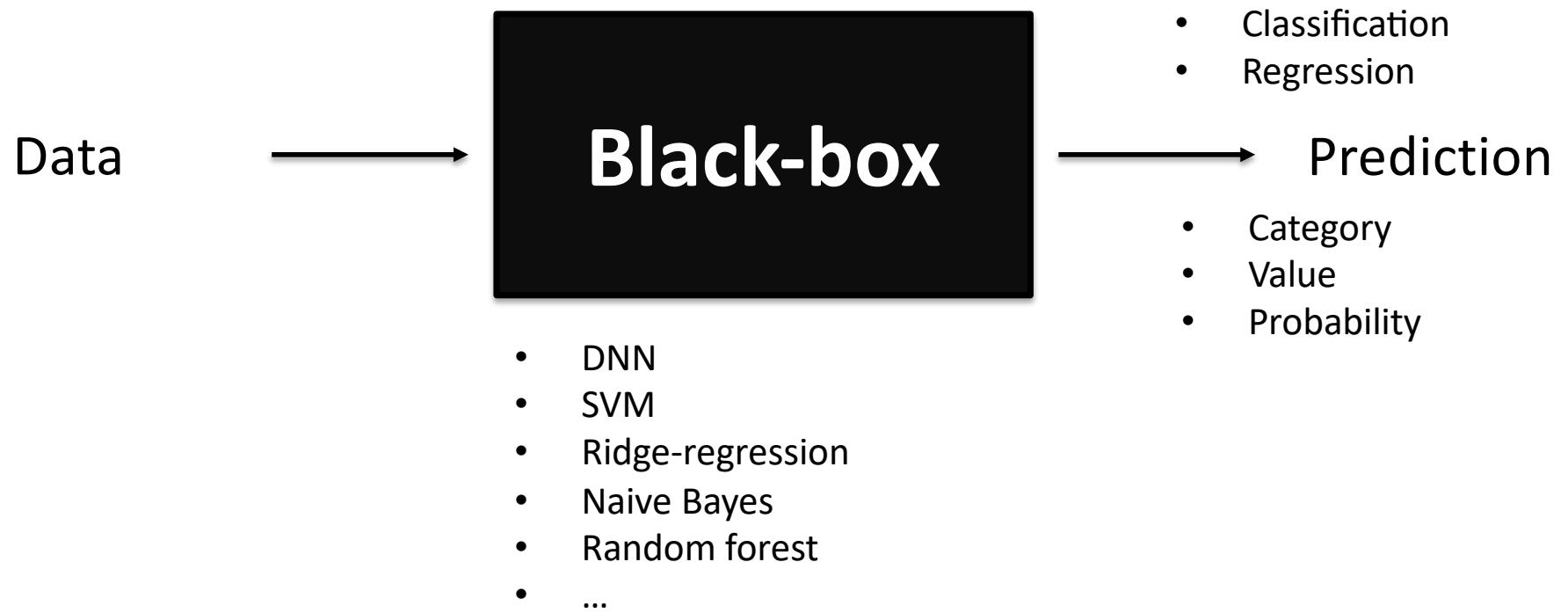


# Explainable AI

xAI

# The problem



**What is the black box doing, using?  
Should we trust it?**

# Why explaining AI?

- Because!
- Ethics
- Security
- Fairness
- Efficiency

# A brief overview (1/4)

- **Purposes of xAI**
  - Create white box interpretable model
  - Explain black-box/complex-model (post-hoc)
  - Enhance fairness of a model
  - Test sensitivity of predictions

## A brief overview (2/4)

- **Model specific vs. Model agnostic**
  - Model-specific methods can only be applied to a single model or a group of models
  - Model agnostic methods can be applied to any kind of model

# A brief overview (3/4)

- **Local vs. Global explanations**
  - **Local method:** explaining a model for a single decision
  - **Global method:** explaining a model globally

# A brief overview (4/4)

- **Data type**

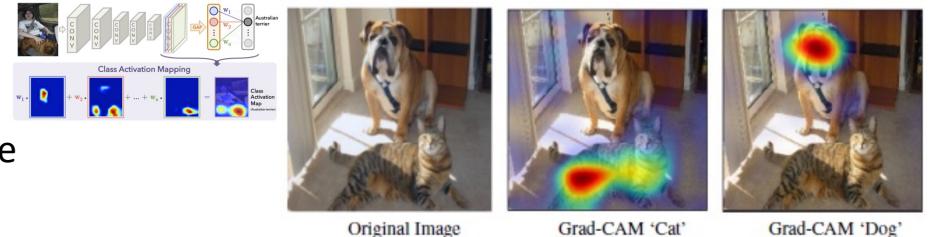
- Image
- Text
- Sound
- ...

# A brief overview (4/4)

## • Grad-cam

Zhou et al. (2016)  
CVPR

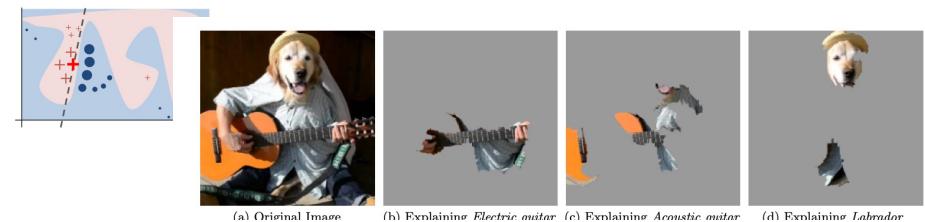
Post-hoc explanation  
Model specific: CNN only  
Local explanation: heatmap on sample  
Images / Sound (spectrogram)



## • LIME

Ribeiro et al. (2016)  
ACM SIGKDD

Post-hoc explanation  
Model agnostic  
Local explanation: parcelation map  
Images / Sound (spectrogram)...

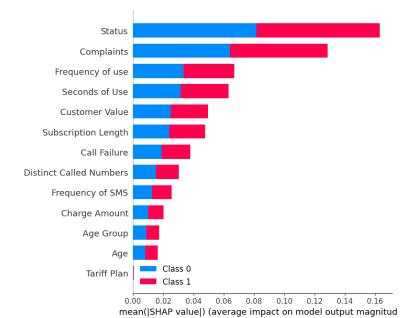


## • SHAP

Shapley (1951)  
The RAND project

Post-hoc explanation  
Model agnostic  
Global explanation: feature importance map  
Any data  
Regression or continuous probability

**Game theory**  
« Coalition between players »



## • RISE

Petsiuk et al. (2018)  
BMVC

Post-hoc explanation  
Model agnostic  
Local explanation: heat map  
Images / Sound (spectrogram)...

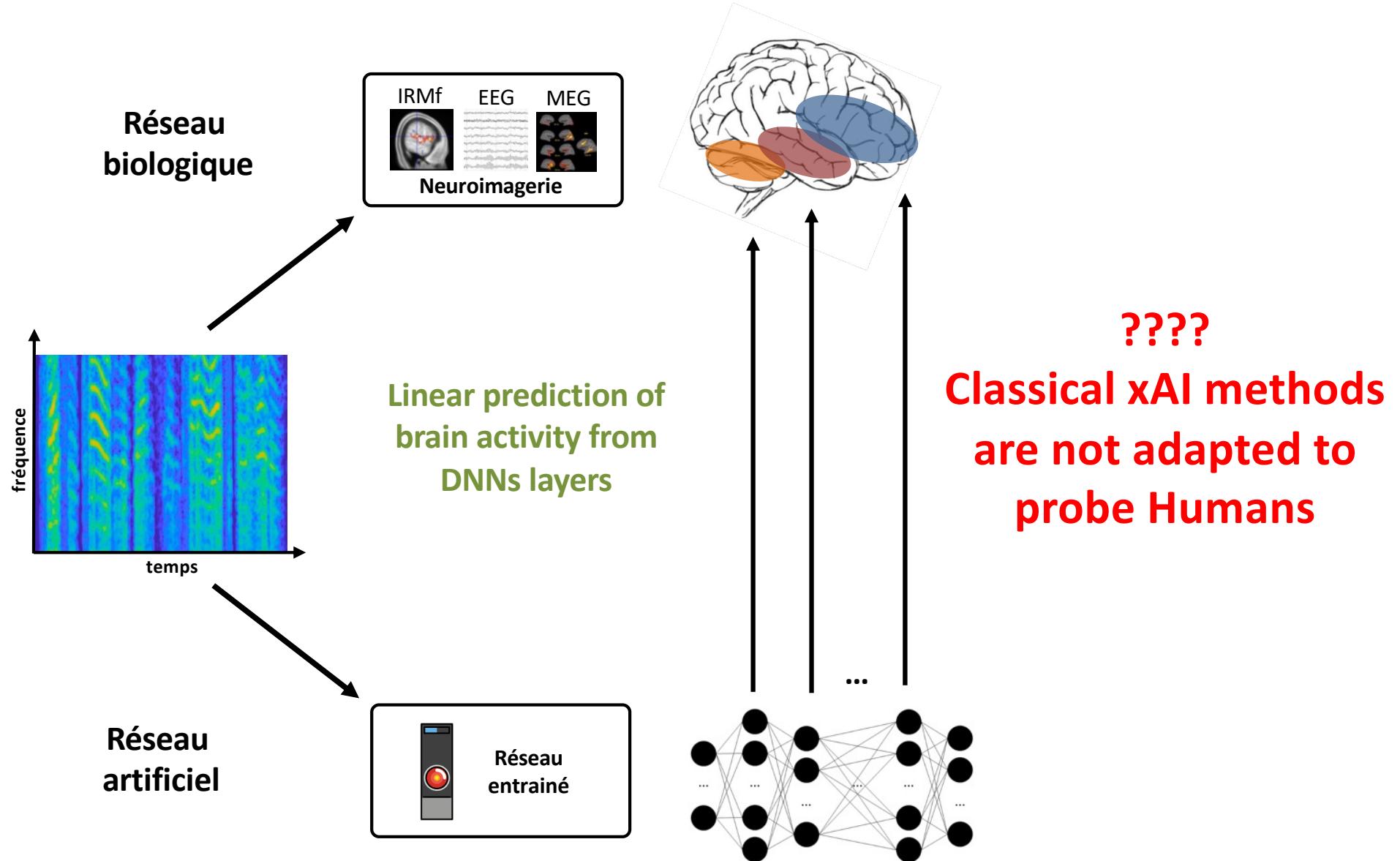


## Limitations (1/2)

- We need samples from the training or the testing set
- No underlying model
- Not necessarily applicable to neuroscientific problems

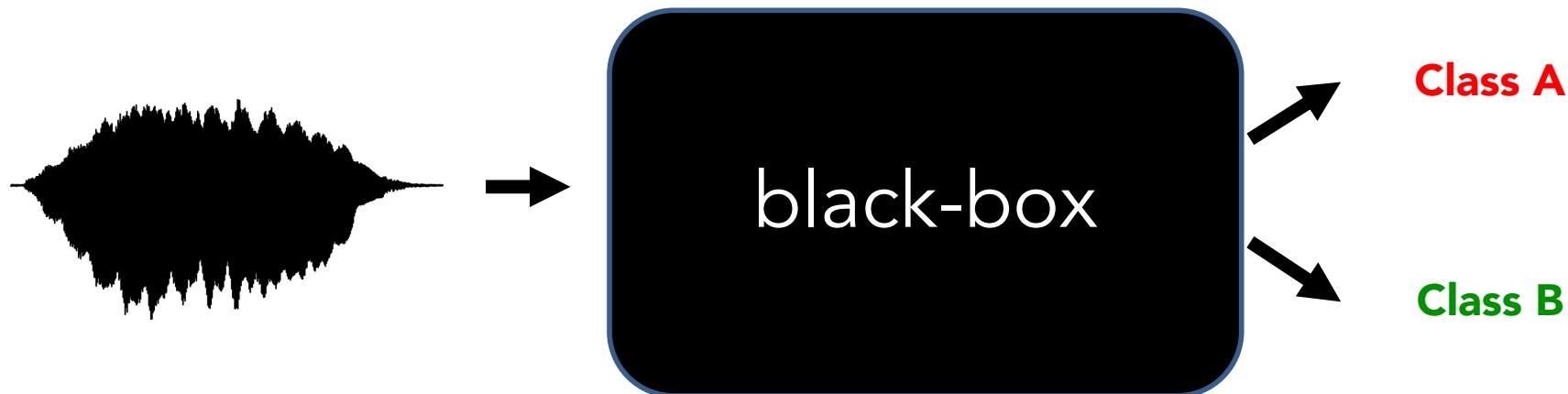
*e.g., plausibility of Artificial Networks*

## Limitations (2/2)



# Explaining AI as we probe cognition

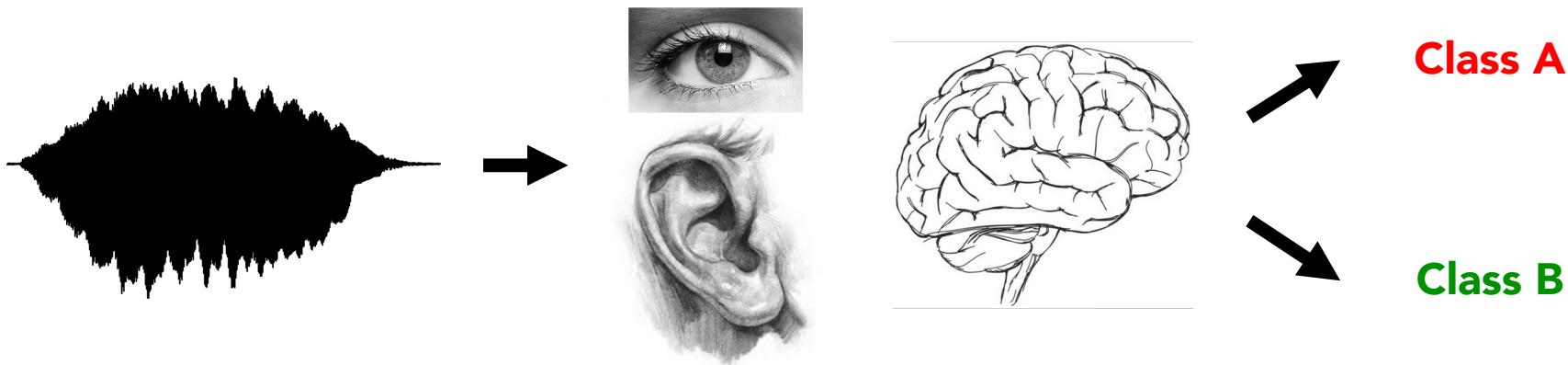
- Motivation



*What is  
the black box  
doing?*

# Explaining AI as we probe cognition

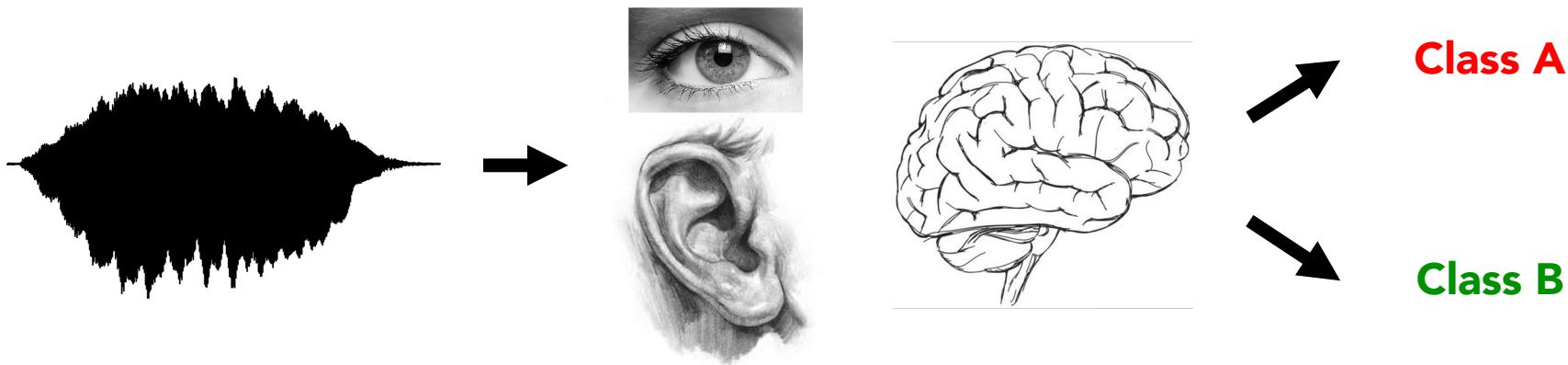
- Motivation



*What is  
the brain  
doing?*

# Explaining AI as we probe cognition

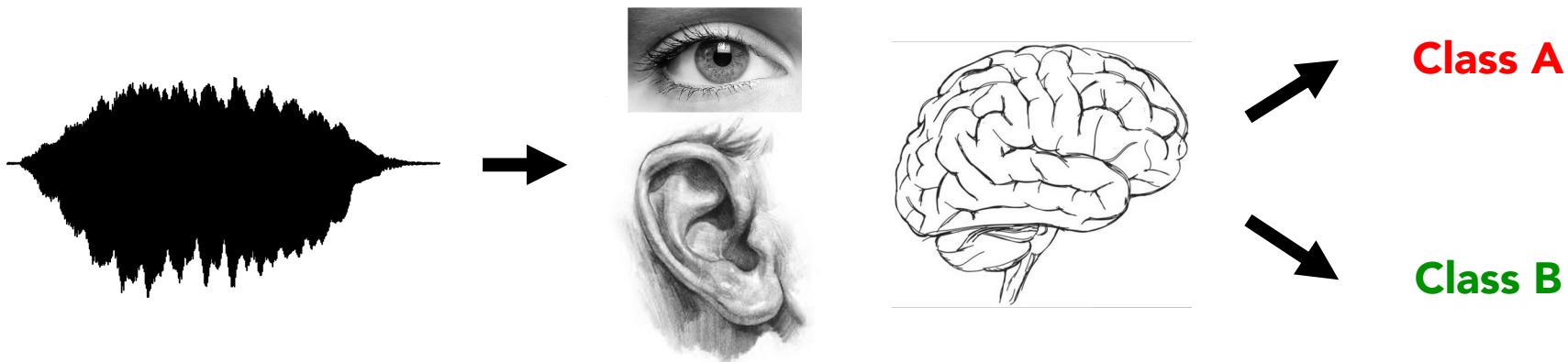
- Motivation



**Psychophysical experiments on human participants**

# Explaining AI as we probe cognition

- Motivation



## Psychophysical experiments **on human participants**

### Bubbles method

Gosselin & Shyns (2001) *Vision Research*

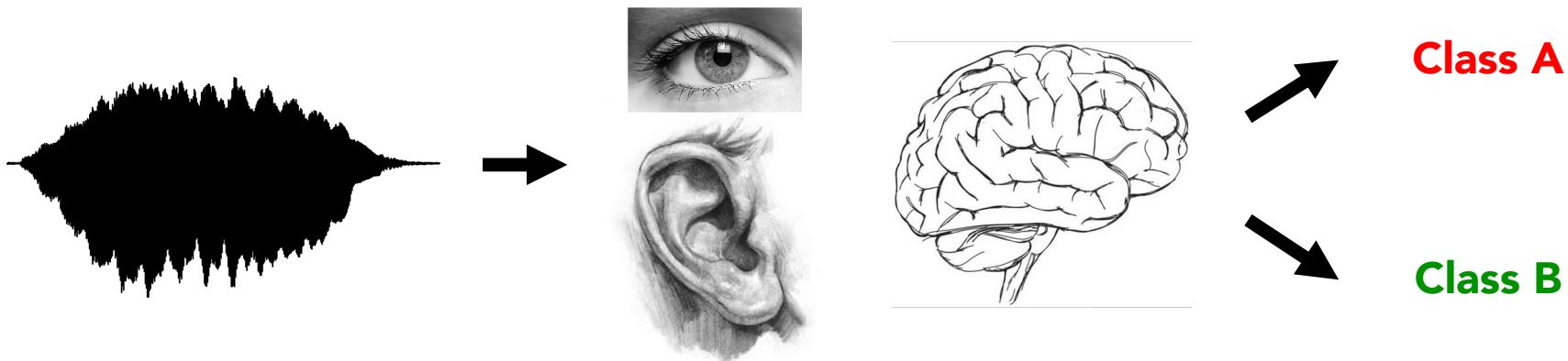
### Reverse correlation

Ahumada & Lovell (1971) *J Acoust Soc Am*

Thoret, E., Andrillon, T., Leger, D., Pressnitzer, D. (2021)  
Probing machine-learning classifiers using noise, bubbles, and reverse correlation  
*J. Neuro Methods*

# Explaining AI as we probe cognition

- **Method**

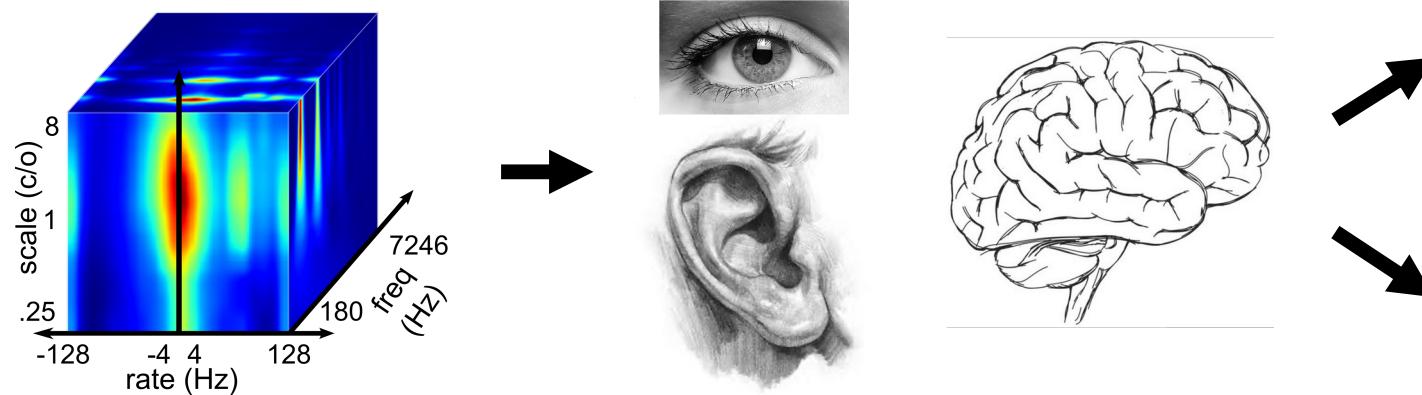


## Idea

- Additive (revcor) or multiplicative (bubbles) noise at the input to fool the subject

Thoret, E., Andrillon, T., Leger, D., Pressnitzer, D. (2021)  
Probing machine-learning classifiers using noise, bubbles, and reverse correlation  
*J. Neuro Methods*

# Explaining AI as we probe cognition

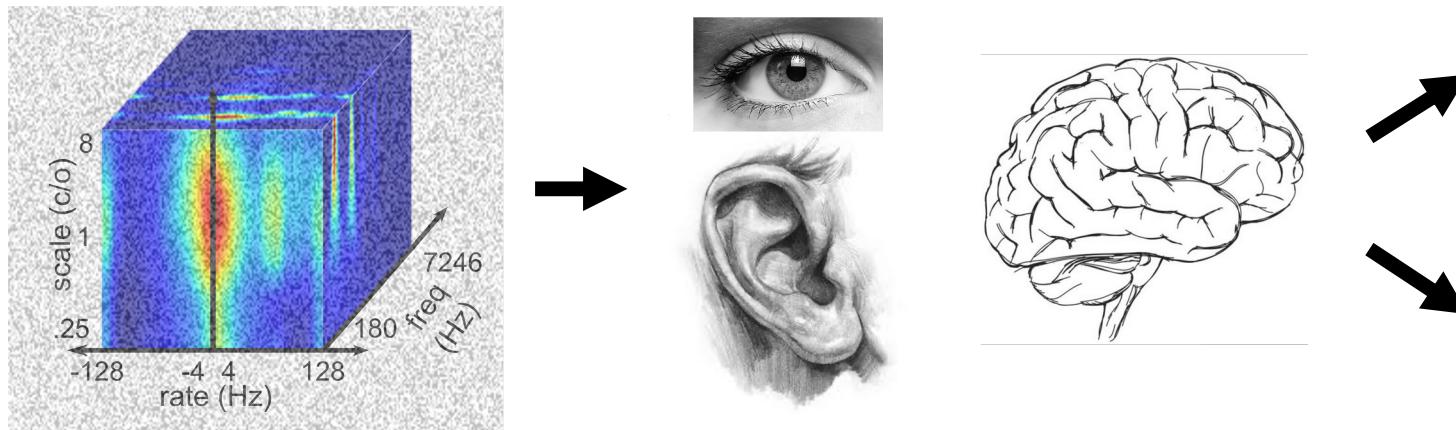


## Idea

- Additive (revcor) or multiplicative (bubbles) noise at the input to fool the subject

Thoret, E., Andrillon, T., Leger, D., Pressnitzer, D. (2021)  
Probing machine-learning classifiers using noise, bubbles, and reverse correlation  
*J. Neuro Methods*

# Explaining AI as we probe cognition

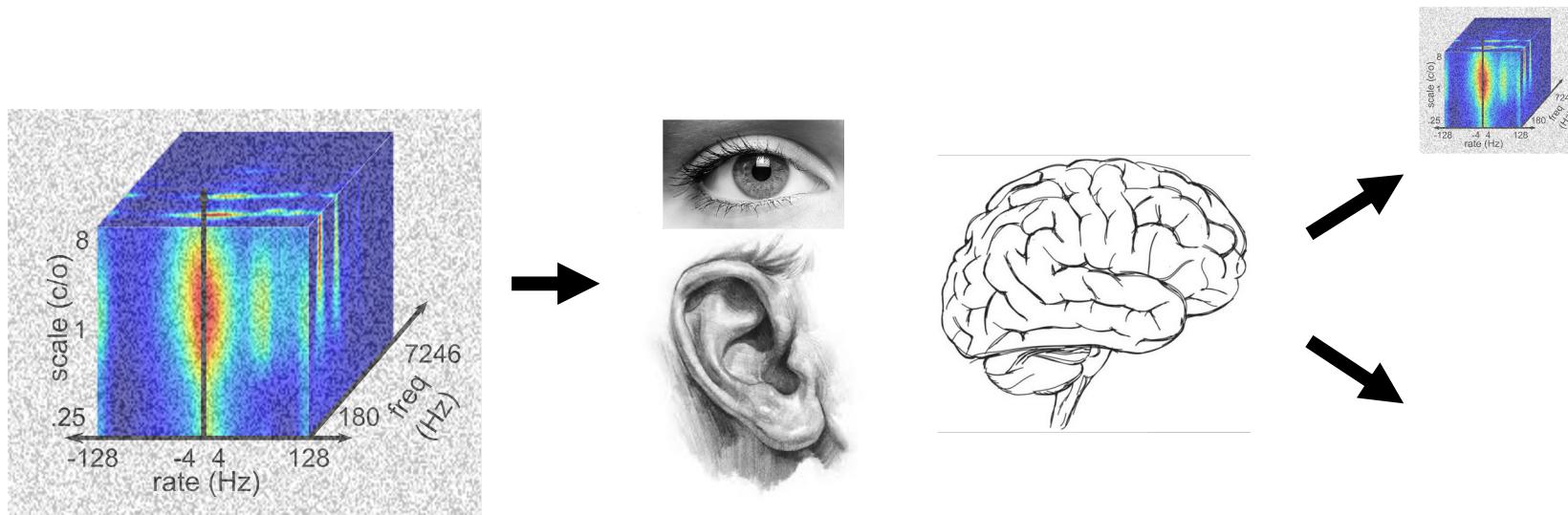


## Idea

- Additive (revcor) or multiplicative (bubbles) noise at the input to fool the subject

Thoret, E., Andrillon, T., Leger, D., Pressnitzer, D. (2021)  
Probing machine-learning classifiers using noise, bubbles, and reverse correlation  
*J. Neuro Methods*

