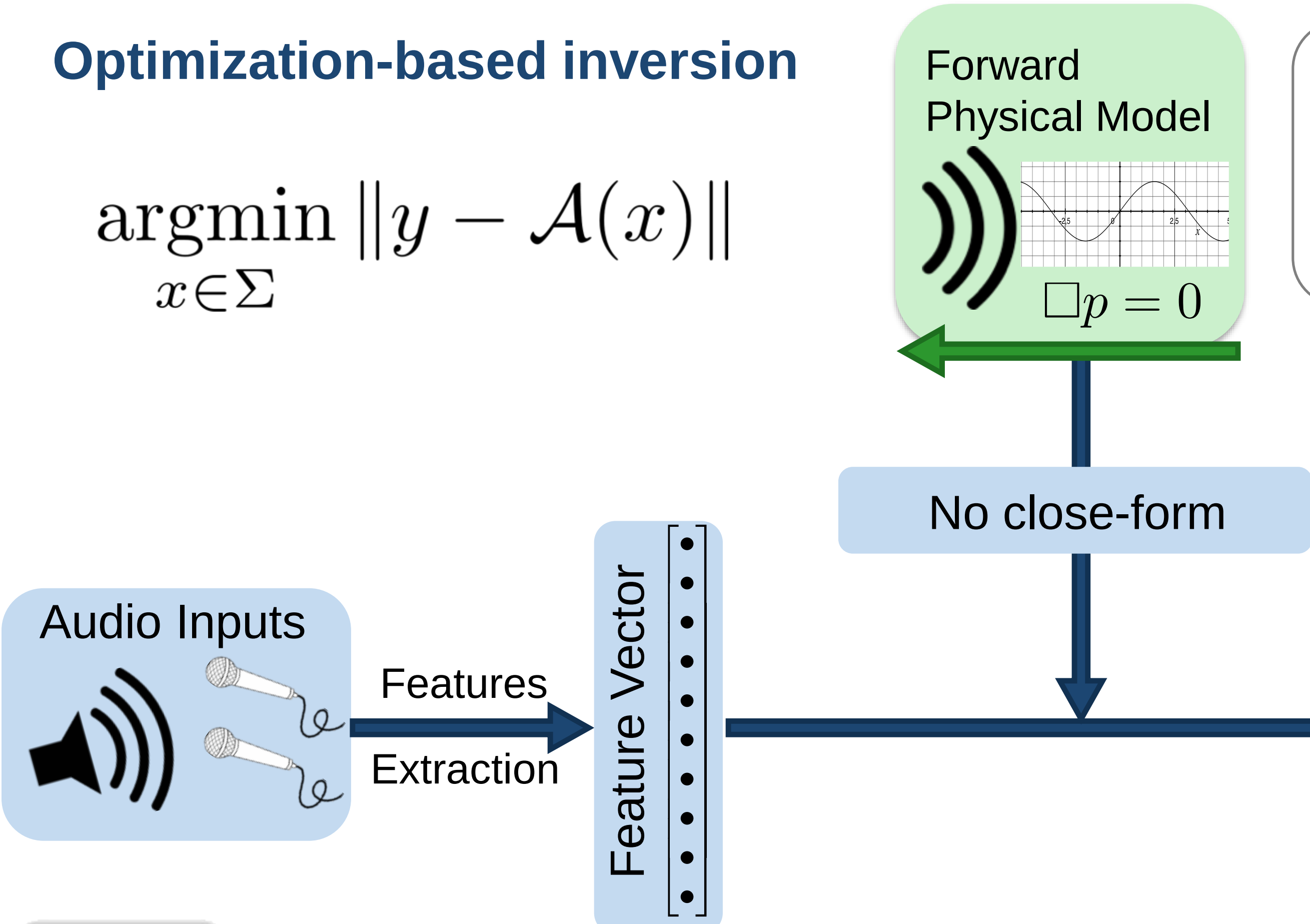
2018–2022

a) Physics-Driven / Traditional Approaches

Optimization-based inversion

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

✓ No training data needed



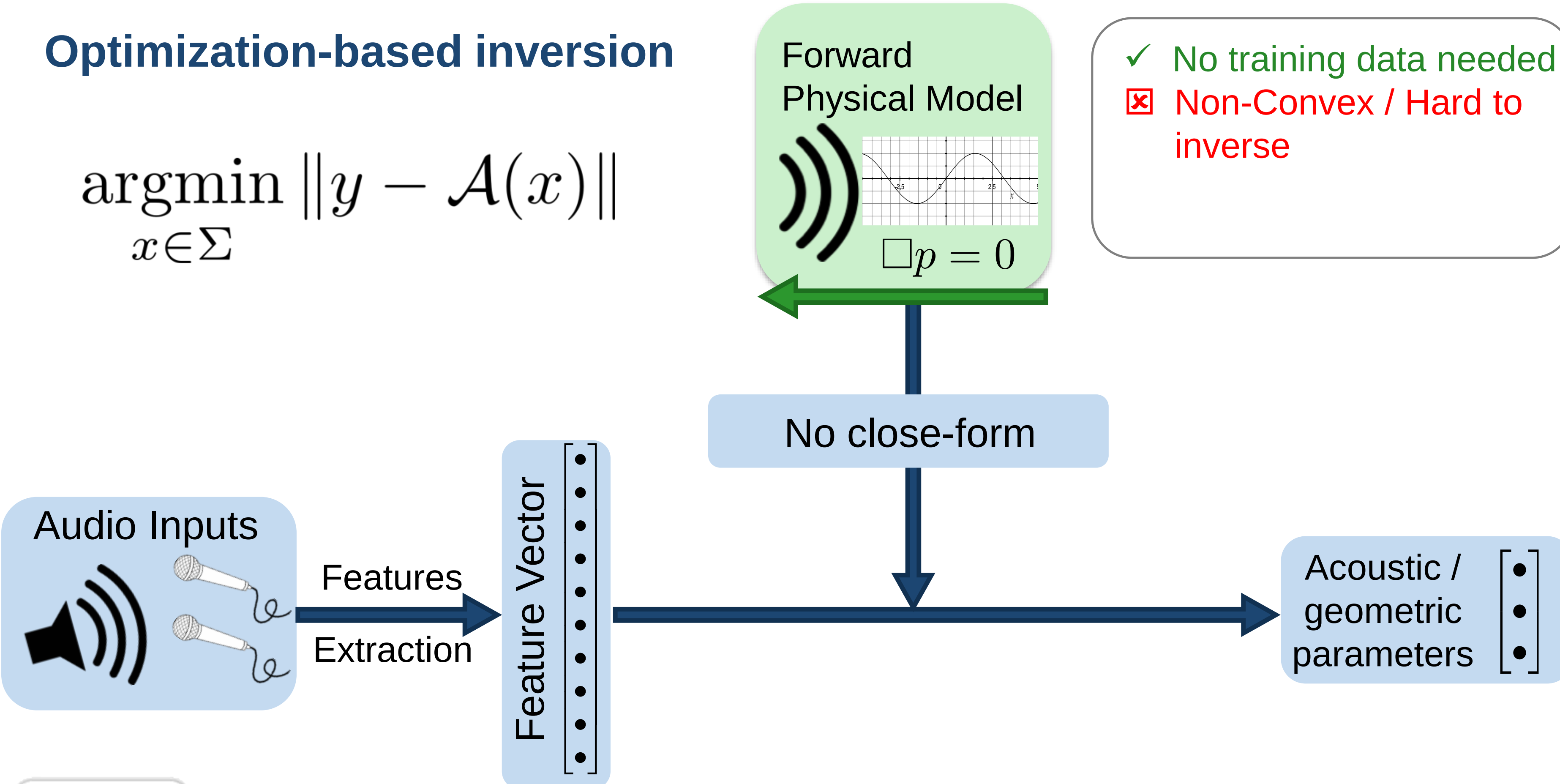
Present time, Inria Nancy

a) Physics-Driven / Traditional Approaches

Optimization-based inversion

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

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



Present time, Inria Nancy

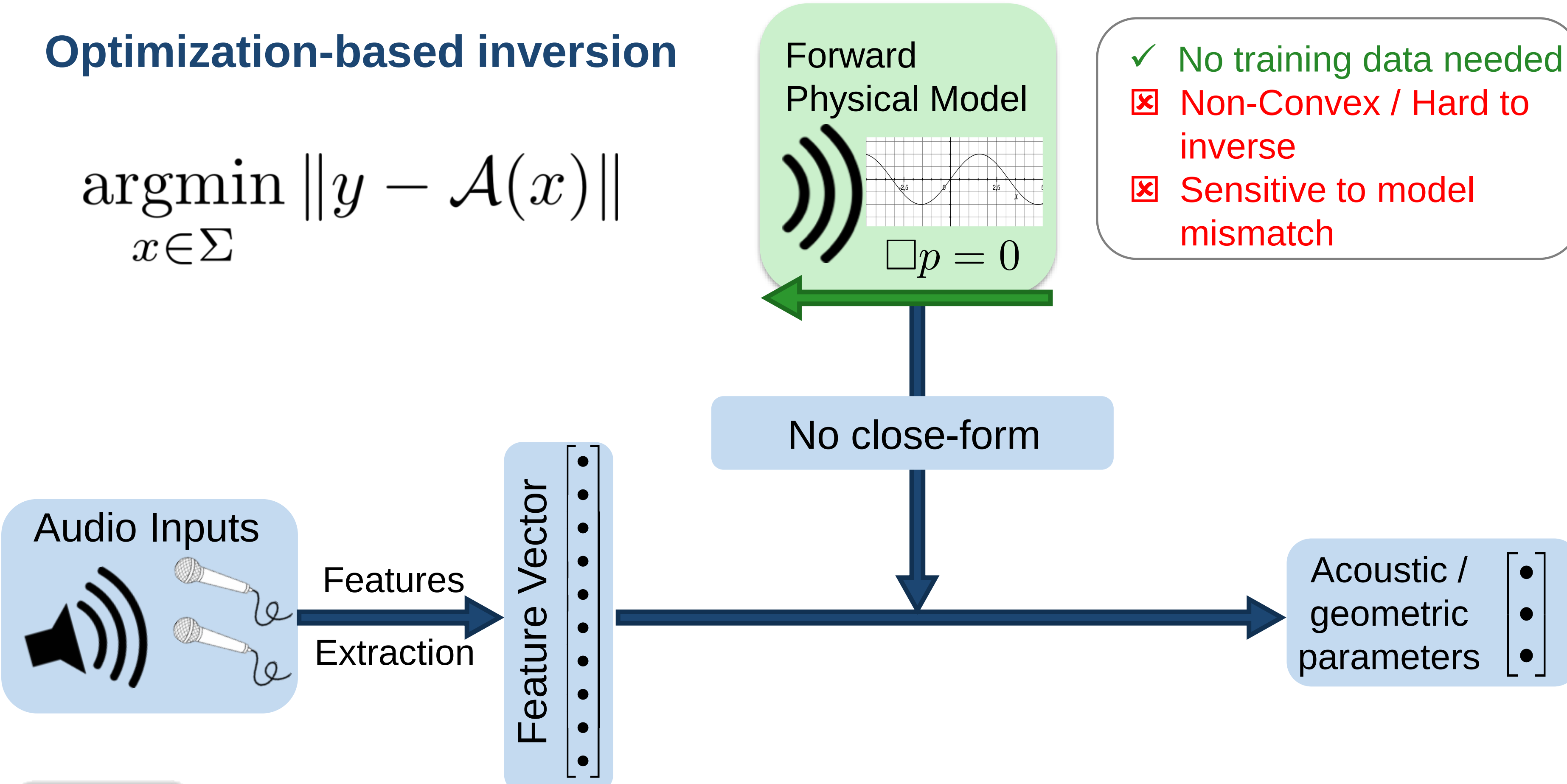
2018–2022

