

# PhD Research + IMVA

Xavier Juanola Molet

Universitat Pompeu Fabra

*xavier.juanola@upf.edu*

MLBrief 2023, third edition

May 30, 2023

# Presentation Overview

## 1 About me

- Background and Interests
- My PhD so far

## 2 About the group

- About IMVA
- Sports Video Analysis
- Generative models. Cross-modal relationships and understanding
- Facial analysis for affective computing
- Remote physiological measurements
- Newborn 3D facial dysmorphology analysis

# About me

## Who am I?

## Who am I?

- Name: Xavier Juanola Molet
- Current Position: PhD Candidate
- Research Group: IMVA (Intelligent Multimodal Vision Analysis)

# About me

## Who am I?

- Name: Xavier Juanola Molet
- Current Position: PhD Candidate
- Research Group: IMVA (Intelligent Multimodal Vision Analysis)

## Educational Background

## Who am I?

- Name: Xavier Juanola Molet
- Current Position: PhD Candidate
- Research Group: IMVA (Intelligent Multimodal Vision Analysis)

## Educational Background

- B.Sc. in Theoretical Physics, University of Barcelona, 2018

## Who am I?

- Name: Xavier Juanola Molet
- Current Position: PhD Candidate
- Research Group: IMVA (Intelligent Multimodal Vision Analysis)

## Educational Background

- B.Sc. in Theoretical Physics, University of Barcelona, 2018
- M.Sc. in Artificial Intelligence, University Pompeu Fabra, 2019

# About me

## Who am I?

- Name: Xavier Juanola Molet
- Current Position: PhD Candidate
- Research Group: IMVA (Intelligent Multimodal Vision Analysis)

## Educational Background

- B.Sc. in Theoretical Physics, University of Barcelona, 2018
- M.Sc. in Artificial Intelligence, University Pompeu Fabra, 2019
- M.Sc. in High Energy Physics, Astrophysics and Cosmology, Autonomous University of Barcelona, 2021

## Who am I?

- Name: Xavier Juanola Molet
- Current Position: PhD Candidate
- Research Group: IMVA (Intelligent Multimodal Vision Analysis)

## Educational Background

- B.Sc. in Theoretical Physics, University of Barcelona, 2018
- M.Sc. in Artificial Intelligence, University Pompeu Fabra, 2019
- M.Sc. in High Energy Physics, Astrophysics and Cosmology, Autonomous University of Barcelona, 2021
- PhD in Sound Source Localization, Universitat Pompeu Fabra, (Started Nov. 2022)

# About me

## Who am I?

- Name: Xavier Juanola Molet
- Current Position: PhD Candidate
- Research Group: IMVA (Intelligent Multimodal Vision Analysis)

## Educational Background

- B.Sc. in Theoretical Physics, University of Barcelona, 2018
- M.Sc. in Artificial Intelligence, University Pompeu Fabra, 2019
- M.Sc. in High Energy Physics, Astrophysics and Cosmology, Autonomous University of Barcelona, 2021
- PhD in Sound Source Localization, Universitat Pompeu Fabra, (Started Nov. 2022)

## Interests

# About me

## Who am I?

- Name: Xavier Juanola Molet
- Current Position: PhD Candidate
- Research Group: IMVA (Intelligent Multimodal Vision Analysis)

## Educational Background

- B.Sc. in Theoretical Physics, University of Barcelona, 2018
- M.Sc. in Artificial Intelligence, University Pompeu Fabra, 2019
- M.Sc. in High Energy Physics, Astrophysics and Cosmology, Autonomous University of Barcelona, 2021
- PhD in Sound Source Localization, Universitat Pompeu Fabra, (Started Nov. 2022)

## Interests

- Computer Vision

# About me

## Who am I?

- Name: Xavier Juanola Molet
- Current Position: PhD Candidate
- Research Group: IMVA (Intelligent Multimodal Vision Analysis)

## Educational Background

- B.Sc. in Theoretical Physics, University of Barcelona, 2018
- M.Sc. in Artificial Intelligence, University Pompeu Fabra, 2019
- M.Sc. in High Energy Physics, Astrophysics and Cosmology, Autonomous University of Barcelona, 2021
- PhD in Sound Source Localization, Universitat Pompeu Fabra, (Started Nov. 2022)