# Explaining AI as we probe cognition

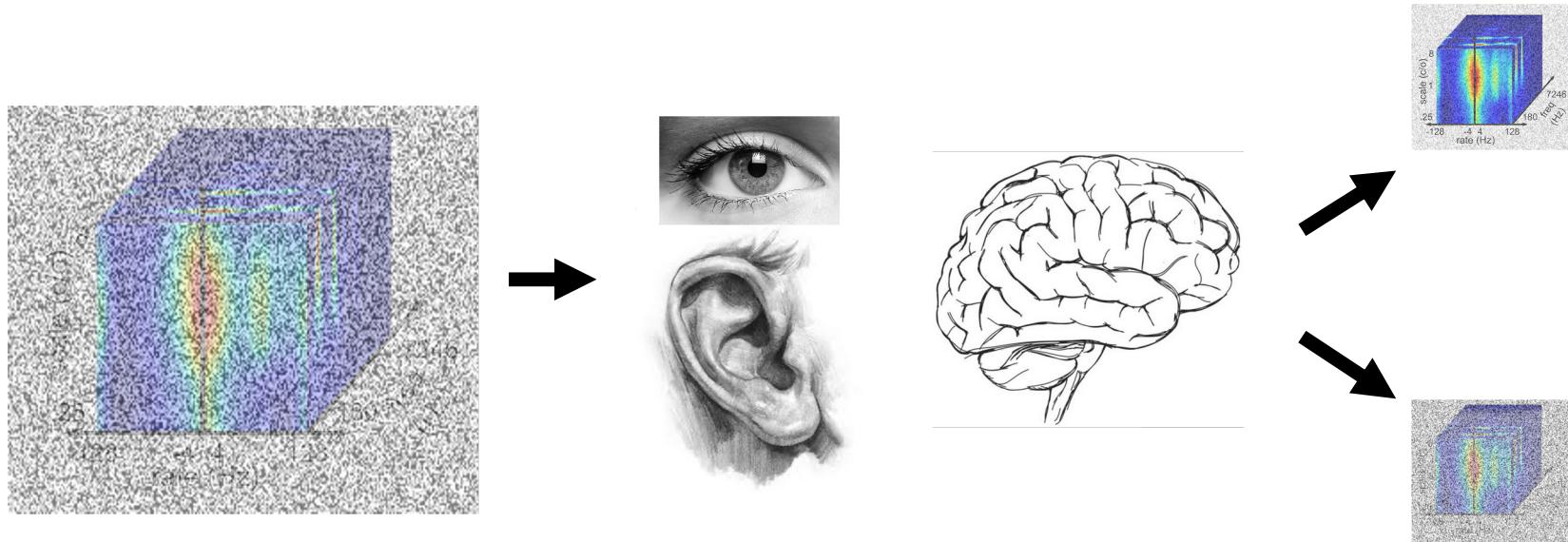


## Idea

- Additive (revcor) or multiplicative (bubbles) noise at the input to fool the subject

Thoret, E., Andrillon, T., Leger, D., Pressnitzer, D. (2021)  
Probing machine-learning classifiers using noise, bubbles, and reverse correlation  
*J. Neuro Methods*

# Explaining AI as we probe cognition

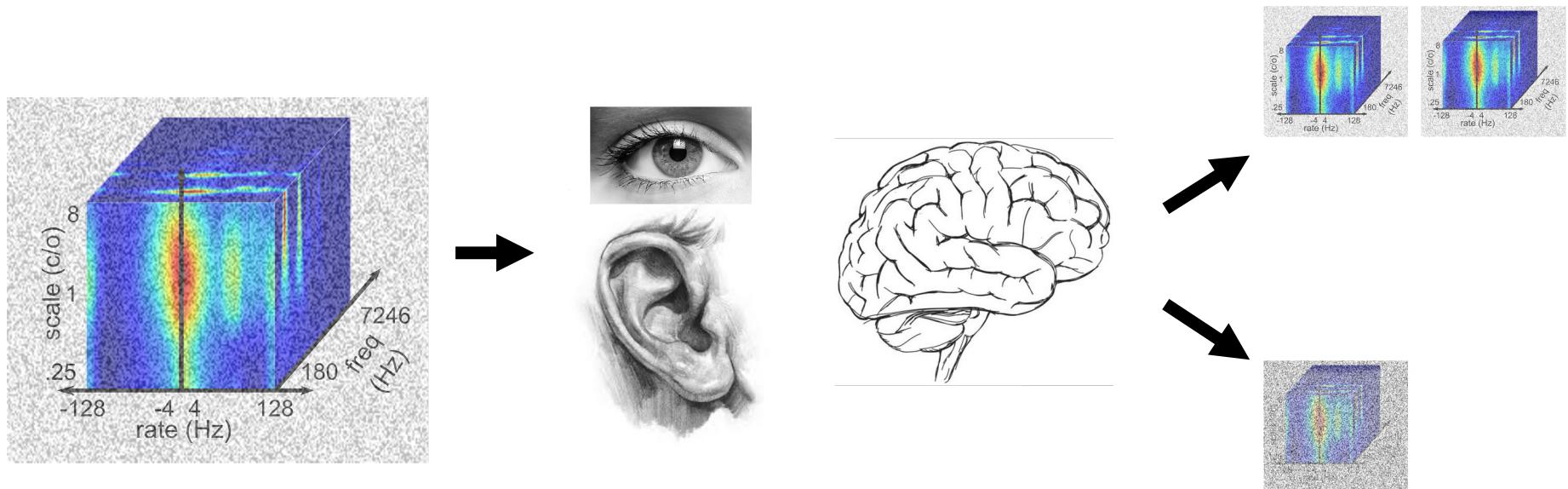


## Idea

- Additive (revcor) or multiplicative (bubbles) noise at the input to fool the subject

Thoret, E., Andrillon, T., Leger, D., Pressnitzer, D. (2021)  
Probing machine-learning classifiers using noise, bubbles, and reverse correlation  
*J. Neuro Methods*

# Explaining AI as we probe cognition

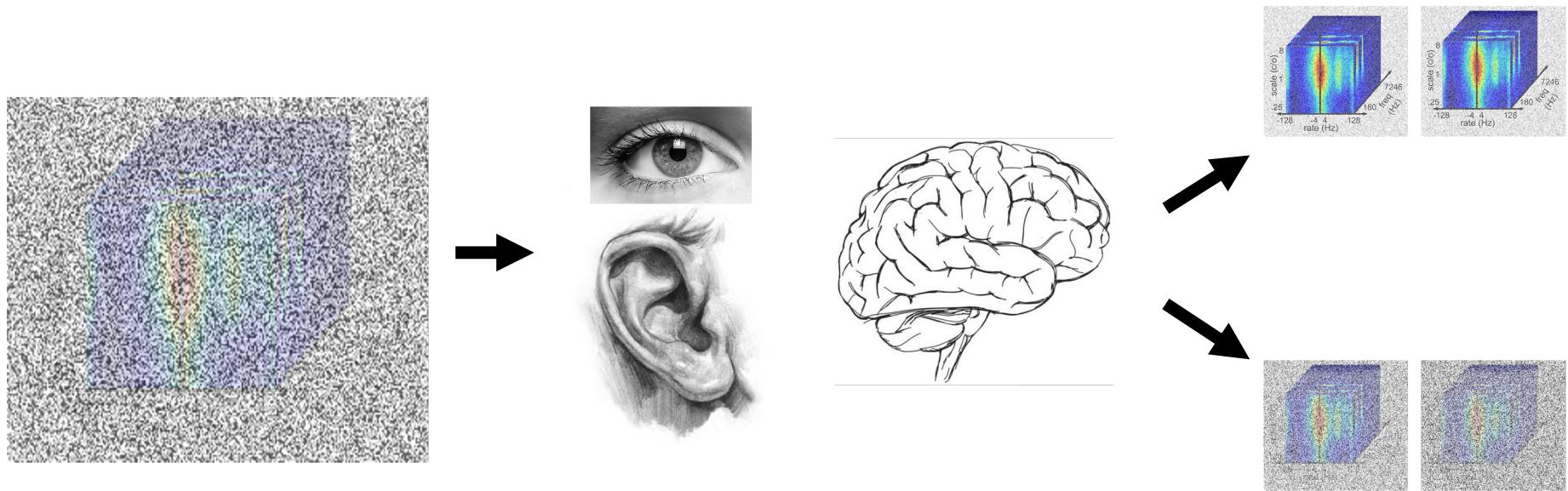


## Idea

- Additive (revcor) or multiplicative (bubbles) noise at the input to fool the subject

Thoret, E., Andrillon, T., Leger, D., Pressnitzer, D. (2021)  
Probing machine-learning classifiers using noise, bubbles, and reverse correlation  
*J. Neuro Methods*

# Explaining AI as we probe cognition

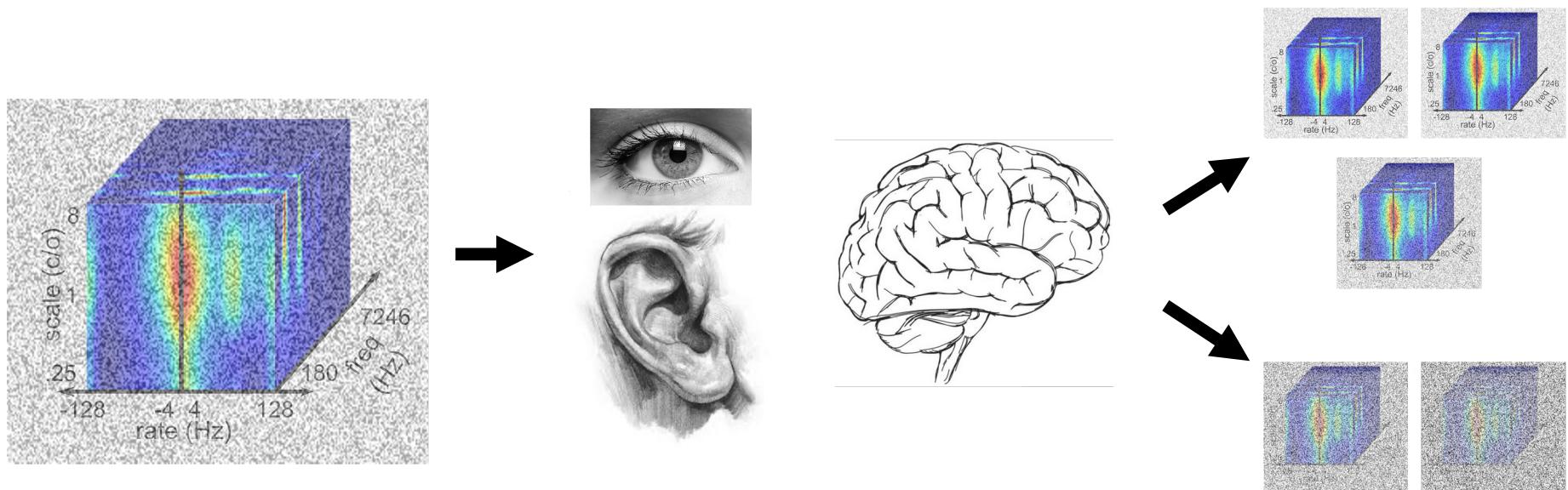


## Idea

- Additive (revcor) or multiplicative (bubbles) noise at the input to fool the subject

Thoret, E., Andrillon, T., Leger, D., Pressnitzer, D. (2021)  
Probing machine-learning classifiers using noise, bubbles, and reverse correlation  
*J. Neuro Methods*

# Explaining AI as we probe cognition



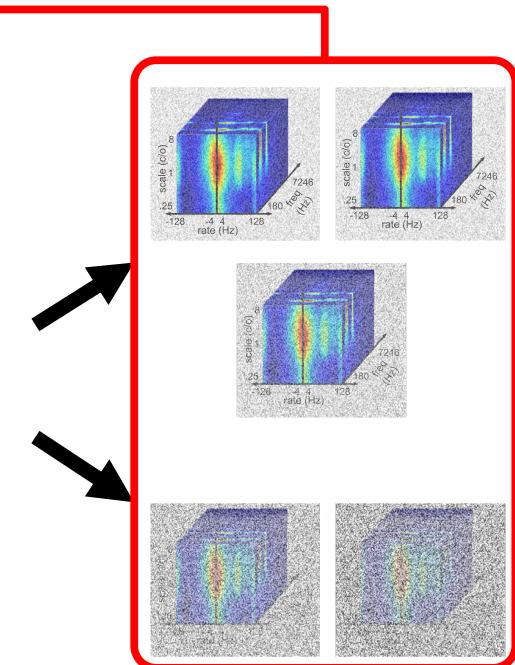
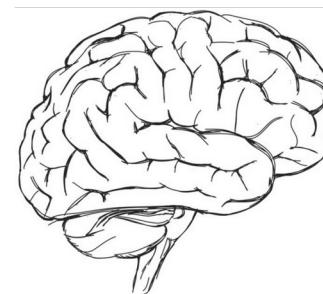
## Idea

- Additive (revcor) or multiplicative (bubbles) noise at the input to fool the subject

Thoret, E., Andrillon, T., Leger, D., Pressnitzer, D. (2021)  
Probing machine-learning classifiers using noise, bubbles, and reverse correlation  
*J. Neuro Methods*

# Explaining AI as we probe cognition

Reverse correlation



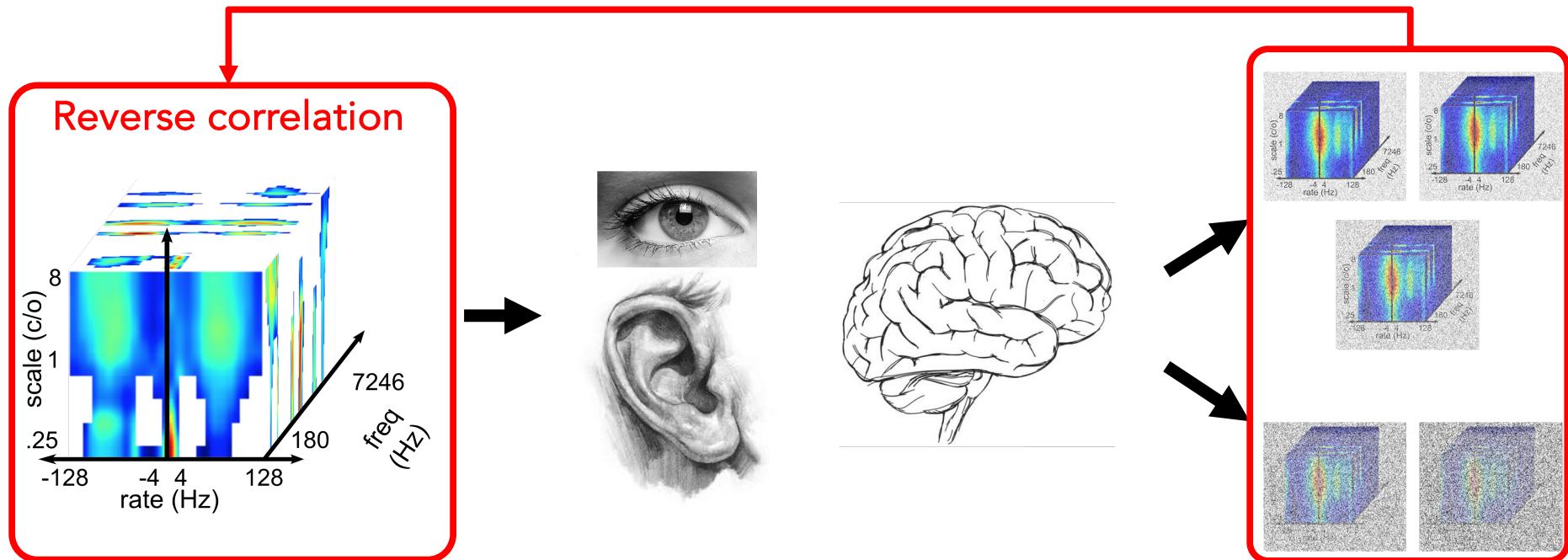
## Idea

- Additive (revcor) or multiplicative (bubbles) noise at the input to fool the subject

## Analysis

- **using the confusions** of the classifier to determine

# Explaining AI as we probe cognition



## Idea

- Additive (revcor) or multiplicative (bubbles) noise at the input to fool the subject

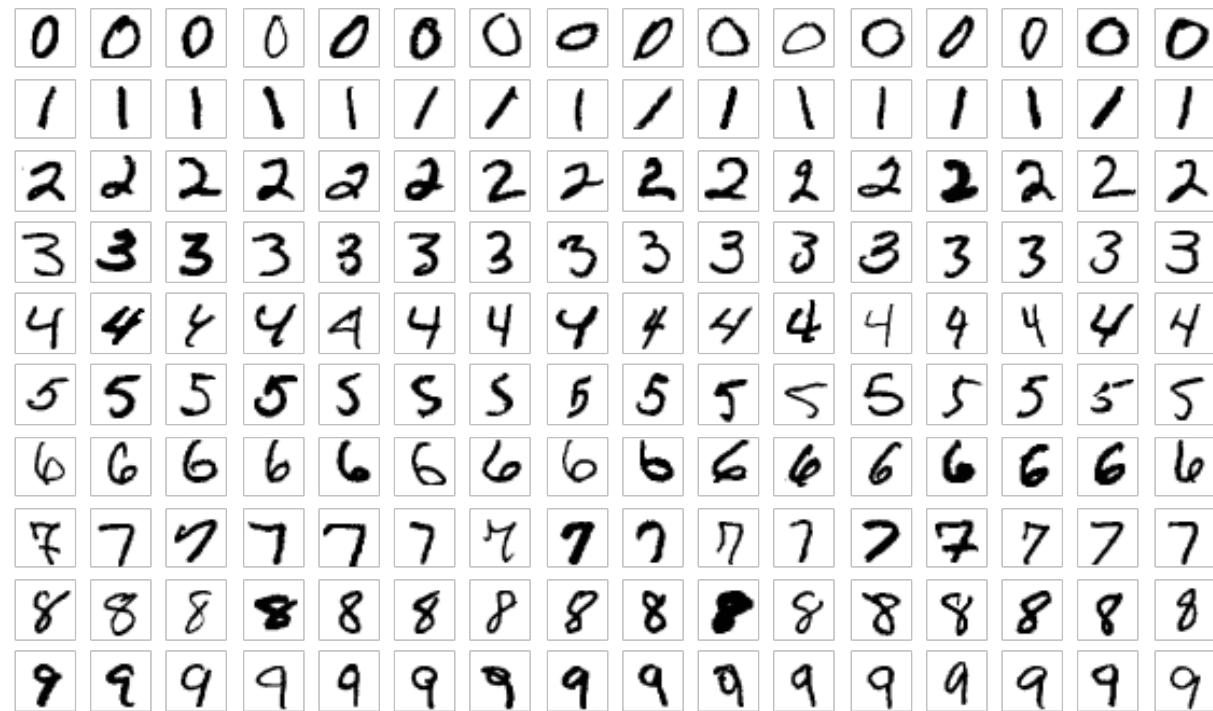
## Analysis

- **using the confusions** of the classifier to determine  
**which parts of the input are the most important**

Thoret, E., Andrillon, T., Leger, D., Pressnitzer, D. (2021)  
Probing machine-learning classifiers using noise, bubbles, and reverse correlation  
*J. Neuro Methods*

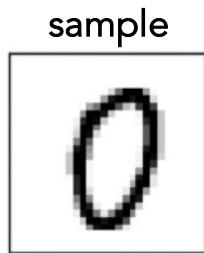
# Two methods: bubbles and reverse correlation

Images from the handwritten digits database  
MNIST



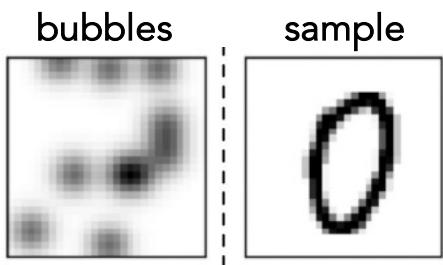
Two methods:  
**bubbles** and reverse correlation  
toy example

## Multiplicative noise



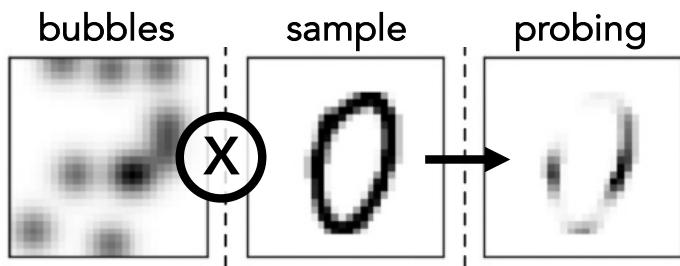
Two methods:  
**bubbles** and reverse correlation  
toy example

## Multiplicative noise



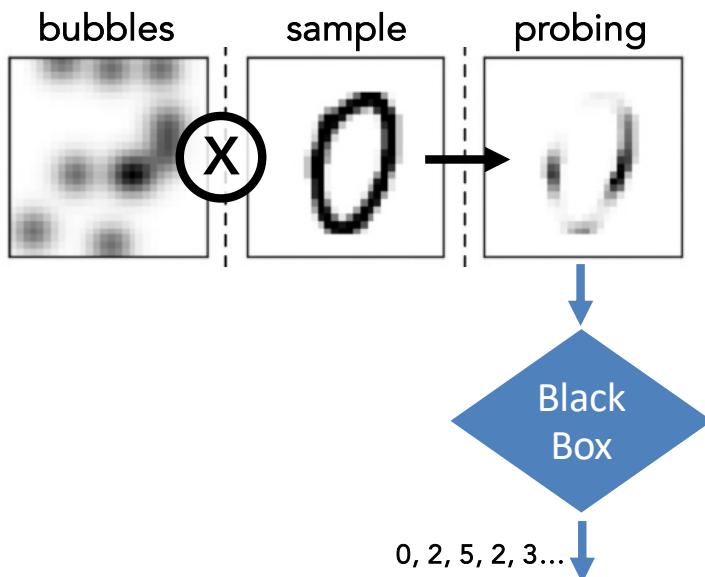
Two methods:  
**bubbles** and reverse correlation  
toy example

## Multiplicative noise



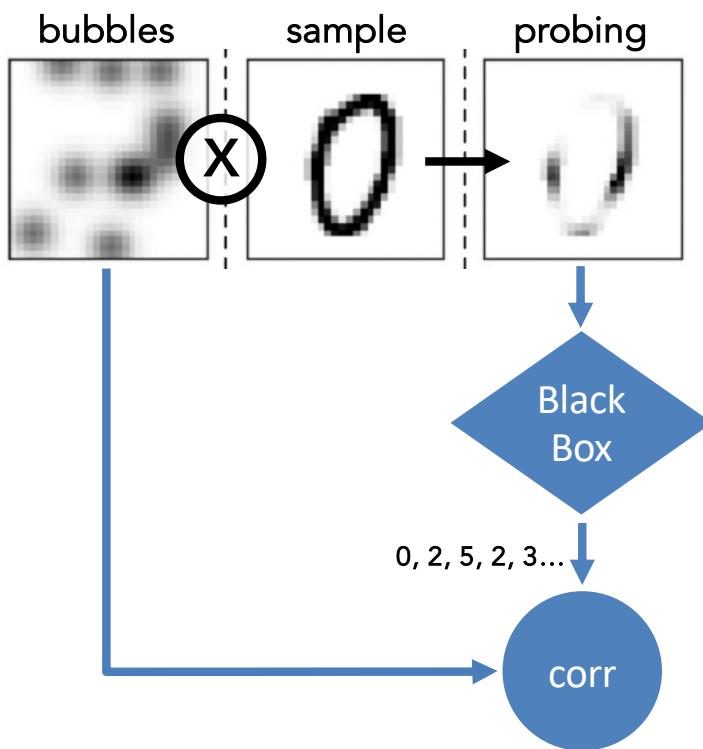
Two methods:  
**bubbles** and reverse correlation  
toy example

## Multiplicative noise

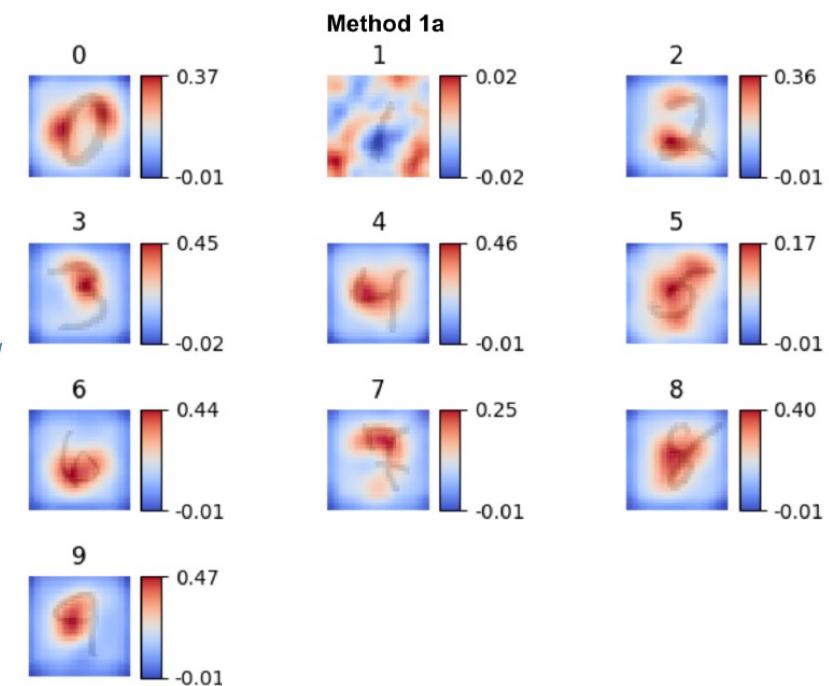
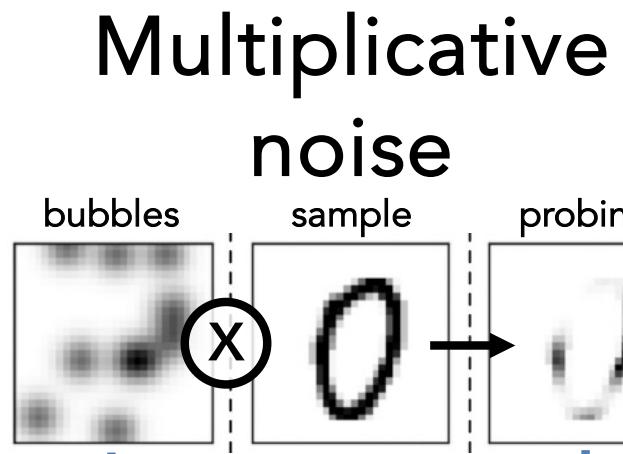


Two methods:  
**bubbles** and reverse correlation  
toy example

## Multiplicative noise

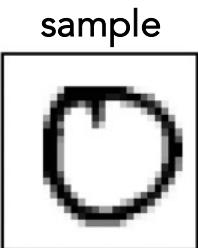


# Two methods: **bubbles** and reverse correlation toy example



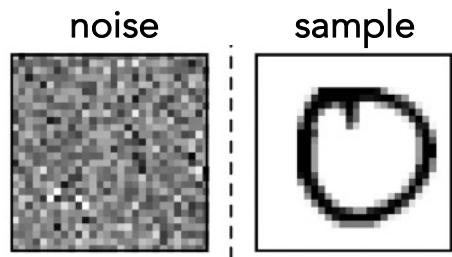
Two methods:  
bubbles and **reverse correlation**  
*toy example*

Additive  
noise



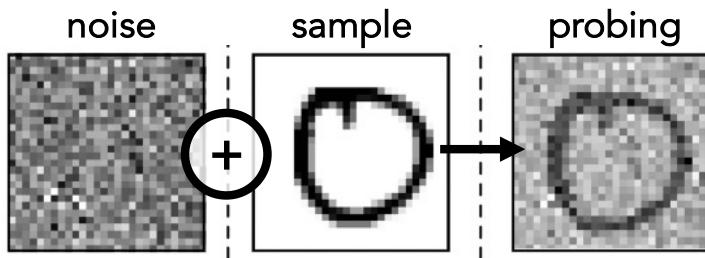
# Two methods: bubbles and **reverse correlation** *toy example*

