Towards the Neuromorphic Implementation of the Auditory Perception in the iCub Robotic Platform

Daniel Gutierrez-Galan, Chiara Bartolozzi, Juan Pedro Dominguez-Morales, Angel Jimenez-Fernandez, Alejandro Linares-Barranco

Neuro-Inspired Computational Elements (NICE) 2022
31-March-2022

robotics & technology of computers lab.

Departamento de Arquitectura y Tecnología de Computadores UNIVERSIDAD DE SEVILLA

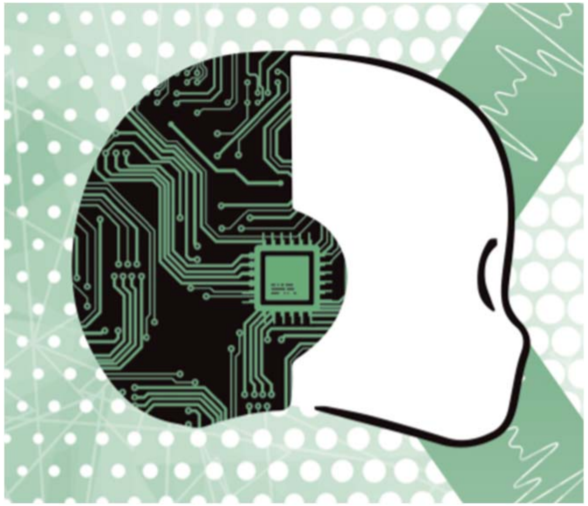
UNIVERSIDAD DE SEVILLA

iit EDPR ISTITUTO ITALIANO DI TECNOLOGIA EVENT-DRIVEN PERCEPTION FOR ROBOTICS

Neuro-Inspired Computational Elements NICE 2022

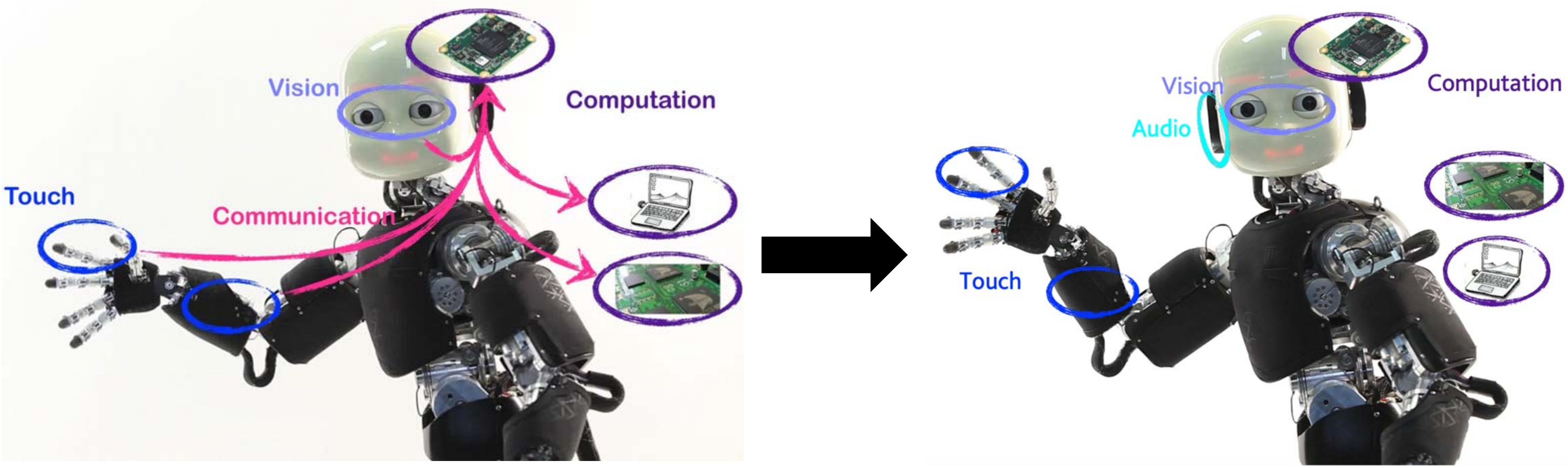
Outline

- **Motivation**
- Auditory ascending pathway
- The iCub robot
- Giving iCub the sense of hearing
- Preliminary results
- Improvements & future works
- ACKs and questions



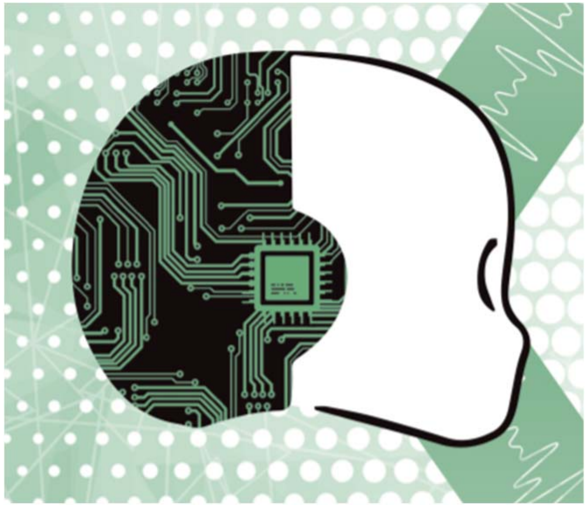
What is the motivation of this work?

- To provide iCub with the hearing sense using neuromorphic sensors for the first time.



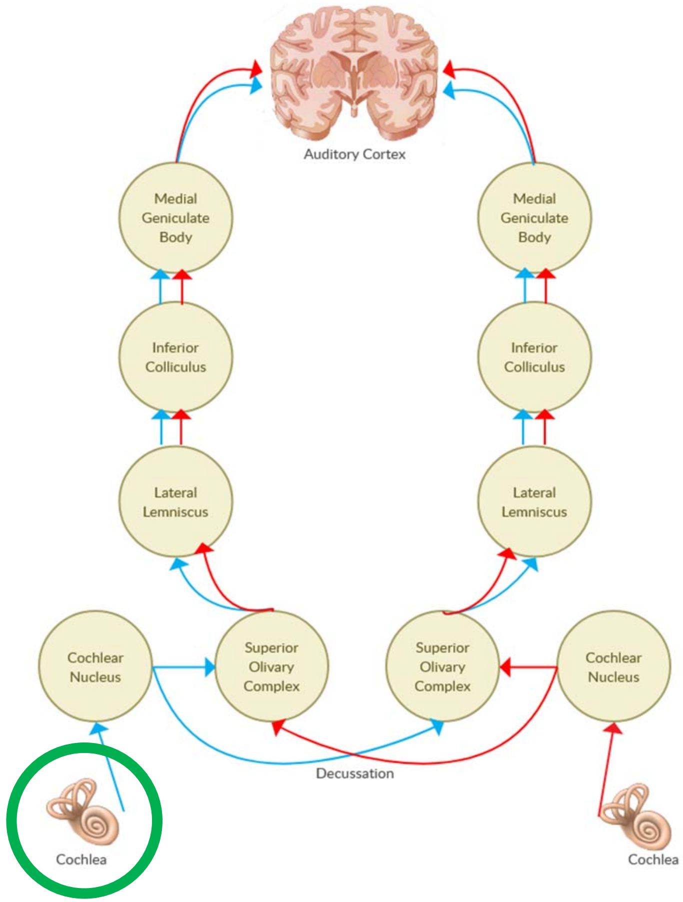
Outline

- Motivation
- **Auditory ascending pathway**
- The iCub robot
- Giving iCub the sense of hearing
- Preliminary results
- Improvements & future works
- ACKs and questions



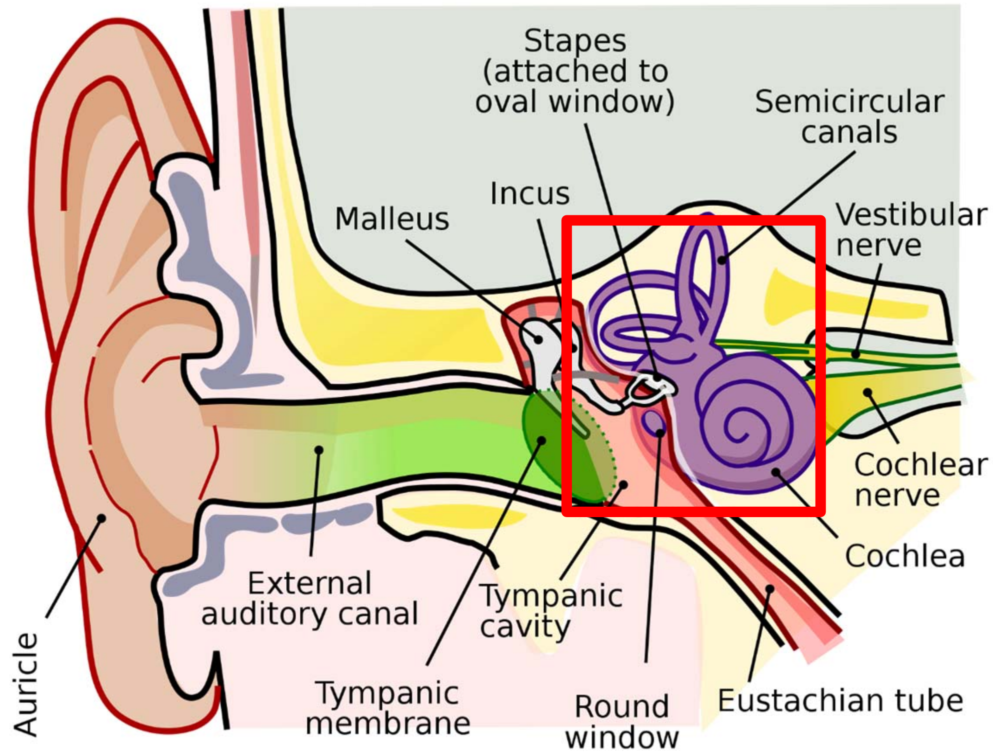
The big picture of the hearing sense

- The ascending auditory pathway:
  - **The cochlea**
  - The superior olivary complex
  - The inferior colliculus
  - The auditory cortex



How do we hear?