## Interests

- Computer Vision
- Audio

# About me

## Who am I?

- Name: Xavier Juanola Molet
- Current Position: PhD Candidate
- Research Group: IMVA (Intelligent Multimodal Vision Analysis)

## Educational Background

- B.Sc. in Theoretical Physics, University of Barcelona, 2018
- M.Sc. in Artificial Intelligence, University Pompeu Fabra, 2019
- M.Sc. in High Energy Physics, Astrophysics and Cosmology, Autonomous University of Barcelona, 2021
- PhD in Sound Source Localization, Universitat Pompeu Fabra, (Started Nov. 2022)

## Interests

- Computer Vision
- Audio
- Deep Learning

# My PhD so far

## What have I done?

# My PhD so far

## What have I done?

- Read SO MANY papers related to Sound Source Localization

# My PhD so far

## What have I done?

- Read SO MANY papers related to Sound Source Localization
- Attend seminars related to publication process, communication skills, etc., and a course on 3D Audio

# My PhD so far

## What have I done?

- Read SO MANY papers related to Sound Source Localization
- Attend seminars related to publication process, communication skills, etc., and a course on 3D Audio
- **Main problem to solve:** Given a video with a mixture of sounds, find the different sound sources in the image

# My PhD so far

## What have I done?

- Read SO MANY papers related to Sound Source Localization
- Attend seminars related to publication process, communication skills, etc., and a course on 3D Audio
- **Main problem to solve:** Given a video with a mixture of sounds, find the different sound sources in the image
  
- Worked in the past for a single sound source

# My PhD so far

## What have I done?

- Read SO MANY papers related to Sound Source Localization
- Attend seminars related to publication process, communication skills, etc., and a course on 3D Audio
- **Main problem to solve:** Given a video with a mixture of sounds, find the different sound sources in the image
  
- Worked in the past for a single sound source
- General case has multiple problems:



# My PhD so far

## What have I done?

- Read SO MANY papers related to Sound Source Localization
- Attend seminars related to publication process, communication skills, etc., and a course on 3D Audio
- **Main problem to solve:** Given a video with a mixture of sounds, find the different sound sources in the image
- Worked in the past for a single sound source
- General case has multiple problems:
  - Mixture of sounds



# My PhD so far

## What have I done?

- Read SO MANY papers related to Sound Source Localization
- Attend seminars related to publication process, communication skills, etc., and a course on 3D Audio
- **Main problem to solve:** Given a video with a mixture of sounds, find the different sound sources in the image
- Worked in the past for a single sound source
- General case has multiple problems:
  - Mixture of sounds
  - Off-screen sounds



# My PhD so far

## What have I done?

- Read SO MANY papers related to Sound Source Localization
- Attend seminars related to publication process, communication skills, etc., and a course on 3D Audio
- **Main problem to solve:** Given a video with a mixture of sounds, find the different sound sources in the image
  
- Worked in the past for a single sound source
- General case has multiple problems:
  - Mixture of sounds
  - Off-screen sounds
  - Silent objects



# My PhD so far

## What have I done?

- Read SO MANY papers related to Sound Source Localization
  - Attend seminars related to publication process, communication skills, etc., and a course on 3D Audio
  - **Main problem to solve:** Given a video with a mixture of sounds, find the different sound sources in the image
- 
- Worked in the past for a single sound source
  - General case has multiple problems:
    - Mixture of sounds
    - Off-screen sounds
    - Silent objects
    - Different objects of the same type in the scene



# Presentation Overview

## 1 About me

- Background and Interests
- My PhD so far

## 2 About the group

- About IMVA
- Sports Video Analysis
- Generative models. Cross-modal relationships and understanding
- Facial analysis for affective computing
- Remote physiological measurements
- Newborn 3D facial dysmorphology analysis

# About IMVA (Intelligent Multimodal Vision Analysis)

- **Goal:** Investigate the automatic analysis and understanding of visual content and address real-world problems such as audio, natural language, ultrasound or magnetic resonance..