## Additive noise



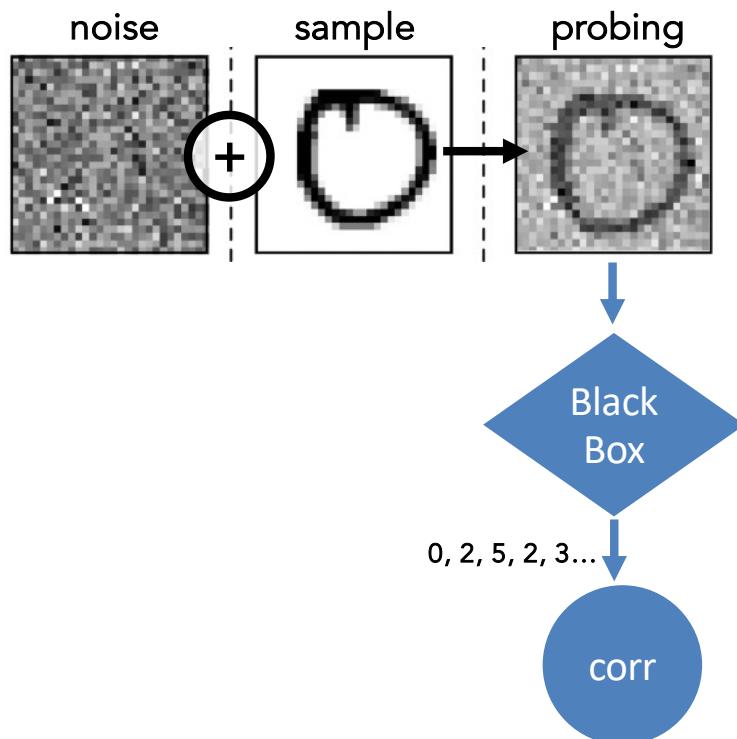
# Two methods: bubbles and **reverse correlation** *toy example*

## Additive noise



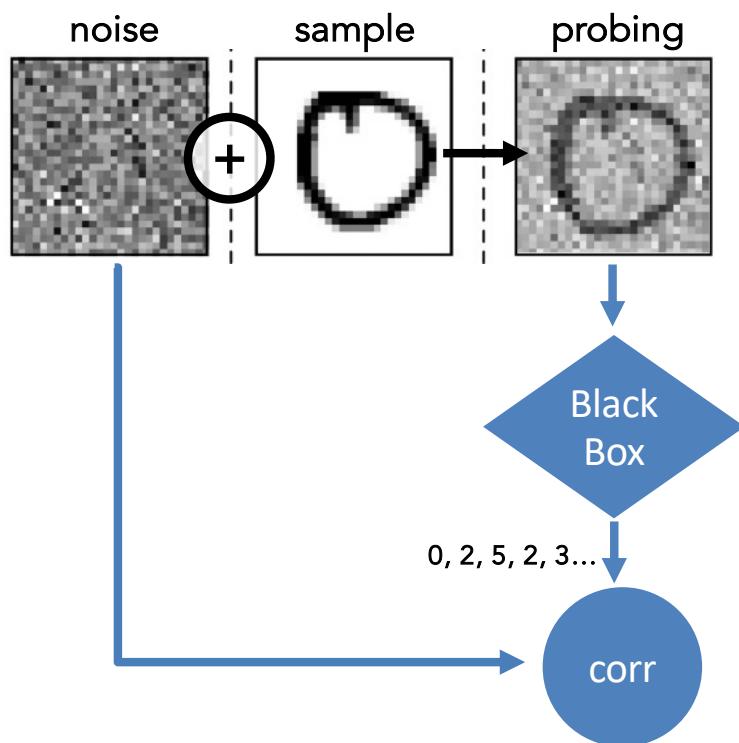
# Two methods: bubbles and **reverse correlation** *toy example*

## Additive noise

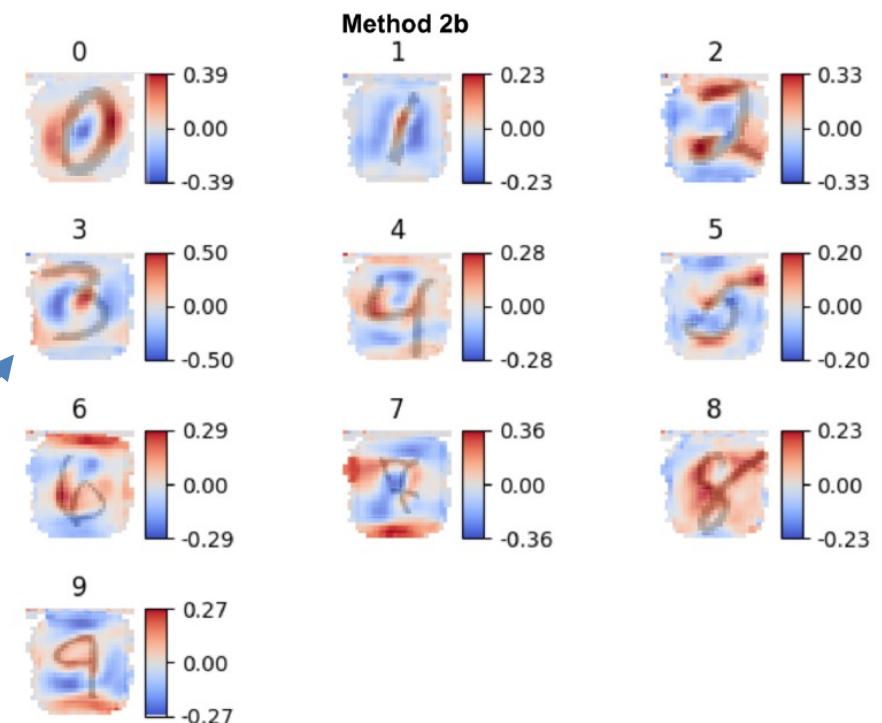
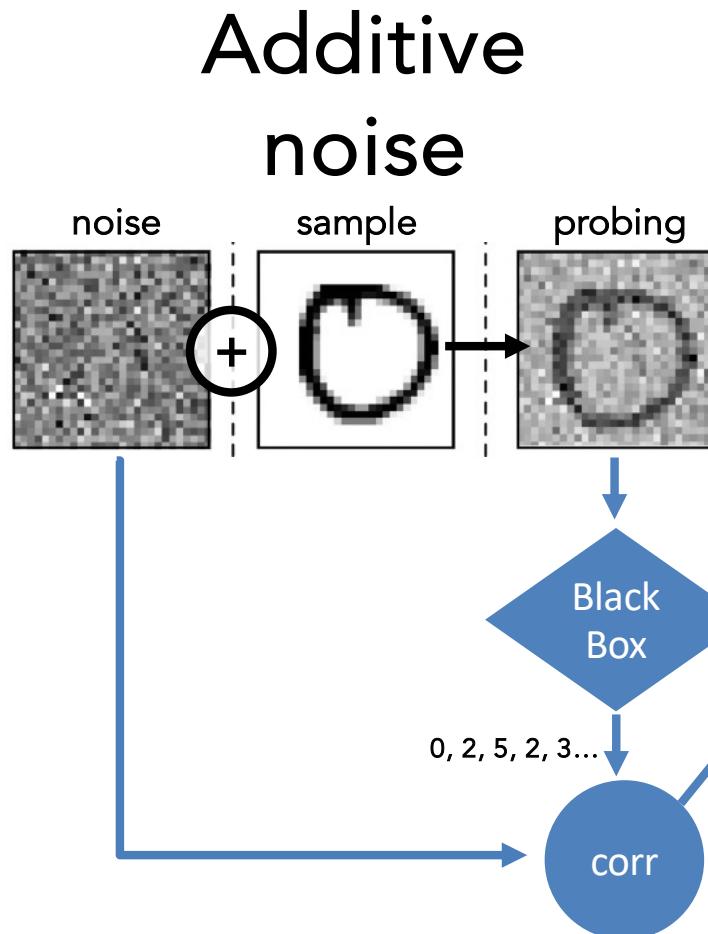


# Two methods: bubbles and **reverse correlation** *toy example*

## Additive noise



# Two methods: bubbles and reverse correlation toy example



« Prototype »

# Bubbles in the jungle

70

Opinion

TRENDS in Cognitive Sciences Vol.6 No.2 February 2002

## RAP: a new framework for visual categorization

Frédéric Gosselin and Philippe G. Schyns

approach that is...  
test its performan...  
conditions of exp...  
discoveries about...  
scene categories,...  
different recognit...  
recognition do no...  
processes of low-l...  
face, object and sc...  
are not always fir...  
principles of early...  
that high- and lov...

Reverse correlation



### Troubles with bubbles

Richard F. Murray<sup>a</sup> Jason M. Gold<sup>b</sup>

Show more

+ Add to Mendeley Share Cite



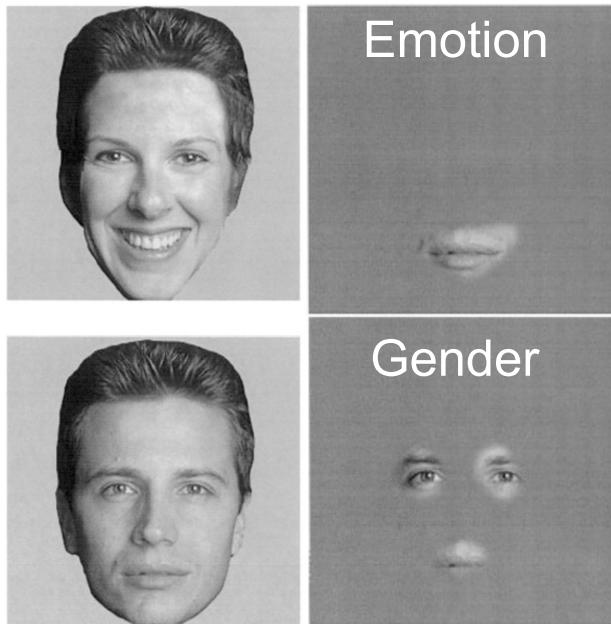
Bubbles

# Explaining AI as we probe cognition

- Example of applications in vision and audition

## Bubbles

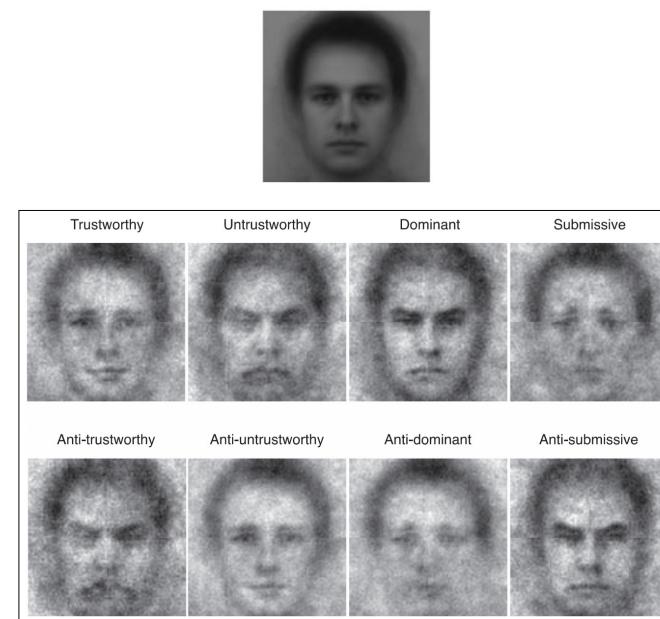
Which information are the most important?



Gosselin, F., & Schyns, P. G. (2001).  
*Vision research*, 41(17), 2261-2271.

## Reverse Correlation

Which information are « stored in memory? »



Dotsch, R., & Todorov, A. (2012).  
*Social Psychological and Personality Science*, 3(5), 562-571.

# Explaining AI as we probe cognition

- **Example of applications in vision and audition**

**Bubbles:** analysis of the spatio-temporal course of visual identification of a face

What visual areas do we focus on to identify a face and how does this change over time?

- “dynamic” bubbles

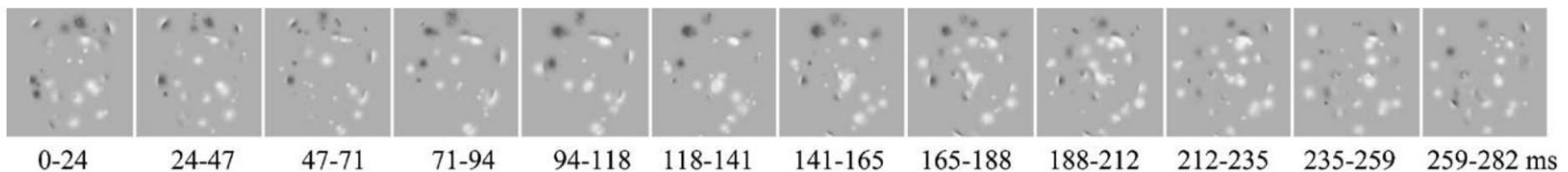
# Explaining AI as we probe cognition

- Example of applications in vision and audition

**Bubbles:** analysis of the spatio-temporal course of visual identification of a face

What visual areas do we focus on to identify a face and how does this change over time?

- “dynamic” bubbles



Cognitive Science 28 (2004) 289–301

COGNITIVE  
SCIENCE

<http://www.elsevier.com/locate/cogsci>

Spatio-temporal dynamics of face recognition  
in a flash: it's in the eyes

Céline Vinette, Frédéric Gosselin <sup>a,\*</sup>, Philippe G. Schyns <sup>b</sup>

<sup>a</sup>Département de psychologie, Université de Montréal, 90 Vincent-d'Indy,  
C.P. 6128 succ. Centre-ville, Montréal, Québec, Canada H3C 3J7

<sup>b</sup>Department of Psychology, University of Glasgow, Glasgow G12 8QB, UK

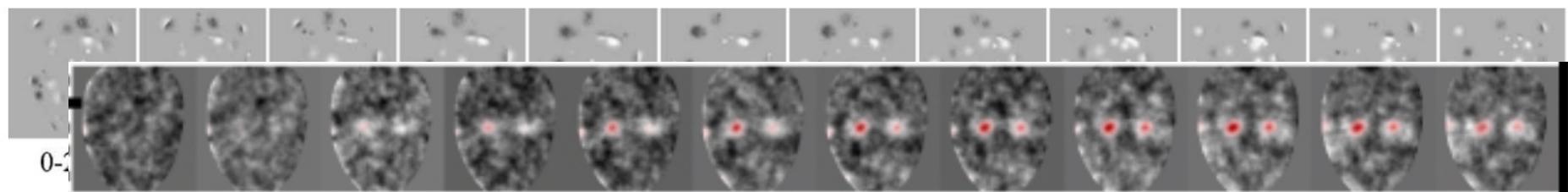
# Explaining AI as we probe cognition

- **Example of applications in vision and audition**

**Bubbles:** analysis of the spatio-temporal course of visual identification of a face

What visual areas do we focus on to identify a face and how does this change over time?

- “dynamic” bubbles



Cognitive Science 28 (2004) 289–301

COGNITIVE  
SCIENCE

<http://www.elsevier.com/locate/cogsci>

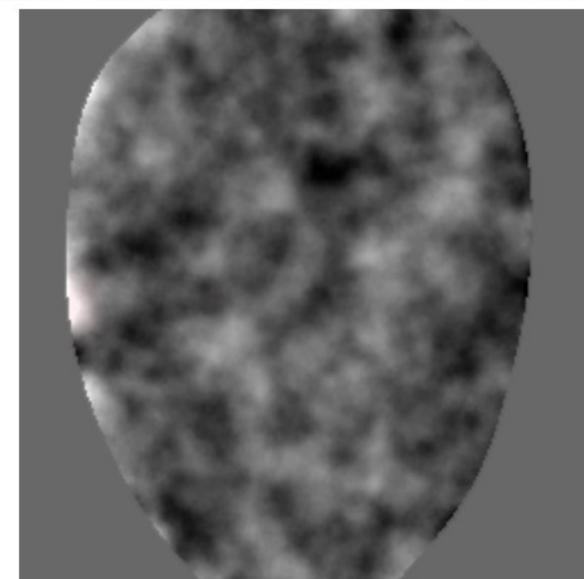
Spatio-temporal dynamics of face recognition  
in a flash: it's in the eyes

Céline Vinette, Frédéric Gosselin <sup>a,\*</sup>, Philippe G. Schyns <sup>b</sup>

<sup>a</sup>Département de psychologie, Université de Montréal, 90 Vincent-d’Indy,

C.P. 6128 succ. Centre-ville, Montréal, Que., Canada H3C 3J7

<sup>b</sup>Department of Psychology, University of Glasgow, Glasgow G12 8QB, UK



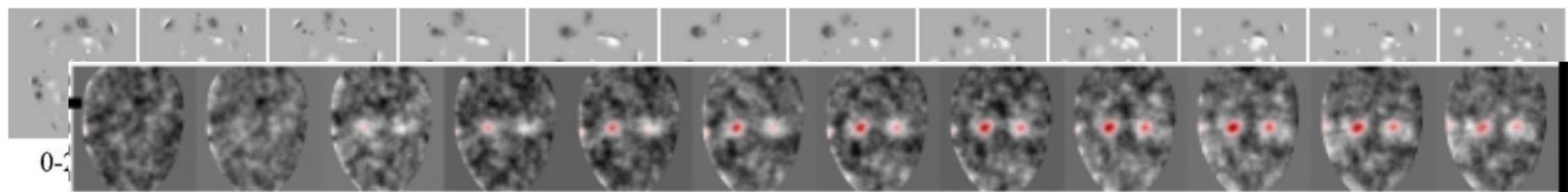
# Explaining AI as we probe cognition

- **Example of applications in vision and audition**

**Bubbles:** analysis of the spatio-temporal course of visual identification of a face

What visual areas do we focus on to identify a face and how does this change over time?

- “dynamic” bubbles



Cognitive Science 28 (2004) 289–301

COGNITIVE  
SCIENCE

<http://www.elsevier.com/locate/cogsci>

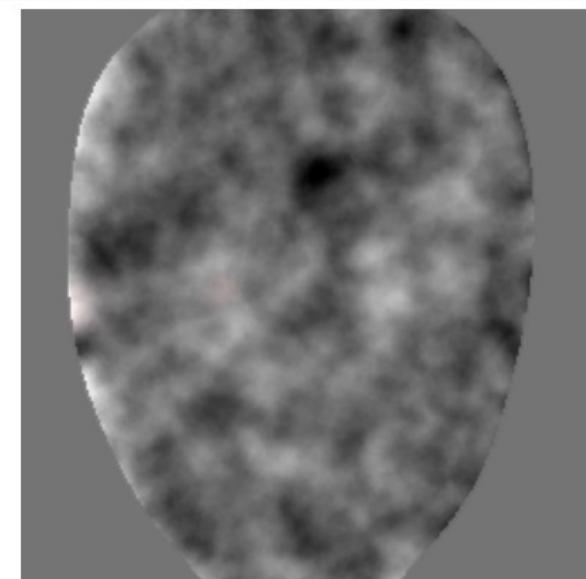
Spatio-temporal dynamics of face recognition  
in a flash: it's in the eyes

Céline Vinette, Frédéric Gosselin <sup>a,\*</sup>, Philippe G. Schyns <sup>b</sup>

<sup>a</sup>Département de psychologie, Université de Montréal, 90 Vincent-d'Indy,

C.P. 6128 succ. Centre-ville, Montréal, Que., Canada H3C 3J7

<sup>b</sup>Department of Psychology, University of Glasgow, Glasgow G12 8QB, UK



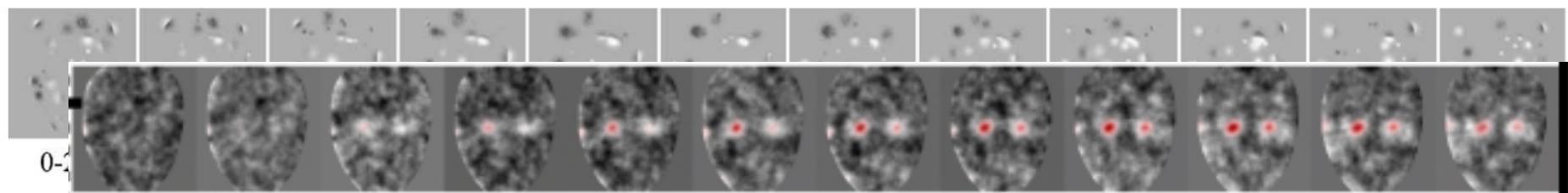
# Explaining AI as we probe cognition

- **Example of applications in vision and audition**

**Bubbles:** analysis of the spatio-temporal course of visual identification of a face

What visual areas do we focus on to identify a face and how does this change over time?

- “dynamic” bubbles



Cognitive Science 28 (2004) 289–301

COGNITIVE  
SCIENCE

<http://www.elsevier.com/locate/cogsci>

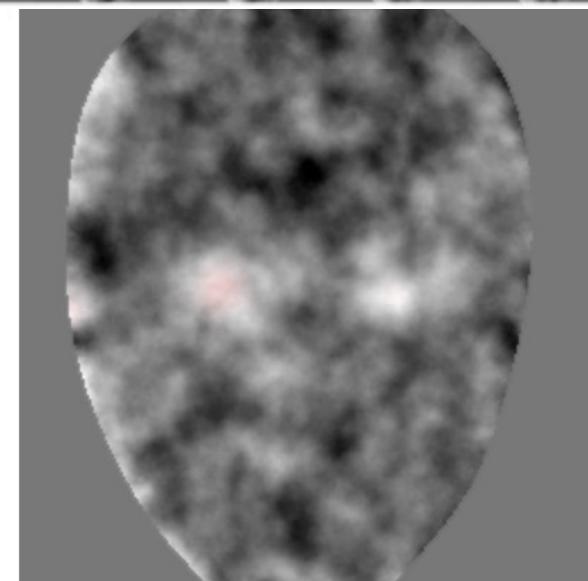
Spatio-temporal dynamics of face recognition  
in a flash: it's in the eyes

Céline Vinette, Frédéric Gosselin <sup>a,\*</sup>, Philippe G. Schyns <sup>b</sup>

<sup>a</sup>Département de psychologie, Université de Montréal, 90 Vincent-d'Indy,

C.P. 6128 succ. Centre-ville, Montréal, Que., Canada H3C 3J7

<sup>b</sup>Department of Psychology, University of Glasgow, Glasgow G12 8QB, UK



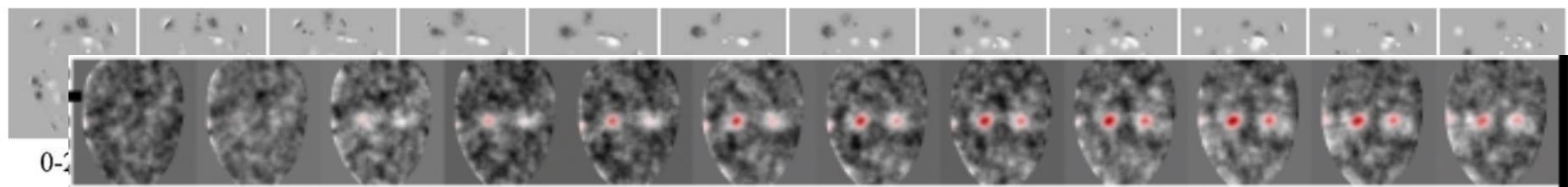
# Explaining AI as we probe cognition

- **Example of applications in vision and audition**

**Bubbles:** analysis of the spatio-temporal course of visual identification of a face

What visual areas do we focus on to identify a face and how does this change over time?

- “dynamic” bubbles



Cognitive Science 28 (2004) 289–301

COGNITIVE  
SCIENCE

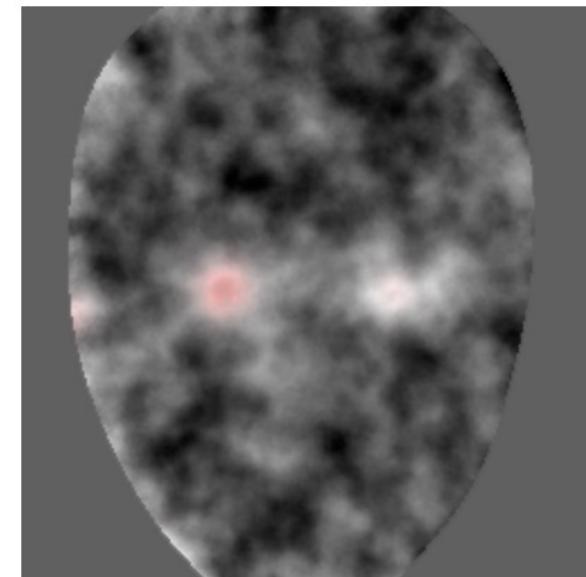
<http://www.elsevier.com/locate/cogsci>

Spatio-temporal dynamics of face recognition  
in a flash: it's in the eyes

Céline Vinette, Frédéric Gosselin <sup>a,\*</sup>, Philippe G. Schyns <sup>b</sup>

<sup>a</sup>Département de psychologie, Université de Montréal, 90 Vincent-d'Indy,  
C.P. 6128 succ. Centre-ville, Montréal, Québec, Canada H3C 3J7

<sup>b</sup>Department of Psychology, University of Glasgow, Glasgow G12 8QB, UK



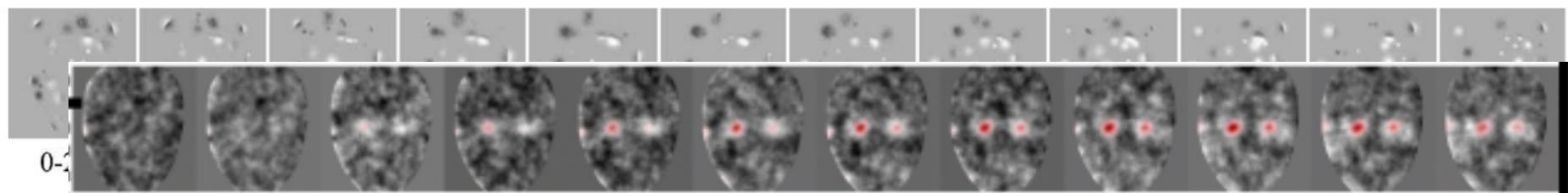
# Explaining AI as we probe cognition

- **Example of applications in vision and audition**

**Bubbles:** analysis of the spatio-temporal course of visual identification of a face

What visual areas do we focus on to identify a face and how does this change over time?