a) Physics-Driven / Traditional Approaches

Optimization-based inversion

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

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



Present time, Inria Nancy

2018–2022

b) Real-Data-Driven Approaches

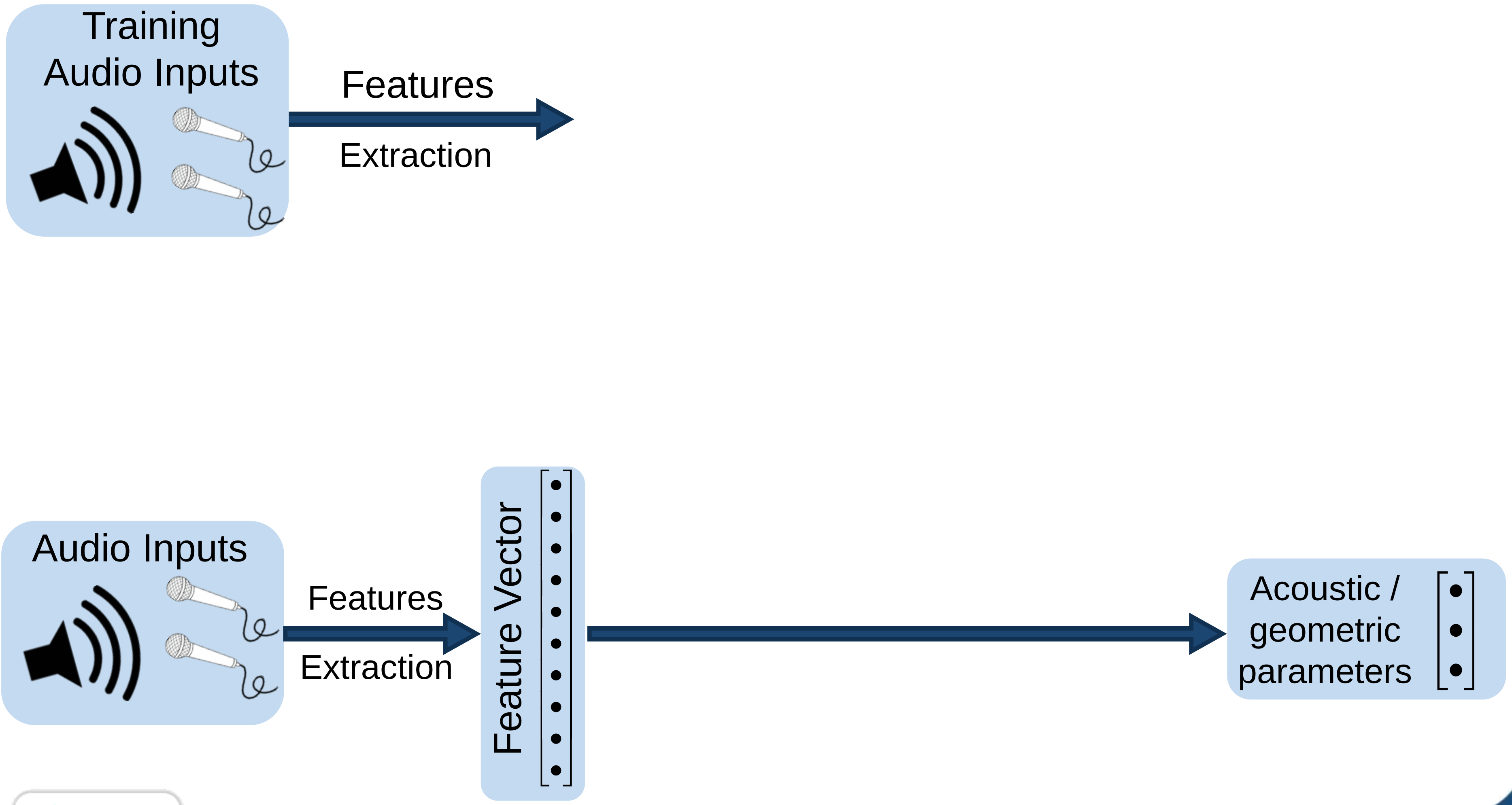
Present time, Inria Nancy

b) Real-Data-Driven Approaches

Present time, Inria Nancy

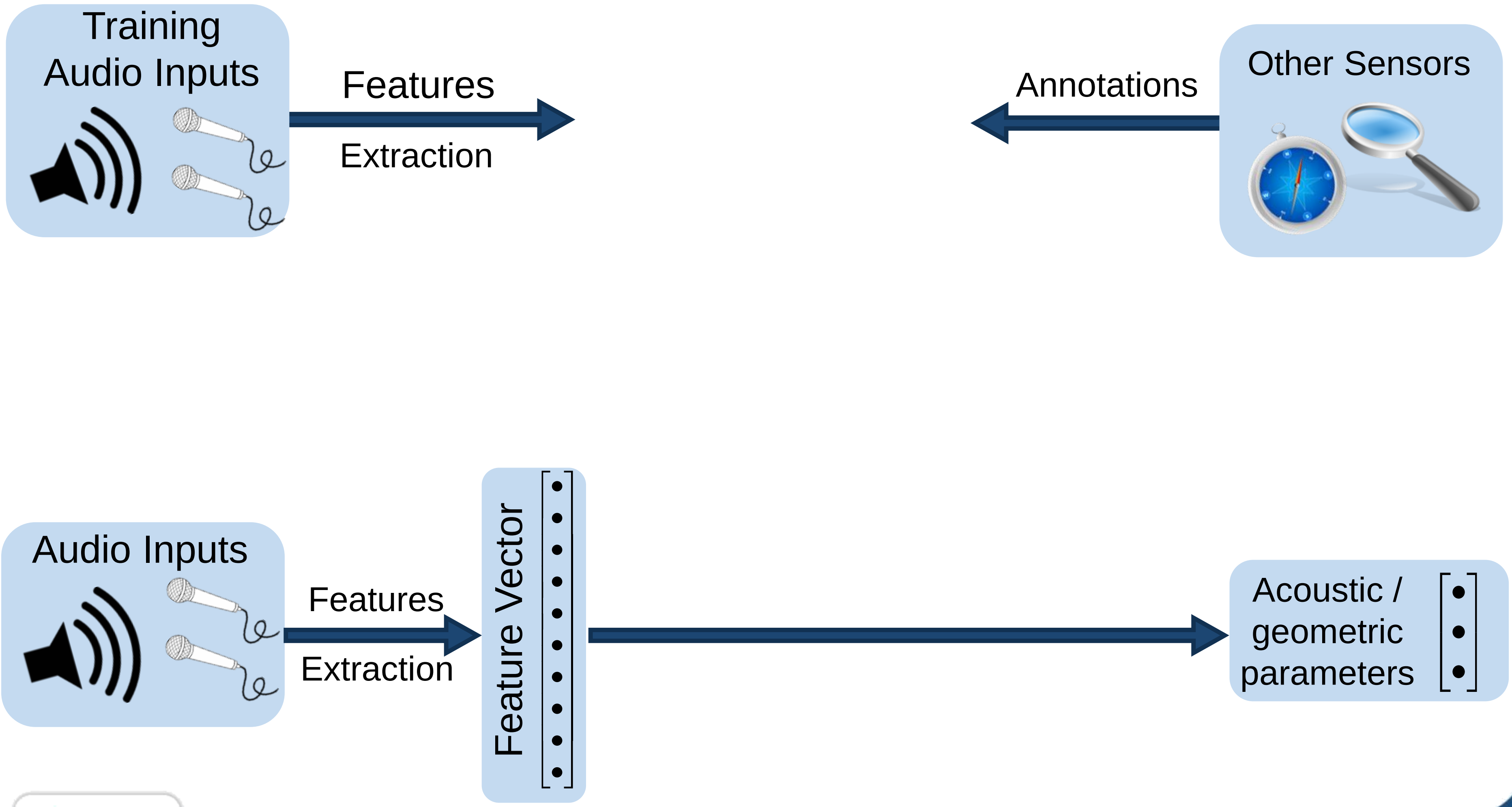
2018–2022

b) Real-Data-Driven Approaches



Present time, Inria Nancy

b) Real-Data-Driven Approaches

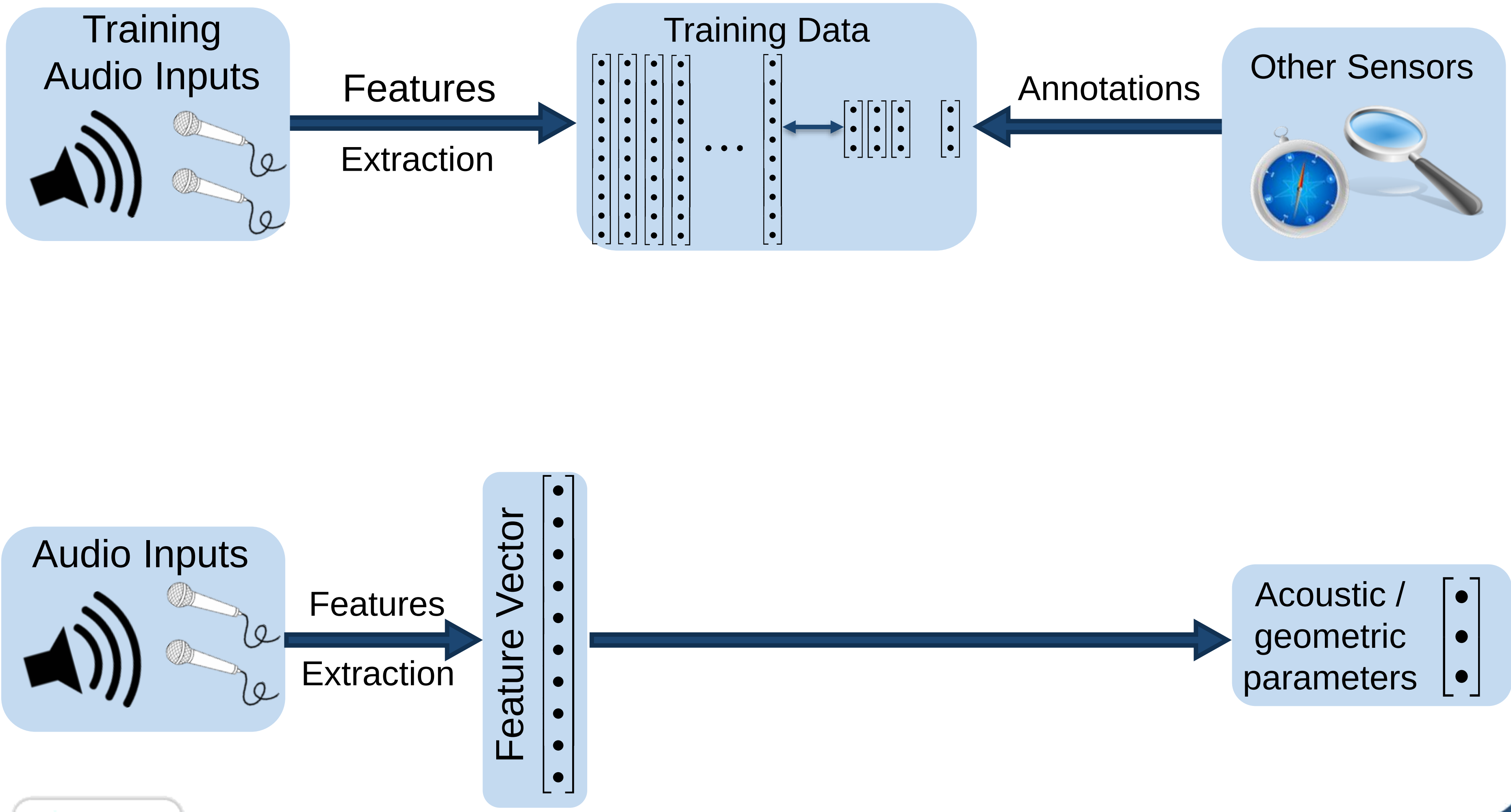