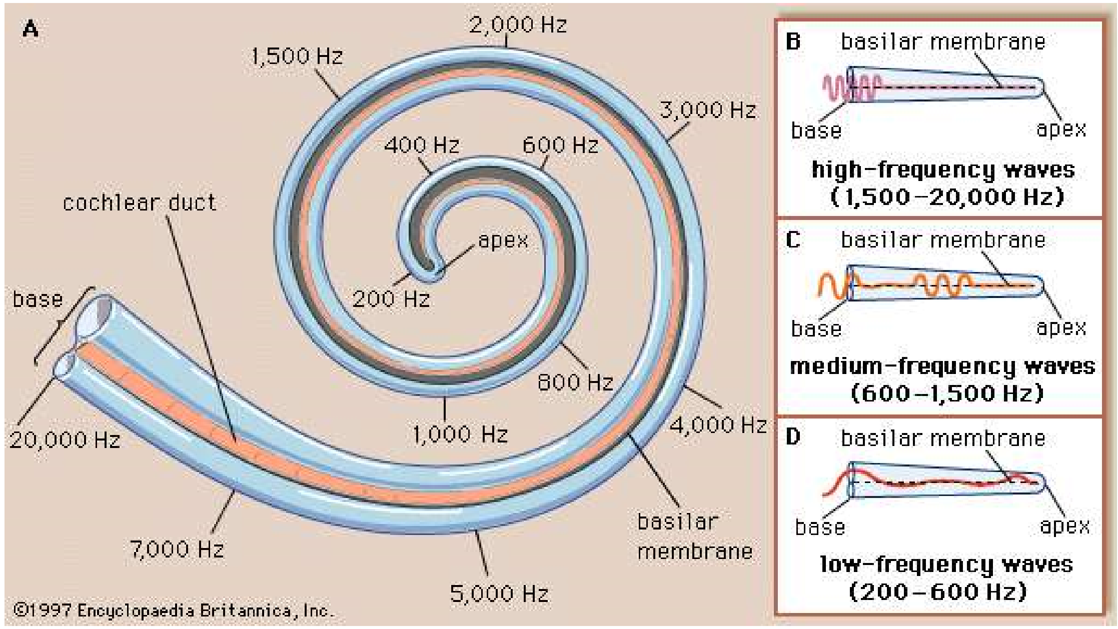
- Three main parts:
  - External ear
  - Middle ear
  - Inner ear



Richard F. Lyon. 2017. Human and Machine Hearing: Extracting Meaning from Sound (1st. ed.). Cambridge University Press, USA.

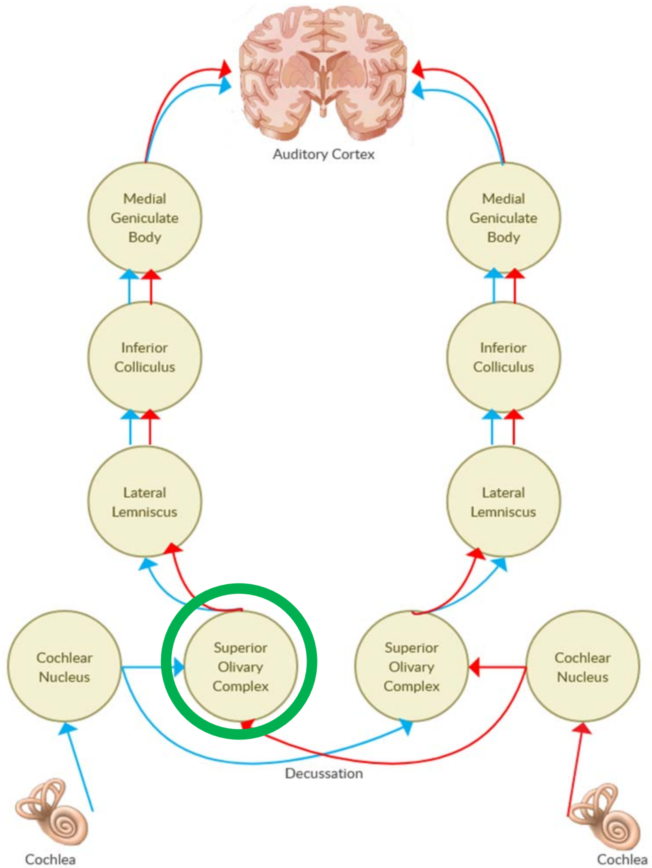
The basilar membrane

- Decompose the sound in its main frequencies.
- Similar to a Fourier transform.
- Humans: from 200 Hz to 22 kHz.



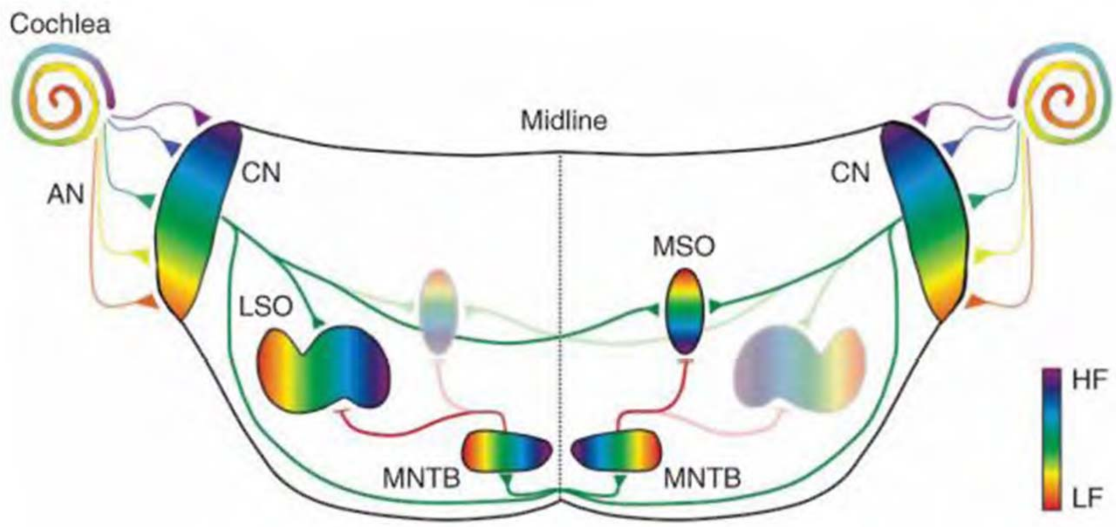
The big picture of the hearing sense

- The ascending auditory pathway:
  - The cochlea
  - **The superior olivary complex**
  - The inferior colliculus
  - The auditory cortex



Superior Olivary Complex (SOC)

- Four main nuclei:
  - Cochlear nucleus (**CN**).
  - Medial Superior Olive (**MSO**).
  - Lateral Superior Olive (**LSO**).
  - Medial Nucleus of the Trapezoid Body (**MNTB**).

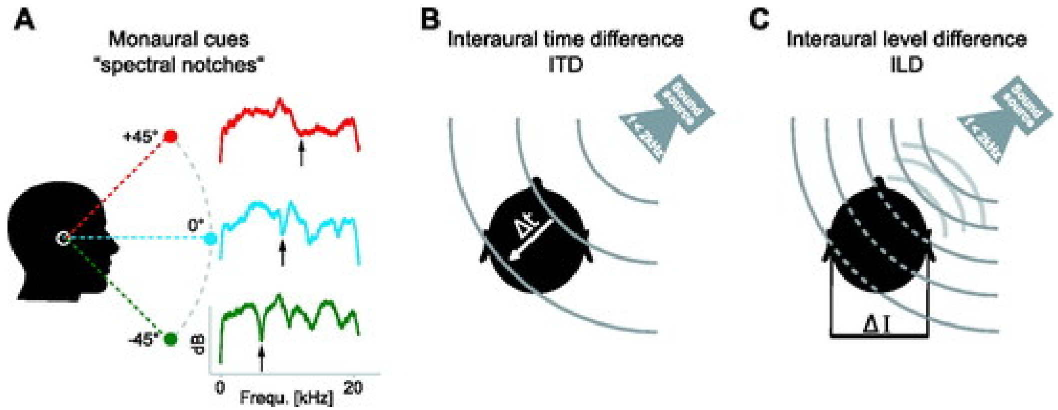


Glackin, B., et al. (2010). A spiking neural network model of the medial superior olive using spike timing dependent plasticity for sound localization. *Frontiers in computational neuroscience*

The Medial and Lateral Superior Olives

- To extract spatial information of the sound sources.

- Binaural cues:
  - Inter-aural Time difference (ITD)
  - Inter-aural Level difference (ILD)

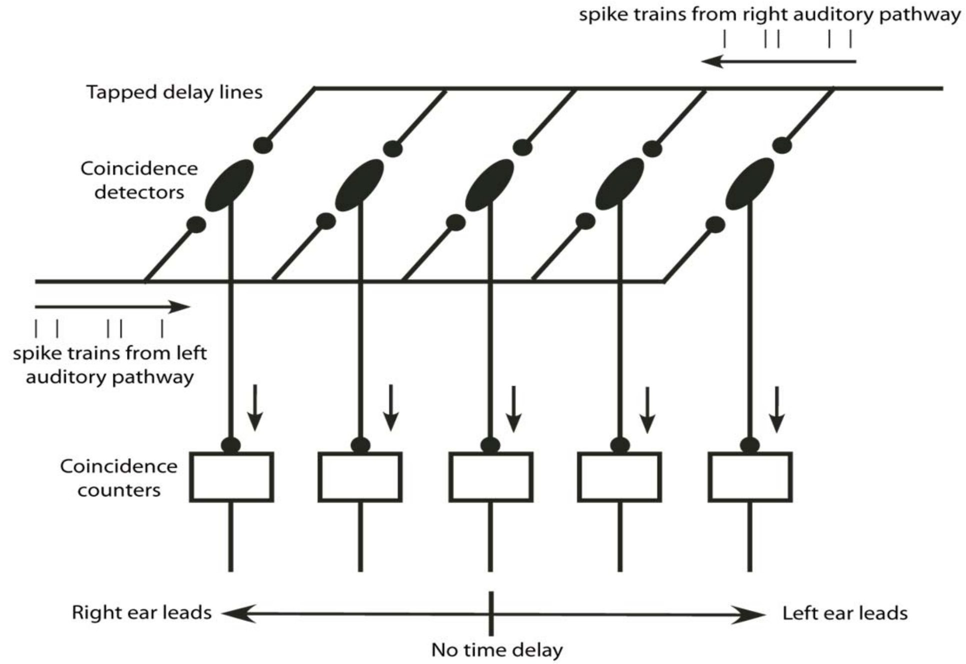


- Spectral cues → Head-Related Transfer Function (HRTF)

Medial Superior Olive

- Extract the time difference of the phase-locked spike pair from both the left and right ear.

- Jeffress model:
  - Array of coincidence detector neurons with different delay lines.