- “dynamic” bubbles



Cognitive Science 28 (2004) 289–301

COGNITIVE  
SCIENCE

<http://www.elsevier.com/locate/cogsci>

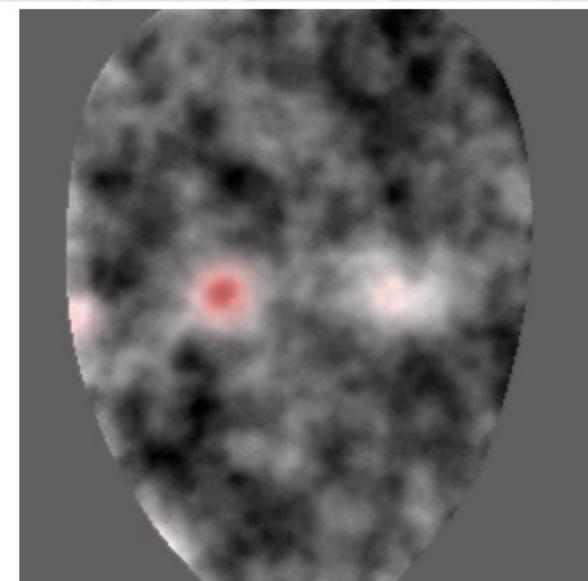
Spatio-temporal dynamics of face recognition  
in a flash: it's in the eyes

Céline Vinette, Frédéric Gosselin <sup>a,\*</sup>, Philippe G. Schyns <sup>b</sup>

<sup>a</sup>Département de psychologie, Université de Montréal, 90 Vincent-d’Indy,

C.P. 6128 succ. Centre-ville, Montréal, Que., Canada H3C 3J7

<sup>b</sup>Department of Psychology, University of Glasgow, Glasgow G12 8QB, UK



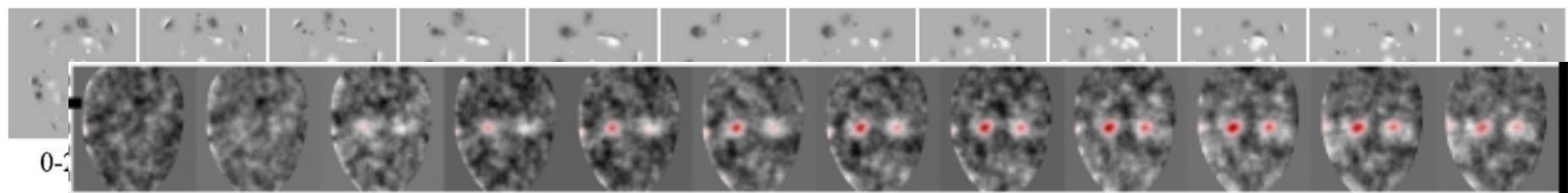
# Explaining AI as we probe cognition

- **Example of applications in vision and audition**

**Bubbles:** analysis of the spatio-temporal course of visual identification of a face

What visual areas do we focus on to identify a face and how does this change over time?

- “dynamic” bubbles



Cognitive Science 28 (2004) 289–301

COGNITIVE  
SCIENCE

<http://www.elsevier.com/locate/cogsci>

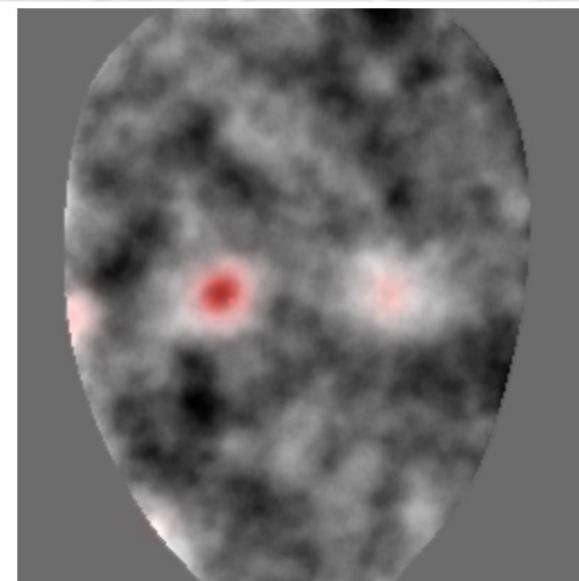
Spatio-temporal dynamics of face recognition  
in a flash: it's in the eyes

Céline Vinette, Frédéric Gosselin <sup>a,\*</sup>, Philippe G. Schyns <sup>b</sup>

<sup>a</sup>Département de psychologie, Université de Montréal, 90 Vincent-d’Indy,

C.P. 6128 succ. Centre-ville, Montréal, Que., Canada H3C 3J7

<sup>b</sup>Department of Psychology, University of Glasgow, Glasgow G12 8QB, UK



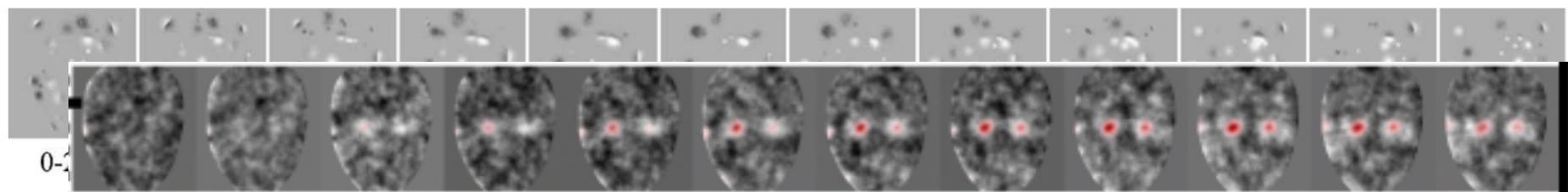
# Explaining AI as we probe cognition

- Example of applications in vision and audition

**Bubbles:** analysis of the spatio-temporal course of visual identification of a face

What visual areas do we focus on to identify a face and how does this change over time?

- “dynamic” bubbles



Cognitive Science 28 (2004) 289–301

COGNITIVE  
SCIENCE

<http://www.elsevier.com/locate/cogsci>

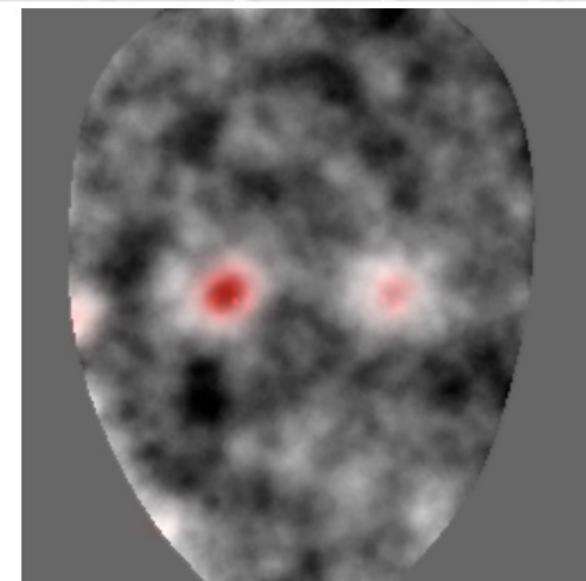
Spatio-temporal dynamics of face recognition  
in a flash: it's in the eyes

Céline Vinette, Frédéric Gosselin <sup>a,\*</sup>, Philippe G. Schyns <sup>b</sup>

<sup>a</sup>Département de psychologie, Université de Montréal, 90 Vincent-d’Indy,

C.P. 6128 succ. Centre-ville, Montréal, Que., Canada H3C 3J7

<sup>b</sup>Department of Psychology, University of Glasgow, Glasgow G12 8QB, UK



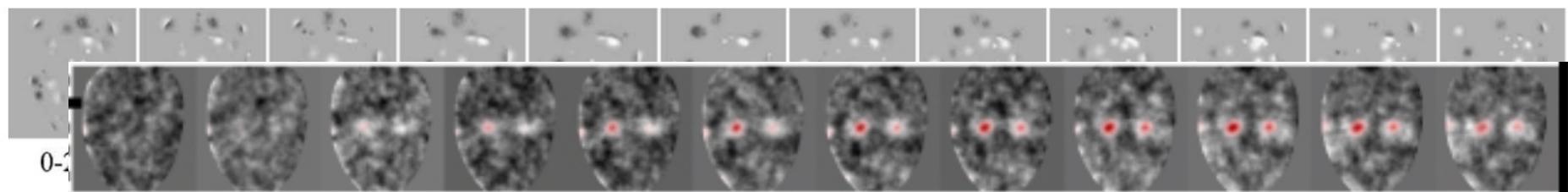
# Explaining AI as we probe cognition

- **Example of applications in vision and audition**

**Bubbles:** analysis of the spatio-temporal course of visual identification of a face

What visual areas do we focus on to identify a face and how does this change over time?

- “dynamic” bubbles



Cognitive Science 28 (2004) 289–301

COGNITIVE  
SCIENCE

<http://www.elsevier.com/locate/cogsci>

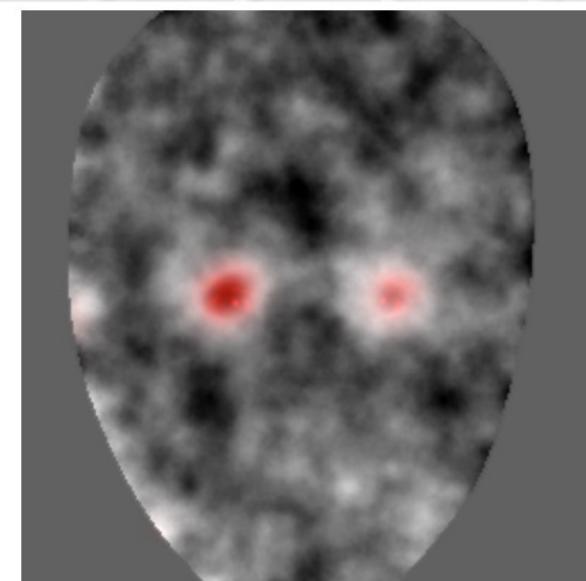
Spatio-temporal dynamics of face recognition  
in a flash: it's in the eyes

Céline Vinette, Frédéric Gosselin <sup>a,\*</sup>, Philippe G. Schyns <sup>b</sup>

<sup>a</sup>Département de psychologie, Université de Montréal, 90 Vincent-d’Indy,

C.P. 6128 succ. Centre-ville, Montréal, Que., Canada H3C 3J7

<sup>b</sup>Department of Psychology, University of Glasgow, Glasgow G12 8QB, UK



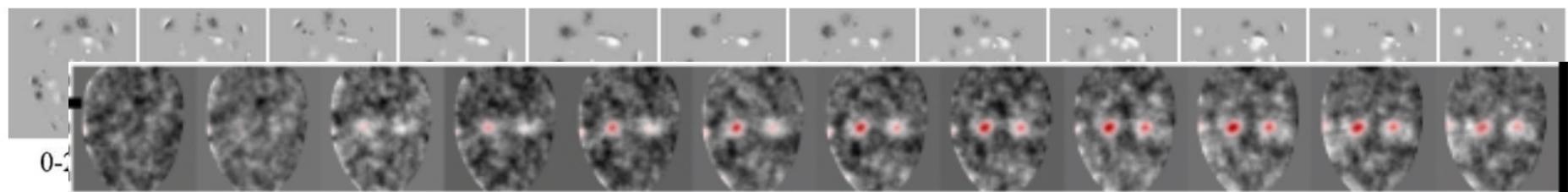
# Explaining AI as we probe cognition

- **Example of applications in vision and audition**

**Bubbles:** analysis of the spatio-temporal course of visual identification of a face

What visual areas do we focus on to identify a face and how does this change over time?

- “dynamic” bubbles



Cognitive Science 28 (2004) 289–301

COGNITIVE  
SCIENCE

<http://www.elsevier.com/locate/cogsci>

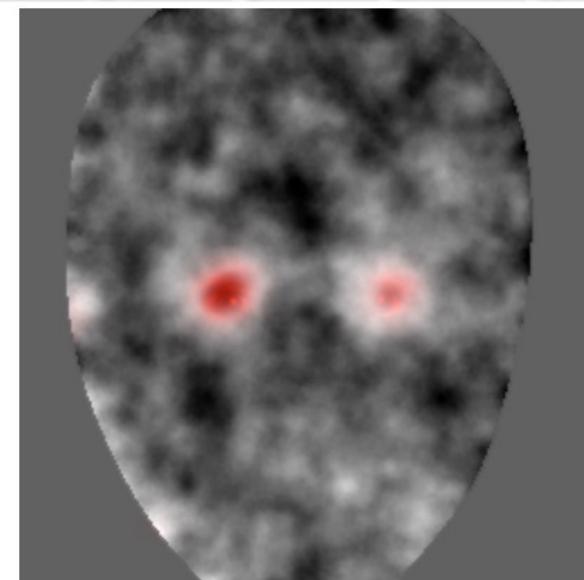
Spatio-temporal dynamics of face recognition  
in a flash: it's in the eyes

Céline Vinette, Frédéric Gosselin <sup>a,\*</sup>, Philippe G. Schyns <sup>b</sup>

<sup>a</sup>Département de psychologie, Université de Montréal, 90 Vincent-d’Indy,

C.P. 6128 succ. Centre-ville, Montréal, Que., Canada H3C 3J7

<sup>b</sup>Department of Psychology, University of Glasgow, Glasgow G12 8QB, UK



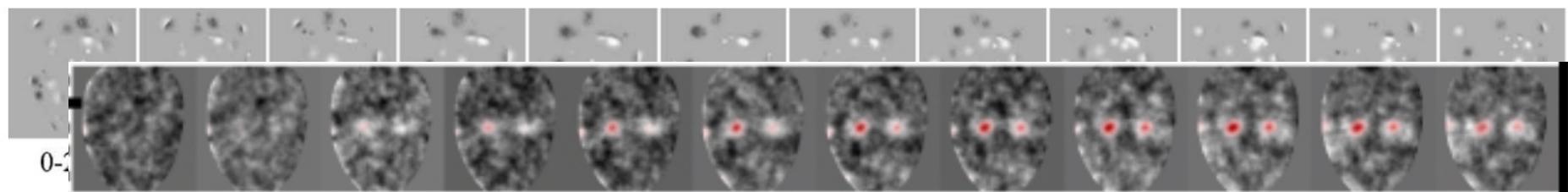
# Explaining AI as we probe cognition

- **Example of applications in vision and audition**

**Bubbles:** analysis of the spatio-temporal course of visual identification of a face

What visual areas do we focus on to identify a face and how does this change over time?

- “dynamic” bubbles



Cognitive Science 28 (2004) 289–301

COGNITIVE  
SCIENCE

<http://www.elsevier.com/locate/cogsci>

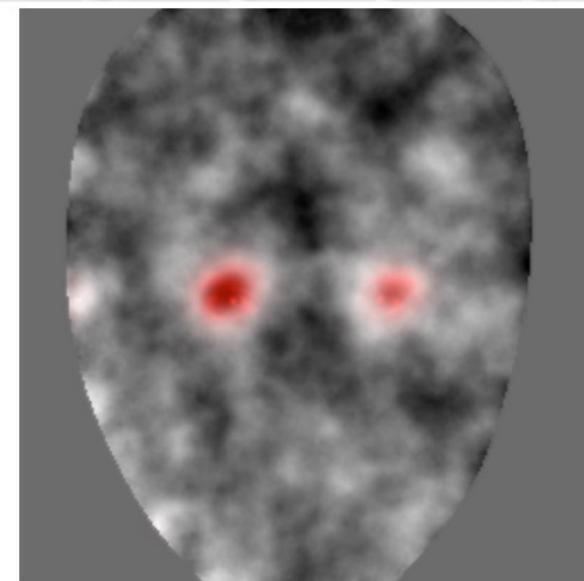
Spatio-temporal dynamics of face recognition  
in a flash: it's in the eyes

Céline Vinette, Frédéric Gosselin <sup>a,\*</sup>, Philippe G. Schyns <sup>b</sup>

<sup>a</sup>Département de psychologie, Université de Montréal, 90 Vincent-d’Indy,

C.P. 6128 succ. Centre-ville, Montréal, Que., Canada H3C 3J7

<sup>b</sup>Department of Psychology, University of Glasgow, Glasgow G12 8QB, UK



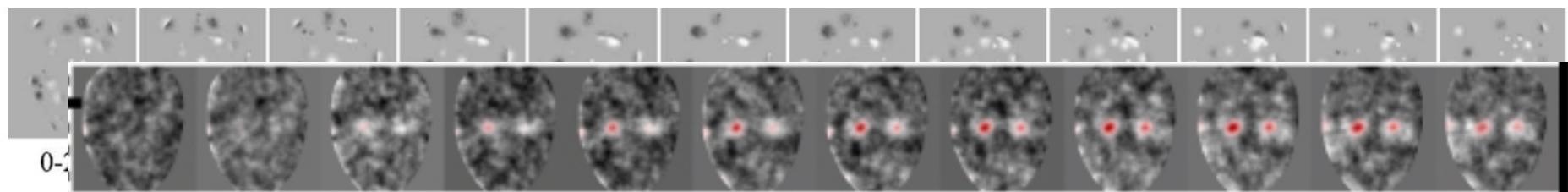
# Explaining AI as we probe cognition

- **Example of applications in vision and audition**

**Bubbles:** analysis of the spatio-temporal course of visual identification of a face

What visual areas do we focus on to identify a face and how does this change over time?

- “dynamic” bubbles



Cognitive Science 28 (2004) 289–301

COGNITIVE  
SCIENCE

<http://www.elsevier.com/locate/cogsci>

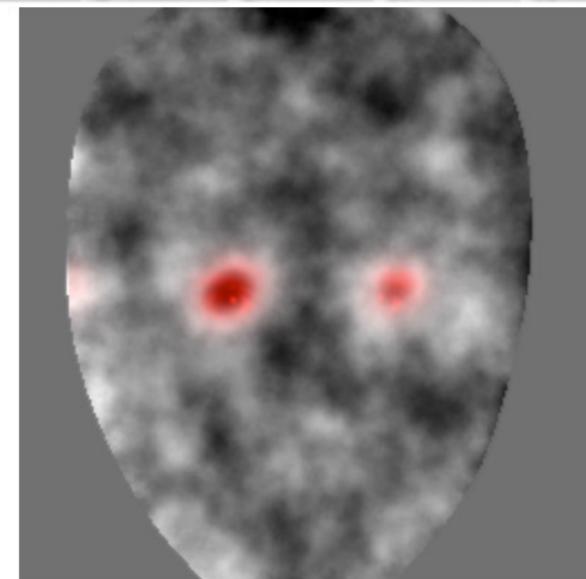
Spatio-temporal dynamics of face recognition  
in a flash: it's in the eyes

Céline Vinette, Frédéric Gosselin <sup>a,\*</sup>, Philippe G. Schyns <sup>b</sup>

<sup>a</sup>Département de psychologie, Université de Montréal, 90 Vincent-d'Indy,

C.P. 6128 succ. Centre-ville, Montréal, Québec, Canada H3C 3J7

<sup>b</sup>Department of Psychology, University of Glasgow, Glasgow G12 8QB, UK



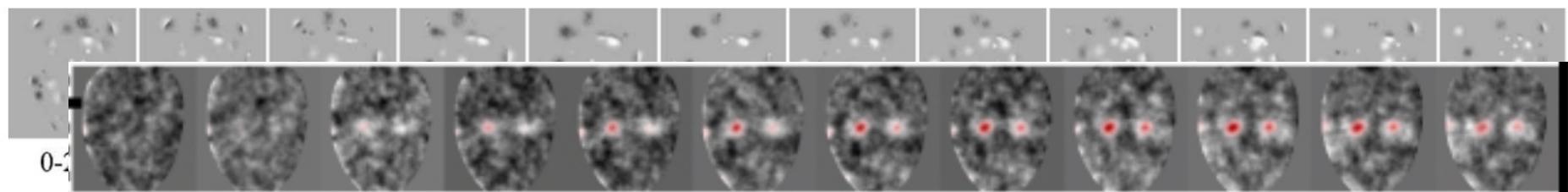
# Explaining AI as we probe cognition

- **Example of applications in vision and audition**

**Bubbles:** analysis of the spatio-temporal course of visual identification of a face

What visual areas do we focus on to identify a face and how does this change over time?

- “dynamic” bubbles



Cognitive Science 28 (2004) 289–301

COGNITIVE  
SCIENCE

<http://www.elsevier.com/locate/cogsci>

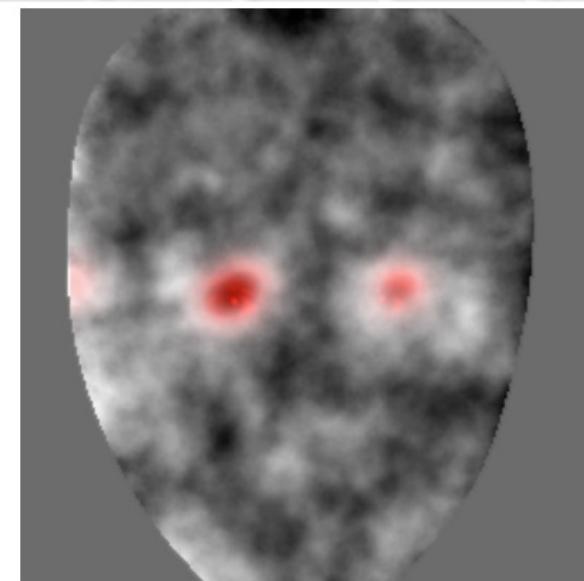
Spatio-temporal dynamics of face recognition  
in a flash: it's in the eyes

Céline Vinette, Frédéric Gosselin <sup>a,\*</sup>, Philippe G. Schyns <sup>b</sup>

<sup>a</sup>Département de psychologie, Université de Montréal, 90 Vincent-d'Indy,

C.P. 6128 succ. Centre-ville, Montréal, Que., Canada H3C 3J7

<sup>b</sup>Department of Psychology, University of Glasgow, Glasgow G12 8QB, UK



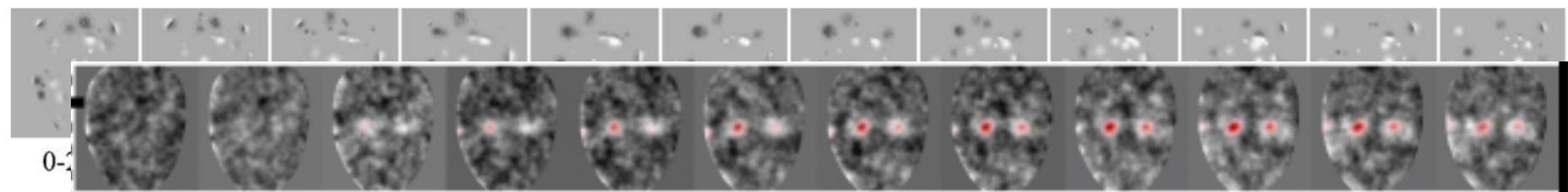
# Explaining AI as we probe cognition

- **Example of applications in vision and audition**

**Bubbles:** analysis of the spatio-temporal course of visual identification of a face

What visual areas do we focus on to identify a face and how does this change over time?

- “dynamic” bubbles



Cognitive Science 28 (2004) 289–301

COGNITIVE  
SCIENCE

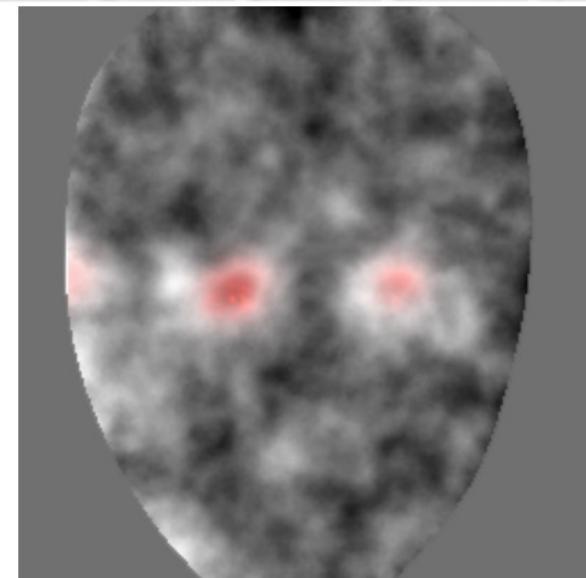
<http://www.elsevier.com/locate/cogsci>

Spatio-temporal dynamics of face recognition  
in a flash: it's in the eyes

Céline Vinette, Frédéric Gosselin <sup>a,\*</sup>, Philippe G. Schyns <sup>b</sup>

<sup>a</sup>Département de psychologie, Université de Montréal, 90 Vincent-d'Indy,  
C.P. 6128 succ. Centre-ville, Montréal, Québec, Canada H3C 3J7

<sup>b</sup>Department of Psychology, University of Glasgow, Glasgow G12 8QB, UK



# Explaining AI as we probe cognition

Journal of Experimental Psychology:  
Animal Behavior Processes  
2005, Vol. 31, No. 3, 376–382

Copyright 2005 by the American Psychological Association  
0897-4036/05/\$12.00 DOI: 10.1037/0897-4033.31.3.376

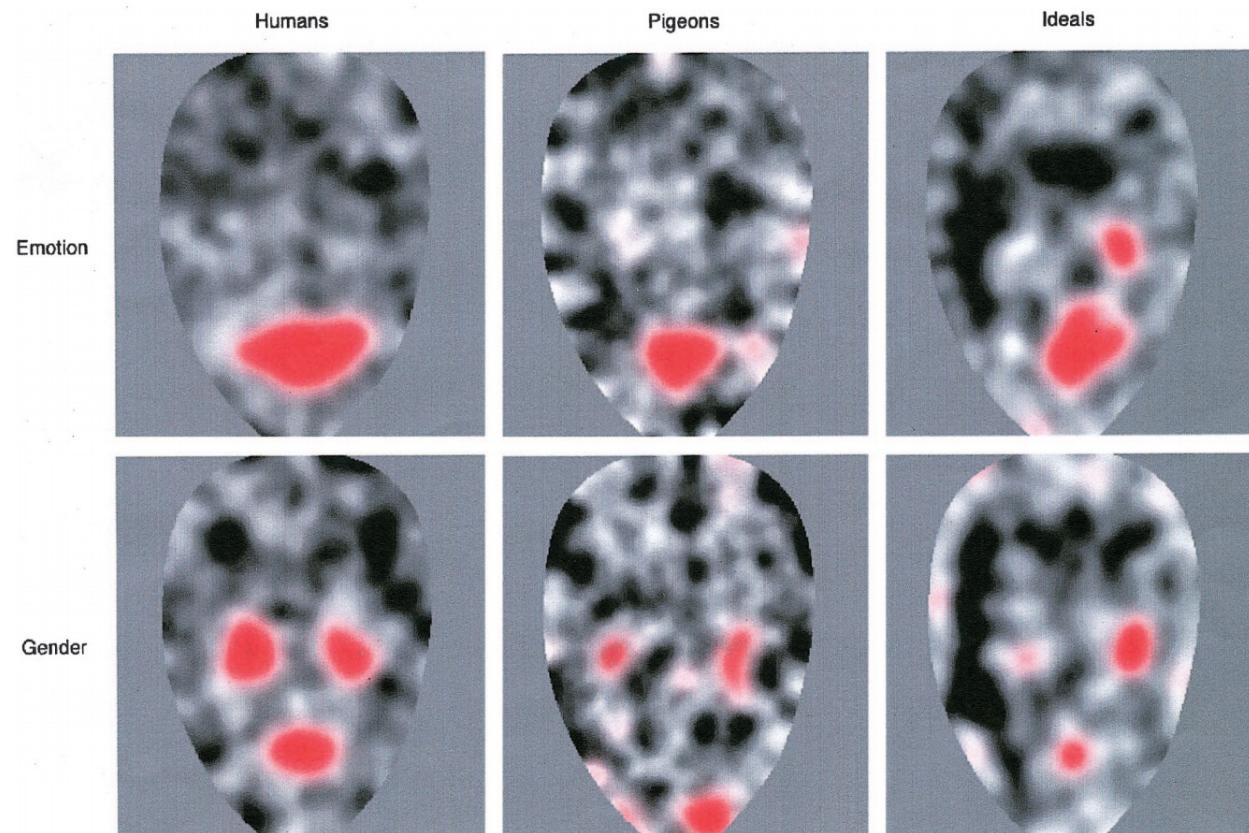
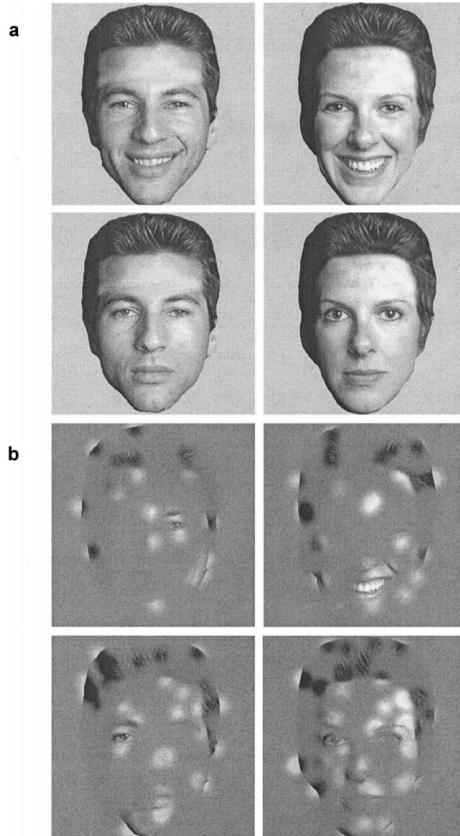
## Applying Bubbles to Localize Features That Control Pigeons' Visual Discrimination Behavior

Brett M. Gibson and Edward A. Wasserman  
The University of Iowa

Frédéric Gosselin  
Université de Montréal

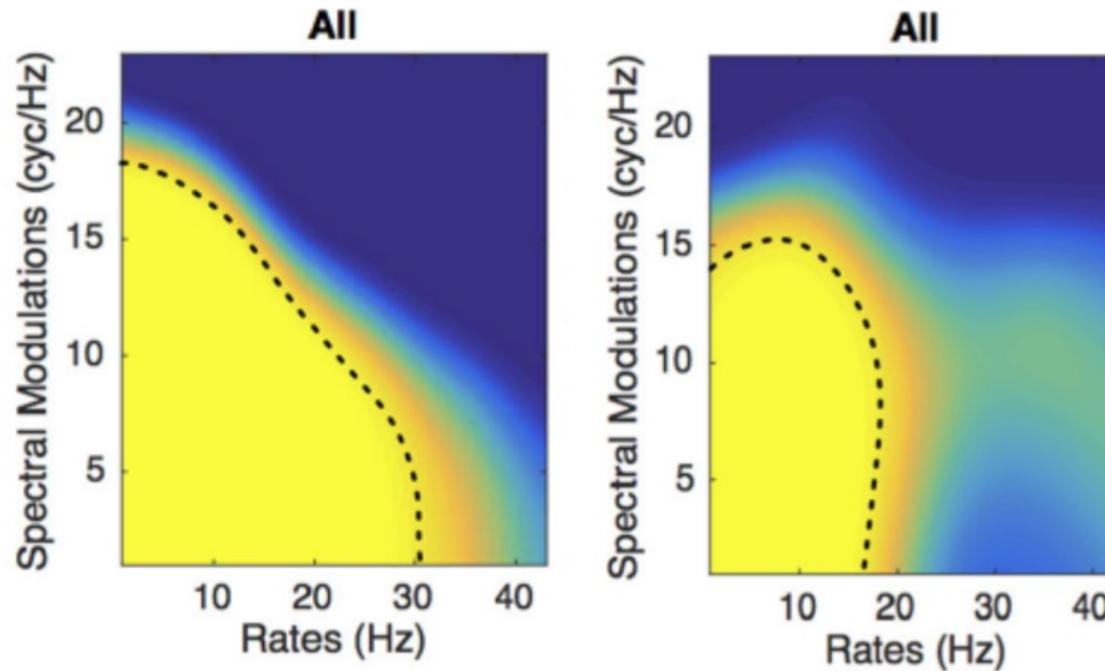
Philippe G. Schyns  
University of Glasgow

Comparison with pigeons!