Antoine.Deleforge@inria.fr

Present time, Inria Nancy

2018–2022

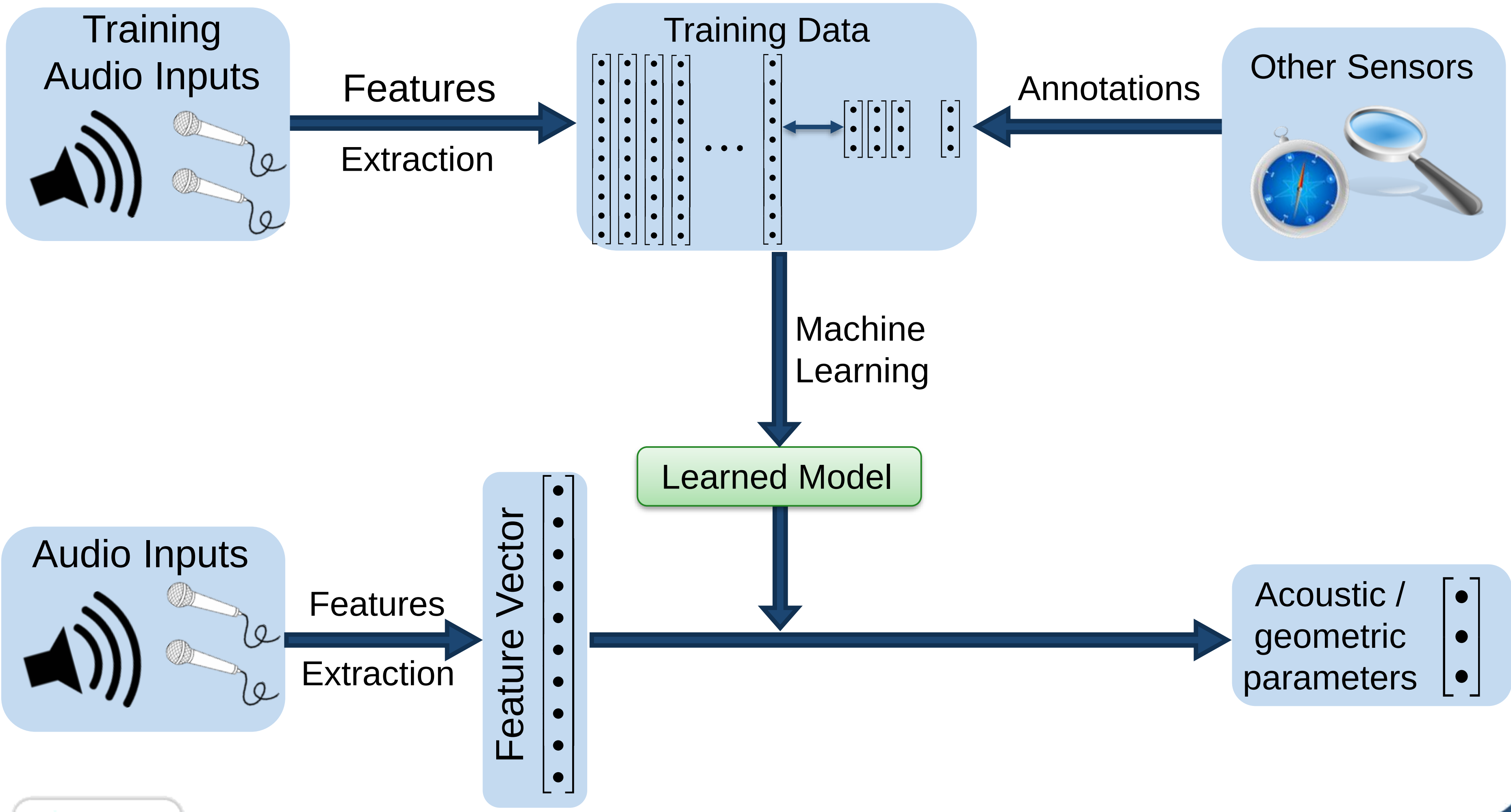
b) Real-Data-Driven Approaches



Present time, Inria Nancy

2018–2022

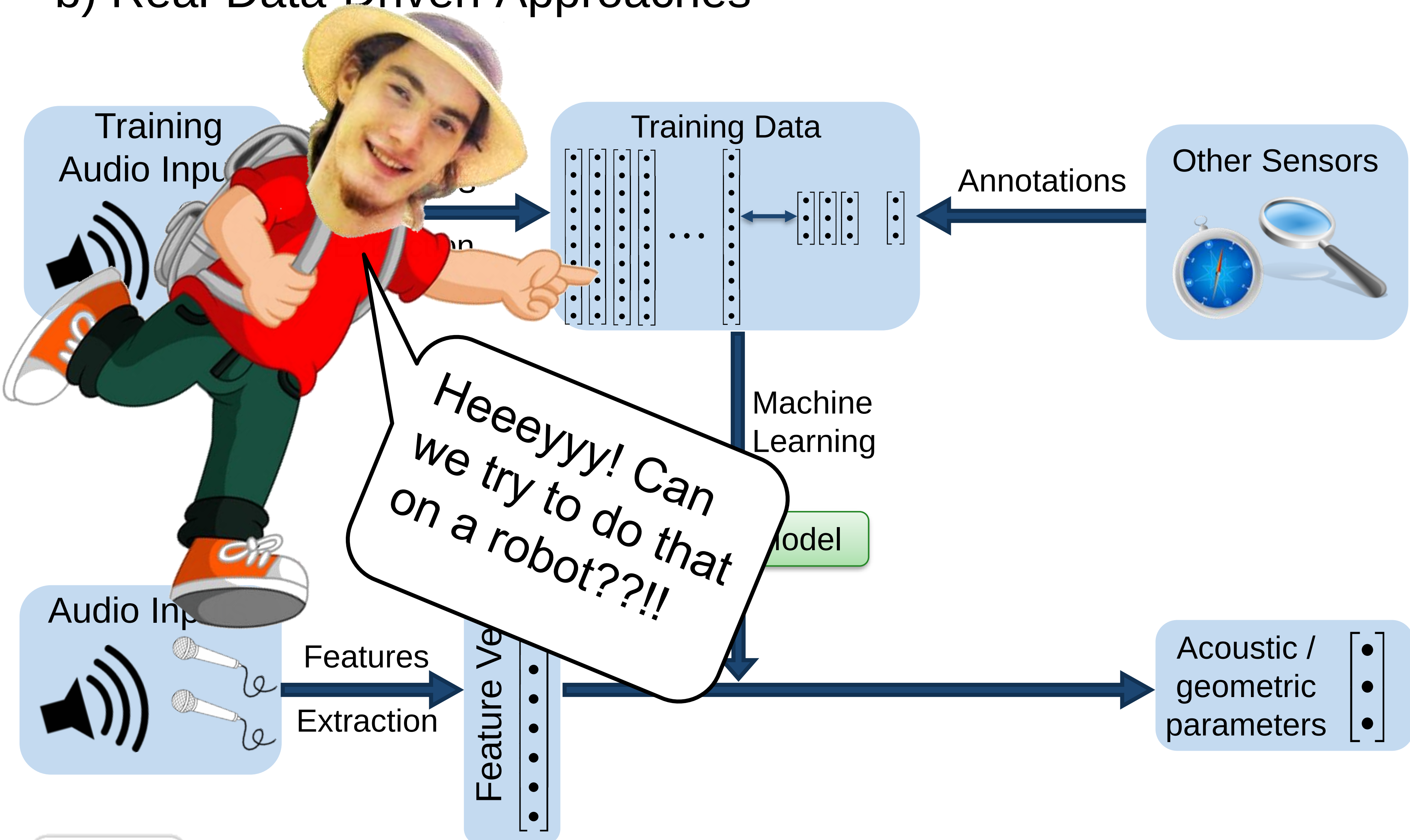
b) Real-Data-Driven Approaches



Present time, Inria Nancy

2018–2022

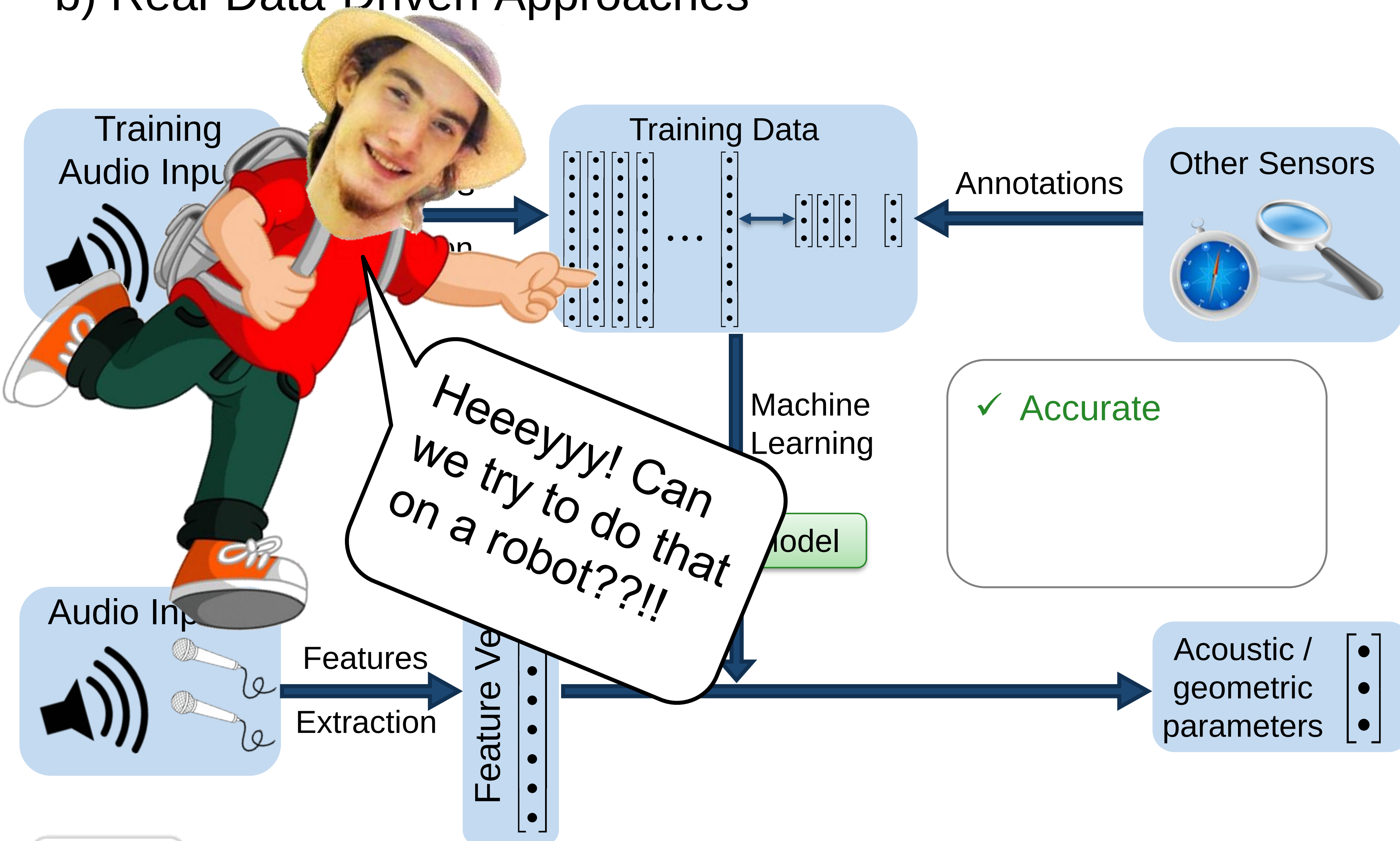
b) Real-Data-Driven Approaches



Present time, Inria Nancy

2018–2022

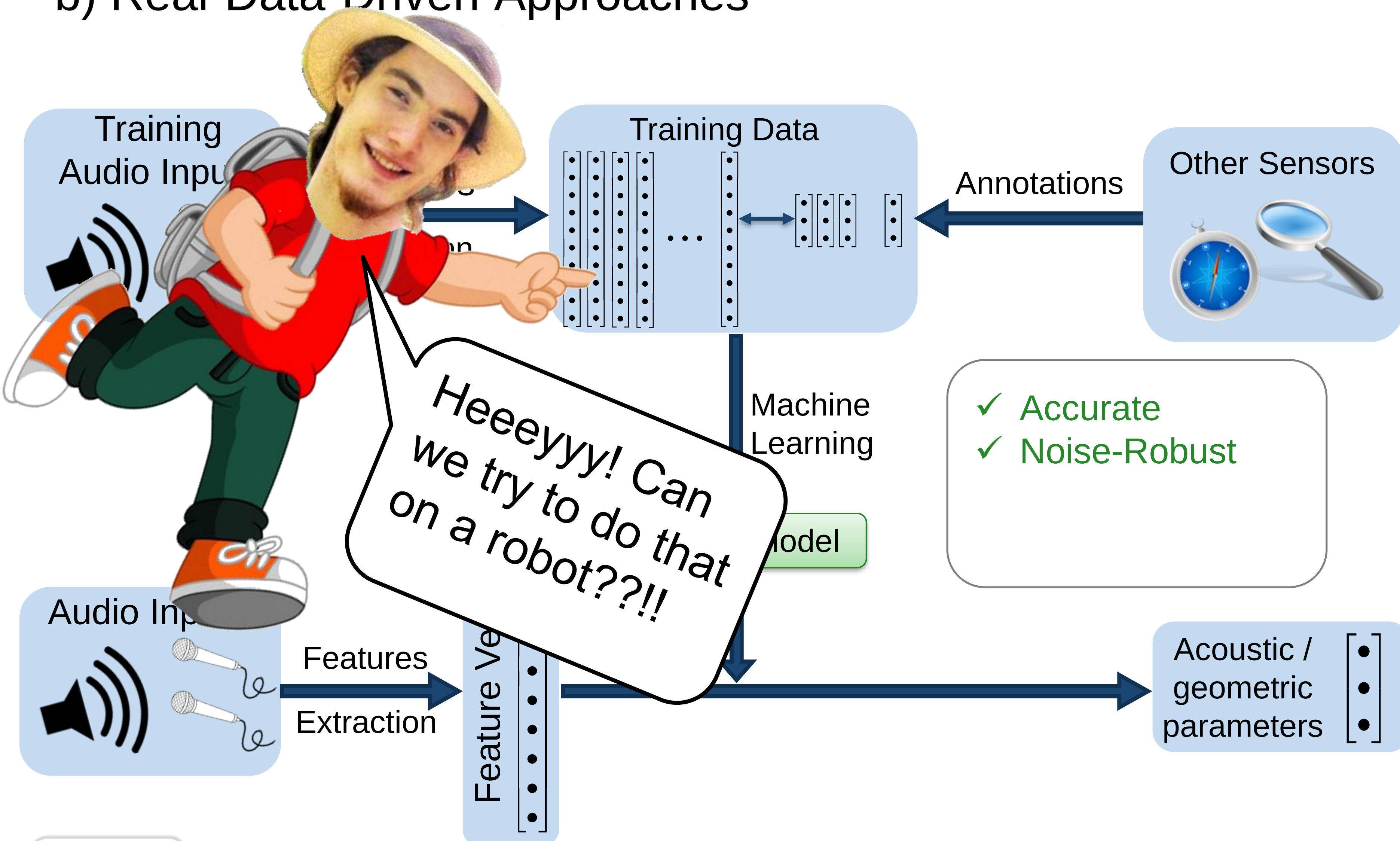
b) Real-Data-Driven Approaches



Present time, Inria Nancy

2018–2022