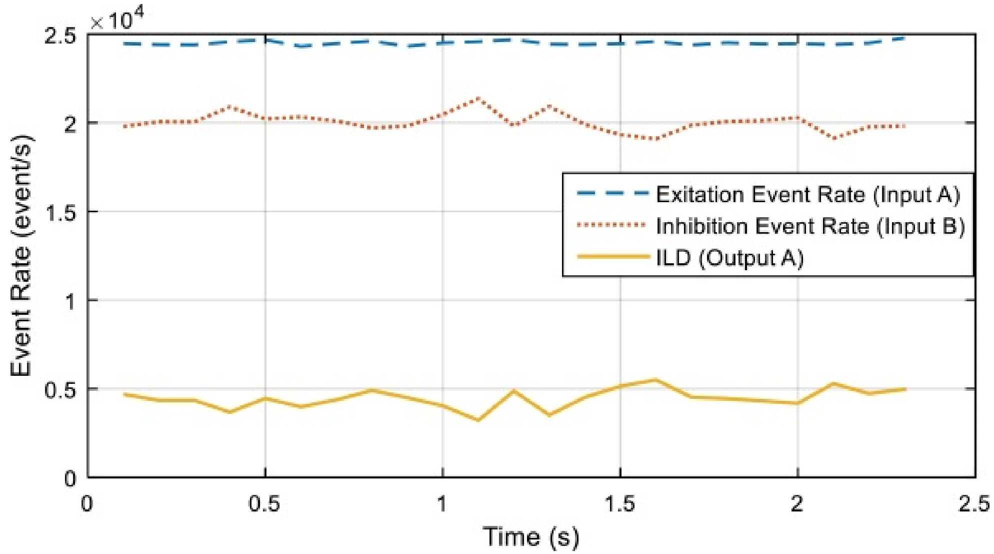


Lloyd A Jeffress. "A place theory of sound localization." Journal of comparative and physiological psychology 41, 1 (1948), 35.

Lateral Superior Olive

- Extract the intensity difference of the spike rates from both the left and right ear.

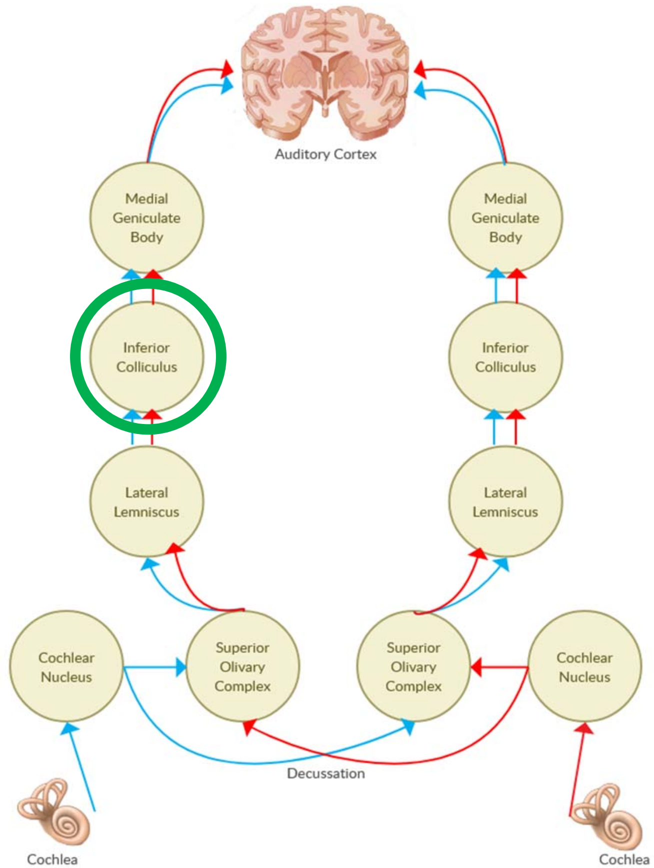
- The output is another stream of spikes.



Cerezuela et al. (2018). Real-time neuro-inspired sound source localization and tracking architecture applied to a robotic platform. *Neurocomputing*

The big picture of the hearing sense

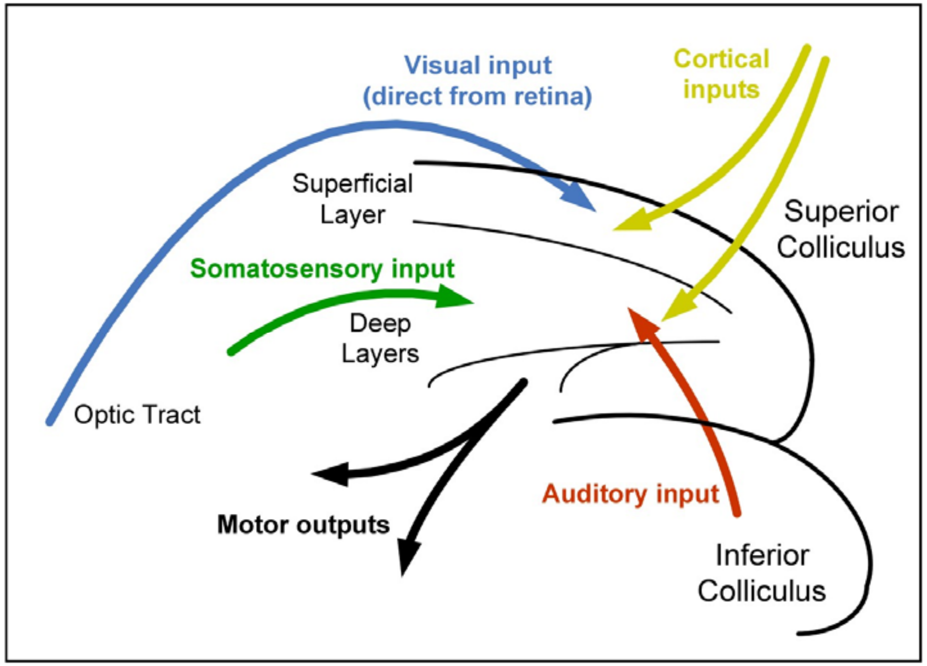
- The ascending auditory pathway:
  - The cochlea
  - The superior olivary complex
  - **The inferior colliculus**
  - The auditory cortex



The inferior and superior colliculus

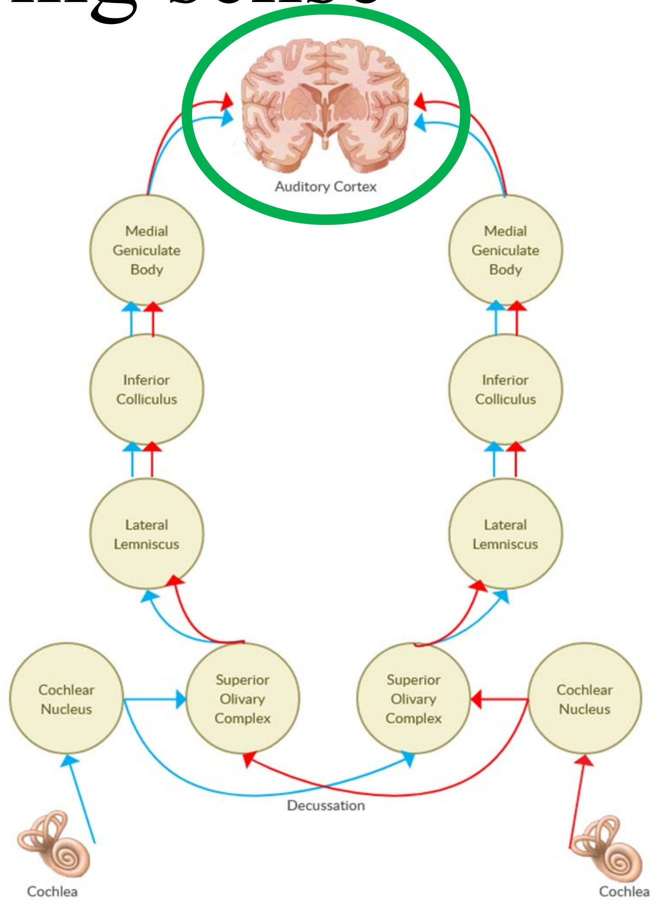
- Inferior colliculus (IC):
  - Combines the auditory information.
  - Involuntary attention.

- Superior colliculus (SC):
  - Combines the sensory information (visual, auditory, touch, …)



The big picture of the hearing sense

- The ascending auditory pathway:
  - The cochlea
  - The superior olivary complex
  - The inferior colliculus
  - **The auditory cortex**

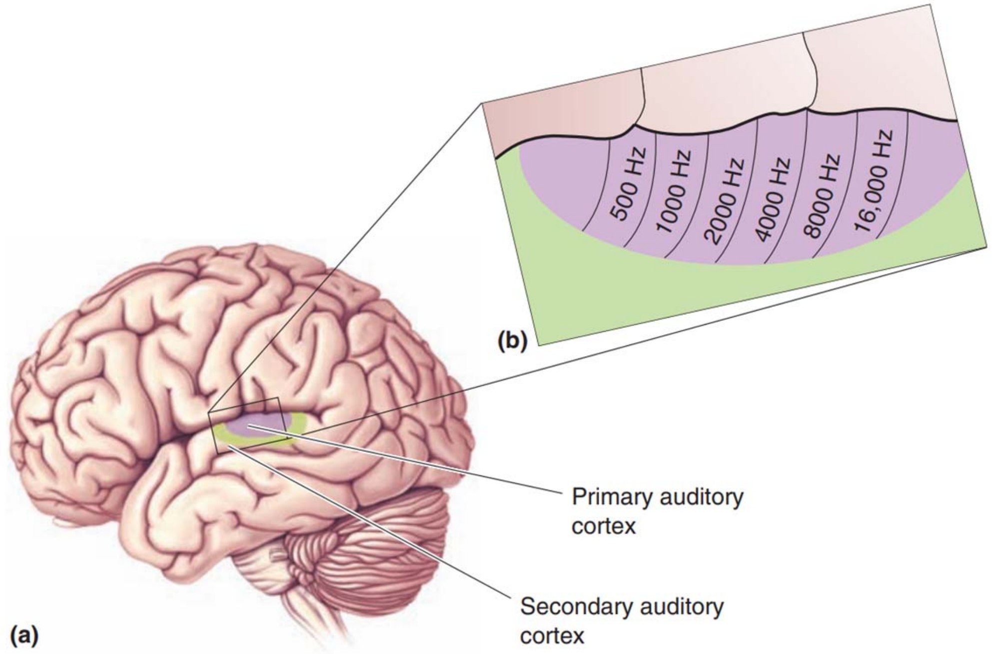


The auditory cortex

- It performs the sound recognition and learning.