# Explaining AI as we probe cognition

Spectro-temporal modulations the most relevant for musical instrument timbre identification

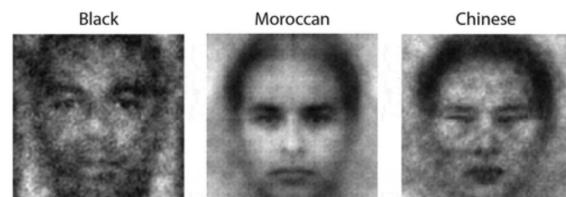
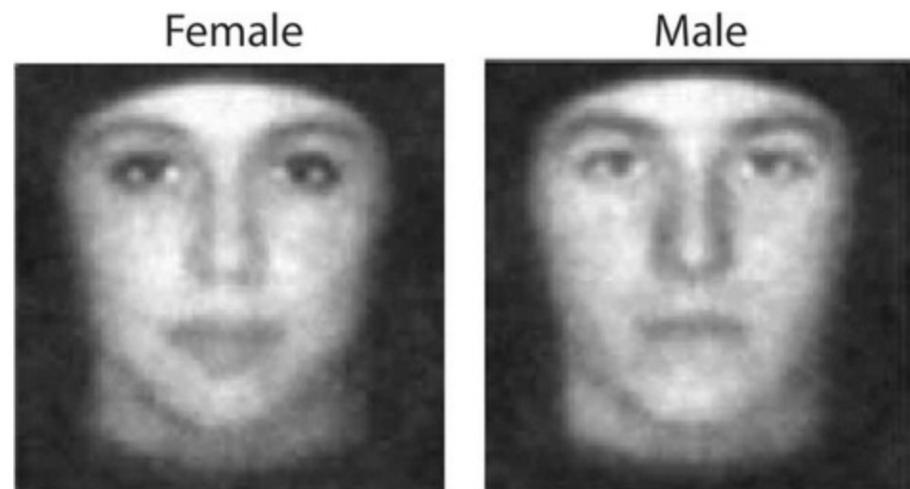
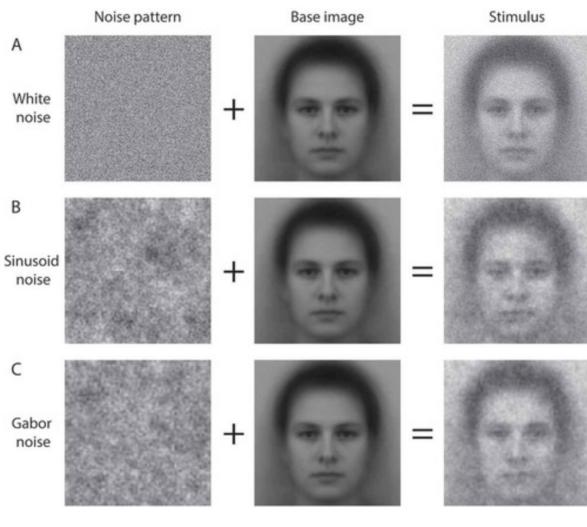


Thoret, Depalle, McAdams (2016) JASA EL  
Thoret, Depalle, McAdams (2017) Frontiers in Psychology

**bubbles**

# Explaining AI as we probe cognition

Social cognition



revcor

# Explaining AI as we probe cognition

revcor



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Vision Research 44 (2004) 1493–1498

**Vision  
Research**

[www.elsevier.com/locate/vires](http://www.elsevier.com/locate/vires)

What makes Mona Lisa smile?

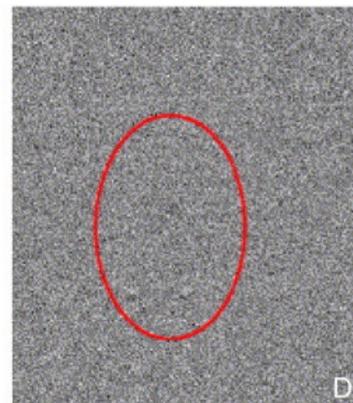
Leonid L. Kontsevich \*, Christopher W. Tyler

*Smith-Kettlewell Eye Research Institute, 2318 Fillmore Street, San Francisco, CA 94115, USA*

Received 19 March 2002; received in revised form 1 October 2003



+



Is Mona Lisa  
smiling?

# Explaining AI as we probe cognition

revcor



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Vision Research 44 (2004) 1493–1498

**Vision  
Research**

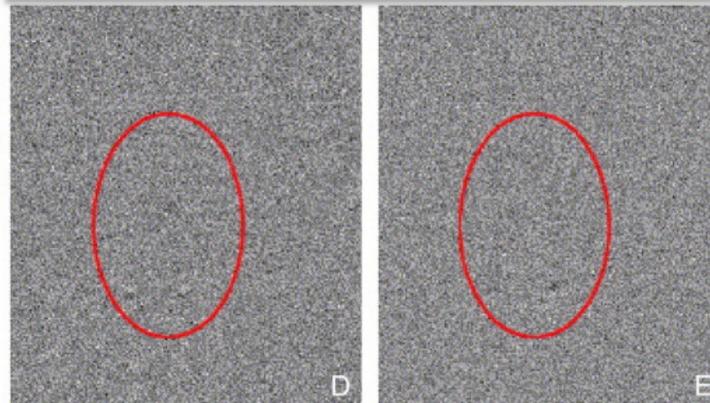
[www.elsevier.com/locate/vires](http://www.elsevier.com/locate/vires)



## Two additive masks

Smiling

Non Smiling



# Explaining AI as we probe cognition



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Vision Research 44 (2004) 1493–1498

**Vision  
Research**

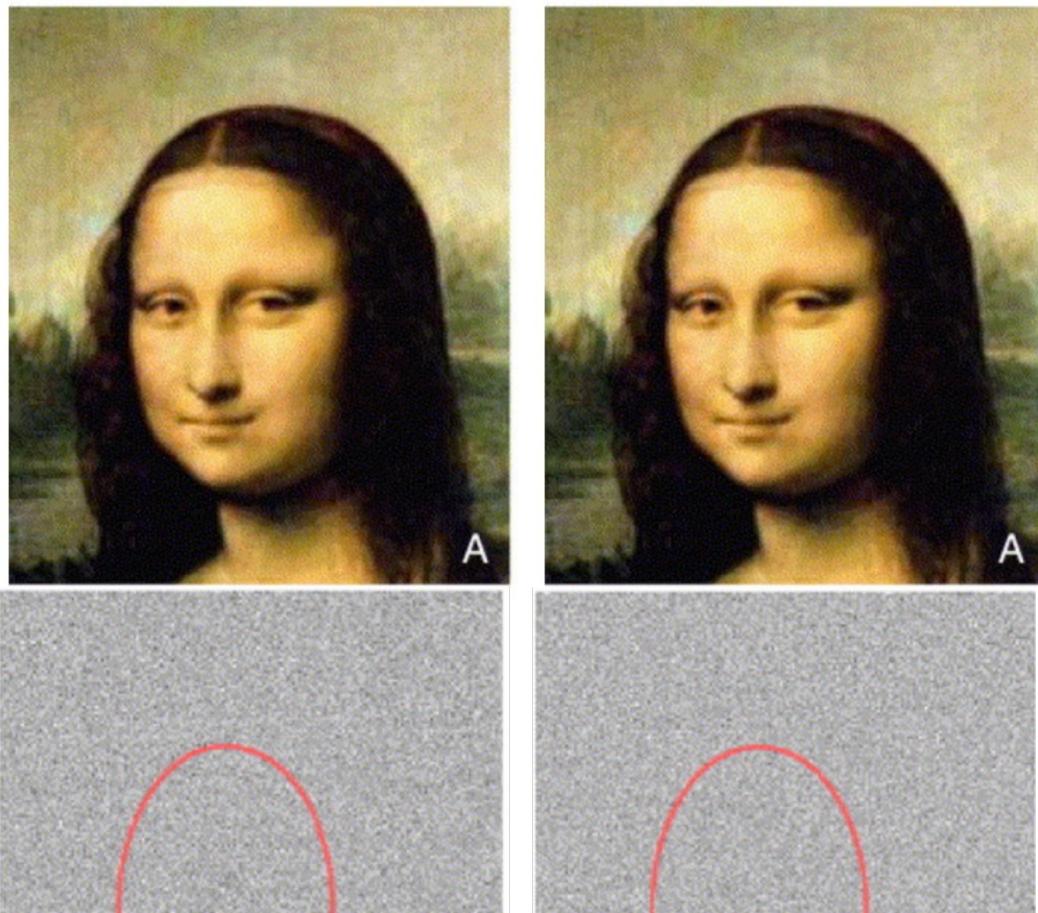
[www.elsevier.com/locate/vires](http://www.elsevier.com/locate/vires)

What makes Mona Lisa smile?

Leonid L. Kontsevich \*, Christopher W. Tyler

*Smith-Kettlewell Eye Research Institute, 2318 Fillmore Street, San Francisco, CA 94115, USA*

Received 19 March 2002; received in revised form 1 October 2003



# Explaining AI as we probe cognition



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Vision Research 44 (2004) 1493–1498

**Vision  
Research**

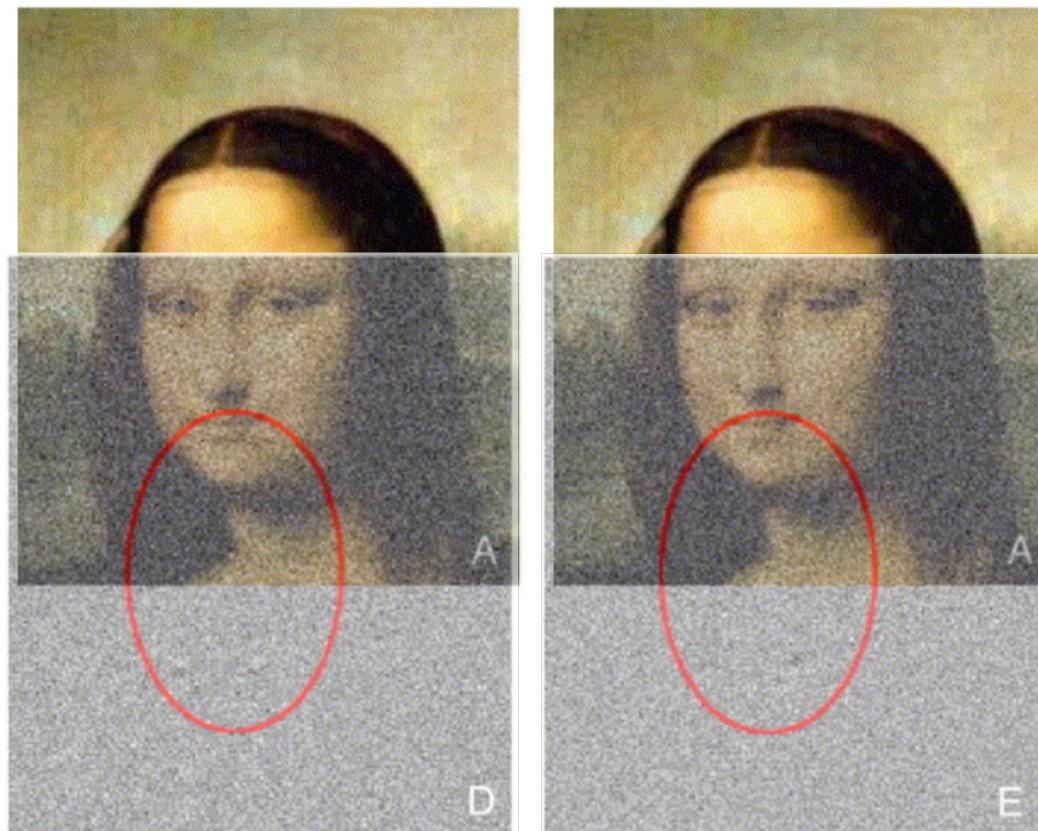
[www.elsevier.com/locate/vires](http://www.elsevier.com/locate/vires)

What makes Mona Lisa smile?

Leonid L. Kontsevich \*, Christopher W. Tyler

*Smith-Kettlewell Eye Research Institute, 2318 Fillmore Street, San Francisco, CA 94115, USA*

Received 19 March 2002; received in revised form 1 October 2003



# Explaining AI as we probe cognition



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Vision Research 44 (2004) 1493–1498

**Vision  
Research**

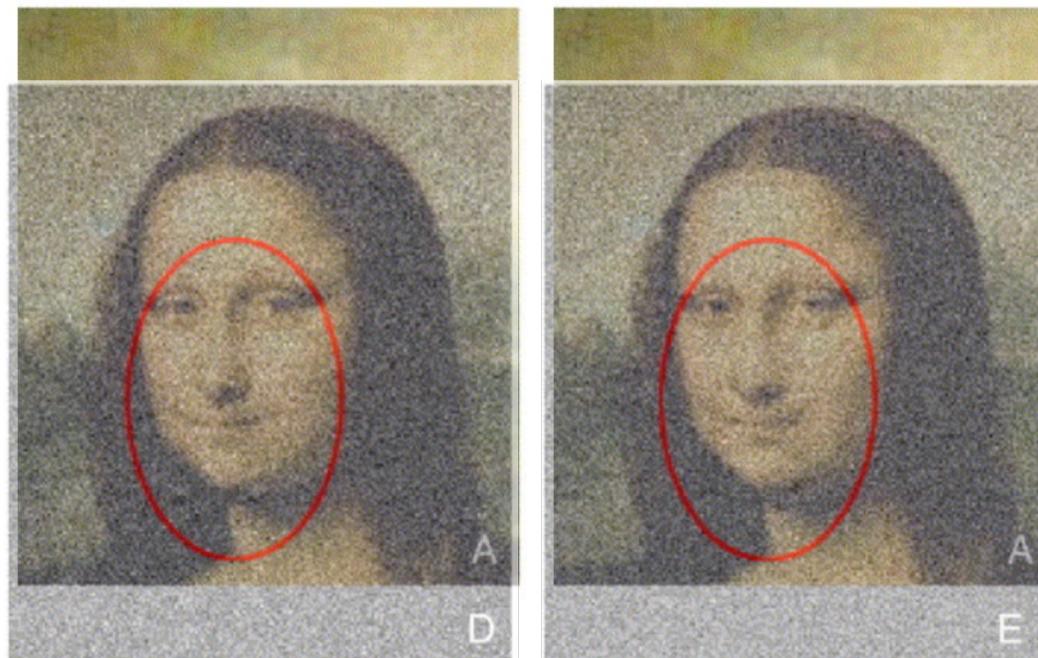
[www.elsevier.com/locate/vires](http://www.elsevier.com/locate/vires)

What makes Mona Lisa smile?

Leonid L. Kontsevich \*, Christopher W. Tyler

*Smith-Kettlewell Eye Research Institute, 2318 Fillmore Street, San Francisco, CA 94115, USA*

Received 19 March 2002; received in revised form 1 October 2003



# Explaining AI as we probe cognition



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Vision Research 44 (2004) 1493–1498

**Vision  
Research**

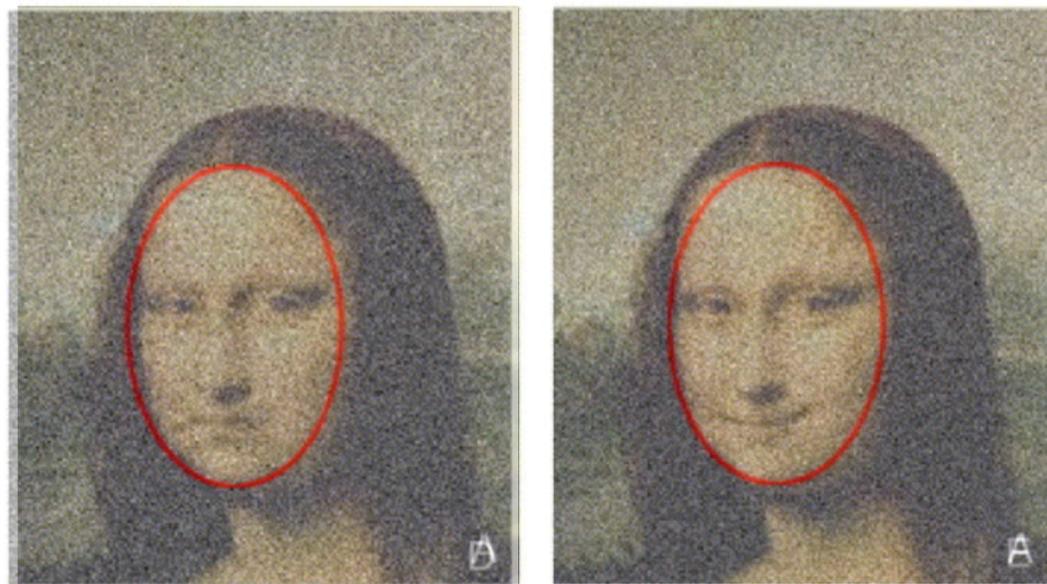
[www.elsevier.com/locate/vires](http://www.elsevier.com/locate/vires)

What makes Mona Lisa smile?

Leonid L. Kontsevich \*, Christopher W. Tyler

*Smith-Kettlewell Eye Research Institute, 2318 Fillmore Street, San Francisco, CA 94115, USA*

Received 19 March 2002; received in revised form 1 October 2003



# Explaining AI as we probe cognition



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Vision Research 44 (2004) 1493–1498

**Vision  
Research**

[www.elsevier.com/locate/vires](http://www.elsevier.com/locate/vires)

What makes Mona Lisa smile?

Leonid L. Kontsevich \*, Christopher W. Tyler

*Smith-Kettlewell Eye Research Institute, 2318 Fillmore Street, San Francisco, CA 94115, USA*

Received 19 March 2002; received in revised form 1 October 2003



# Explaining AI as we probe cognition



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Vision Research 44 (2004) 1493–1498

**Vision  
Research**

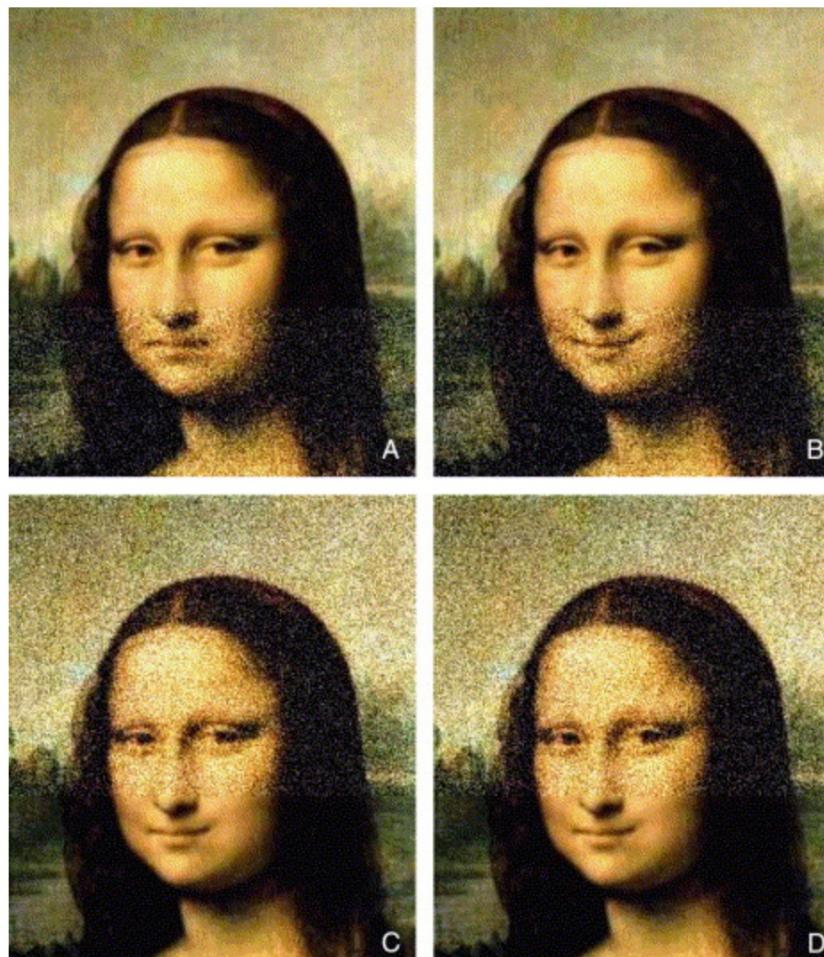
[www.elsevier.com/locate/vires](http://www.elsevier.com/locate/vires)

What makes Mona Lisa smile?

Leonid L. Kontsevich \*, Christopher W. Tyler

*Smith-Kettlewell Eye Research Institute, 2318 Fillmore Street, San Francisco, CA 94115, USA*

Received 19 March 2002; received in revised form 1 October 2003

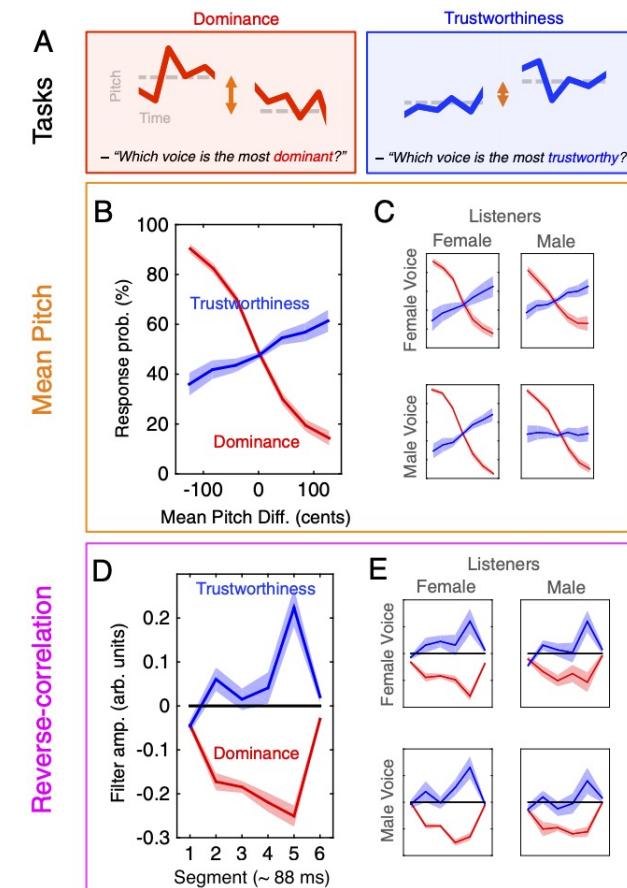
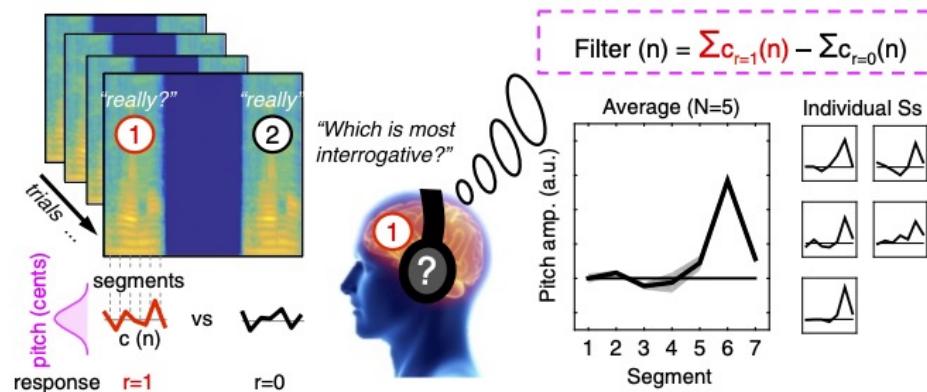


# Explaining AI as we probe cognition

## Cracking the social code of speech prosody using reverse correlation

Emmanuel Ponsot<sup>a,b,c,1</sup>, Juan José Burred<sup>d</sup>, Pascal Belin<sup>e,f,g,2</sup>, and Jean-Julien Aucouturier<sup>a,h,2</sup>

<sup>a</sup>Sciences et Technologies de la Musique et du Son, UMR 9912, Institut de Recherche et Coordination Acoustique/Musique, CNRS and Sorbonne Université, 75004 Paris, France; <sup>b</sup>Laboratoire des Systèmes Perceptifs, UMR 8248, Ecole Normale Supérieure, Paris Sciences et Lettres Research University, 75005 Paris, France; <sup>c</sup>Département d'Études Cognitives, Ecole Normale Supérieure, Paris Sciences et Lettres Research University, 75005 Paris, France; <sup>d</sup>Independent Researcher, 75013 Paris, France; <sup>e</sup>Institut de Neurosciences de la Timone, UMR 7289, Centre National de la Recherche Scientifique and Aix-Marseille Université, 13007 Marseille, France; <sup>f</sup>Institute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QQ, United Kingdom; <sup>g</sup>Département