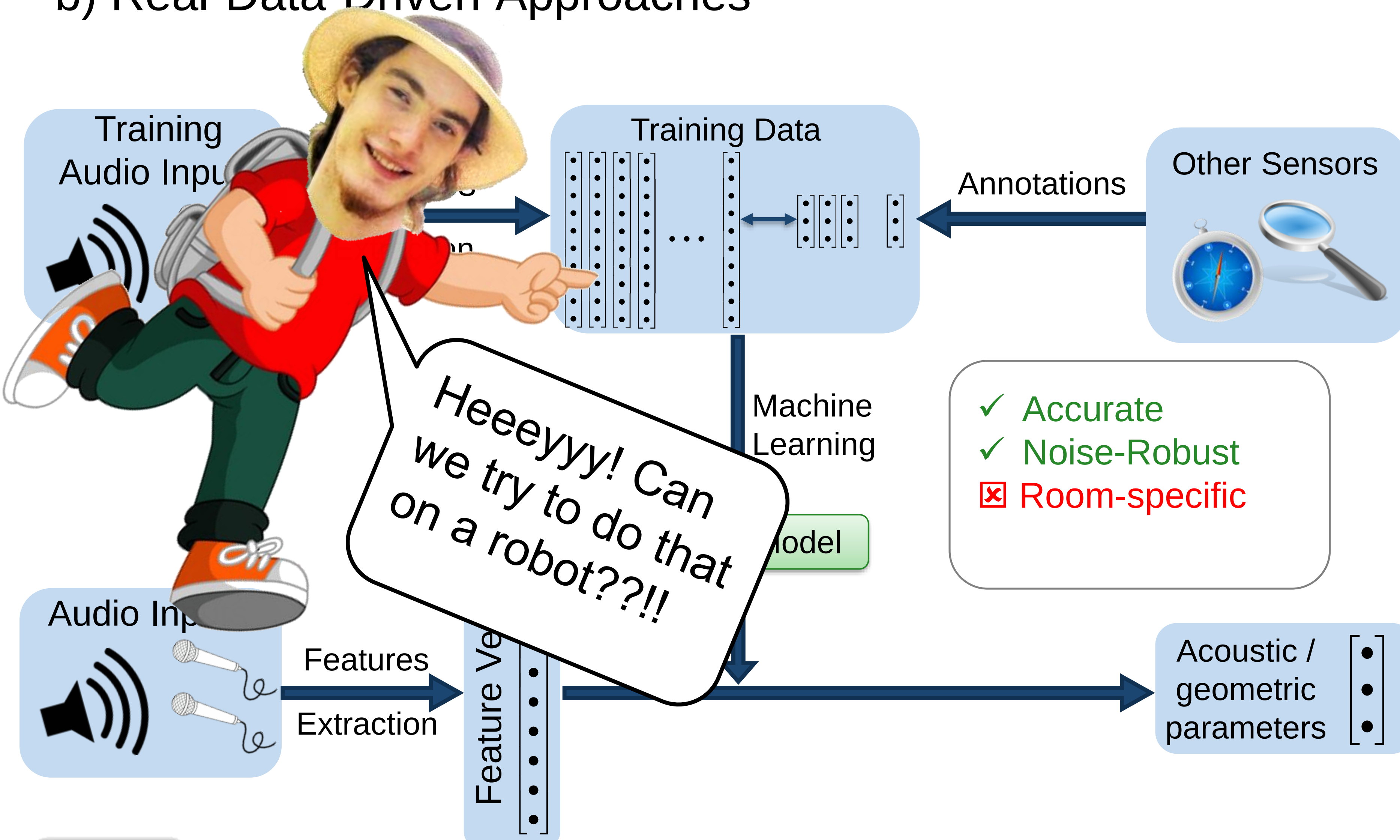
b) Real-Data-Driven Approaches



Present time, Inria Nancy

2018–2022

b) Real-Data-Driven Approaches

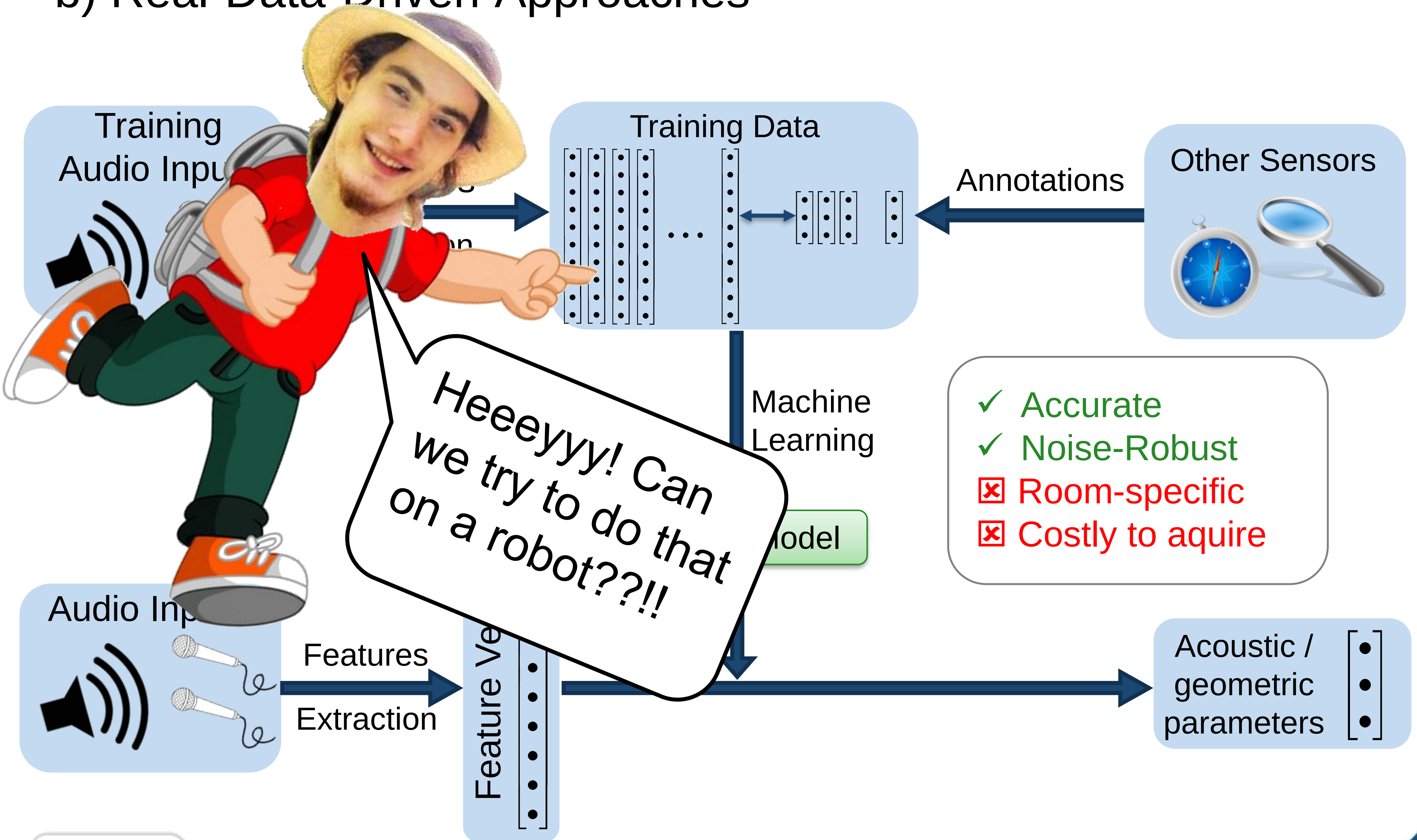


- ✓ Accurate
- ✓ Noise-Robust
- ☒ Room-specific

Present time, Inria Nancy

2018–2022

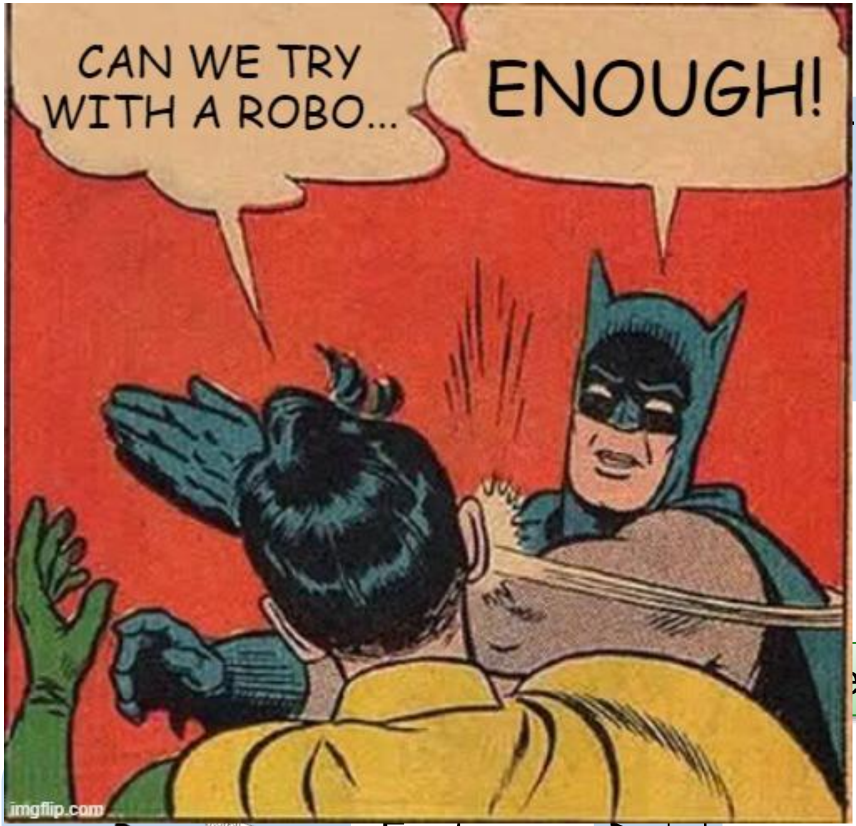
b) Real-Data-Driven Approaches

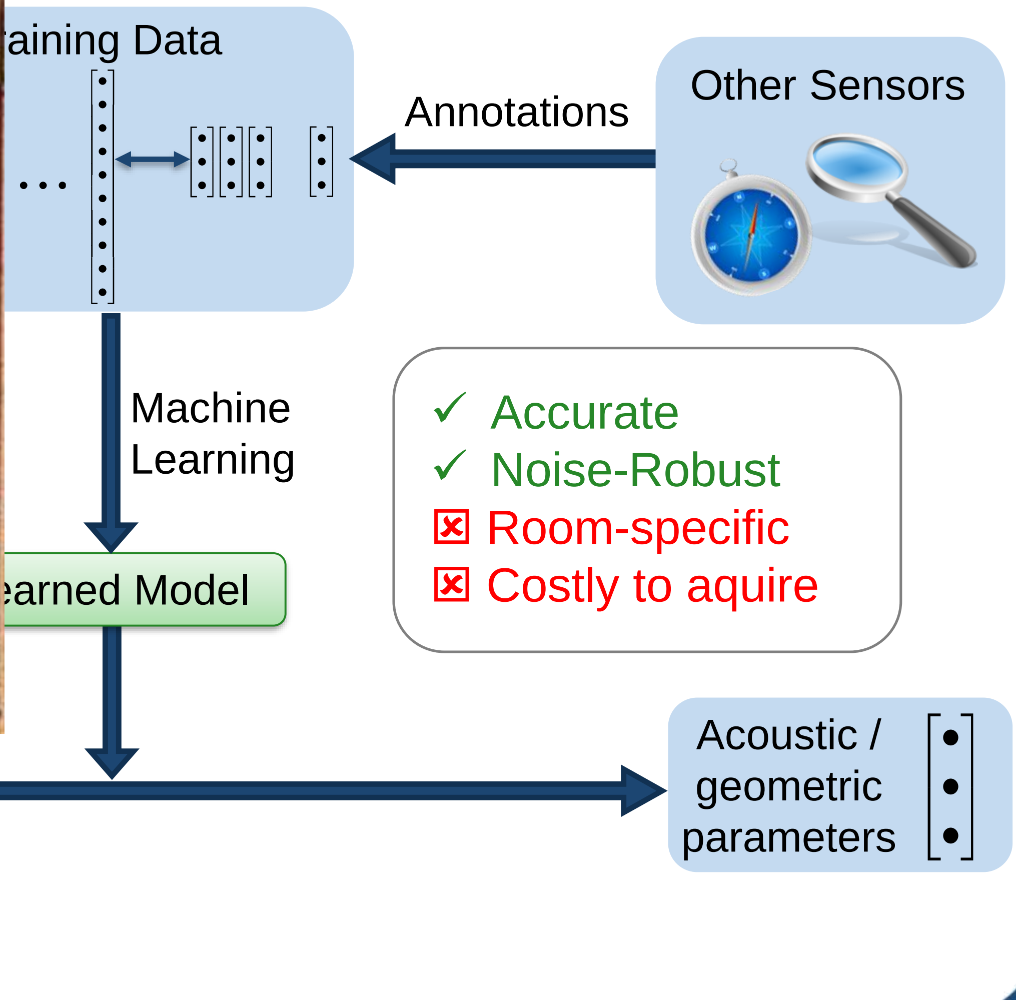


Present time, Inria Nancy

2018–2022

b) Real-Data-Driven Approaches





Present time, Inria Nancy

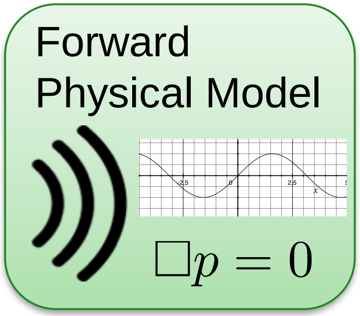
2018–2022