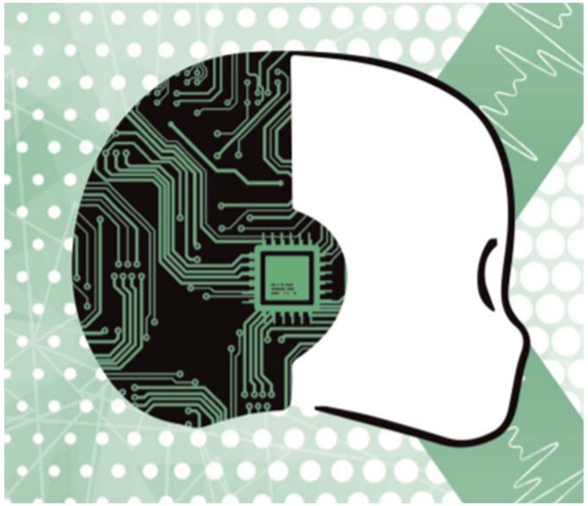
- Also, the voluntary attention.

- It is tonotopically organized.



Outline

- Motivation
- Ascending auditory pathway
- **The iCub robot**
- Giving iCub the sense of hearing
- Preliminary results
- Improvements & future works
- ACKs and questions



What is iCub?

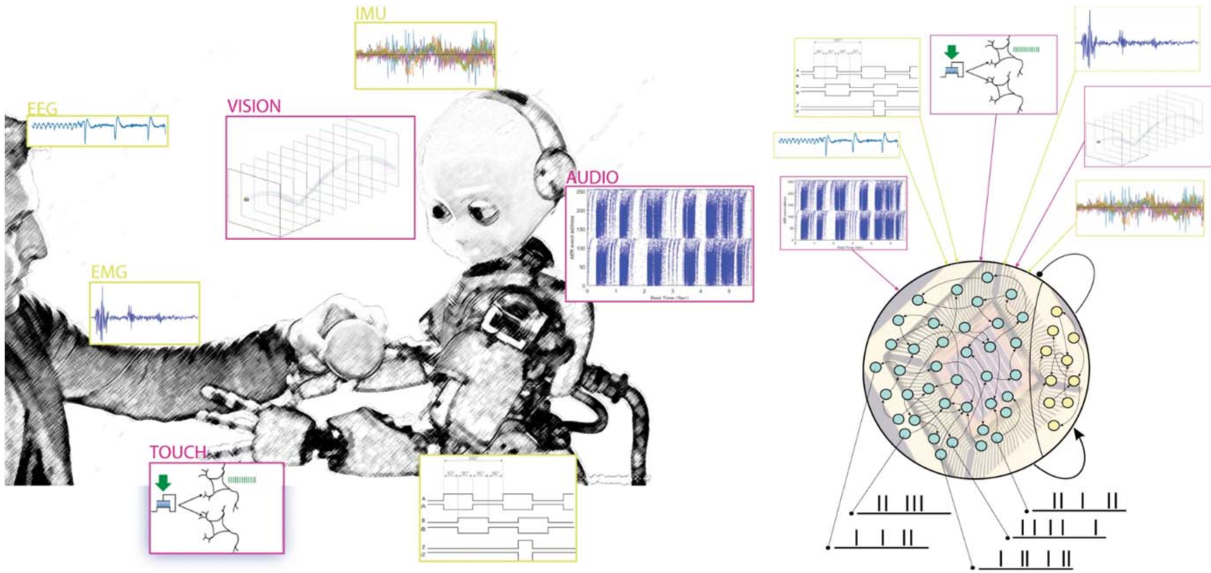
- Humanoid robot looking like a child.
- Developed by the Italian Institute of Technology (IIT).
- Cameras as eyes, microphones as ears, skin and fingertips, motors.

The neuromorphic iCub

- Neuromorphic retinas (ATIS) as eyes and event-based skin.

- Direct interface with SpiNNaker and Loihi.

- But traditional sound processing system!



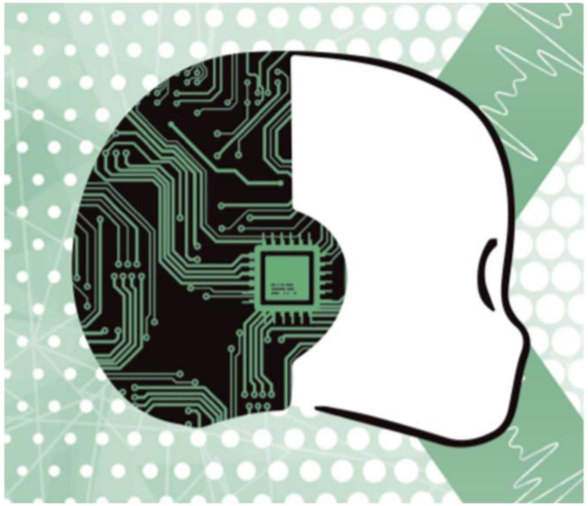
Bartolozzi, C., Indiveri, G. & Donati, E. Embodied neuromorphic intelligence. *Nat Commun* **13,** 1024 (2022).

How can we use iCub?

- By using the YARP framework:
  - Yet Another Robotic Platform (https://www.yarp.it/latest/)
- Based on C++.
- Interface between the hardware devices.
- It is modular → new functionalities just need new modules.

Outline

- Motivation
- Ascending auditory pathway
- The iCub robot
- **Giving iCub the sense of hearing**
- Preliminary results
- Improvements & future works
- ACKs and questions



Main contributions of this work

- The design and implementation of a digital event-based MSO model for extracting the ITD from the cochlea's output events.

- The integration of the MSO with the cochlea model (called NAC).

- The integration in hardware of the NAC within the iCub robot.

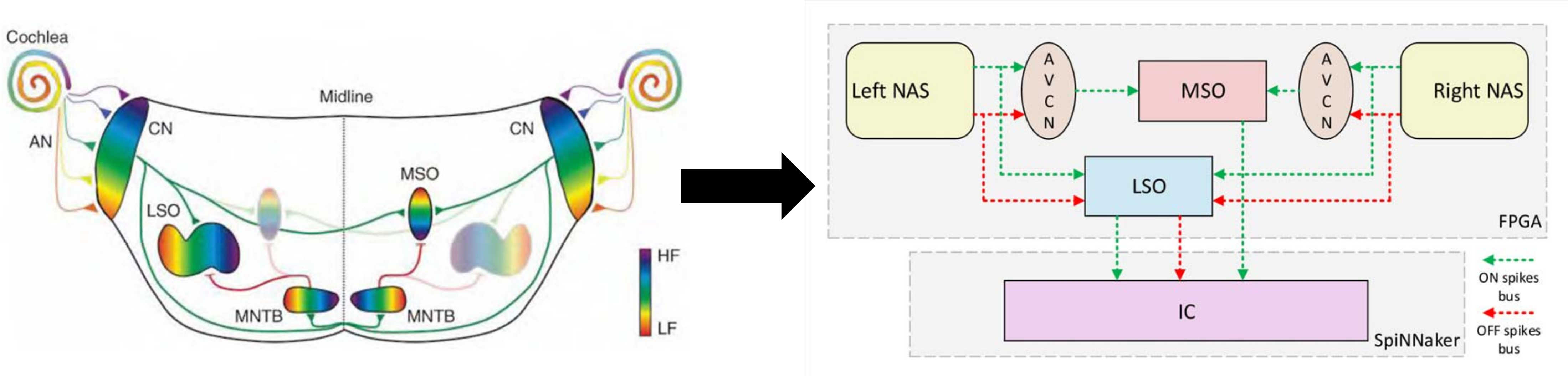
- The integration in software of the NAC within YARP.

Main contributions of this work

- The dataset collection: pure tones, speech commands, speech numbers, static and dynamic sound sources.
- The implementation of the IC model by using SNNs on SpiNNaker.
- The implementation of an open-loop sound source localization demo in real-time moving the iCub's head towards the sound.

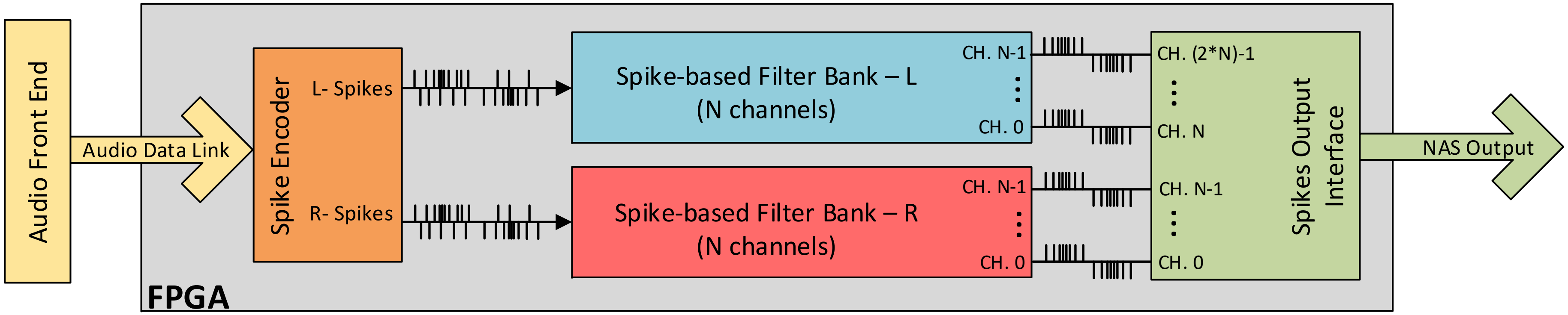
Neuromorphic approach of the hearing sense

- Cochlea and Superior Olivary Complex: on FPGA.
- Inferior colliculus and auditory cortex: on SpiNNaker.