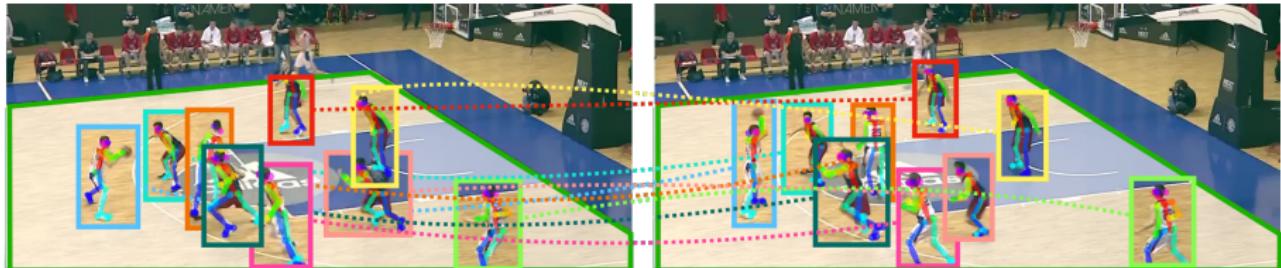
# About IMVA (Intelligent Multimodal Vision Analysis)

- **Goal:** Investigate the automatic analysis and understanding of visual content and address real-world problems such as audio, natural language, ultrasound or magnetic resonance..
- **Collaborations:** IMVA collaborates with academia, industry, and research institutions worldwide to foster knowledge exchange and accelerate research advancements.

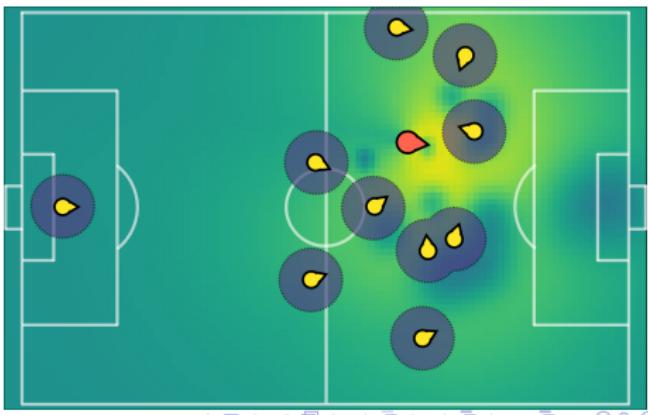
# About IMVA (Intelligent Multimodal Vision Analysis)

- **Goal:** Investigate the automatic analysis and understanding of visual content and address real-world problems such as audio, natural language, ultrasound or magnetic resonance..
- **Collaborations:** IMVA collaborates with academia, industry, and research institutions worldwide to foster knowledge exchange and accelerate research advancements.
- **Applications:**
  - Accessibility of people with visual, hearing or reading impairment to multimedia content
  - Analysis of the human face both in terms of its morphology and its dynamics (potential for disciplines such as psychology, linguistics, neuroscience, health or developmental biology)
  - Separation of the different audio sources that make up the audio mixture of a particular video
  - Understanding and the exploitation of the correlations and complementations among different modalities

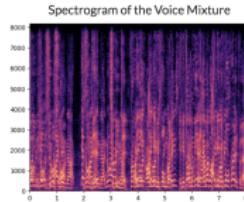
# IMVA / Sports Video Analysis



- Player detection and classification
- Player tracking
- Estimation of player orientation
- Estimation of pass feasibility
- Action spotting

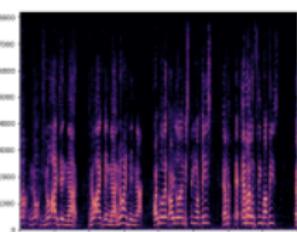
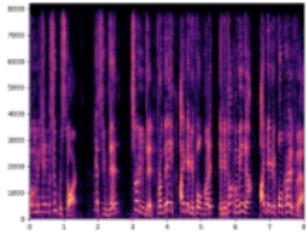


# IMVA / Audio-visual analysis

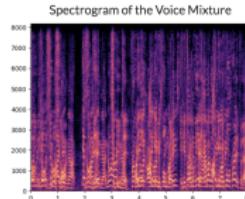


- Sound source separation (speech, singing voice, music)
- Demos:** (VoViT, Acappella)

Target Speaker

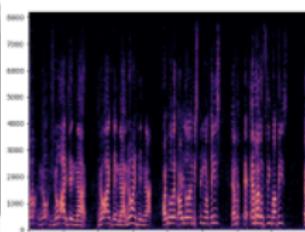
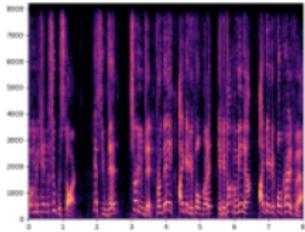