c) ***Virtually Supervised Learning***

Present time, Inria Nancy

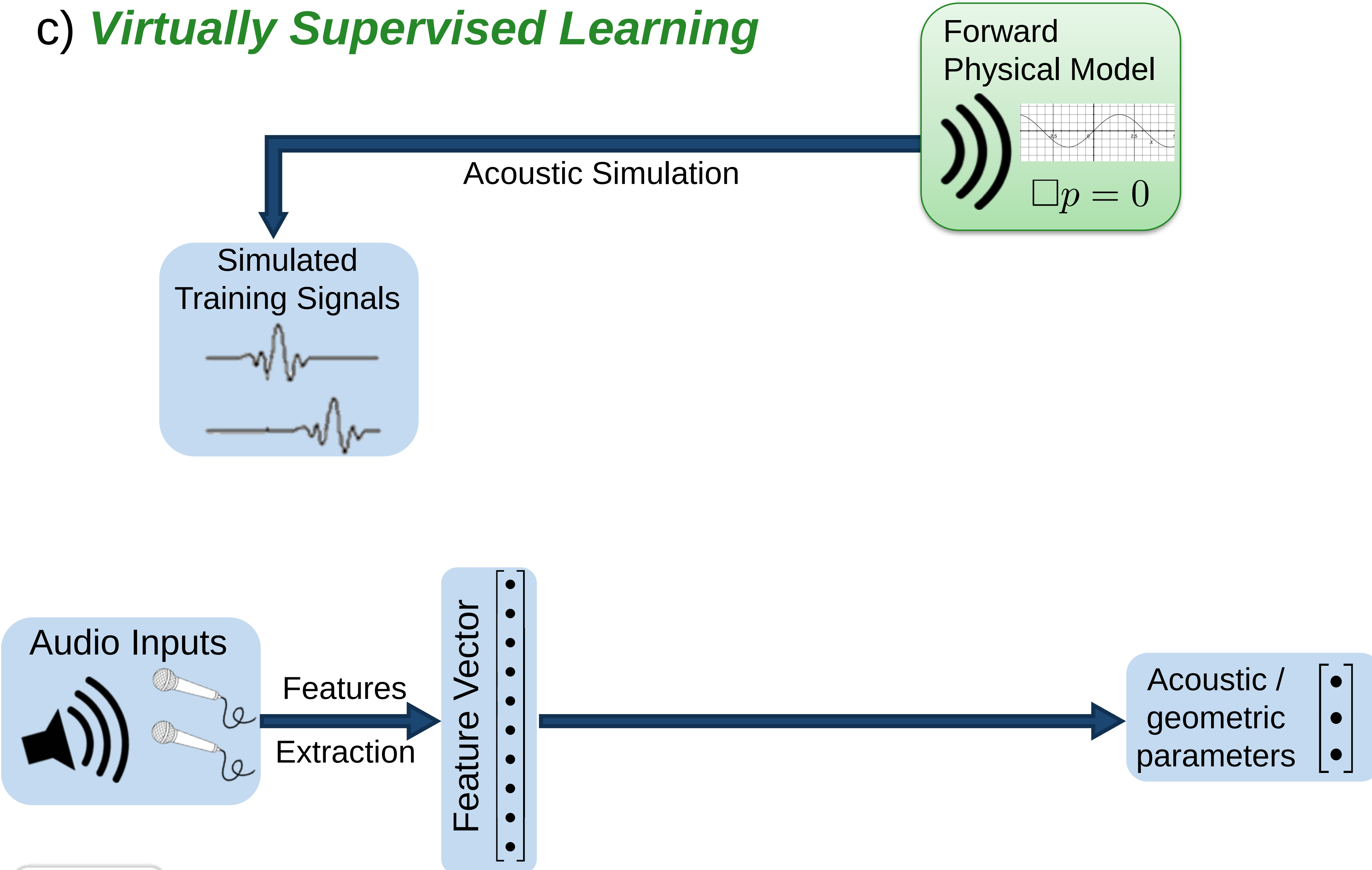
2018–2022

c) ***Virtually Supervised Learning***



Present time, Inria Nancy

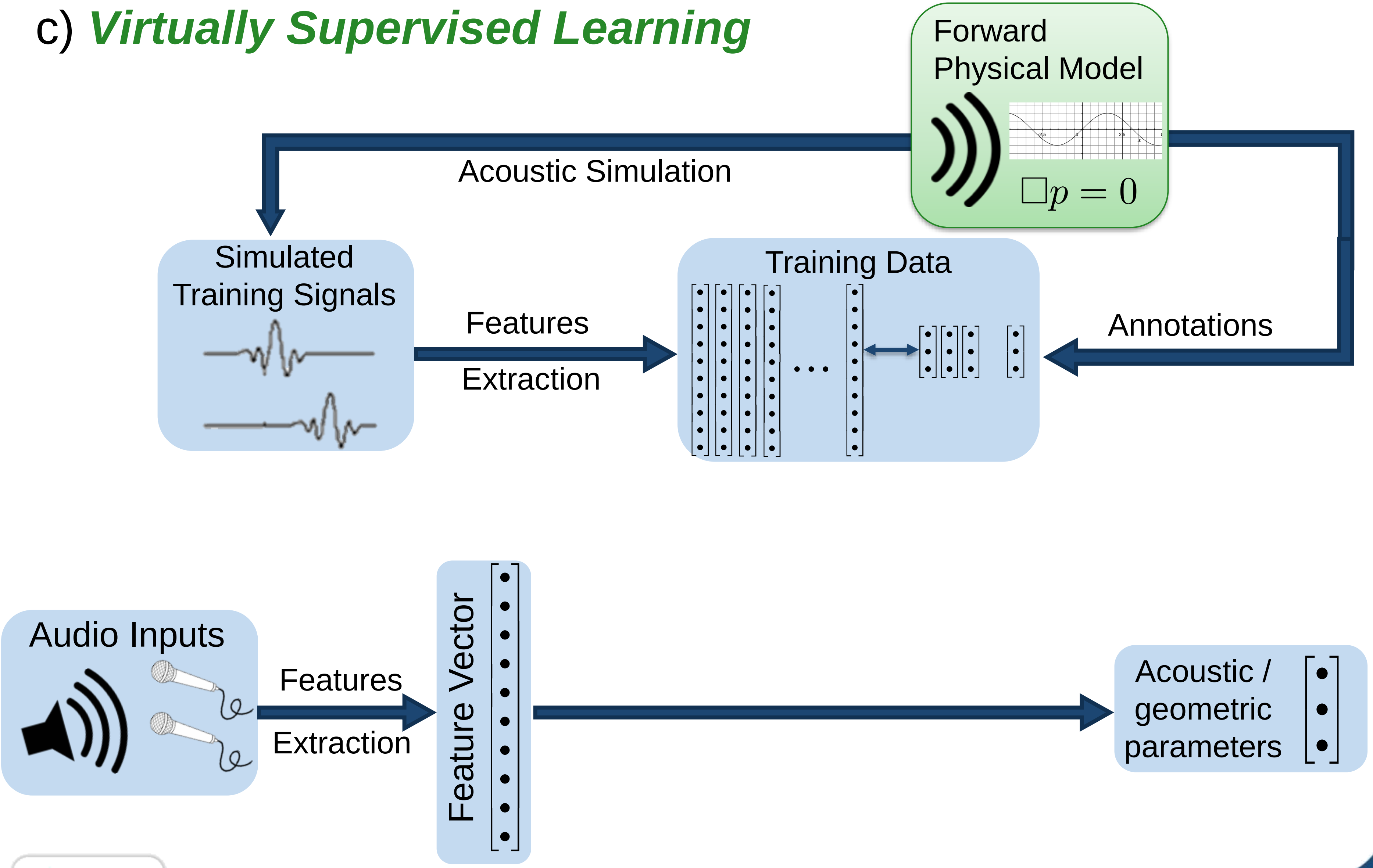
c) *Virtually Supervised Learning*



Present time, Inria Nancy

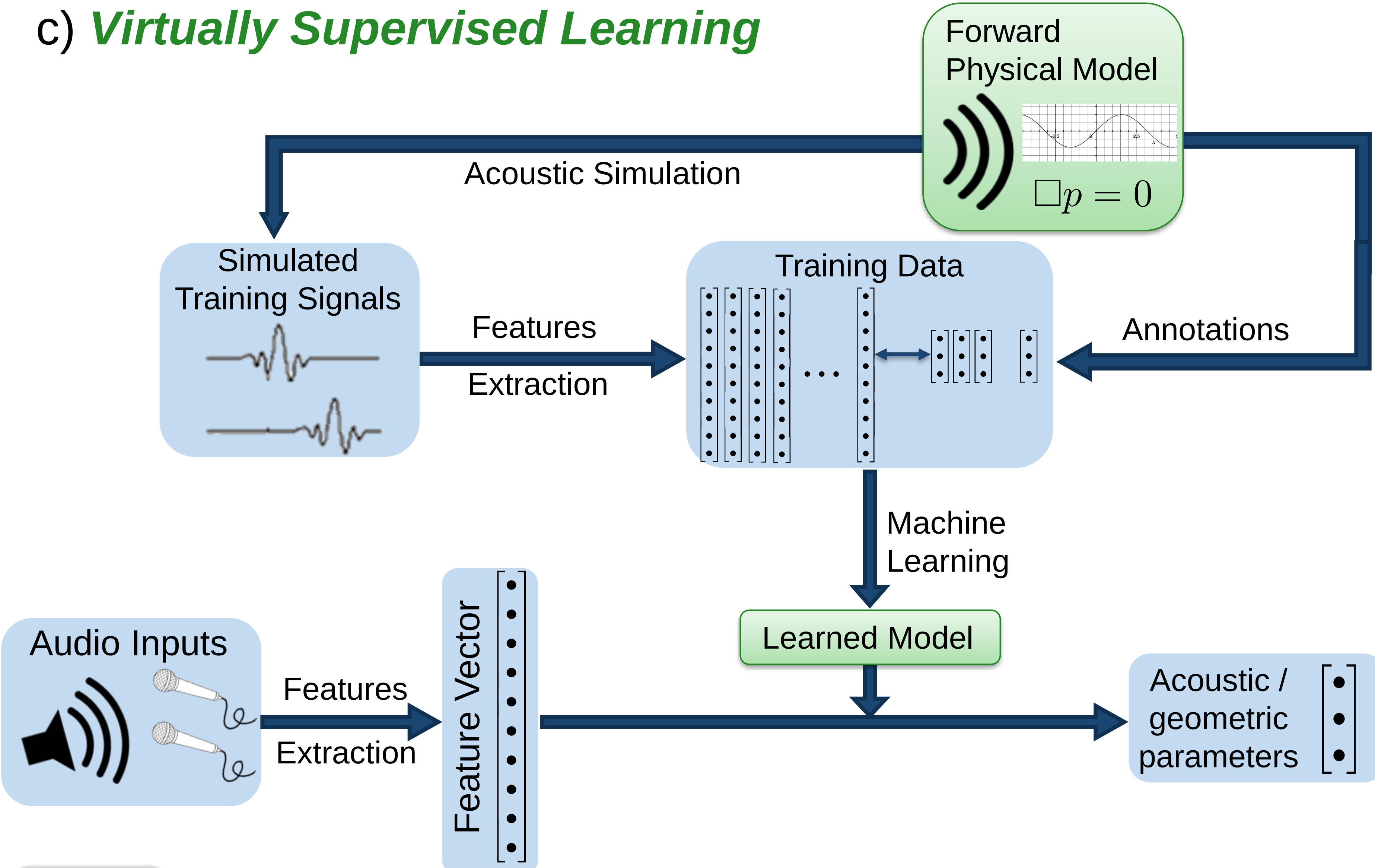
2018–2022

c) *Virtually Supervised Learning*



Present time, Inria Nancy

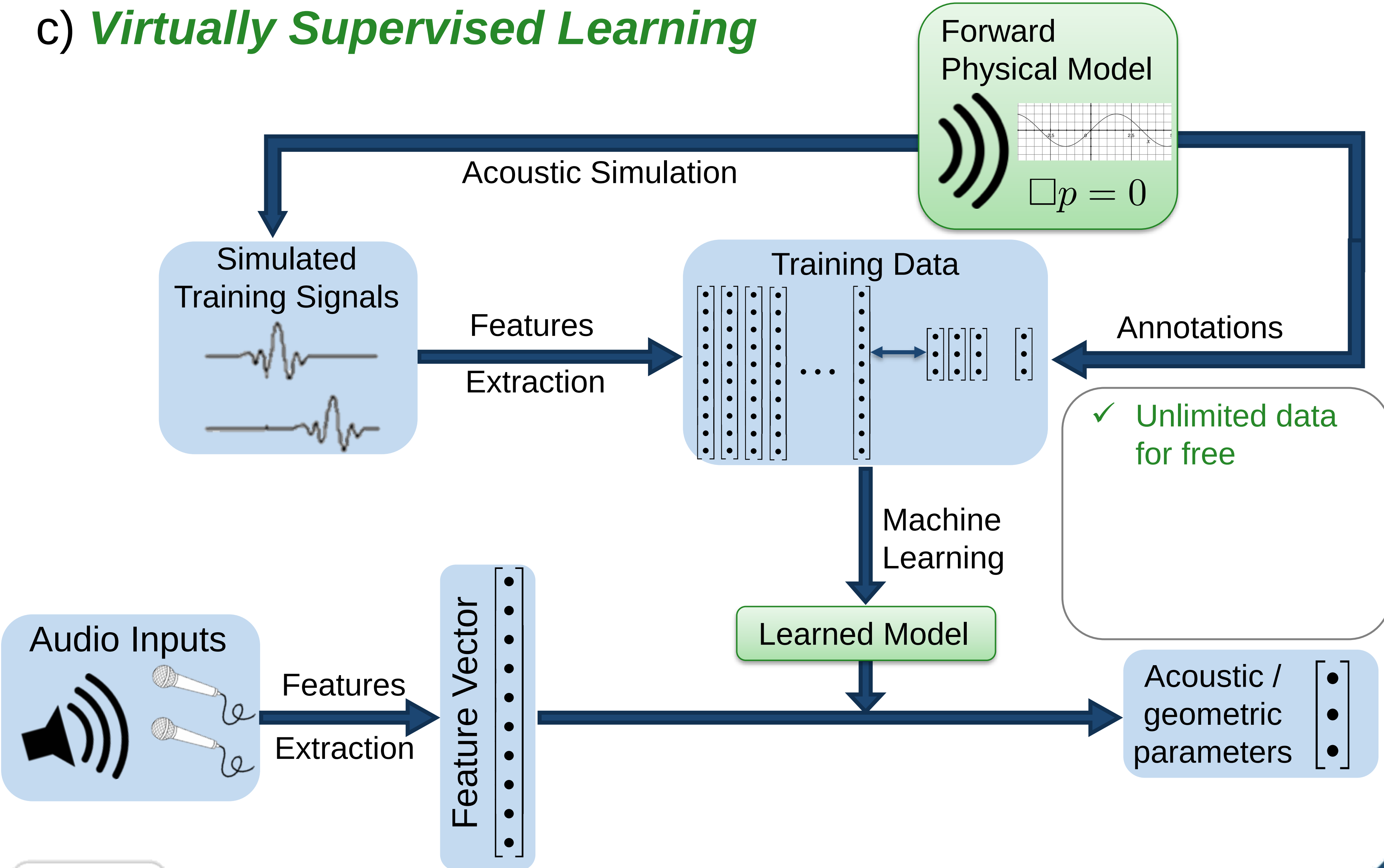
c) *Virtually Supervised Learning*



Present time, Inria Nancy

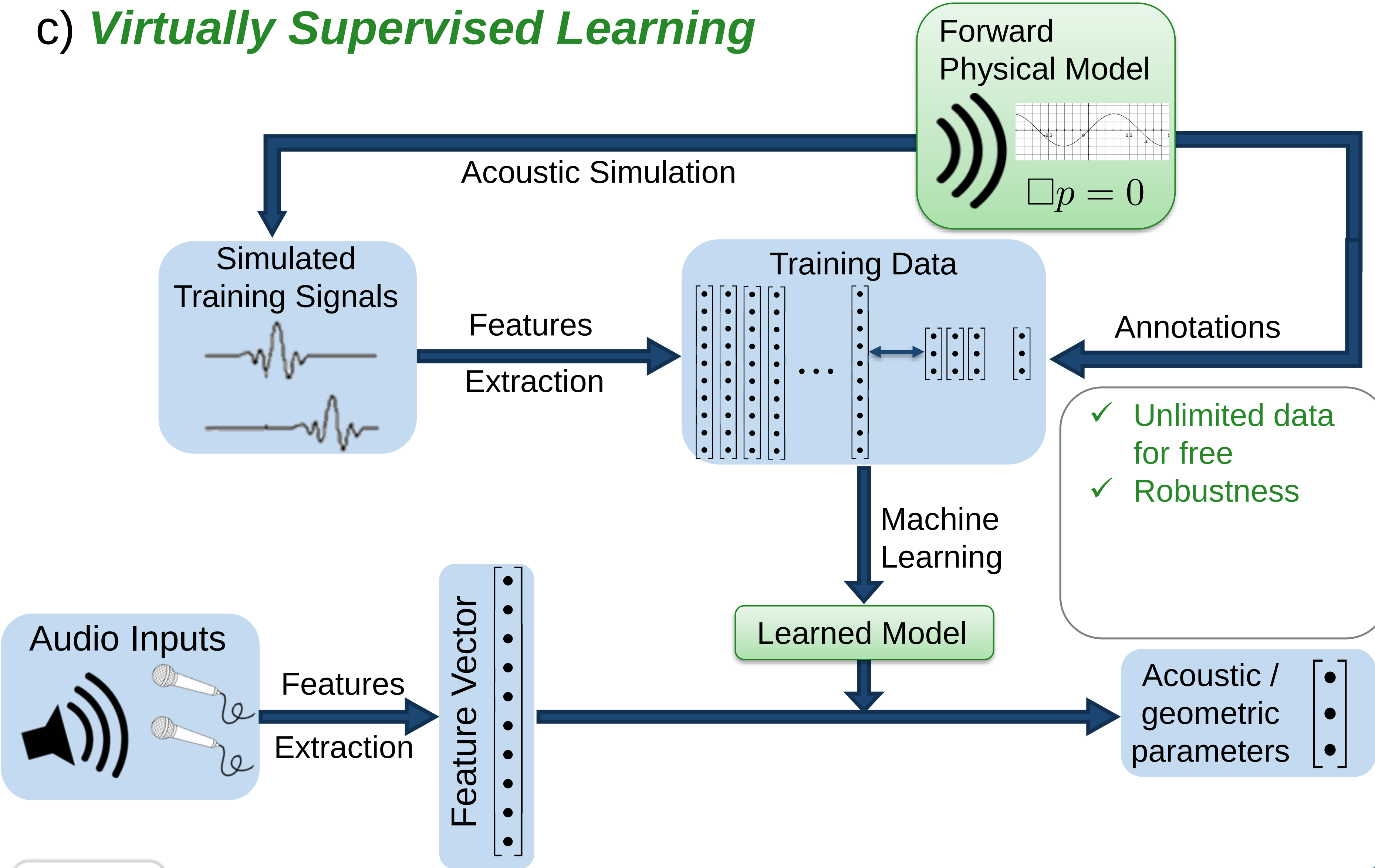