# Explaining AI as we probe cognition

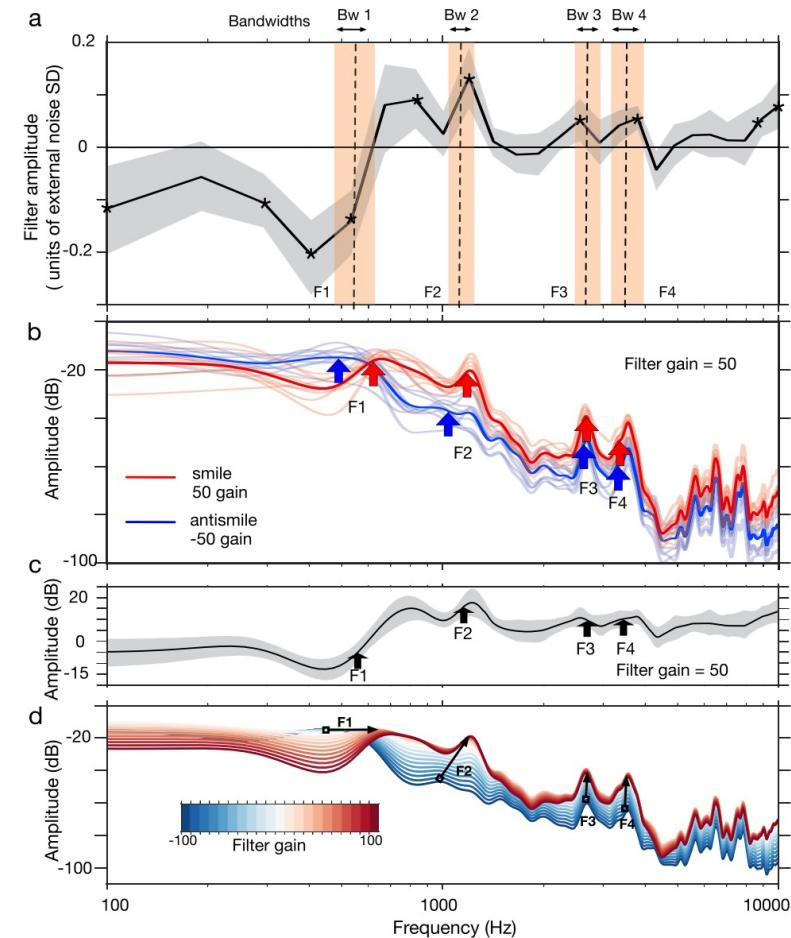
Ponsot et al.: JASA Express Letters

<https://doi.org/10.1121/1.5020989>

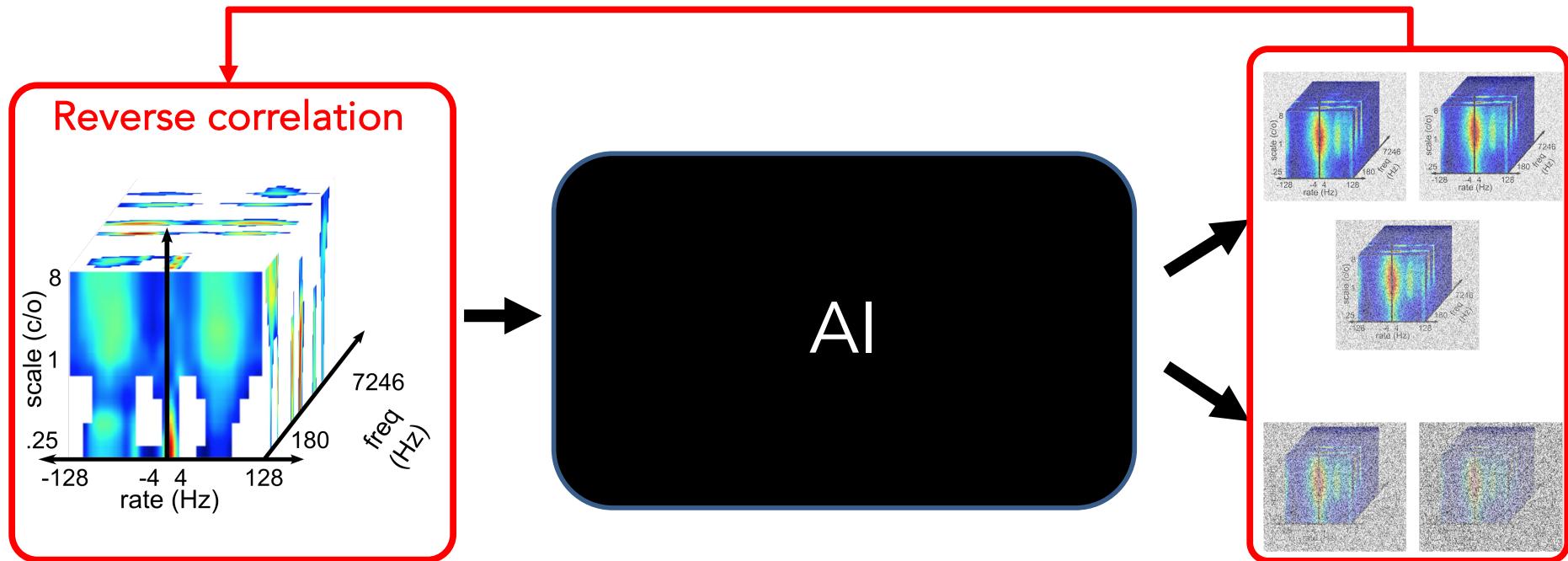
Published Online 19 January 2018

## Uncovering mental representations of smiled speech using reverse correlation

Emmanuel Ponsot,<sup>a)</sup> Pablo Arias, and Jean-Julien Aucouturier  
STMS (Sciences et Technologies de la Musique et du Son) Lab (Ircam/CNRS/UPMC),  
1 place Igor Stravinsky, Paris, France  
ponsot@ircam.fr, arias@ircam.fr, aucouturier@ircam.fr



# Explaining AI as we probe cognition

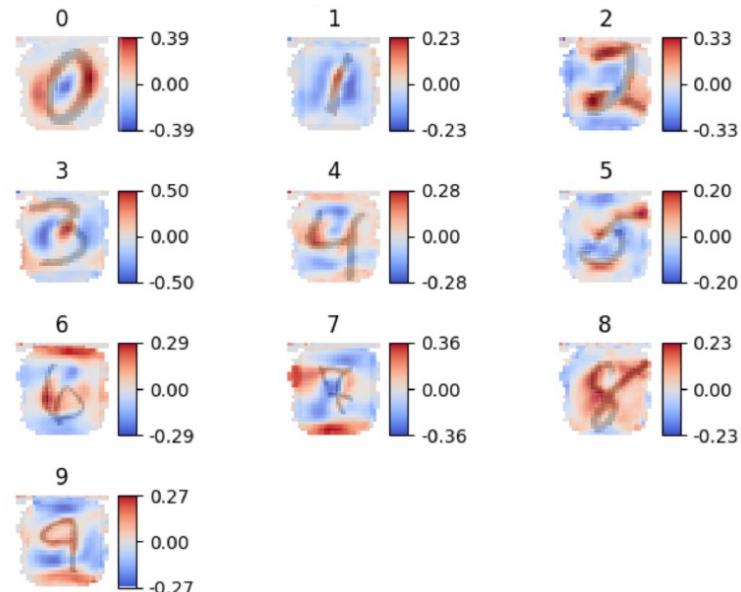


Same idea with classifiers

Thoret, E., Andrillon, T., Leger, D., Pressnitzer, D. (2021)  
Probing machine-learning classifiers using noise, bubbles, and reverse correlation  
*J. Neuro Methods*

# A versatile approach for AI

## Images classification



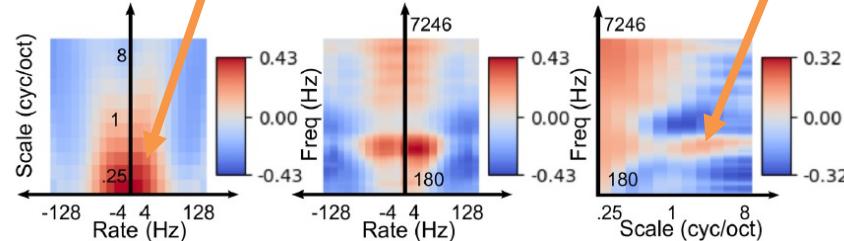
IMAGES

Thoret, E., Andrillon, T., Leger, D., Pressnitzer, D. (2021)  
Probing machine-learning classifiers using noise, bubbles, and reverse correlation  
*J. Neuro Methods*

# A versatile approach for AI

Images classification

speech/music  
characterization  
Prosody

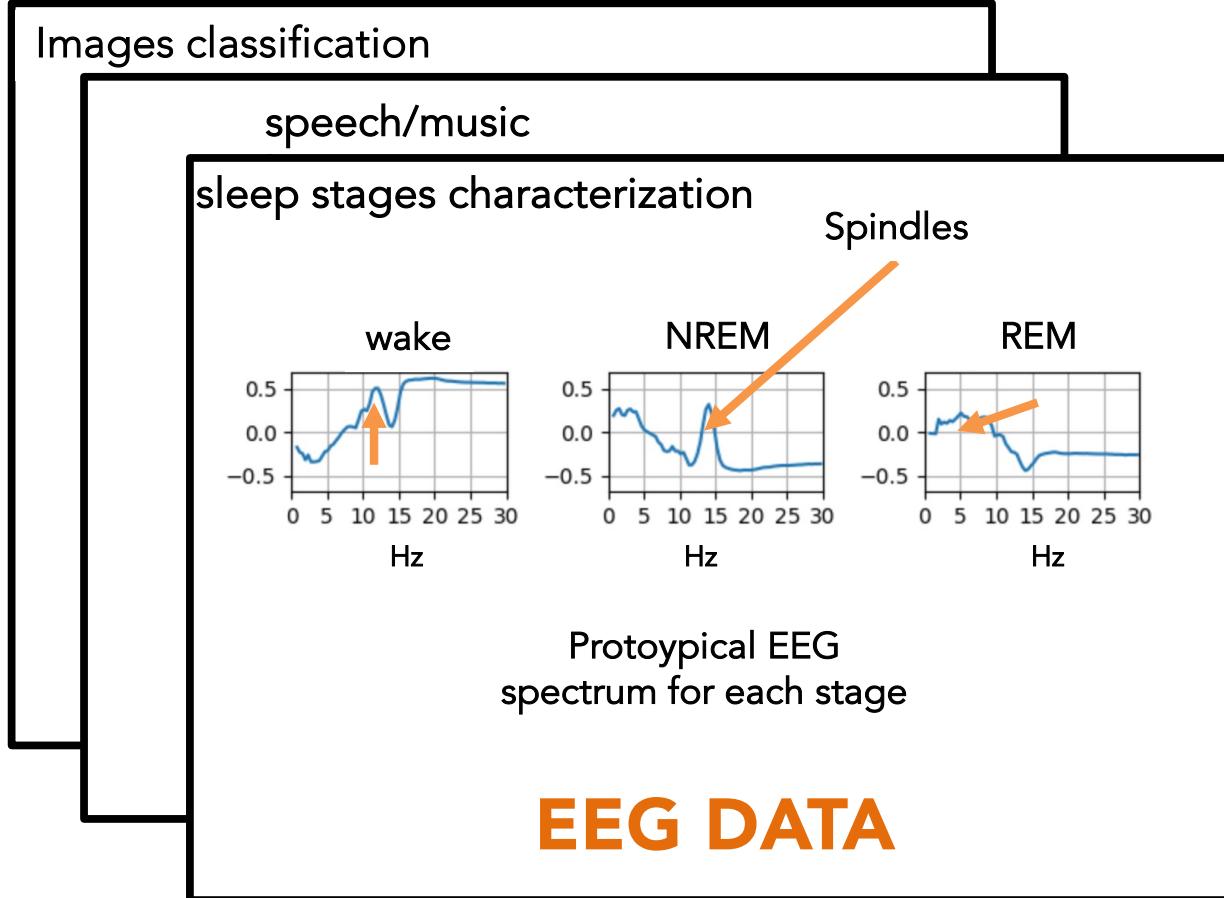


Speech prototypical  
STRF

**ACOUSTIC  
REPRESENTATIONS**

Thoret, E., Andrillon, T., Leger, D., Pressnitzer, D. (2021)  
Probing machine-learning classifiers using noise, bubbles, and reverse correlation  
*J. Neuro Methods*

# A versatile approach for AI



Thoret, E., Andrillon, T., Leger, D., Pressnitzer, D. (2021)  
Probing machine-learning classifiers using noise, bubbles, and reverse correlation  
*J. Neuro Methods*

# **Sleep deprivation detected by voice analysis**

# Sleep deprivation detected by voice analysis



Health

AI can tell if you are sleep deprived by listening to your voice



Lecture à  
voix haute

Detecting sleepiness in voice



PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

## Sleep deprivation detected by voice analysis

Etienne Thoret<sup>1,2,3\*</sup>, Thomas Andrilhon<sup>4,5,6</sup>, Caroline Gauriau<sup>5,6</sup>, Damien Léger<sup>5,6†</sup>, Daniel Pressnitzer<sup>1,‡</sup>

**1** Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France, **2** Aix-Marseille University, CNRS, Institut de Neurosciences de la Timone (INT) UMR7289, Perception Representation Image Sound Music (PRISM) UMR7061, Laboratoire d'Informatique et Systèmes (LIS) UMR7020, Marseille, France, **3** Institute of Language Communication and the Brain, Aix-Marseille University, Marseille, France, **4** Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Mov'it team, Inserm, CNRS, Paris, France, **5** Université Paris Cité, VIFASOM, ERC 7330, Vigilance Fatigue Sommeil et santé publique, Paris, France, **6** APHP, Hôtel-Dieu, Centre du Sommeil et de la Vigilance, Paris, France

\* These authors joint last first authorship on this work.

† etienne.thoret@univ-amu.fr



Explaining AI to create  
vocal biomarkers

# Sleep deprivation detected by voice analysis



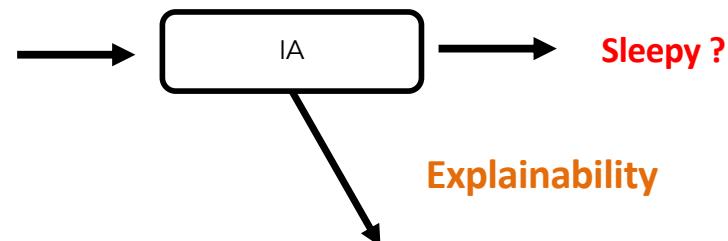
Health

AI can tell if you are sleep deprived by listening to your voice

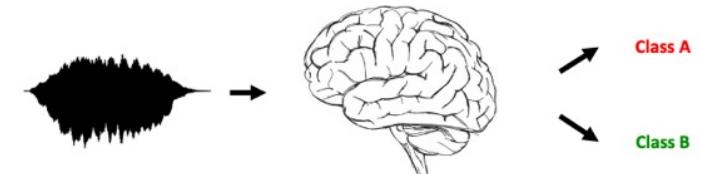


Lecture à  
voix haute

Detecting sleepiness in voice



Reverse correlation:  
« probing AI as we probe brain »



PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

## Sleep deprivation detected by voice analysis

Etienne Thoret<sup>1,2,3\*</sup>, Thomas Andrilhon<sup>4,5,6</sup>, Caroline Gauriau<sup>5,6</sup>, Damien Léger<sup>5,6†</sup>, Daniel Pressnitzer<sup>1,‡</sup>

**1** Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France, **2** Aix-Marseille University, CNRS, Institut de Neurosciences de la Timone (INT) UMR7289, Perception Representation Image Sound Music (PRISM) UMR7061, Laboratoire d'Informatique et Systèmes (LIS) UMR7202, Marseille, France, **3** Institute of Language Communication and the Brain, Aix-Marseille University, Marseille, France, **4** Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Mov'itteam, Inserm, CNRS, Paris, France, **5** Université Paris Cité, VIFASOM, ERC 7330, Vigilance Fatigue Sommeil et santé publique, Paris, France, **6** APHP, Hôtel-Dieu, Centre du Sommeil et de la Vigilance, Paris, France

\* These authors joint last first authorship on this work.

† etienne.thoret@univ-amu.fr



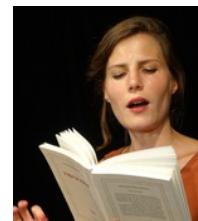
Explaining AI to create  
vocal biomarkers

# Sleep deprivation detected by voice analysis



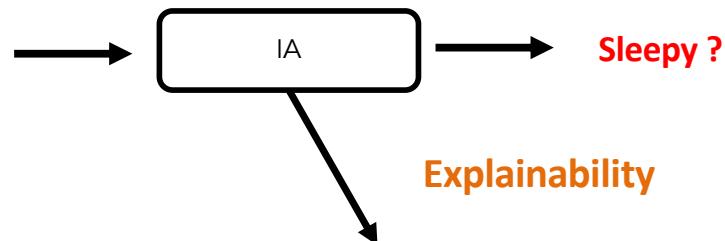
Health

AI can tell if you are sleep deprived by listening to your voice



Lecture à  
voix haute

Detecting sleepiness in voice



Reverse correlation:  
« probing AI as we probe brain »



PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

## Sleep deprivation detected by voice analysis

Etienne Thoret<sup>1,2,3\*</sup>, Thomas Andrilhon<sup>4,5,6</sup>, Caroline Gauriau<sup>5,6</sup>, Damien Léger<sup>5,6†</sup>, Daniel Pressnitzer<sup>1,‡</sup>

**1** Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France, **2** Aix-Marseille University, CNRS, Institut de Neurosciences de la Timone (INT) UMR7289, Perception Representation Image Sound Music (PRISM) UMR7061, Laboratoire d'Informatique et Systèmes (LIS) UMR7202, Marseille, France, **3** Institute of Language Communication and the Brain, Aix-Marseille University, Marseille, France, **4** Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Mov'itteam, Inserm, CNRS, Paris, France, **5** Université Paris Cité, VIFASOM, ERC 7330, Vigilance Fatigue Sommeil et santé publique, Paris, France, **6** APHP, Hôtel-Dieu, Centre du Sommeil et de la Vigilance, Paris, France

\* These authors joint last first authorship on this work.

† etienne.thoret@univ-amu.fr



Explaining AI to create  
vocal biomarkers

# Sleep deprivation detected by voice analysis



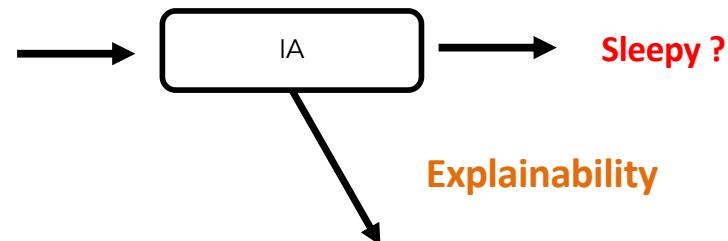
Health

AI can tell if you are sleep deprived by listening to your voice



Lecture à  
voix haute

Detecting sleepiness in voice



Reverse correlation:  
« probing AI as we probe brain »



PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

## Sleep deprivation detected by voice analysis

Etienne Thoret<sup>1,2,3\*</sup>, Thomas Andrilhon<sup>4,5,6</sup>, Caroline Gauriau<sup>5,6</sup>, Damien Léger<sup>5,6†</sup>, Daniel Pressnitzer<sup>1,‡</sup>

**1** Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France, **2** Aix-Marseille University, CNRS, Institut de Neurosciences de la Timone (INT) UMR7289, Perception Representation Image Sound Music (PRISM) UMR7061, Laboratoire d'Informatique et Systèmes (LIS) UMR7202, Marseille, France, **3** Institute of Language Communication and the Brain, Aix-Marseille University, Marseille, France, **4** Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Mov'it team, Inserm, CNRS, Paris, France, **5** Université Paris Cité, VIFASOM, ERC 7330, Vigilance Fatigue Sommeil et santé publique, Paris, France, **6** APHP, Hôtel-Dieu, Centre du Sommeil et de la Vigilance, Paris, France

\* These authors joint last first authorship on this work.

† etienne.thoret@univ-amu.fr



Explaining AI to create  
vocal biomarkers

# Sleep deprivation detected by voice analysis



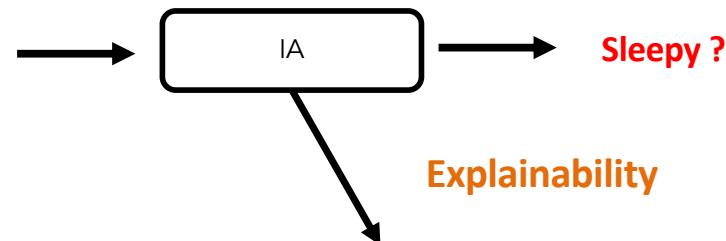
Health

AI can tell if you are sleep deprived by listening to your voice

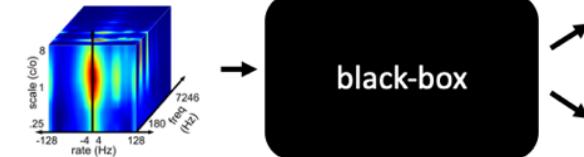


Lecture à  
voix haute

Detecting sleepiness in voice



Reverse correlation:  
« probing AI as we probe brain »



PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

## Sleep deprivation detected by voice analysis

Etienne Thoret<sup>1,2,3\*</sup>, Thomas Andrilhon<sup>4,5,6</sup>, Caroline Gauriau<sup>5,6</sup>, Damien Léger<sup>5,6†</sup>, Daniel Pressnitzer<sup>1,‡</sup>

**1** Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France, **2** Aix-Marseille University, CNRS, Institut de Neurosciences de la Timone (INT) UMR7289, Perception Representation Image Sound Music (PRISM) UMR7061, Laboratoire d'Informatique et Systèmes (LIS) UMR7202, Marseille, France, **3** Institute of Language Communication and the Brain, Aix-Marseille University, Marseille, France, **4** Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Mov'it team, Inserm, CNRS, Paris, France, **5** Université Paris Cité, VIFASOM, ERC 7330, Vigilance Fatigue Sommeil et santé publique, Paris, France, **6** APHP, Hôtel-Dieu, Centre du Sommeil et de la Vigilance, Paris, France

\* These authors joint last first authorship on this work.

† etienne.thoret@univ-amu.fr



Explaining AI to create  
vocal biomarkers

# Sleep deprivation detected by voice analysis



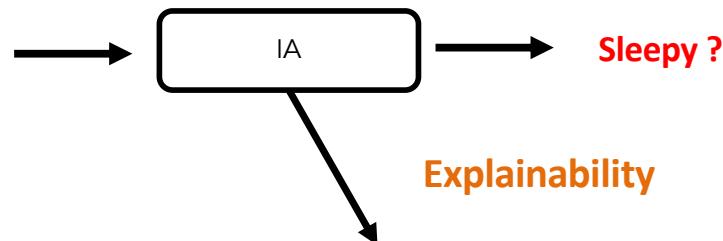
Health

AI can tell if you are sleep deprived by listening to your voice

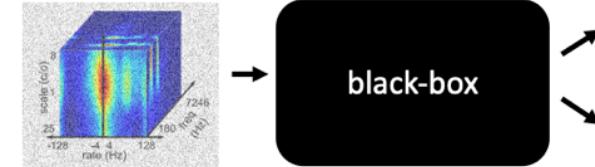


Lecture à  
voix haute

Detecting sleepiness in voice



Reverse correlation:  
« probing AI as we probe brain »



PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

## Sleep deprivation detected by voice analysis

Etienne Thoret<sup>1,2,3\*</sup>, Thomas Andrilhon<sup>4,5,6</sup>, Caroline Gauriau<sup>5,6</sup>, Damien Léger<sup>5,6†</sup>, Daniel Pressnitzer<sup>1,‡</sup>

**1** Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France, **2** Aix-Marseille University, CNRS, Institut de Neurosciences de la Timone (INT) UMR7289, Perception Representation Image Sound Music (PRISM) UMR7061, Laboratoire d'Informatique et Systèmes (LIS) UMR7020, Marseille, France, **3** Institute of Language Communication and the Brain, Aix-Marseille University, Marseille, France, **4** Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Mov'it team, Inserm, CNRS, Paris, France, **5** Université Paris Cité, VIFASOM, ERC 7330, Vigilance Fatigue Sommeil et santé publique, Paris, France, **6** APHP, Hôpital-Dieu, Centre du Sommeil et de la Vigilance, Paris, France

\* These authors joint last first authorship on this work.

† etienne.thoret@univ-amu.fr



Explaining AI to create  
vocal biomarkers

# Sleep deprivation detected by voice analysis



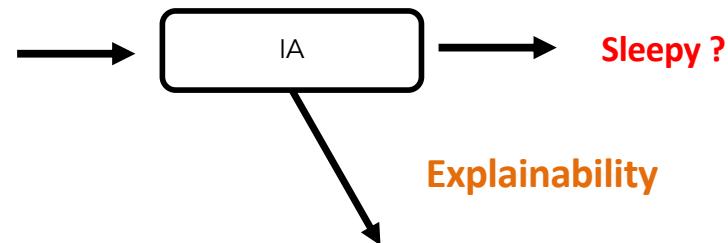
Health

AI can tell if you are sleep deprived by listening to your voice

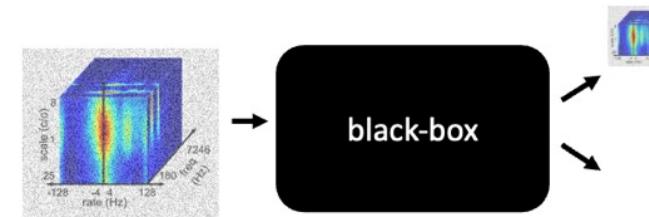


Lecture à  
voix haute

Detecting sleepiness in voice



Reverse correlation:  
« probing AI as we probe brain »



PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

## Sleep deprivation detected by voice analysis

Etienne Thoret<sup>1,2,3\*</sup>, Thomas Andrilhon<sup>4,5,6</sup>, Caroline Gauriau<sup>5,6</sup>, Damien Léger<sup>5,6†</sup>, Daniel Pressnitzer<sup>1,‡</sup>

**1** Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France, **2** Aix-Marseille University, CNRS, Institut de Neurosciences de la Timone (INT) UMR7289, Perception Representation Image Sound Music (PRISM) UMR7061, Laboratoire d'Informatique et Systèmes (LIS) UMR7020, Marseille, France, **3** Institute of Language Communication and the Brain, Aix-Marseille University, Marseille, France, **4** Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Mov'it team, Inserm, CNRS, Paris, France, **5** Université Paris Cité, VIFASOM, ERC 7330, Vigilance Fatigue Sommeil et santé publique, Paris, France, **6** APHP, Hôtel-Dieu, Centre du Sommeil et de la Vigilance, Paris, France

\* These authors joint last first authorship on this work.

† etienne.thoret@univ-amu.fr



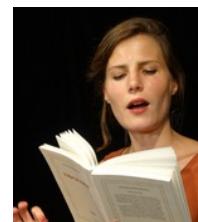
Explaining AI to create  
vocal biomarkers

# Sleep deprivation detected by voice analysis



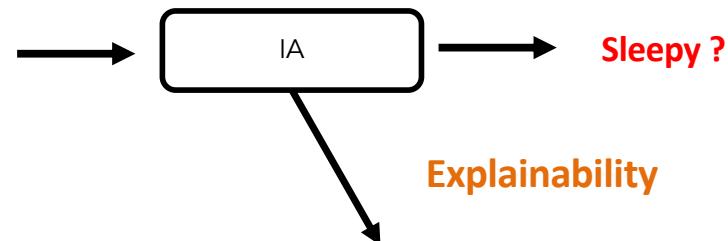
Health

AI can tell if you are sleep deprived by listening to your voice

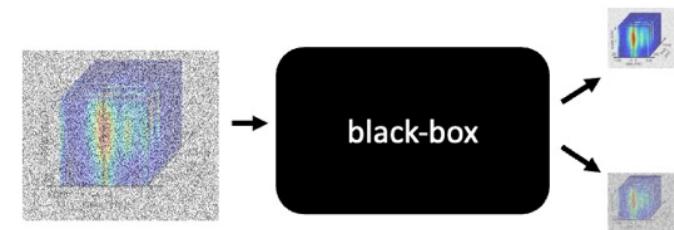


Lecture à  
voix haute

Detecting sleepiness in voice



Reverse correlation:  
« probing AI as we probe brain »



PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

## Sleep deprivation detected by voice analysis

Etienne Thoret<sup>1,2,3\*</sup>, Thomas Andrilhon<sup>4,5,6</sup>, Caroline Gauriau<sup>5,6</sup>, Damien Léger<sup>5,6†</sup>, Daniel Pressnitzer<sup>1,‡</sup>

**1** Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France, **2** Aix-Marseille University, CNRS, Institut de Neurosciences de la Timone (INT) UMR7289, Perception Representation Image Sound Music (PRISM) UMR7061, Laboratoire d'Informatique et Systèmes (LIS) UMR7020, Marseille, France, **3** Institute of Language Communication and the Brain, Aix-Marseille University, Marseille, France, **4** Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Mov'itteam, Inserm, CNRS, Paris, France, **5** Université Paris Cité, VIFASOM, ERC 7330, Vigilance Fatigue Sommeil et santé publique, Paris, France, **6** APHP, Hôpital-Dieu, Centre du Sommeil et de la Vigilance, Paris, France

\* These authors joint last first authorship on this work.