The Neuromorphic Auditory Sensor (NAS)

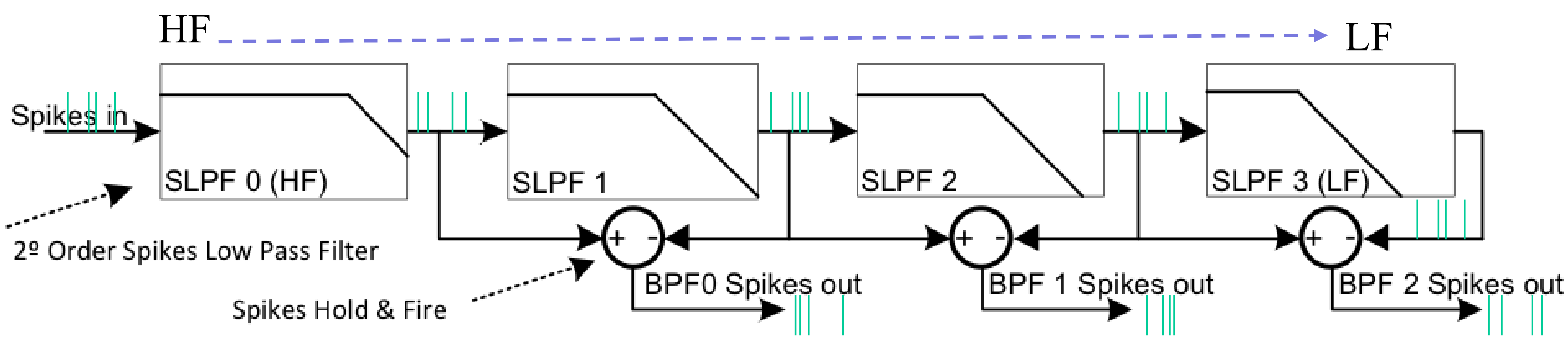
- Digital spike-based cochlea model with cascade architecture.
- Spikes are directly processed using spike-based filter: no audio → no input spikes → no processing.



A. Jiménez-Fernández *et al.*, "A Binaural Neuromorphic Auditory Sensor for FPGA: A Spike Signal Processing Approach," in *IEEE Transactions on Neural Networks and Learning Systems*

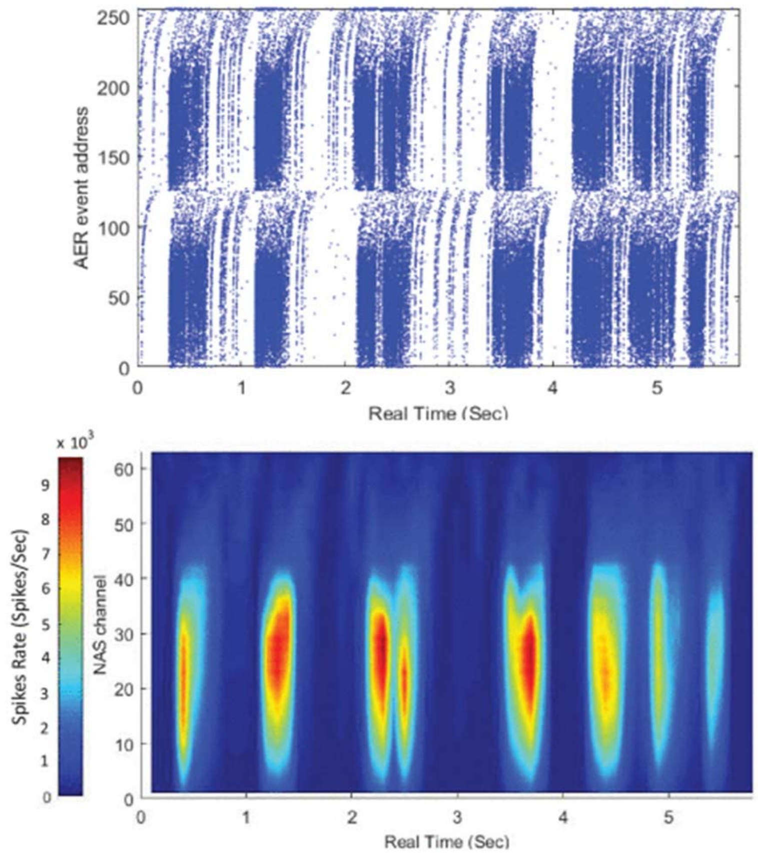
NAS' filter bank

- Cascade of N+1 spike-based low-pass filter.
- Open source: OpenNAS tool for automatically generate the VHDL files and project. https://github.com/RTC-research-group/OpenNAS



NAS’ response

- Binaural 64 frequency channels NAS output.
- “En un lugar de la Mancha”
- Analyzed using pyNAVIS (https://github.com/jpdominguez/pyNAVIS)

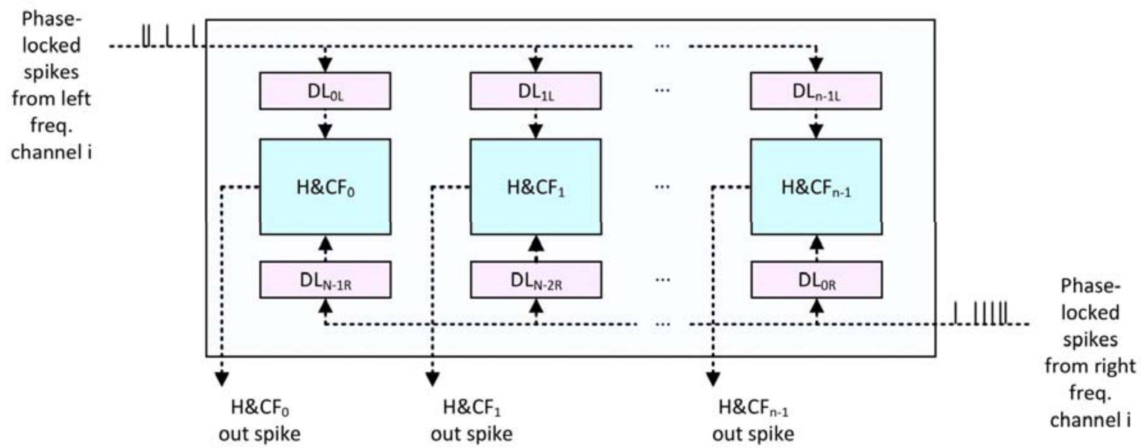


Event-based MSO for FPGA

- Following the Jeffress model:
  - Coincidence detector neuron → **Hold&Coincidence Fire neuron** (H&CF).
  - Delay lines → **Hold & Fire model** (H&F).

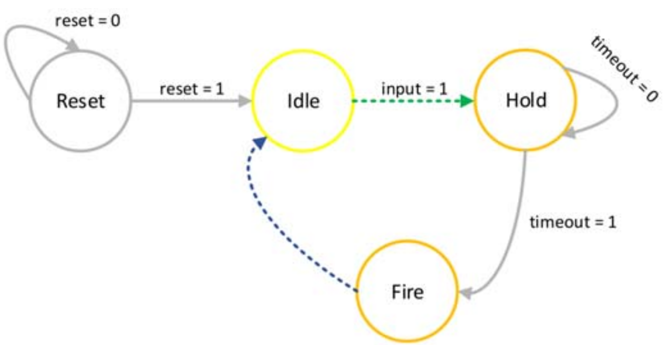
- Array of H&CF as network.

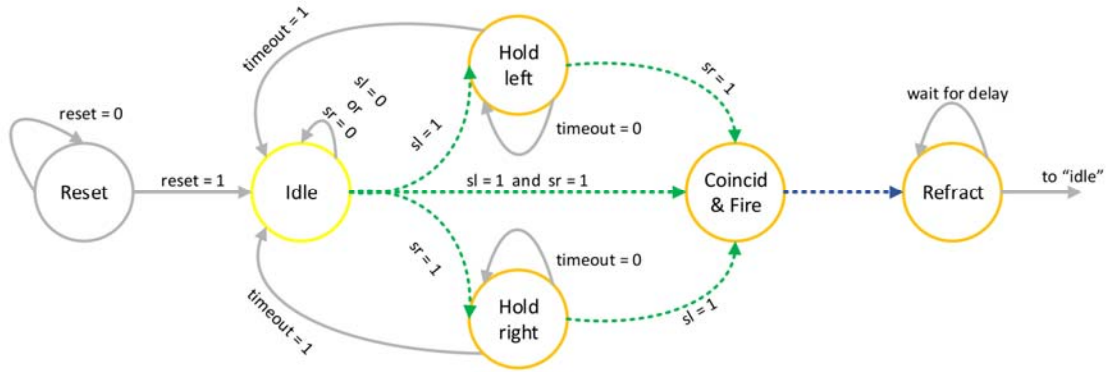
- Two H&F for each neuron.



Event-based Jeffress model for FPGA

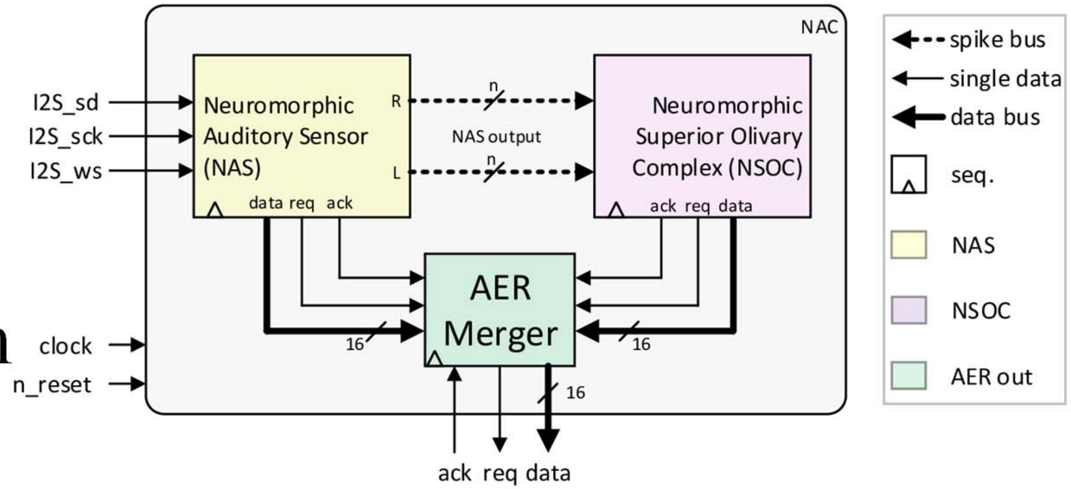
- FPGA resources:
  - Delay line: **0.01% registers and 0.03% LUTs.**
  - Coincidence detector neuron: **0.01% registers and 0.03% LUTs.**