2018–2022

c) *Virtually Supervised Learning*



Present time, Inria Nancy

c) *Virtually Supervised Learning*

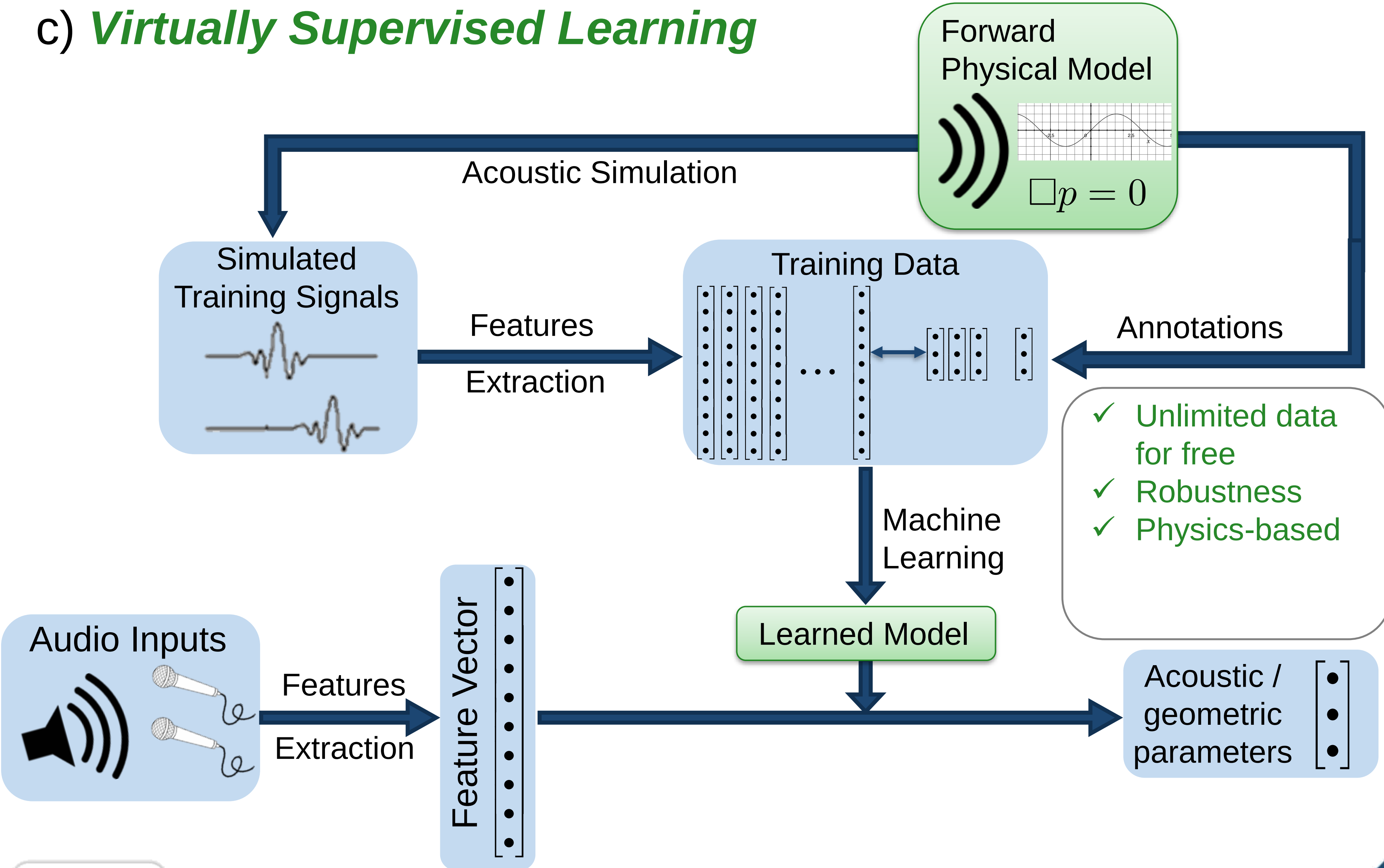


✓ Unlimited data for free
✓ Robustness

Present time, Inria Nancy

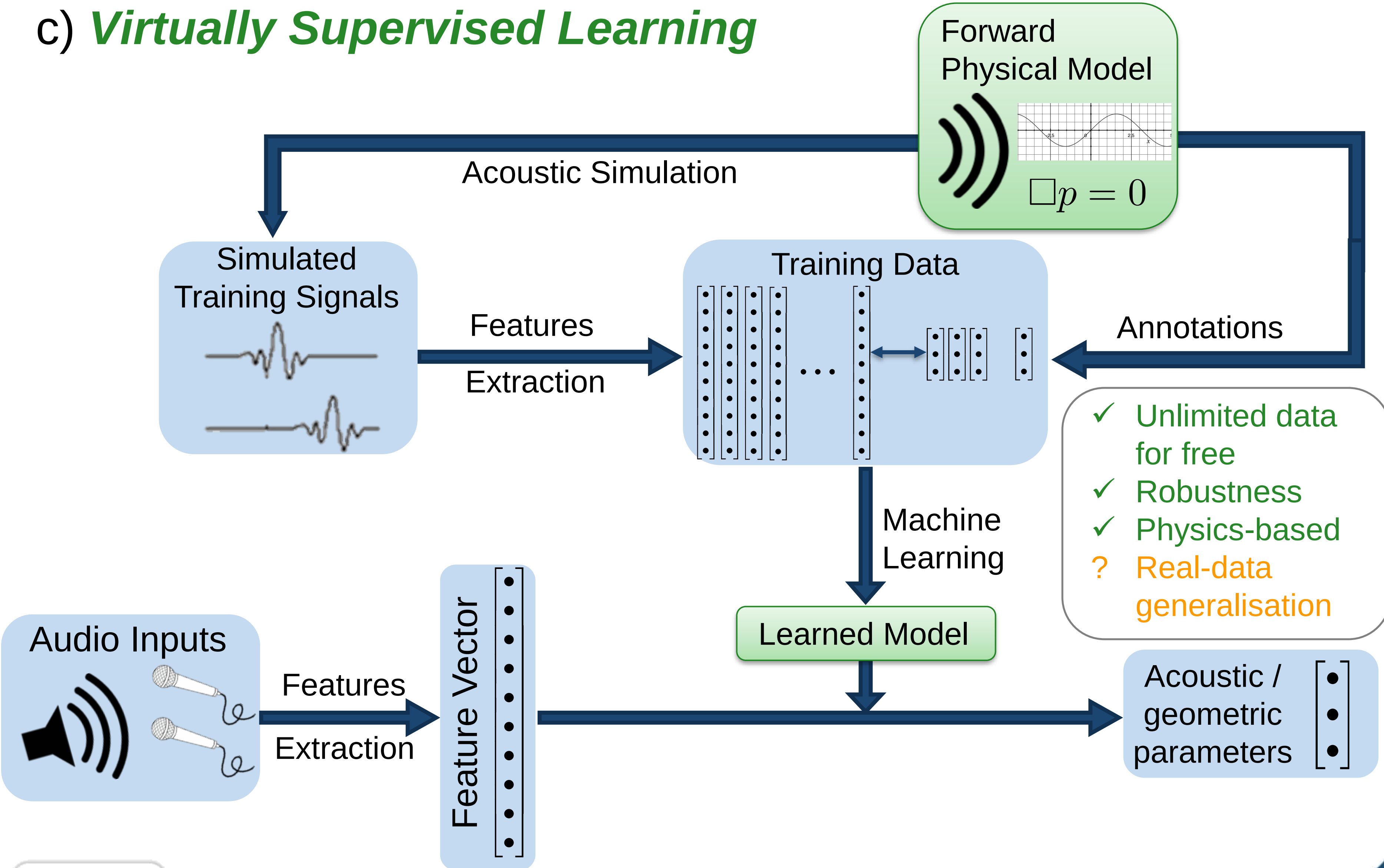
2018–2022

c) *Virtually Supervised Learning*



Present time, Inria Nancy

c) *Virtually Supervised Learning*



A new preprint

Oct. 2022

HOW TO (VIRTUALLY) TRAIN YOUR SOUND SOURCE LOCALIZER

Prerak Srivastava[1], *Antoine Deleforge*[1], *Archontis Politis*[2], *Emmanuel Vincent*[1]