# IMVA / Audio-visual analysis

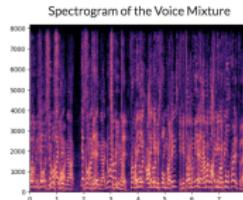


- Sound source separation (speech, singing voice, music)  
**Demos:** (VoViT, Acappella)
- Audio-visual speech inpainting **Demo:** (AV Inpainting)

Target Speaker

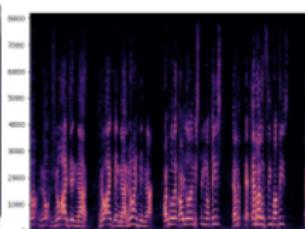
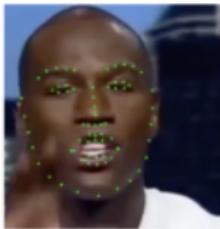
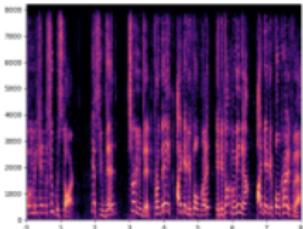


# IMVA / Audio-visual analysis

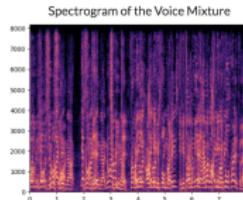


- Sound source separation (speech, singing voice, music)  
**Demos:** (VoViT, Acappella)
- Audio-visual speech inpainting **Demo:** (AV Inpainting)
- Audio-visual synchronization **Demo:** (AV Synchronization)

Target Speaker

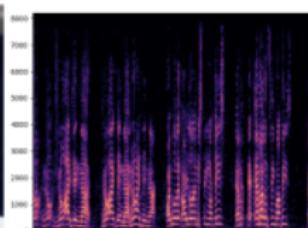
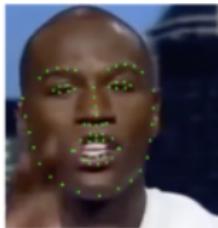
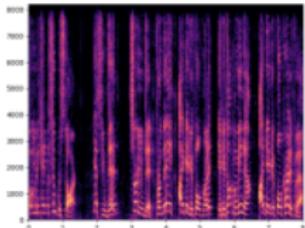


# IMVA / Audio-visual analysis



- Sound source separation (speech, singing voice, music)  
**Demos:** (VoViT, Acappella)
- Audio-visual speech inpainting **Demo:** (AV Inpainting)
- Audio-visual synchronization **Demo:** (AV Synchronization)
- Sound source localization in image/video

Target Speaker



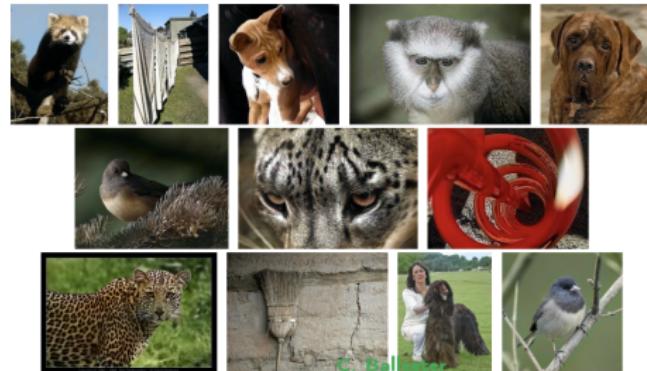
# IMVA / Generative models. Cross-modal relationships and understanding



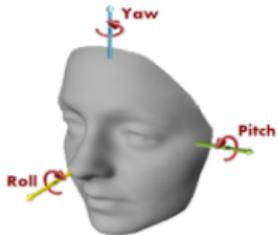
**Generative models and its use  
for solving inverse problems in  
imaging**