† etienne.thoret@univ-amu.fr



Explaining AI to create  
vocal biomarkers

# Sleep deprivation detected by voice analysis



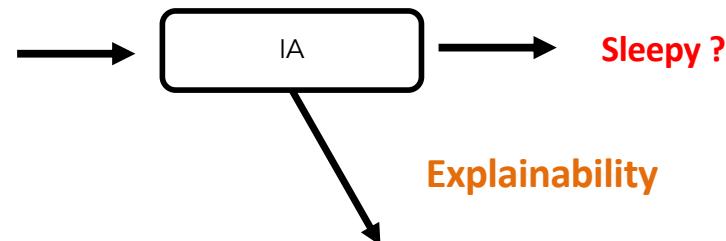
Health

AI can tell if you are sleep deprived by listening to your voice

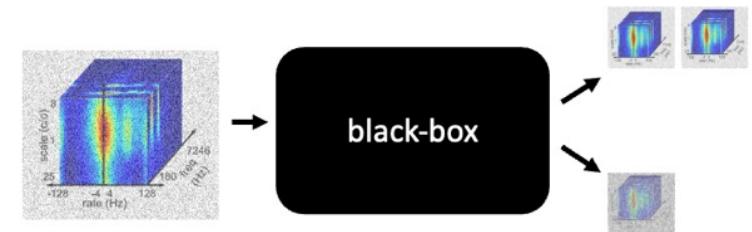


Lecture à  
voix haute

Detecting sleepiness in voice



Reverse correlation:  
« probing AI as we probe brain »



PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

## Sleep deprivation detected by voice analysis

Etienne Thoret<sup>1,2,3\*</sup>, Thomas Andrilhon<sup>4,5,6</sup>, Caroline Gauriau<sup>5,6</sup>, Damien Léger<sup>5,6†</sup>, Daniel Pressnitzer<sup>1,‡</sup>

1 Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France, 2 Aix-Marseille University, CNRS, Institut de Neurosciences de la Timone (INT) UMR7289, Perception Representation Image Sound Music (PRISM) UMR7061, Laboratoire d'Informatique et Systèmes (LIS) UMR7020, Marseille, France, 3 Institute of Language Communication and the Brain, Aix-Marseille University, Marseille, France, 4 Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Mov'it team, Inserm, CNRS, Paris, France, 5 Université Paris Cité, VIFASOM, ERC 7330, Vigilance Fatigue Sommeil et santé publique, Paris, France, 6 APHP, Hôtel-Dieu, Centre du Sommeil et de la Vigilance, Paris, France

\* These authors joint last first authorship on this work.

† etienne.thoret@univ-amu.fr



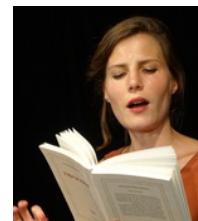
Explaining AI to create  
vocal biomarkers

# Sleep deprivation detected by voice analysis



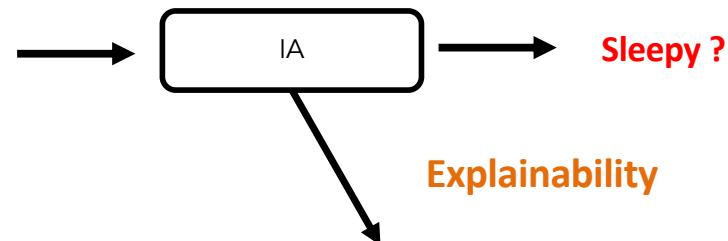
Health

AI can tell if you are sleep deprived by listening to your voice

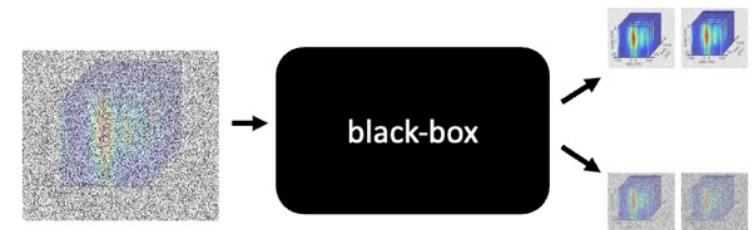


Lecture à  
voix haute

Detecting sleepiness in voice



Reverse correlation:  
« probing AI as we probe brain »



PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

## Sleep deprivation detected by voice analysis

Etienne Thoret<sup>1,2,3\*</sup>, Thomas Andrilhon<sup>4,5,6</sup>, Caroline Gauriau<sup>5,6</sup>, Damien Léger<sup>5,6†</sup>, Daniel Pressnitzer<sup>1,‡</sup>

**1** Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France, **2** Aix-Marseille University, CNRS, Institut de Neurosciences de la Timone (INT) UMR7289, Perception Representation Image Sound Music (PRISM) UMR7061, Laboratoire d'Informatique et Systèmes (LIS) UMR7020, Marseille, France, **3** Institute of Language Communication and the Brain, Aix-Marseille University, Marseille, France, **4** Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Mov'itteam, Inserm, CNRS, Paris, France, **5** Université Paris Cité, VIFASOM, ERC 7330, Vigilance Fatigue Sommeil et santé publique, Paris, France, **6** APHP, Hôtel-Dieu, Centre du Sommeil et de la Vigilance, Paris, France

\* These authors joint last first authorship on this work.

† etienne.thoret@univ-amu.fr



Explaining AI to create  
vocal biomarkers

# Sleep deprivation detected by voice analysis



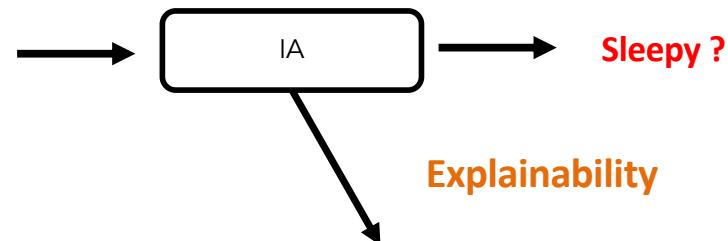
Health

AI can tell if you are sleep deprived by listening to your voice

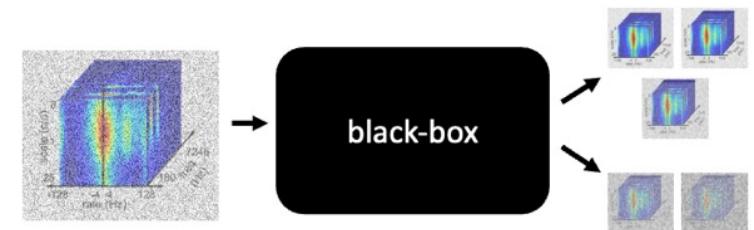


Lecture à  
voix haute

Detecting sleepiness in voice



Reverse correlation:  
« probing AI as we probe brain »



PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

## Sleep deprivation detected by voice analysis

Etienne Thoret<sup>1,2,3\*</sup>, Thomas Andrilhon<sup>4,5,6</sup>, Caroline Gauriau<sup>5,6</sup>, Damien Léger<sup>5,6†</sup>, Daniel Pressnitzer<sup>1,‡</sup>

1 Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France, 2 Aix-Marseille University, CNRS, Institut de Neurosciences de la Timone (INT) UMR7289, Perception Representation Image Sound Music (PRISM) UMR7061, Laboratoire d'Informatique et Systèmes (LIS) UMR7020, Marseille, France, 3 Institute of Language Communication and the Brain, Aix-Marseille University, Marseille, France, 4 Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Mov'itteam, Inserm, CNRS, Paris, France, 5 Université Paris Cité, VIFASOM, ERC 7330, Vigilance Fatigue Sommeil et santé publique, Paris, France, 6 APHP, Hôtel-Dieu, Centre du Sommeil et de la Vigilance, Paris, France

\* These authors joint last first authorship on this work.

† etienne.thoret@univ-amu.fr



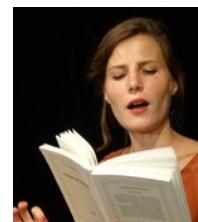
Explaining AI to create  
vocal biomarkers

# Sleep deprivation detected by voice analysis



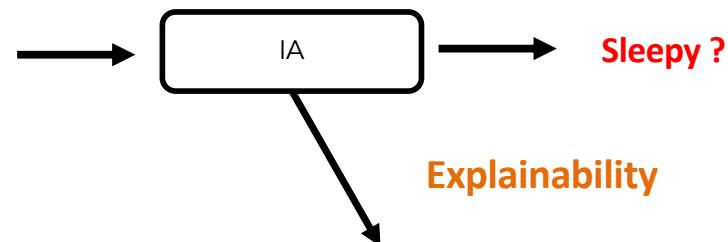
Health

AI can tell if you are sleep deprived by listening to your voice



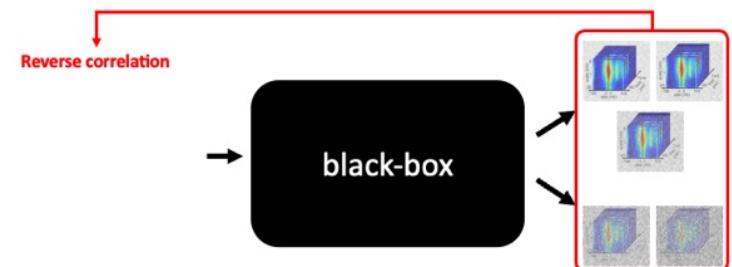
Lecture à  
voix haute

Detecting sleepiness in voice



Explainability

Reverse correlation:  
« probing AI as we probe brain »



PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

## Sleep deprivation detected by voice analysis

Etienne Thoret<sup>1,2,3\*</sup>, Thomas Andrilhon<sup>4,5,6</sup>, Caroline Gauriau<sup>5,6</sup>, Damien Léger<sup>5,6†</sup>, Daniel Pressnitzer<sup>1,‡</sup>

1 Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France, 2 Aix-Marseille University, CNRS, Institut de Neurosciences de la Timone (INT) UMR7289, Perception Representation Image Sound Music (PRISM) UMR7061, Laboratoire d'Informatique et Systèmes (LIS) UMR7020, Marseille, France, 3 Institute of Language Communication and the Brain, Aix-Marseille University, Marseille, France, 4 Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Mov'it team, Inserm, CNRS, Paris, France, 5 Université Paris Cité, VIFASOM, ERC 7330, Vigilance Fatigue Sommeil et santé publique, Paris, France, 6 APHP, Hôtel-Dieu, Centre du Sommeil et de la Vigilance, Paris, France

\* These authors joint last first authorship on this work.

† etienne.thoret@univ-amu.fr



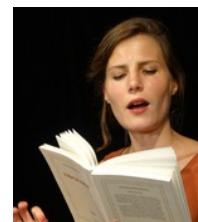
Explaining AI to create  
vocal biomarkers

# Sleep deprivation detected by voice analysis



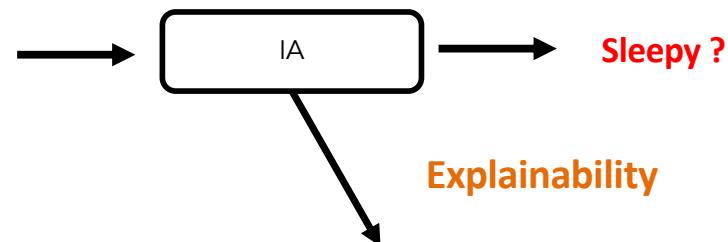
Health

AI can tell if you are sleep deprived by listening to your voice

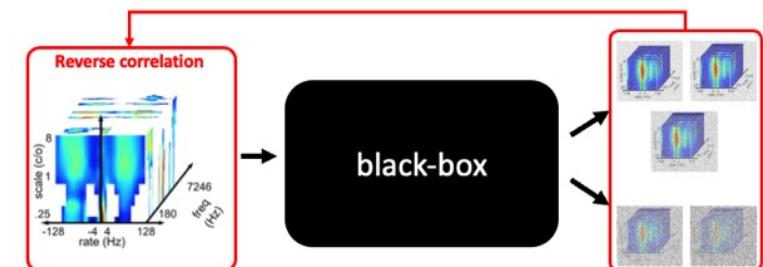


Lecture à  
voix haute

Detecting sleepiness in voice



Reverse correlation:  
« probing AI as we probe brain »



PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

## Sleep deprivation detected by voice analysis

Etienne Thoret<sup>1,2,3\*</sup>, Thomas Andrilhon<sup>4,5,6</sup>, Caroline Gauriau<sup>5,6</sup>, Damien Léger<sup>5,6†</sup>, Daniel Pressnitzer<sup>1,‡</sup>

**1** Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France, **2** Aix-Marseille University, CNRS, Institut de Neurosciences de la Timone (INT) UMR7289, Perception Representation Image Sound Music (PRISM) UMR7061, Laboratoire d'Informatique et Systèmes (LIS) UMR7020, Marseille, France, **3** Institute of Language Communication and the Brain, Aix-Marseille University, Marseille, France, **4** Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Mov'it team, Inserm, CNRS, Paris, France, **5** Université Paris Cité, VIFASOM, ERC 7330, Vigilance Fatigue Sommeil et santé publique, Paris, France, **6** APHP, Hôtel-Dieu, Centre du Sommeil et de la Vigilance, Paris, France

\* These authors joint last first authorship on this work.

† etienne.thoret@univ-amu.fr



Explaining AI to create  
vocal biomarkers

# Sleep deprivation detected by voice analysis



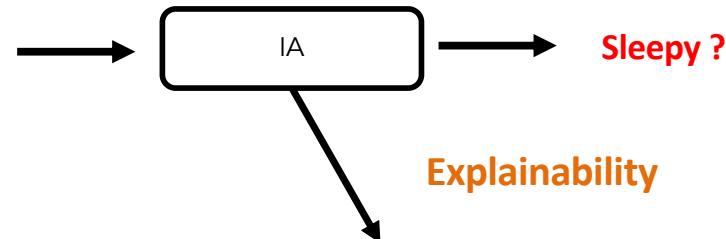
Health

AI can tell if you are sleep deprived by listening to your voice



Lecture à  
voix haute

Detecting sleepiness in voice



PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

## Sleep deprivation detected by voice analysis

Etienne Thoret<sup>1,2,3\*</sup>, Thomas Andrilhon<sup>4,5,6</sup>, Caroline Gauriau<sup>5,6</sup>, Damien Léger<sup>5,6†</sup>, Daniel Pressnitzer<sup>1,‡</sup>

**1** Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France, **2** Aix-Marseille University, CNRS, Institut de Neurosciences de la Timone (INT) UMR7289, Perception Representation Image Sound Music (PRISM) UMR7061, Laboratoire d'Informatique et Systèmes (LIS) UMR7020, Marseille, France, **3** Institute of Language Communication and the Brain, Aix-Marseille University, Marseille, France, **4** Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Mov'it team, Inserm, CNRS, Paris, France, **5** Université Paris Cité, VIFASOM, ERC 7330, Vigilance Fatigue Sommeil et santé publique, Paris, France, **6** APHP, Hôtel-Dieu, Centre du Sommeil et de la Vigilance, Paris, France

\* These authors joint last first authorship on this work.

† etienne.thoret@univ-amu.fr



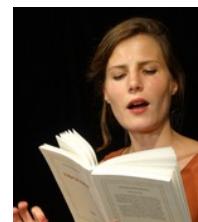
Explaining AI to create  
vocal biomarkers

# Sleep deprivation detected by voice analysis



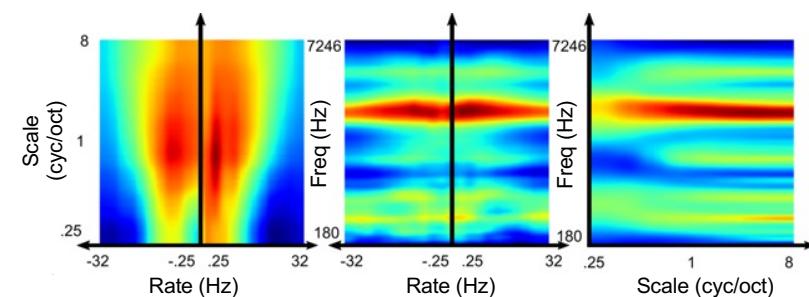
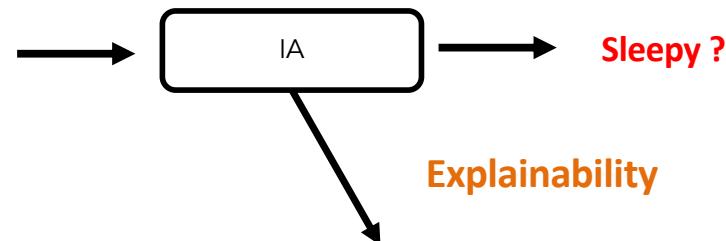
Health

AI can tell if you are sleep deprived by listening to your voice



Sleepiness  
vocal biomarker

Detecting sleepiness in voice



PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

## Sleep deprivation detected by voice analysis

Etienne Thoret<sup>1,2,3\*</sup>, Thomas Andrilhon<sup>4,5,6</sup>, Caroline Gauriau<sup>5,6</sup>, Damien Léger<sup>5,6†</sup>, Daniel Pressnitzer<sup>1,‡</sup>

**1** Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France, **2** Aix-Marseille University, CNRS, Institut de Neurosciences de la Timone (INT) UMR7289, Perception Representation Image Sound Music (PRISM) UMR7061, Laboratoire d'Informatique et Systèmes (LIS) UMR7020, Marseille, France, **3** Institute of Language Communication and the Brain, Aix-Marseille University, Marseille, France, **4** Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Mov'itteam, Inserm, CNRS, Paris, France, **5** Université Paris Cité, VIFASOM, ERC 7330, Vigilance Fatigue Sommeil et santé publique, Paris, France, **6** APHP, Hôtel-Dieu, Centre du Sommeil et de la Vigilance, Paris, France

\* These authors joint last first authorship on this work.

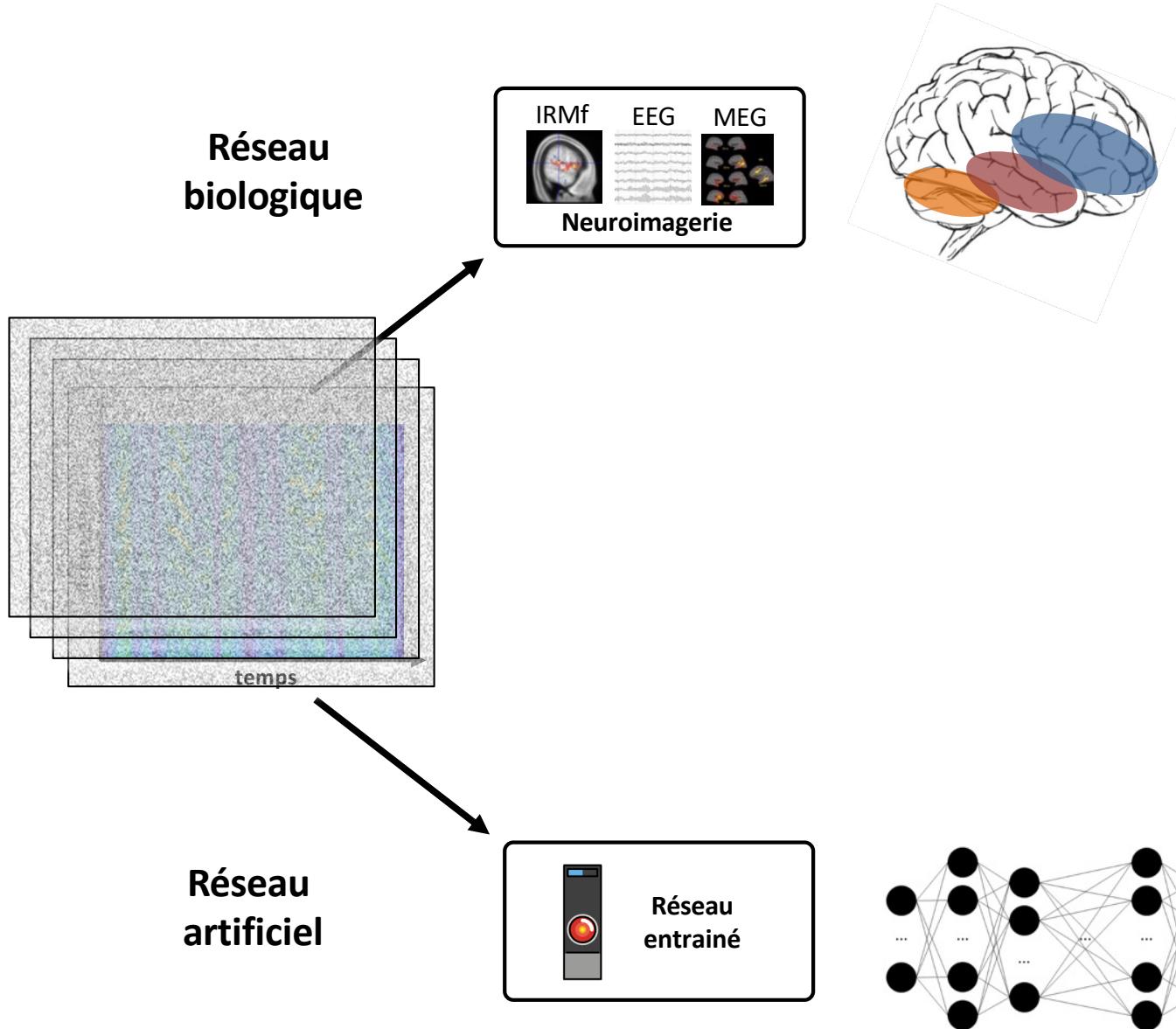
† etienne.thoret@univ-amu.fr



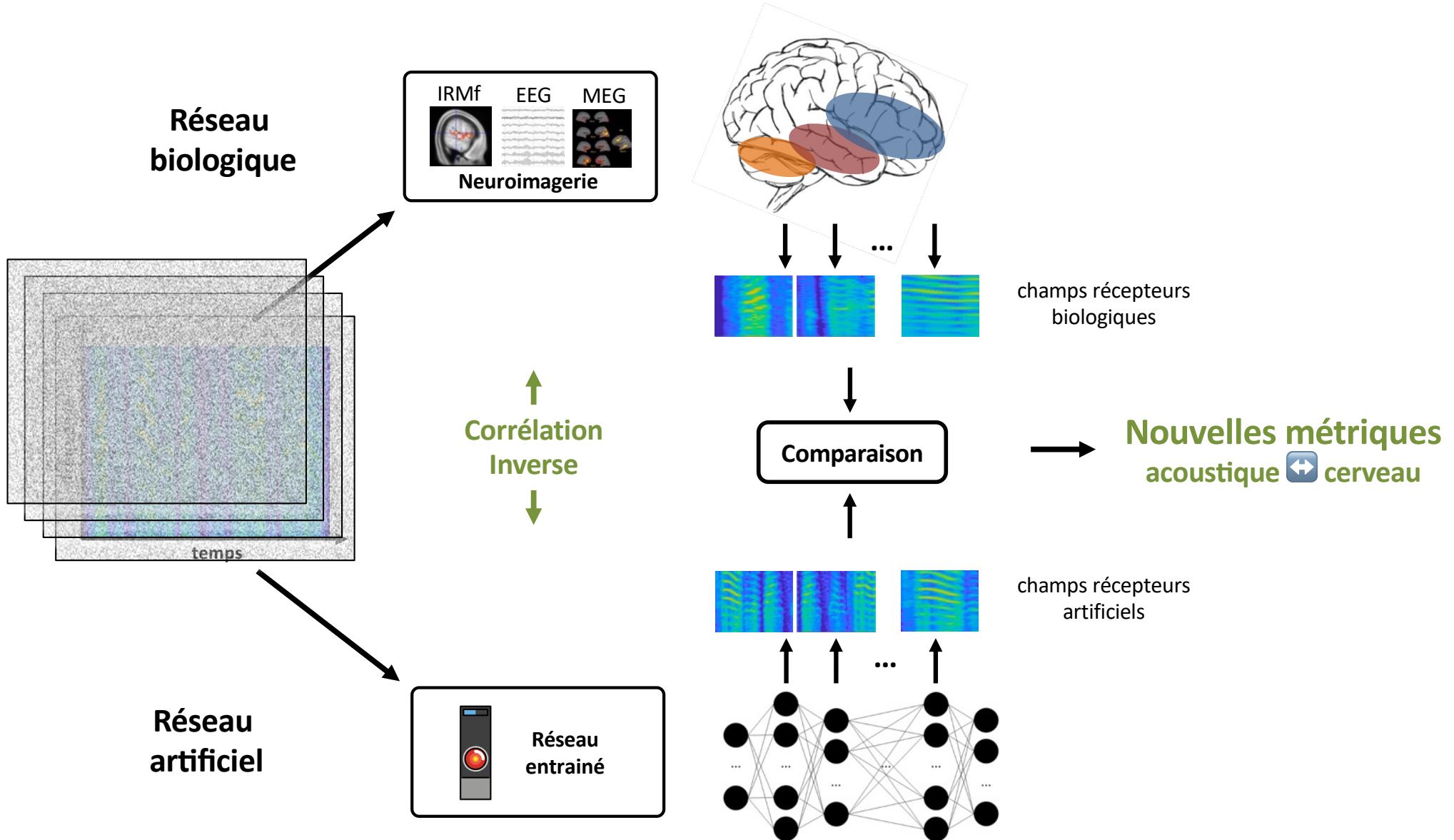
Explaining AI to create  
vocal biomarkers

# **How to apply it to DNNs?**

# How to apply it to DNNs?



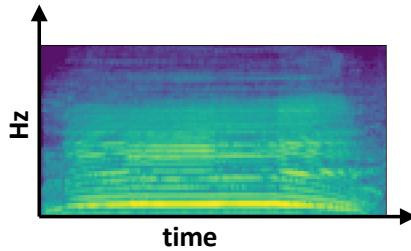
# How to apply it to DNNs?