A new preprint

Oct. 2022

HOW TO (VIRTUALLY) TRAIN YOUR SOUND SOURCE LOCALIZER

Prerak Srivastava[1], Antoine Deleforge[1], Archontis Politis[2], Emmanuel Vincent[1]

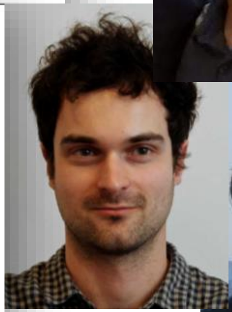
July 2022

JASA REVIEW-ARTICLE

CrossMark

A survey of sound source localization with deep learning methods

Pierre-Amaury Grumiaux,[1,a)] Srđan Kitić,[2] Laurent Girin,[3] and Alexandre Guérin[2]
[1]*Nantes Université, École Centrale Nantes, CNRS, LS2N, 2 chemin de la Houssinière, F-44332 Nantes, France*
[2]*Orange Labs, 4 Rue du Clos Courtel, 35510 Cesson-Sévigné, France*
[3]*Univ. Grenoble Alpes, Grenoble-INP, GIPSA-lab, 11 Rue des Mathématiques, 38400 Saint-Martin-d'Hères, France*



A new preprint

Oct. 2022

HOW

Prerak

JA

A surv
methoc

Pierre-Amau
[1]Nantes Unive
[2]Orange Labs,
[3]Univ. Grenob

							Sources			Data							
										Train				Test			
Author	Year	Architecture	Type	Learn.	Input features	Output	NoS	Kno.	Mov.	SA	RA	SR	RR	SA	RA	SR	RR
Adavanne *et al.* (2021)	2021	CRNN + SA	R	S	FOA Mel spectrograms + intensity + GCC-PHAT	x,y,z	2	✓	✓	✗	✗	✗	✓	✗	✗	✗	✓
Bai *et al.* (2021)	2021	Res. CRNN	R	S	Log-Mel spectrograms + intensity	x, y, z	1	✓	✓	✗	✗	✗	✓	✗	✗	✗	✓
Bianco *et al.* (2021)	2021	VAE	C	SS	RTF	θ	1	✓	✗	✗	✗	✓	✗	✗	✗	✓	✓
Bohlender *et al.* (2021)	2021	CNN/CRNN	C	S	Phase map	θ	1-3	✓	✗	✗	✗	✓	✗	✗	✗	✗	✓
Bologni *et al.* (2021)	2021	CNN	C	S	Waveforms	θ, d	1	✓	✗	✗	✗	✓	✗	✗	✗	✓	✗
Cao *et al.* (2021)	2021	SA	R	S	Log-Mel spectrograms + intensity	x, y, z	0-2	✗	✓	✗	✗	✗	✓	✗	✗	✗	✓
Castellini *et al.* (2021)	2021	MLP	R	S	real + imaginary CPS	x, y	1-3	✓	✗	✓	✗	✗	✗	✗	✗	✗	✓
Diaz-Guerra *et al.* (2021b)	2021	CNN	R	S	SRP-PHAT power map	x, y, z	1	✓	✓	✗	✗	✓	✗	✗	✗	✓	✓
Emmanuel *et al.* (2021)	2021	CNN + SA	R	S	Log-spectrograms + intensity	ACCDOA	1	✓	✓	✗	✗	✗	✓	✗	✗	✗	✓
Gelderblom *et al.* (2021)	2021	MLP	C/R	S	GCC-PHAT	θ	2	✓	✗	✗	✗	✓	✗	✗	✗	✗	✓
Gonçalves Pinto *et al.* (2021)	2021	CNN	R	S	Magnitude CPS	x, y	1-10	✗	✗	✓	✗	✗	✗	✓	✗	✗	✗
Grumiaux *et al.* (2021a)	2021	CRNN	C	S	Intensity	θ, ϕ	1-3	✓	✓	✗	✗	✓	✗	✗	✗	✓	✓
Grumiaux *et al.* (2021b)	2021	CNN + SA	C	S	Intensity	θ, ϕ	1-3	✓	✗	✗	✗	✓	✗	✗	✗	✓	✓
Guirguis *et al.* (2020)	2021	TCN	R	S	Magnitude + phase spectrograms	x, y, z	1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Hammer *et al.* (2021)	2021	U-net	C	S	Phase map of the RTF between each mic pair	θ	∞	✗	✓	✗	✗	✓	✗	✗	✗	✗	✓
He *et al.* (2021a)	2021	Res. CNN	C	WS	Magnitude + phase spectrograms	θ	1-4	✓/✗	✗	✗	✗	✓	✓	✗	✗	✗	✓
He *et al.* (2021b)	2021	CNN	R	S	Waveforms	x, y, z	1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Huang and Perez (2021)	2021	Res. CNN + SA	R	S	Waveforms	ACCDOA	1	✓	✓	✗	✗	✗	✓	✗	✗	✗	✓
Komatsu *et al.* (2020)	2021	CRNN	R	S	FOA magnitude + phase spectrograms	θ, ϕ	1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Krause *et al.* (2020a)	2021	CNN	R	S	Magnitude + phase spectrograms	x, y, z	1	✓	✗	✗	✗	✓	✗	✗	✗	✓	✗
Krause *et al.* (2020b)	2021	CRNN	R	S	Misc.	θ, ϕ	1	✓	✗	✗	✗	✗	✓	✗	✗	✗	✓
Lee *et al.* (2021a)	2021	U-Net	R	S	SRP power map	x,y	1-3	✗	✗	✓	✗	✗	✗	✗	✓	✗	✗
Lee *et al.* (2021b)	2021	CNN + attention	C	S	Log-Mel spectrograms + intensity	θ	1	✓	✓	✗	✗	✓	✗	✗	✗	✓	✓
Liu *et al.* (2021)	2021	CNN	C	S	Intensity	θ	1	✓	✗	✗	✗	✓	✗	✗	✗	✓	✓
Naranjo-Alcazar *et al.* (2021)	2021	Res. CRNN	R	S	Log-Mel spectrograms + GCC-PHAT	ACCDOA	1	✓	✓	✗	✗	✗	✓	✗	✗	✗	✓
Nguyen *et al.* (2021a)	2021	CRNN	C	S	Intensity/GCC-PHAT	θ, ϕ	1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Nguyen *et al.* (2021b)	2021	CNN + RNN/SA	R	S	Log-spectrograms + DRR + SCM eigenvectors	ACCDOA	1	✓	✓	✗	✗	✗	✓	✗	✗	✗	✓
Park *et al.* (2021a)	2021	SA	R	S	log-Mel spectrograms + intensity	x, y, z	1	✓	✓	✗	✗	✗	✓	✗	✗	✗	✓
Poschadel *et al.* (2021a)	2021	CRNN	C	S	HOA magnitude + phase spectrograms	θ, ϕ	1	✓	✗	✗	✗	✓	✗	✗	✗	✓	✓
Poschadel *et al.* (2021b)	2021	CRNN	C	S	HOA magnitude + phase spectrograms	θ, ϕ	2-3	✓	✗	✗	✗	✓	✗	✗	✗	✓	✓
Pujol *et al.* (2021)	2021	Res. CNN	R	S	Waveforms	θ, ϕ	1	✓	✗	✗	✗	✓	✗	✗	✓	✓	✓
Rho *et al.* (2021)	2021	CRNN + SA	R	S	Log-Mel spectrograms + intensity	θ, ϕ	1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Schymura *et al.* (2021)	2021	CNN + SA	R	S	Magnitude + phase spectrograms	θ, ϕ	1	✓	✗	✓	✗	✓	✓	✓	✗	✓	✓
Schymura *et al.* (2020)	2021	CNN + AE + attent.	R	S	FOA magnitude + phase spectrograms	θ, ϕ	1	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓
Shimada *et al.* (2021)	2021	Res. CRNN + SA	R	S	IPD	ACCDOA	1	✓	✓	✗	✗	✗	✓	✗	✗	✗	✓
Subramanian *et al.* (2021a)	2021	CRNN	C/R	S	Phase spectrogram	θ	2	✓	✗	✗	✗	✓	✗	✗	✗	✓	✗
Subramanian *et al.* (2021b)	2021	CRNN	C		Phase spectrograms, IPD	θ	2	✓	✗	✗	✗	✓	✗	✗	✗	✓	✗
Sudarsanam *et al.* (2021)	2021	SA	R	S	Log-Mel spectrograms + intensity	ACCDOA	1	✓	✓	✗	✗	✗	✓	✗	✗	✗	✓

A new preprint

Oct. 2022

HOW

Prerak

JA

A surv
metho

Pierre-Amau
[1]*Nantes Unive*
[2]*Orange Labs,*
[3]*Univ. Grenob*

							Sources			Data							
										Train				Test			
Author	Year	Architecture	Type	Learn.	Input features	Output	NoS	Kno.	Mov.	SA	RA	SR	RR	SA	RA	SR	RR
Adavanne *et al.* (2021)	2021	CRNN + SA	R	S	FOA Mel spectrograms + intensity + GCC-PHAT	x,y,z	2	✓	✓	✗	✗	✗	✓	✗	✗	✗	✓
Bai *et al.* (2021)	2021	Res. CRNN	R	S	Log-Mel spectrograms + intensity	x, y, z	1	✓	✓	✗	✗	✗	✓	✗	✗	✗	✓
Bianco *et al.* (2021)	2021	VAE	C	SS	RTF	θ	1	✓	✗	✗	✗	✓	✗	✗	✗	✓	✓
Bohlender *et al.* (2021)	2021	CNN/CRNN	C	S	Phase map	θ	1-3	✓	✗	✗	✗	✓	✗	✗	✗	✗	✓
Bologni *et al.* (2021)	2021	CNN	C	S	Waveforms	θ, d	1	✓	✗	✗	✗	✓	✗	✗	✗	✓	✗
Cao *et al.* (2021)	2021	SA	R	S	Log-Mel spectrograms + intensity	x, y, z	0-2	✗	✓	✗	✗	✗	✓	✗	✗	✗	✓
Castellini *et al.* (2021)	2021	MLP	R	S	real + imaginary CPS	x, y	1-3	✓	✗	✓	✗	✗	✗	✗	✗	✗	✓
Diaz-Guerra *et al.* (2021b)	2021	CNN	R	S	SRP-PHAT power map	x, y, z	1	✓	✓	✗	✗	✓	✗	✗	✗	✓	✓
Emmanuel *et al.* (2021)	2021	CNN + SA	R	S	Log-spectrograms + intensity	ACCDOA	1	✓	✓	✗	✗	✗	✓	✗	✗	✗	✓
Gelderblom *et al.* (2021)	2021	MLP	C/R	S	GCC-PHAT	θ	2	✓	✗	✗	✗	✓	✗	✗	✗	✗	✓
Gonçalves Pinto *et al.* (2021)	2021	CNN	R	S	Magnitude CPS	x, y	1-10	✗	✗	✓	✗	✗	✗	✓	✗	✗	✗
Grumiaux *et al.* (2021a)	2021	CRNN	C	S	Intensity	θ, ϕ	1-3	✓	✓	✗	✗	✓	✗	✗	✗	✓	✓
Grumiaux *et al.* (2021b)	2021	CNN + SA	C	S	Intensity	θ, ϕ	1-3	✓	✗	✗	✗	✓	✗	✗	✗	✓	✓
Guirguis *et al.* (2020)	2021	TCN	R	S	Magnitude + phase spectrograms	x, y, z	1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Hammer *et al.* (2021)	2021	U-net	C	S	Phase map of the RTF between each mic pair	θ	∞	✗	✓	✗	✗	✓	✗	✗	✗	✗	✓
He *et al.* (2021a)	2021	Res. CNN	C	WS	Magnitude + phase spectrograms	θ	1-4	✓/✗	✗	✗	✗	✓	✓	✗	✗	✗	✓
He *et al.* (2021b)	2021	CNN	R	S	Waveforms	x, y, z	1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Huang and Perez (2021)	2021	Res. CNN + SA	R	S	Waveforms	ACCDOA	1	✓	✓	✗	✗	✗	✓	✗	✗	✗	✓
Komatsu *et al.* (2020)	2021	CRNN	R	S	FOA magnitude + phase spectrograms	θ, ϕ	1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Krause *et al.* (2020a)	2021	CNN	R	S	Magnitude + phase spectrograms	x, y, z	1	✓	✗	✗	✗	✓	✗	✗	✗	✓	✗
Krause *et al.* (2020b)	2021	CRNN	R	S	Misc.	θ, ϕ	1	✓	✗	✗	✗	✗	✓	✗	✗	✗	✓
Lee *et al.* (2021a)	2021	U-Net	R	S	SRP power map	x,y	1-3	✗	✗	✓	✗	✗	✗	✗	✓	✗	✗
Lee *et al.* (2021b)	2021	CNN + attention	C	S	Log-Mel spectrograms + intensity	θ	1	✓	✓	✗	✗	✓	✗	✗	✗	✓	✓
Liu *et al.* (2021)	2021	CNN	C	S	Intensity	θ	1	✓	✗	✗	✗	✓	✗	✗	✗	✓	✓
Naranjo-Alcazar *et al.* (2021)	2021	Res. CRNN	R	S	Log-Mel spectrograms + GCC-PHAT	ACCDOA	1	✓	✓	✗	✗	✗	✓	✗	✗	✗	✓
Nguyen *et al.* (2021a)	2021	CRNN	C	S	Intensity/GCC-PHAT	θ, ϕ	1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Nguyen *et al.* (2021b)	2021	CNN + RNN/SA	R	S	Log-spectrograms + DRR + SCM eigenvectors	ACCDOA	1	✓	✓	✗	✗	✗	✓	✗	✗	✗	✓
Park *et al.* (2021a)	2021	SA	R	S	log-Mel spectrograms + intensity	x, y, z	1	✓	✓	✗	✗	✗	✓	✗	✗	✗	✓
Poschadel *et al.* (2021a)	2021	CRNN	C	S	HOA magnitude + phase spectrograms	θ, ϕ	1	✓	✗	✗	✗	✓	✗	✗	✗	✓	✓
Poschadel *et al.* (2021b)	2021	CRNN	C	S	HOA magnitude + phase spectrograms	θ, ϕ	2-3	✓	✗	✗	✗	✓	✗	✗	✗	✓	✓
Pujol *et al.* (2021)	2021	Res. CNN	R	S	Waveforms	θ, ϕ	1	✓	✗	✗	✗	✓	✗	✗	✓	✓	✓
Rho *et al.* (2021)	2021	CRNN + SA	R	S	Log-Mel spectrograms + intensity	θ, ϕ	1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Schymura *et al.* (2021)	2021	CNN + SA	R	S	Magnitude + phase spectrograms	θ, ϕ	1	✓	✗	✓	✗	✓	✓	✓	✗	✓	✓
Schymura *et al.* (2020)	2021	CNN + AE + attent.	R	S	FOA magnitude + phase spectrograms	θ, ϕ	1	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓
Shimada *et al.* (2021)	2021	Res. CRNN + SA	R	S	IPD	ACCDOA	1	✓	✓	✗	✗	✗	✓	✗	✗	✗	✓
Subramanian *et al.* (2021a)	2021	CRNN	C/R	S	Phase spectrogram	θ	2	✓	✗	✗	✗	✓	✗	✗	✗	✓	✗
Subramanian *et al.* (2021b)	2021	CRNN	C		Phase spectrograms, IPD	θ	2	✓	✗	✗	✗	✓	✗	✗	✗	✓	✗
Sudarsanam *et al.* (2021)	2021	SA	R	S	Log-Mel spectrograms + intensity	ACCDOA	1	✓	✓	✗	✗	✗	✓	✗	✗	✗	✓