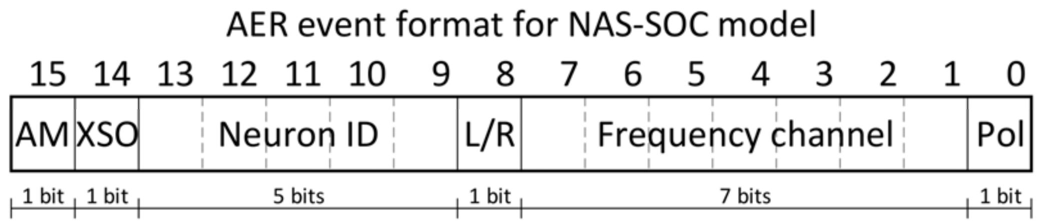




Event-based Neuromorphic Auditory Complex

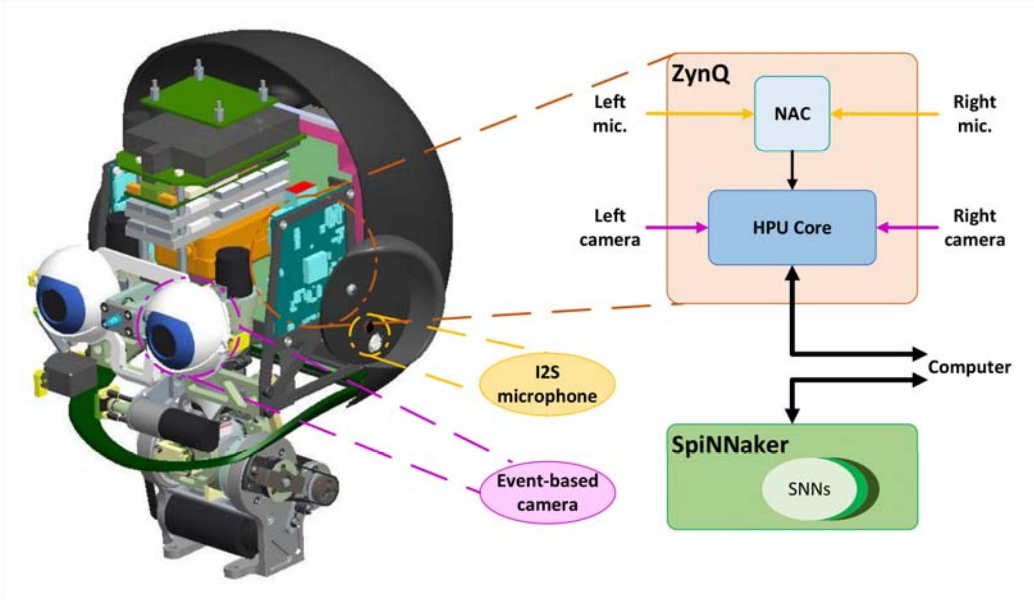
- Combination of the NAS and SOC models.
- Input data: digital samples from the microphones.
- Output data: AER events.





Integrating the NAC within the iCub FPGA

- The NAC was integrated as a new IP module.
- It receives the samples from the I2S MEMS microphones.
- The AER output is sent to the HPU core module.



Integrating the NAC within the iCub FPGA

- NAC features:
  - NAS stereo with **32** frequency channels.
  - Input interface: **I2S** microphones; Output interface: **AER**.

  - MSO model with **4** networks: channels 13 to 16 (592 Hz to1166 Hz)
  - **16** coincidence detector neurons.
  - **700** µs of detection time with **10** µs of overlapping.

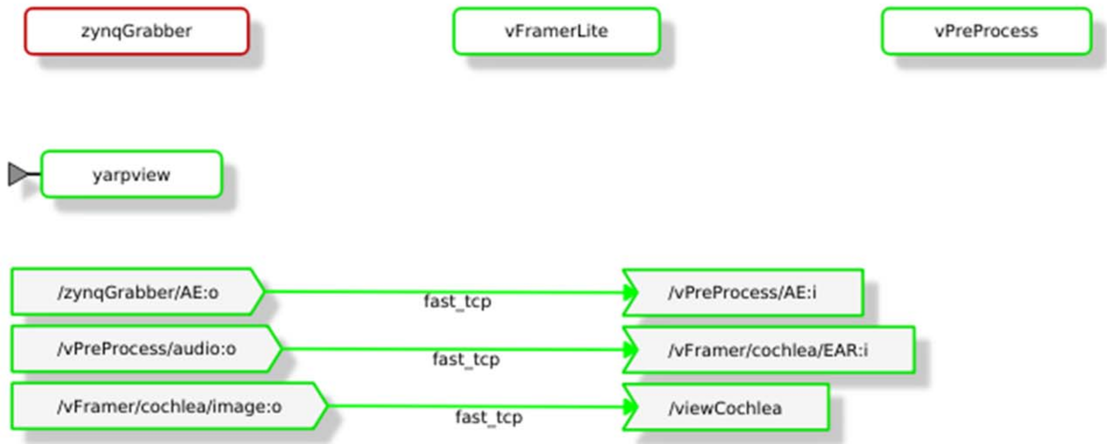
  - No LSO model integrated yet.

Integrating the NAC within the iCub FPGA

- Resources consumption:
  - Entire design: 83.40% LUTs.
    - NAC: 52.18% LUTs.
    - HPU core: 10.41% LUTs.
    - Other modules: 20.07 % LUTs.
  - Resources available for the LSO model: ~15% LUTs.

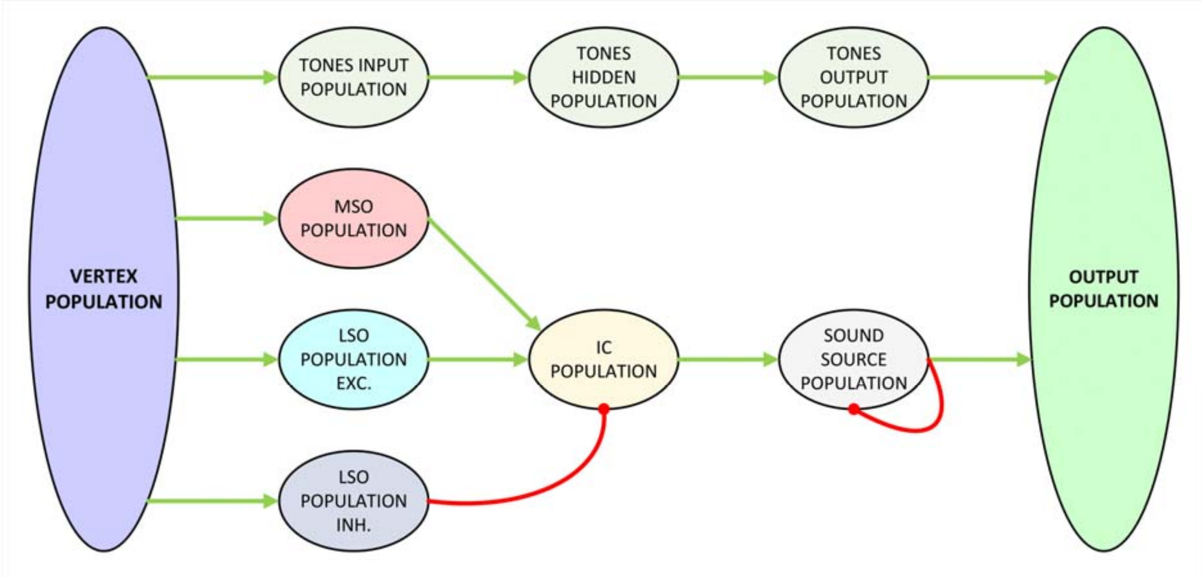
NAC in YARP

- Creation of a new type of event type: EAR.
- Creation of a new decoding process.
- Creating a new set of visualizers.



SNN models on SpiNNaker

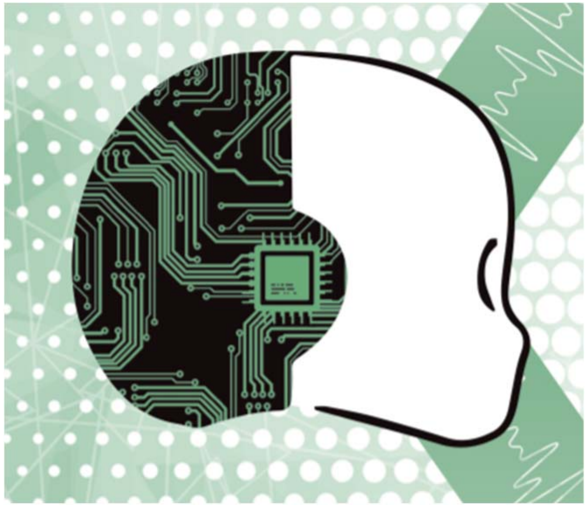
- Two main networks:
  - Sound recognition.
  - Sound source localization.
- The combination of both will lead to an attentional model.



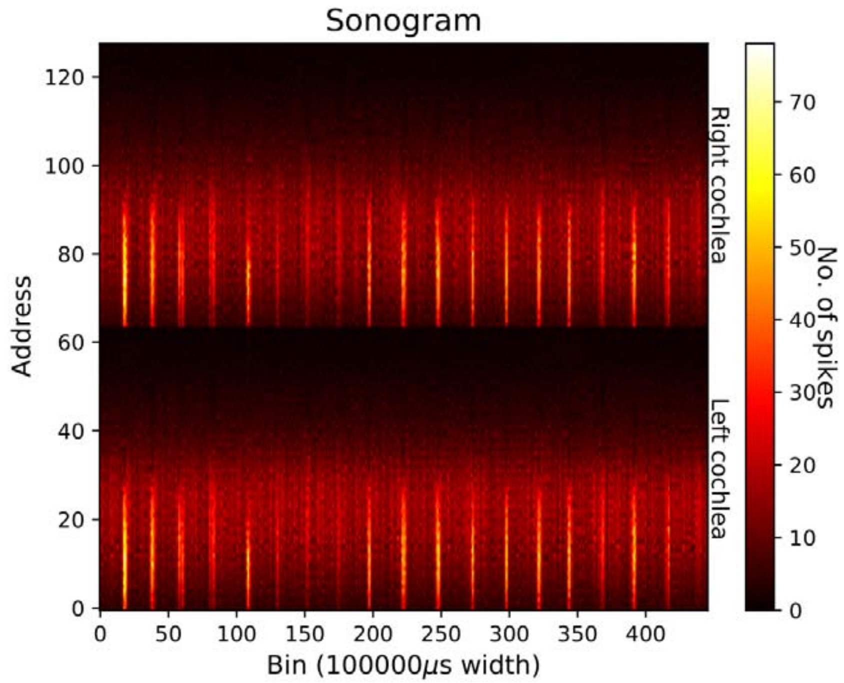
Dominguez-Morales,, et al. "Multilayer spiking neural network for audio samples classification using SpiNNaker." *International conference on artificial neural networks*. 2016.