# How to apply it to DNNs?

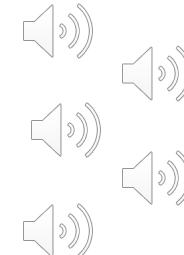
## Database

- Making sense of sounds database (CVSSP)
- *Sounds labeled by humans in 5 categories*
- 5 x 300 audio samples of 5 seconds cuted in 2 seconds
- Logmel spectrogram : 64x129 features
- 2250 training samples
- 750 testing samples



## 5 categories

- Human
- Urban
- Nature
- Music
- Effects



## Training performances:

- 100 epochs
- Training accuracy: .97
- Testing accuracy: .75

## Deepnet architecture

- CNN with 6 convolutional layers
  - Conv2d (9,9), BN, ReLU, BN, MaxPool2D
  - Conv2d (6,6), BN, ReLU, BN, MaxPool2D
  - Conv2d (3,3), BN, ReLU, BN, MaxPool2D
  - GlobalMaxPool2D

## Interpretability of the global CNN and of the layers

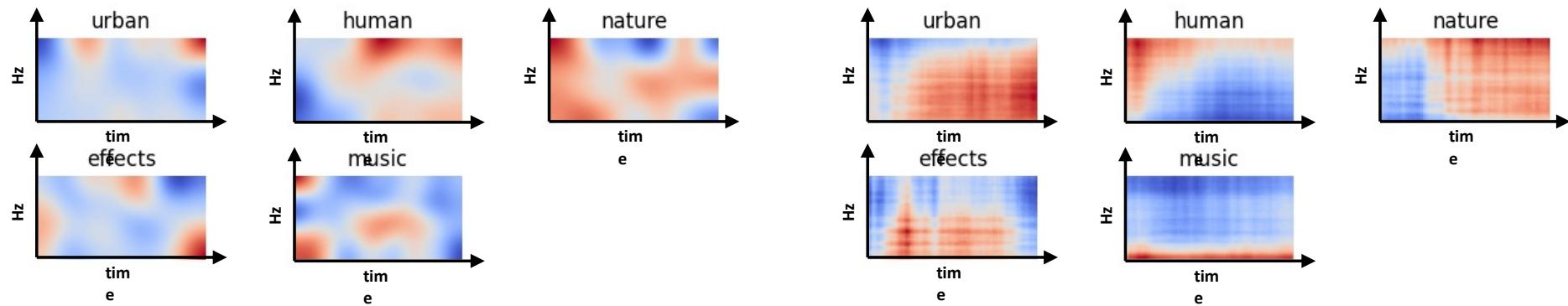
- one classifier at the output of each layers
- Bubbles method
- Reverse correlation

# How to apply it to DNNs?

For each layer, we interpret:

- The most relevant areas
- The prototypical encoded information

Layer #1



Potent information

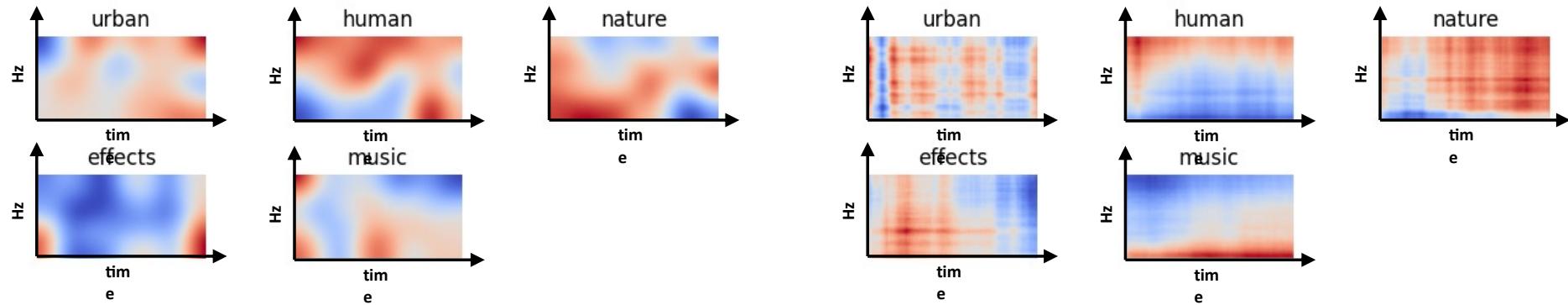
Represented information

# How to apply it to DNNs?

For each layer, we interpret:

- The most relevant areas
- The prototypical encoded information

Layer #2



Potent information

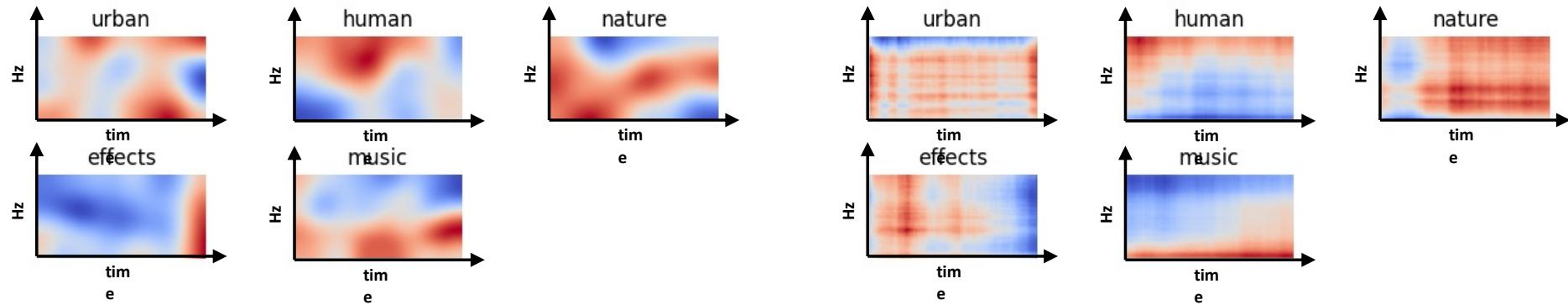
Represented information

# How to apply it to DNNs?

For each layer, we interpret:

- The most relevant areas
- The prototypical encoded information

Layer #3



Potent information

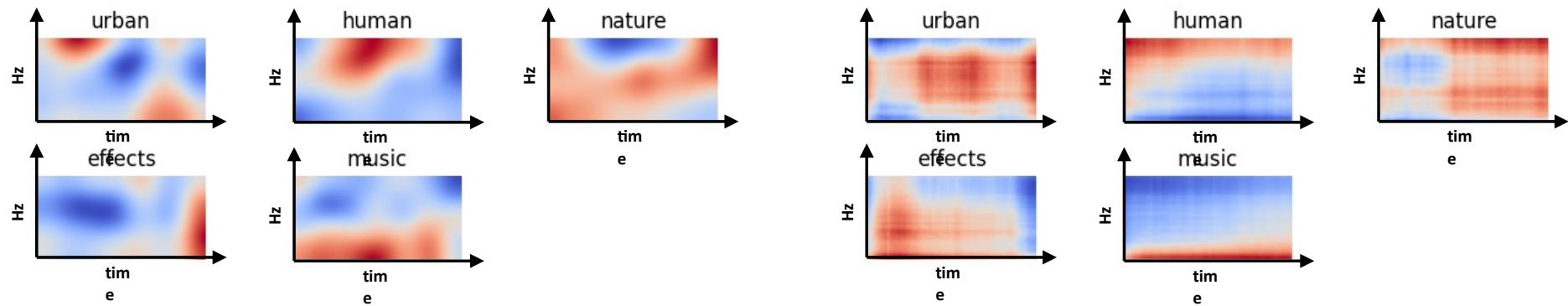
Represented information

# How to apply it to DNNs?

For each layer, we interpret:

- The most relevant areas
- The prototypical encoded information

Layer #4



Potent information

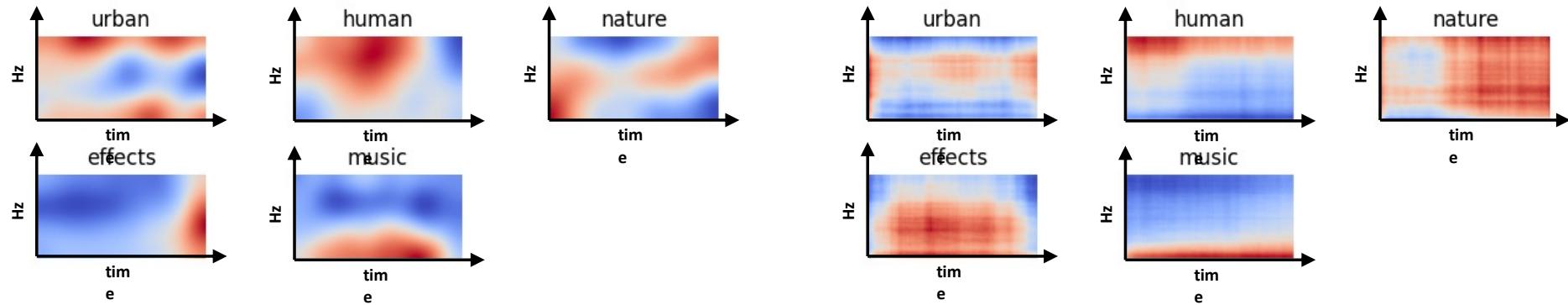
Represented information

# How to apply it to DNNs?

For each layer, we interpret:

- The most relevant areas
- The prototypical encoded information

Layer #5



Potent information

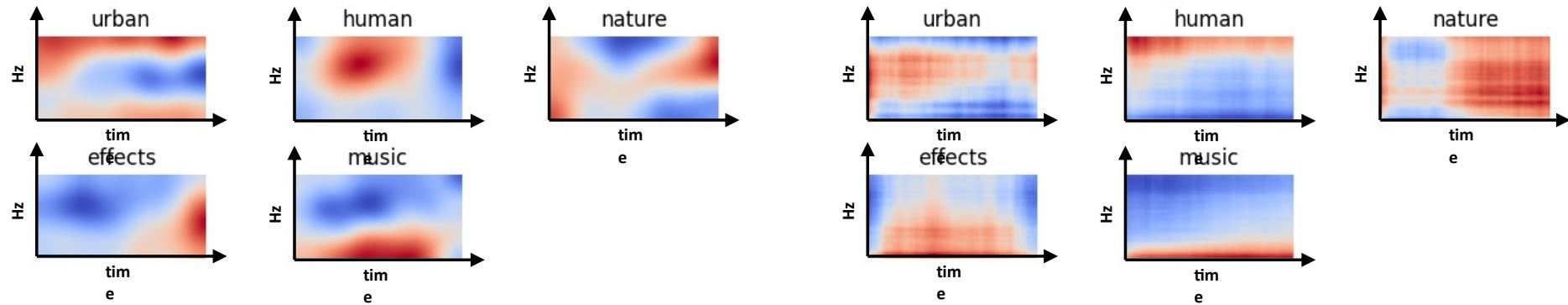
Represented information

# How to apply it to DNNs?

For each layer, we interpret:

- The most relevant areas
- The prototypical encoded information

Layer #6



Potent information

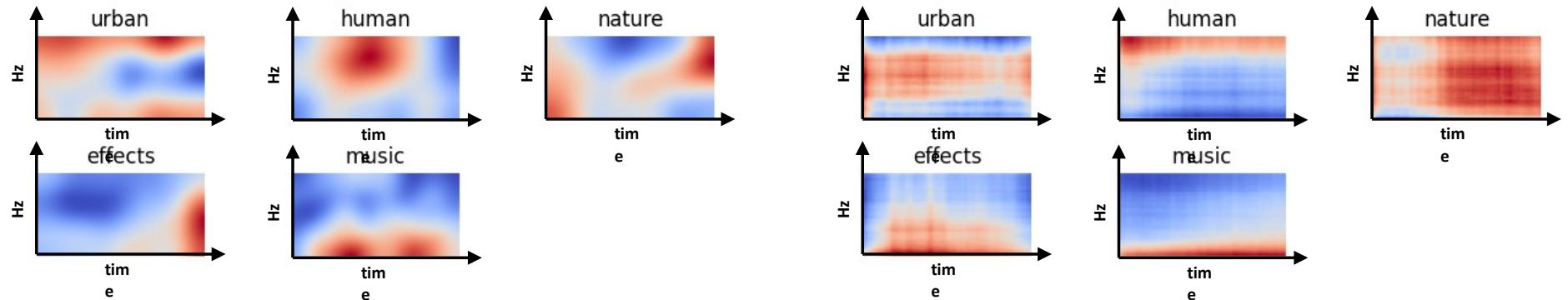
Represented information

# How to apply it to DNNs?

For each layer, we interpret:

- The most relevant areas
- The prototypical encoded information

Globally



Potent information

Represented information

*Does it make sense?*

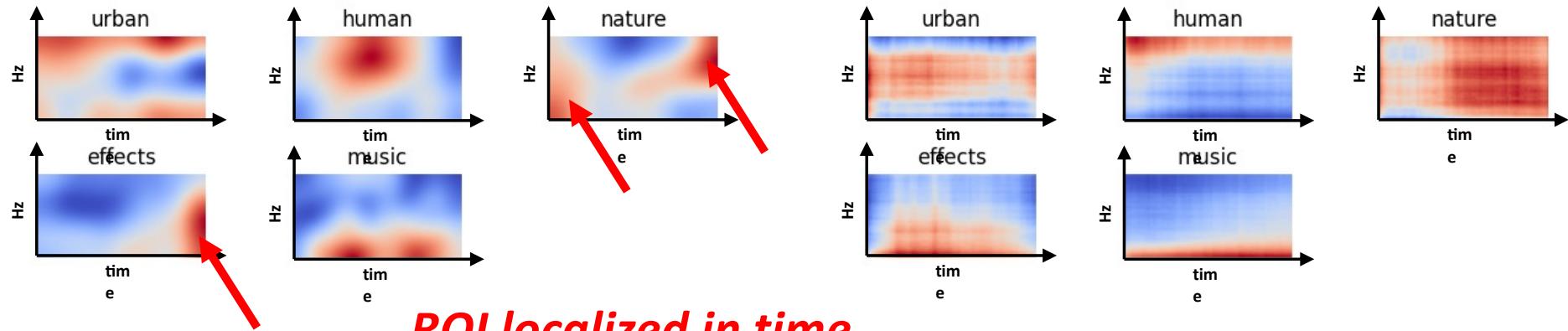
*acoustic vs. perception*

# How to apply it to DNNs?

**For each layer, we interpret:**

- The most relevant areas
  - The prototypical encoded information

**Globally**



# *ROI localized in time, plausible?*

## Potent information

## **Represented information**

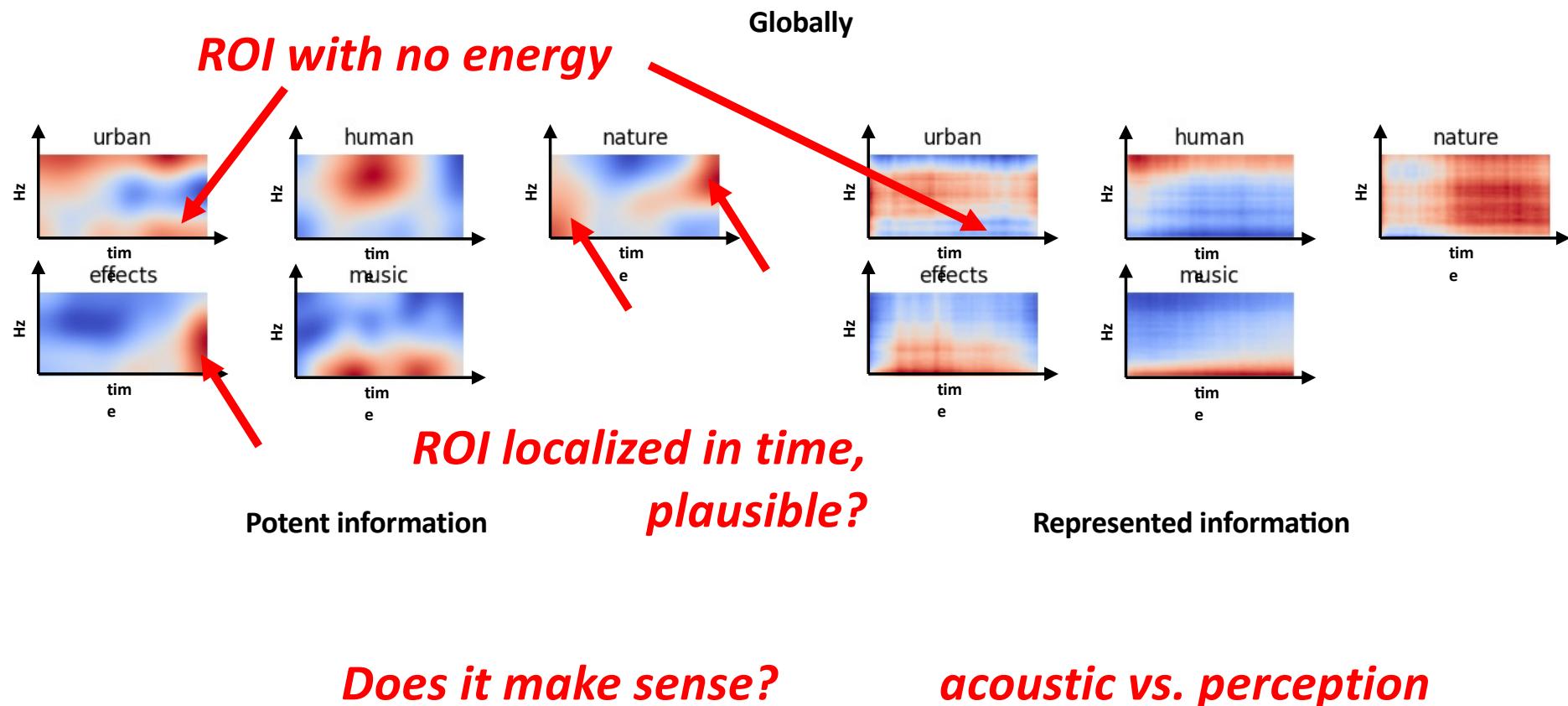
# *Does it make sense?*

# *acoustic vs. perception*

# How to apply it to DNNs?

For each layer, we interpret:

- The most relevant areas
- The prototypical encoded information

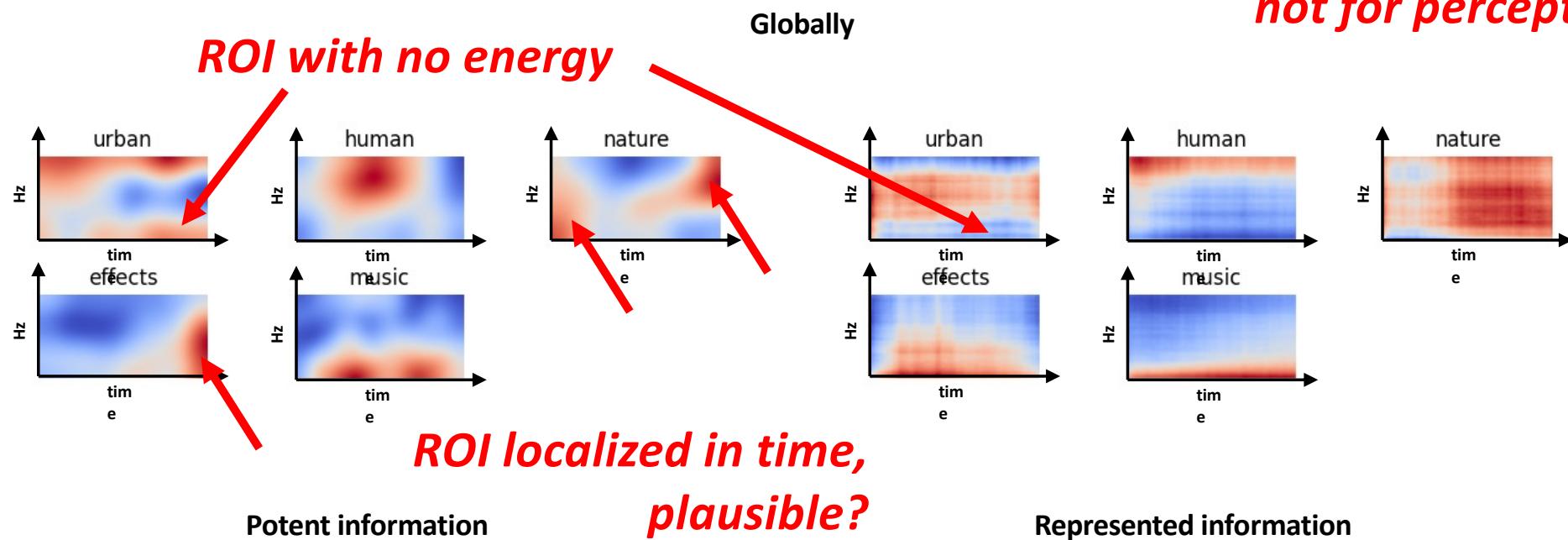


# How to apply it to DNNs?

For each layer, we interpret:

- The most relevant areas
- The prototypical encoded information

*CNN optimizes the discrimination for separability, not for perception*



*Does it make sense?*

*acoustic vs. perception*

# *Hands on!!*

<https://github.com/EtienneTho/neuroschool-phd-program-xAI-2024>

**1. Training a classifier**

**2. Random perturbations: additive or multiplicative**

**3. Bubbles**

**4. Reverse correlation**

# *Hands on!!*

<https://github.com/EtienneTho/neuroschool-phd-program-xAI-2024>

- **Changing bubbles & revcor parameters**
- **Changing the classifier**
- **Pseudo-random noise**



# ***Hands on!!***

# Pseudo random noise

# ***Hands on!!***

# Pseudo random noise

A 5x10 grid of handwritten digits from 0 to 9, arranged in five rows and ten columns. The digits are written in a cursive style and are mostly black on a white background.

## PCA

# Latent space

$x_{1,1}$

X<sub>2,1</sub>

$$x_{3,1}$$

•

•

•

AN

# ***Hands on!!***

# Pseudo random noise

A 5x10 grid of handwritten digits from 0 to 9, arranged in five rows and ten columns. The digits are written in a cursive style and are mostly black on a white background.

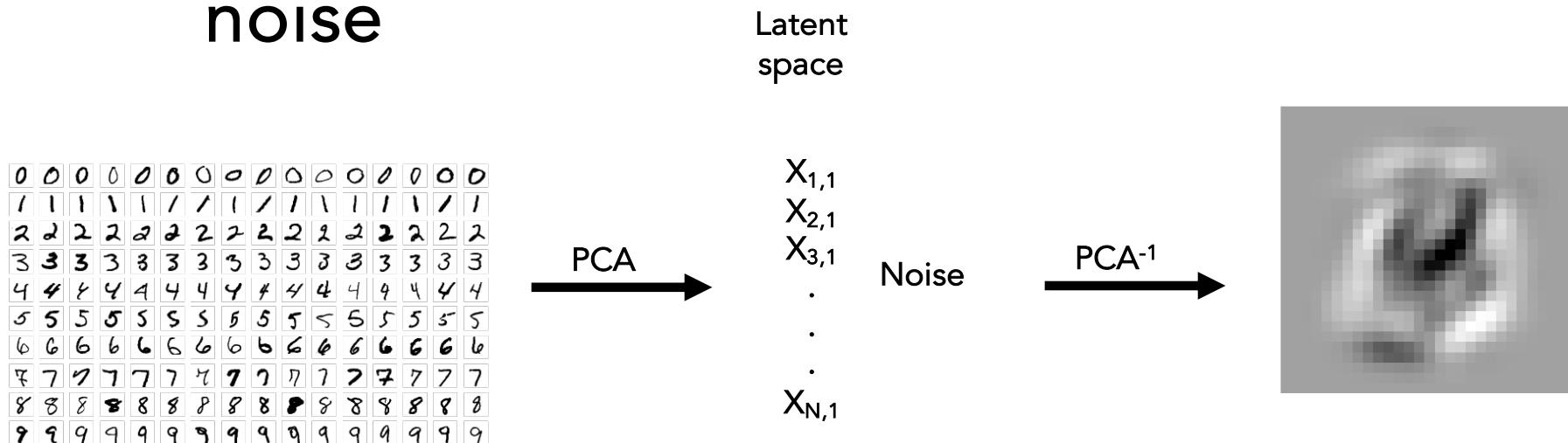
PCA

## Latent space

$X_{1,1}$   
 $X_{2,1}$   
 $X_{3,1}$   
 .  
 .  
 .  
 $X_{N,1}$

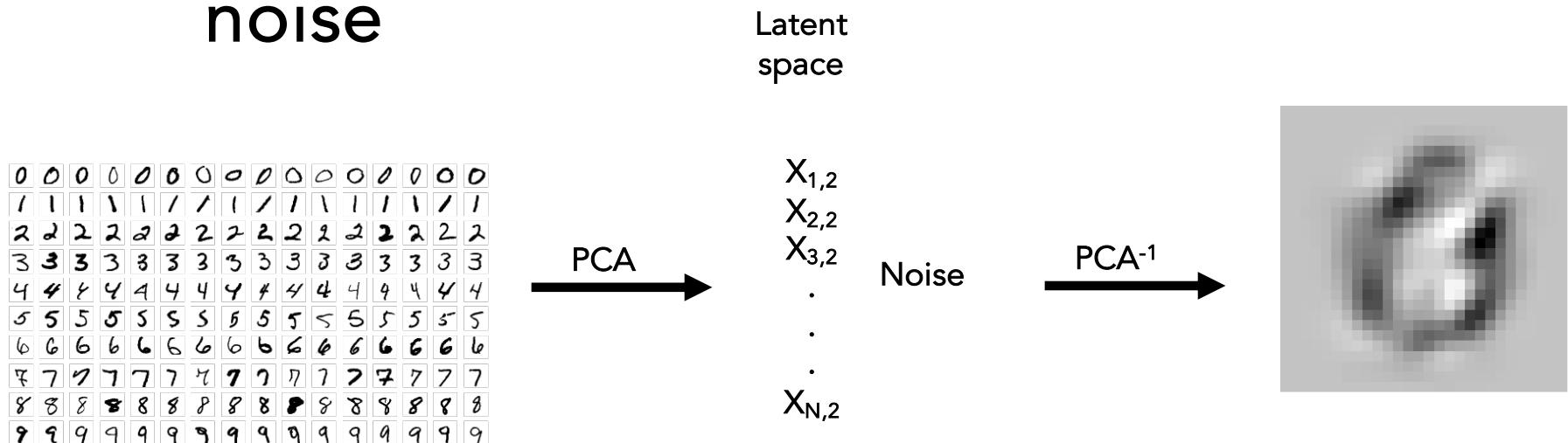
# ***Hands on!!***

# Pseudo random noise



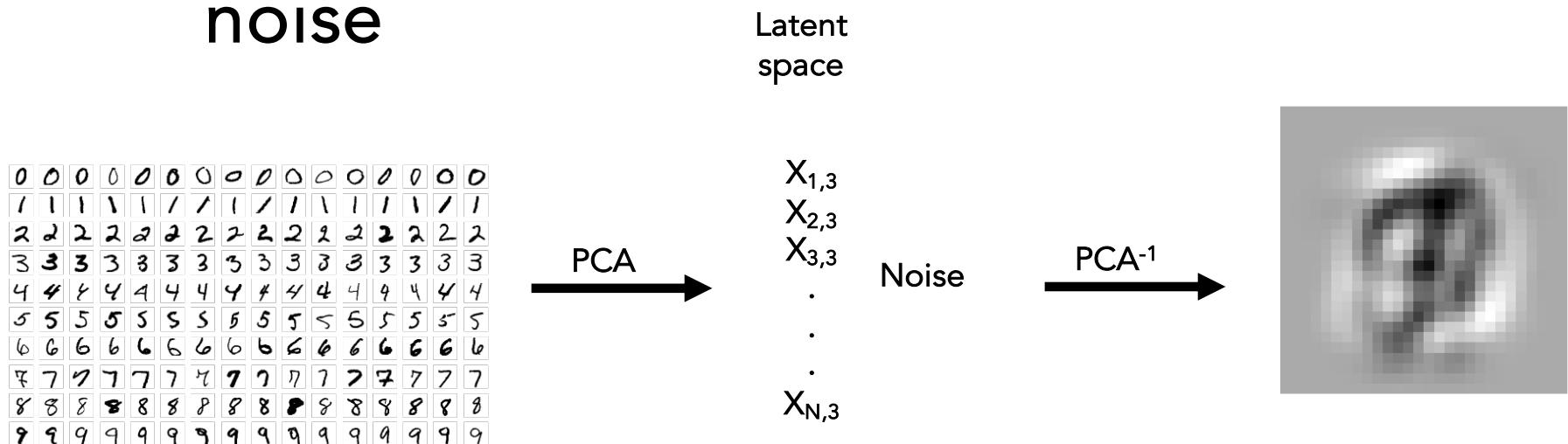
# ***Hands on!!***

# Pseudo random noise



# ***Hands on!!***

# Pseudo random noise



# ***Hands on!!***

# Pseudo random noise