A new preprint

Oct. 2022

3 layers of **acoustic simulation realism** at train time

Wall Realism

Mic. Realism

Source Realism

A new preprint

Oct. 2022

3 layers of **acoustic simulation realism** at train time

Wall Realism

Mic. Realism

Source Realism

Naive: identical, frequency-independent walls

A new preprint

3 layers of **acoustic simulation realism** at train time

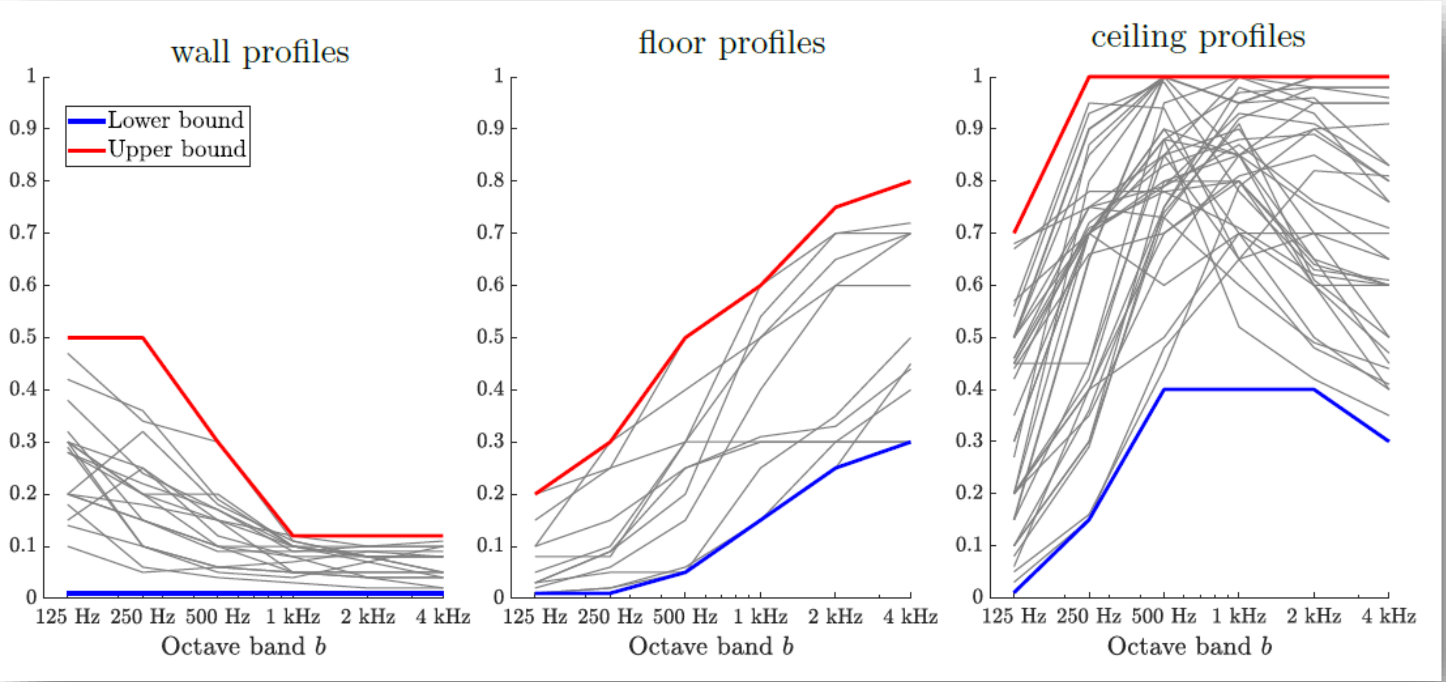
Wall Realism

Mic. Realism

Source Realism

Naive: identical, frequency-independent walls

Advanced: Based on real material absorption databases



A new preprint

Oct. 2022

3 layers of **acoustic simulation realism** at train time

Wall Realism

Mic. Realism

Source Realism

Naive: identical, frequency-independent walls

Advanced: Based on real material absorption databases

Naive: omnidirectional, frequency-independent microphones and sources

A new preprint

Oct. 2022

3 layers of **acoustic simulation realism** at train time

Wall Realism

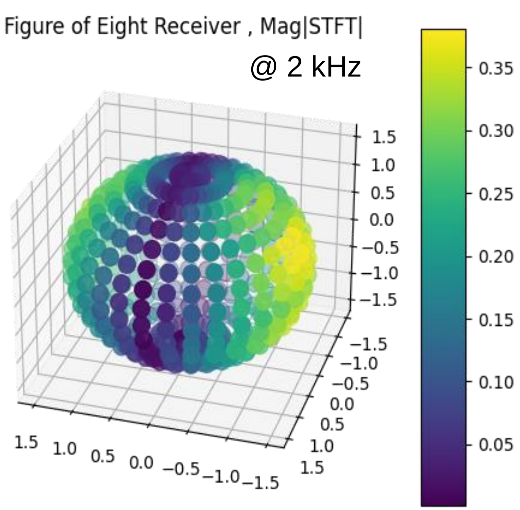
Mic. Realism

Source Realism

Naive: identical, frequency-independent walls

Naive: omnidirectional, frequency-independent microphones and sources

Advanced: Based on real material absorption databases



Advanced: Based on real measured directivity profiles

A new preprint

Oct. 2022

3 **Real** Test Sets

DIRHA (2017) | **VoiceHome2 (2018)** | **STARSS (2022)**

- > 20 real Human speakers
- > 15 real rooms (living room, kitchen, classroom)
- 3 arrays of 2 microphones

---

95 minutes of speech annotated with sound source direction

A new preprint

Oct. 2022

Results

- Advanced simulation significantly improves sound source localization results across all test sets
- Every added layer of realism contributes to these results

Real Test Sets	**VoiceHome-2** [24]		**DIRHA** [25]		**STARS22** [26]	
Methods	$\uparrow$ Recall	$\downarrow$ MAE ($^\circ$)	$\uparrow$ Recall	$\downarrow$ MAE ($^\circ$)	$\uparrow$ Recall	$\downarrow$ MAE ($^\circ$)
SRP-PHAT	70%	9.9 ± 1.5	61%	15.0 ± 2.3	45%	14.9 ± 0.6
Naive Training	78%	7.6 ± 1.2	77%	8.4 ± 1.4	57%	12.9 ± 0.6
Advanced Training	**85%**	$\mathbf{5.8 \pm 0.8}$	**84%**	$\mathbf{6.3 \pm 1.0}$	**61%**	$\mathbf{11.4 \pm 0.5}$
Ablation study						
w/o wall realism	83%	6.2 ± 0.8	81%	7.5 ± 1.4	59%	12.1 ± 0.6
w/o source realism	82%	7.1 ± 1.1	80%	7.8 ± 1.2	**63%**	$\mathbf{11.4 \pm 0.6}$
w/o receiver realism	N/A	N/A	78%	8.3 ± 1.5	53%	13.4 ± 0.6

Looking 9 years back…

Sept. 2013

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Mathmatiques et Informatique**

Arrêté ministériel :

Présentée par

Antoine Deleforge

Thèse dirigée par **Radu Horaud**

préparée au sein **de l'Université Joseph Fourier, de l'INRIA Grenoble Rhône-Alpes**
et de **L'École Doctorale de Mathématiques, Sciences et Technologies de l'Information, Informatique**

Acoustic Space Mapping

A Machine Learning Approach to Sound Source Separation and Localization

Thèse soutenue publiquement le ,
devant le jury composé de :

Pr. Jonathon Chambers
Loughborough University, Rapporteur
Pr. Rémi Gribonval
INRIA Rennes, Rapporteur
Pr. Florence Forbes
INRIA Grenoble Rhône-Alpes, Examinatrice
Pr. Geoff MacLachlan
University of Queensland, Examinateur
Pr. Laurent Girin
GIPSA-lab, Grenoble, Examinateur
Pr. Radu Horaud
INRIA Grenoble Rhône-Alpes, Directeur de thèse

Looking 9 years back…

Sept. 2013

u

THÈSE

Pour obtenir le grade

DOCTEUR DE

Spécialité : **Mathmat**

Arrêté ministériel :

Présentée par

Antoine Delef

Thèse dirigée par **Ra**

préparée au sein **de**
Rhône-Alpes
et de **L'École Docto**
de l'Information, Inf

Acoustic

A Machine
Source Sepa

Thèse soutenue publ
devant le jury compo

Pr. Jonathon Cha
Loughborough Univers
Pr. Rémi Gribonva
INRIA Rennes, Rappor
Pr. Florence Forb
INRIA Grenoble Rhône
Pr. Geoff MacLacl
University of Queensla
Pr. Laurent Girin
GIPSA-lab, Grenoble,
Pr. Radu Horaud
INRIA Grenoble Rhône-Alpes,

is possible. This could include live music recording on a stage or concert hall, speech localization, diarisation or enhancement in a meeting or conference room, or hearing aid devices (the system could be calibrated for a specific wearer).

6.2 Direction for Future Research

Rather than an end, we like to view this thesis as a starting point for fascinating future research topics. We propose here a non-exhaustive list of possible follow-ups.

- An important direction is to study more thoroughly the influence of changes in experimental conditions on binaural manifolds. What happens when changing the position of the recording setup? Moving to another room? What is the influence of the sound source distance and directivity? What happens when the HRTF change? While the PLOM model is a possible direction to improve robustness to such situations, other methods such as *transfer learning* [Pan 10] could be envisioned. A more ambitious idea would be to learn acoustic spaces in virtual environments, using a room simulator such as Roomsim [Campbell 05]. One could imagine learning many different models in different room configurations, *e.g.*, microphones position, room size, reverberations. When dealing with real world data, the most appropriate model could be selected from virtually learned one using, *e.g.*, model se
- In our view, the surprisingly good results obtai
open the doors to a new category of bi
deeper understanding. First of al

What's next?

2023 +

What's next?

2023 +

What's next?

👏

What's next?

2023 +

What’s next?

« What is the shape of the room? »

Antoine.Deleforge@inria.fr

What's next?

« What is the shape of the room? »

« Is the floor made of tiles or carpet? »

What's next?

« What is the shape of the room? »

« Is the floor made of tiles or carpet? »

- Room acoustic diagnosis
- Audio augmented reality
- Echo-aware audio signal enhancement

Antoine.Deleforge@inria.fr

What's next?

« What is the shape of the room? »

« Is the floor made of tiles or carpet? »

- Room acoustic diagnosis
- Audio augmented reality
- Echo-aware audio signal enhancement
- Plenty of interesting open inverse problems

Thank You Radu

for teaching me
the joys of research!

Thank You Radu

for teaching me
the joys of research!