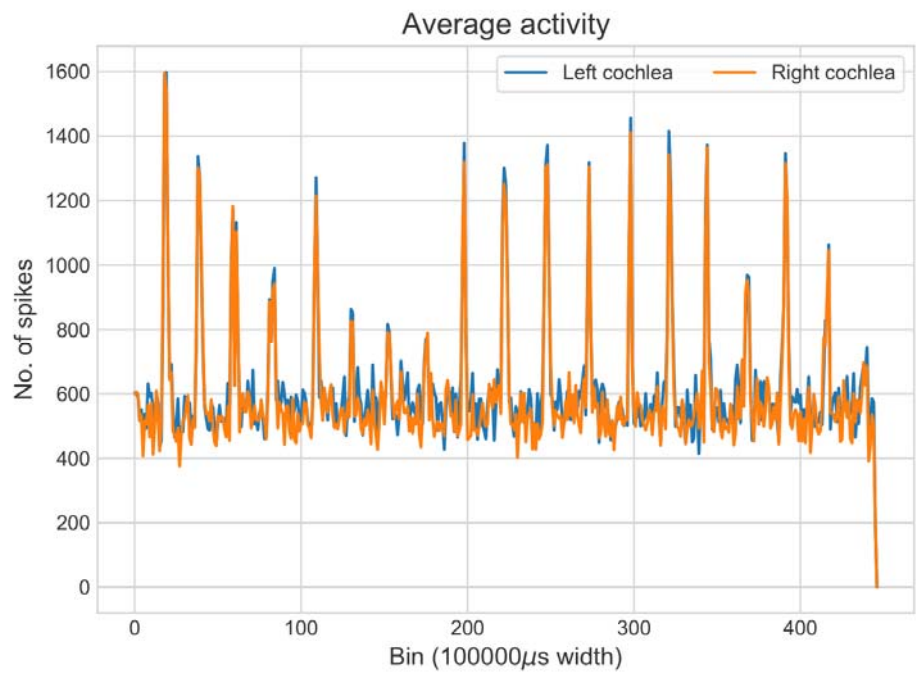
Outline

- Motivation
- Ascending auditory pathway
- The iCub robot
- Giving iCub the sense of hearing
- **Preliminary results**
- Improvements & future works
- ACKs and questions



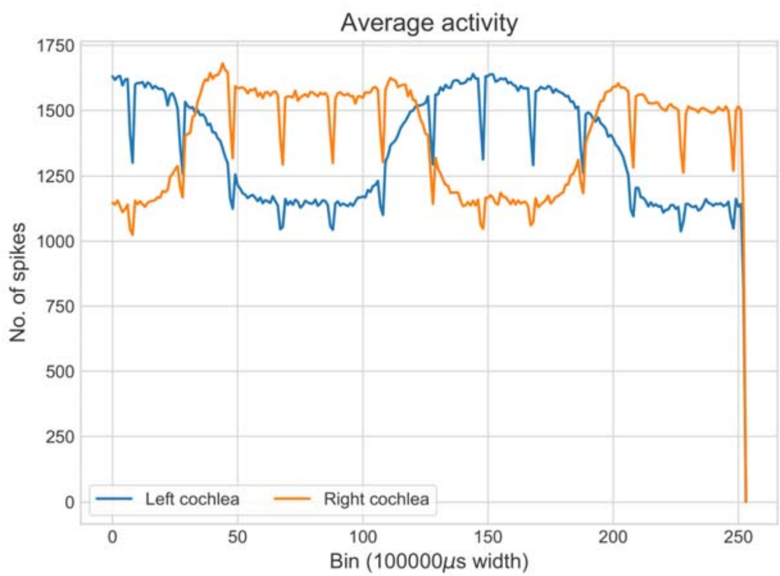
NAS output for the speech commands dataset

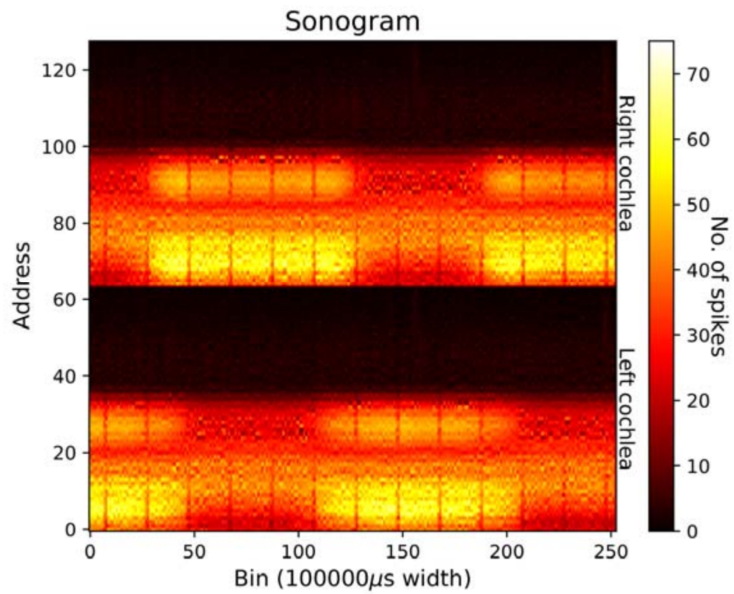




MSO output with dynamic source

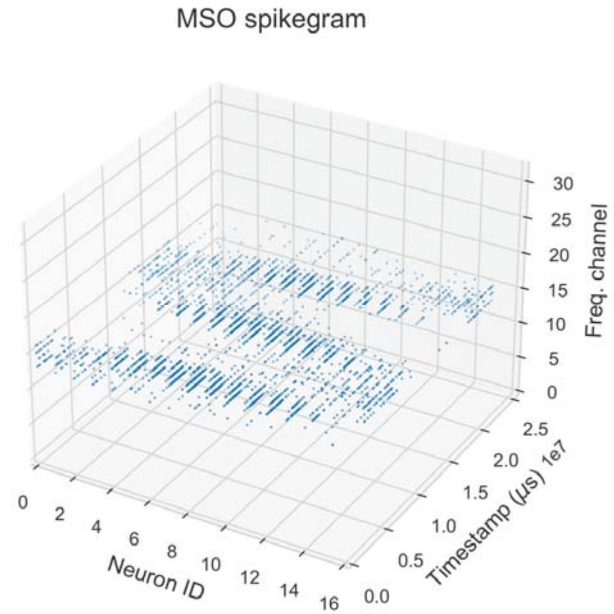
- Sound source localization of a 500 Hz pure tone with the iCub moving its head from left to right.

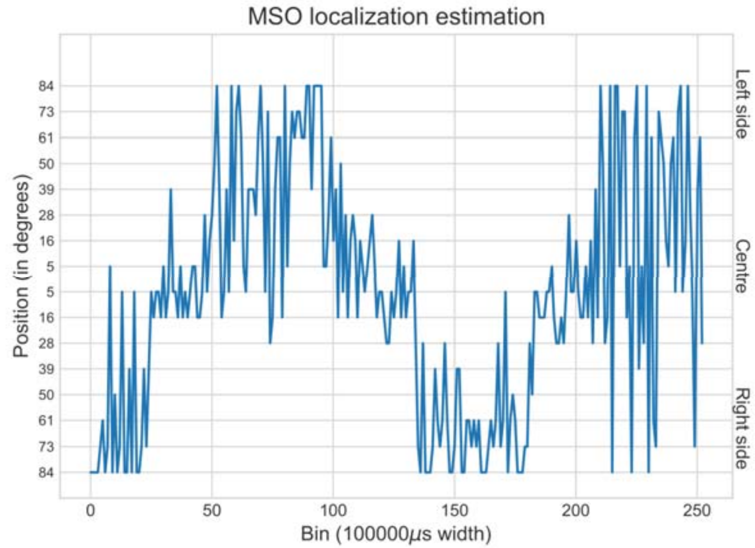




MSO output with dynamic source

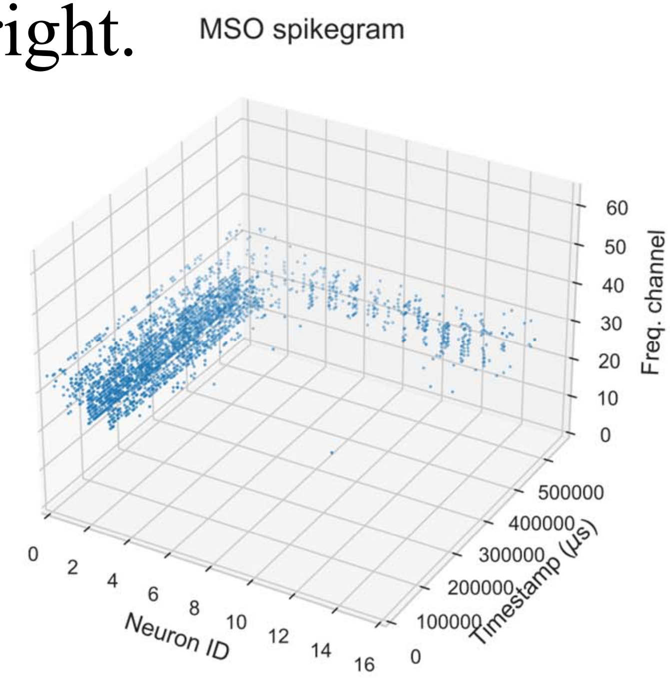
- Sound source localization of a 500 Hz pure tone with the iCub moving its head from left to right.