- Inpainting
- Colorization
- Blind video restoration
- Anomaly detection

**Cross-modal relationships and  
understanding: visual, audio  
and natural language data**



# IMVA / Facial analysis for affective computing



Head Pose  
Estimation

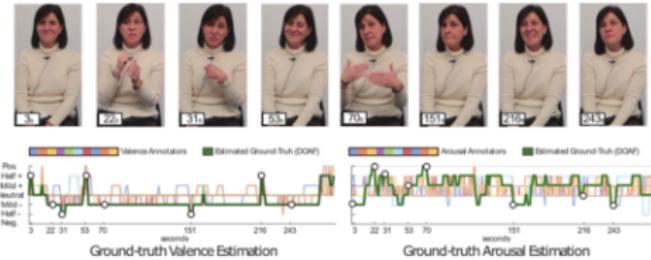


Facial  
expressions

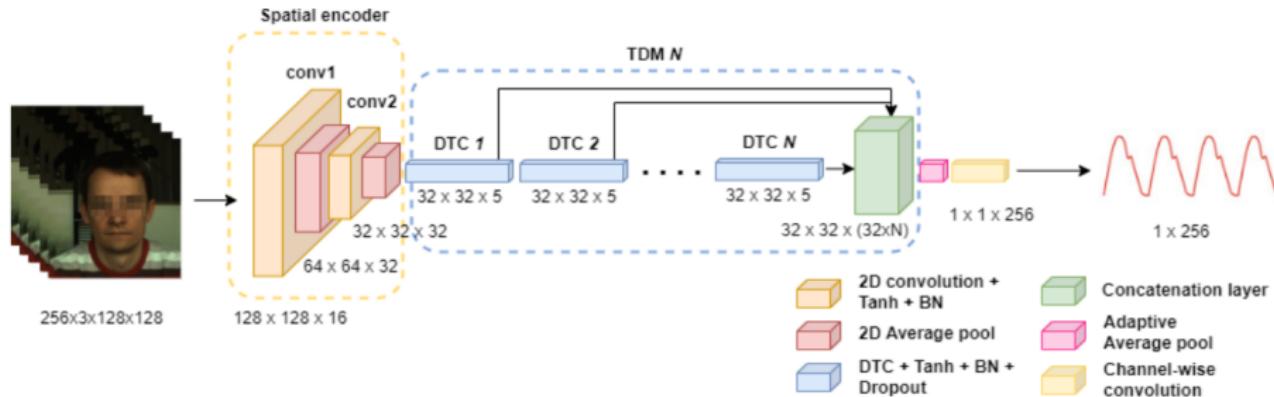


Automatic  
Lip-Reading

- Automatic facial analysis in 2D 3D
- 3D head pose estimation
- Emotion recognition: Discrete labels; Valence arousal
- Automatic lip reading



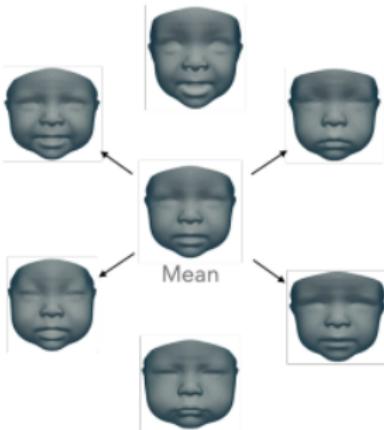
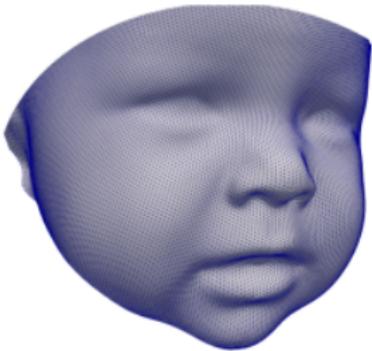
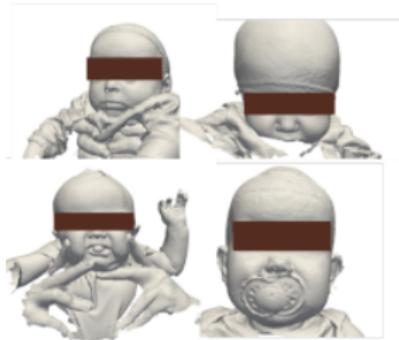
# IMVA / Remote physiological measurements



- **Estimation of cardiac frequency from facial videos**

- Through the amplification of color changes due to blood flow
- Contactless physiological sensing
- Cutting edge deep neural networks
- Challenging input signal requirements

# IMVA / Newborn 3D facial dysmorphology analysis



- **3D morphological analysis of the newborn face**
  - Early screening of genetic developmental disorders
  - First 3D Morphable Model of the facial geometry of newborns
  - Deep learning-based reconstruction of 3D geometry from 2D facial pictures