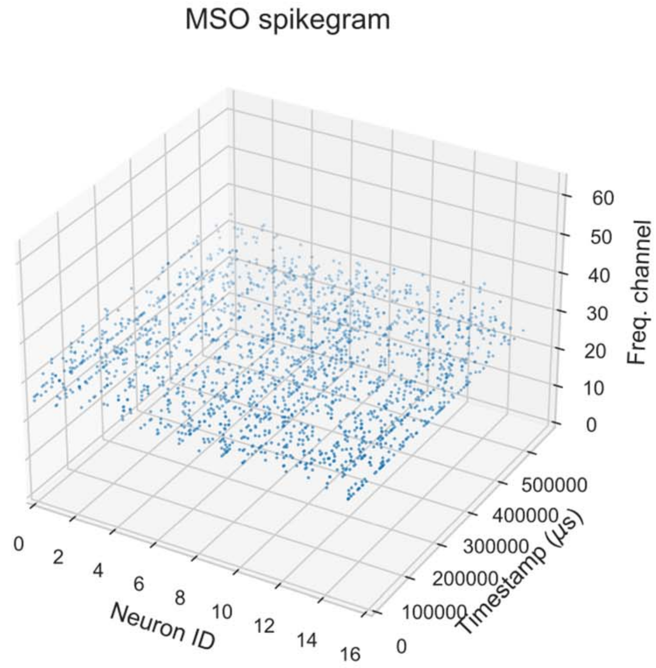




Validating the MSO model

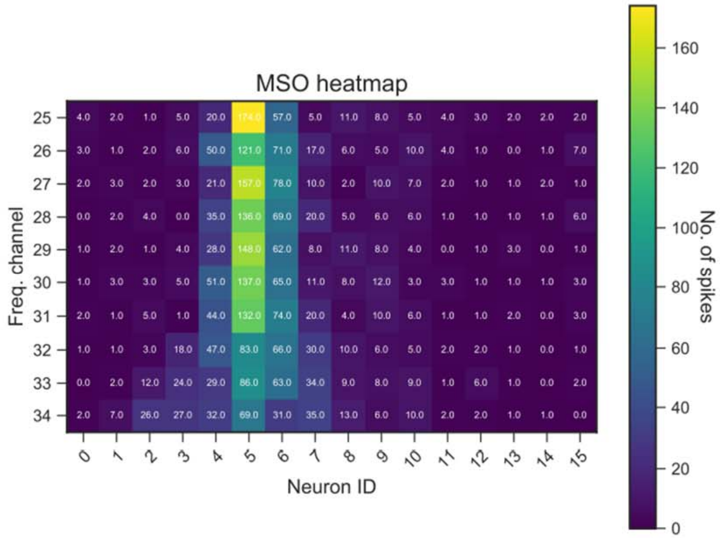
- How does the frequency of the sound affect the model? Sound source placed at right.
  - 500 Hz (left)
  - 1550 Hz (right)

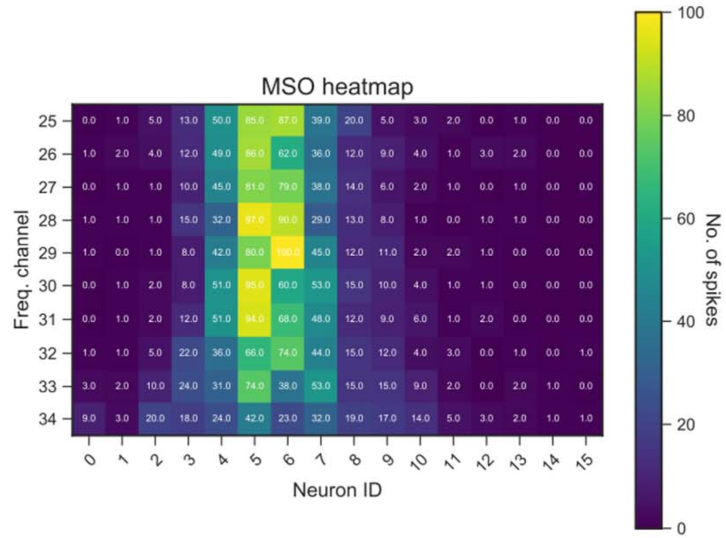




Validating the MSO model

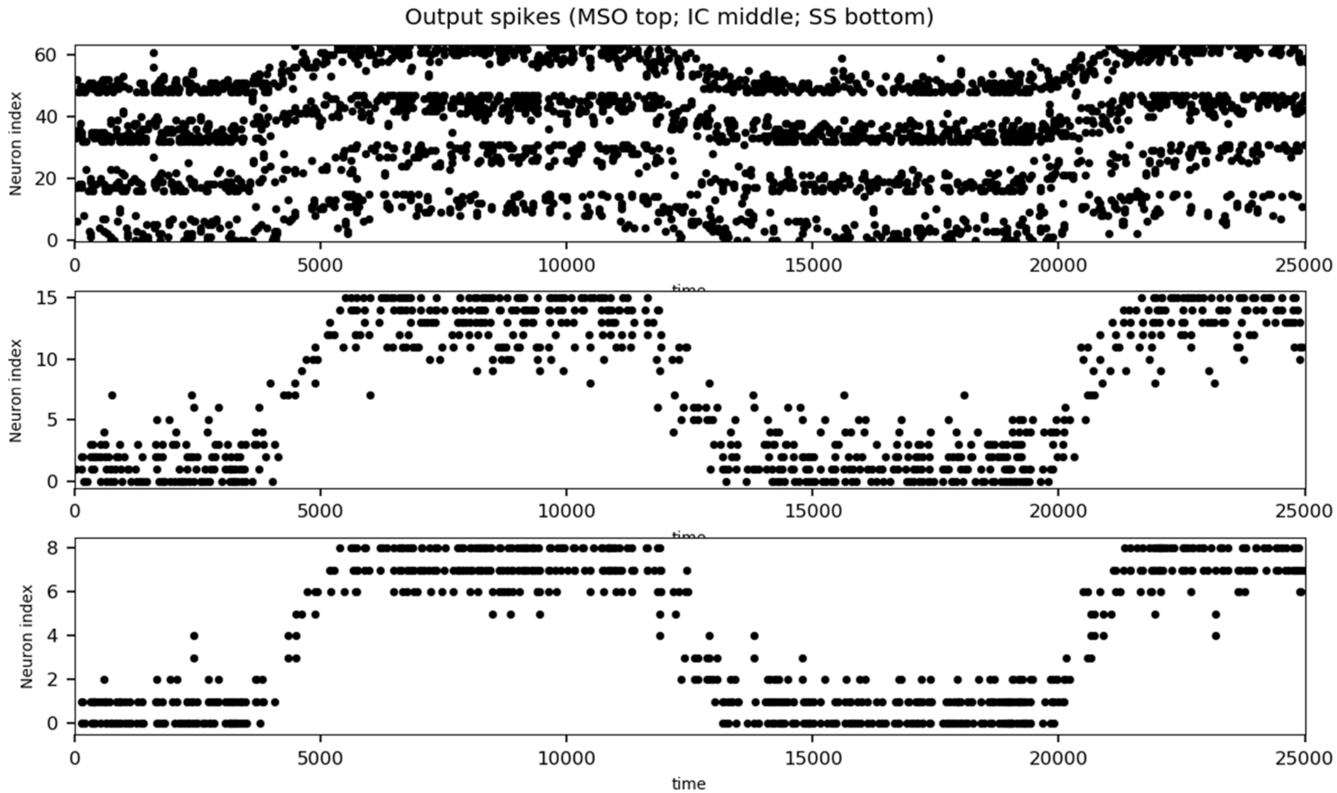
- How does the distance of the sound affect the model? Sound source placed at 45º on the right.
  - 0.5 m (left)
  - 3.0 m (right)





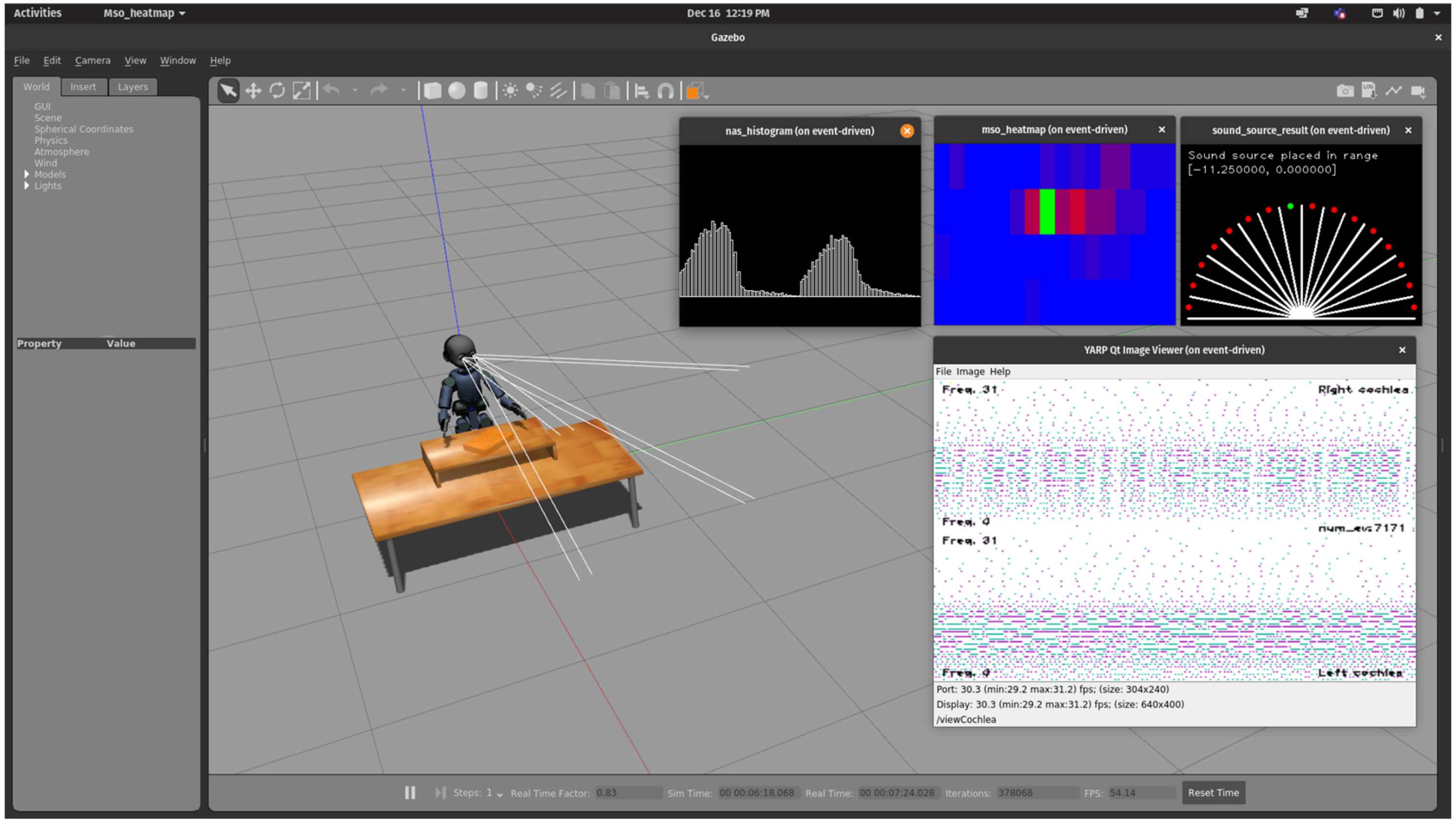
IC output from SpiNNaker

- Output spikes from the inferior colliculus SNN on SpiNNaker for the head sweep.



Live demo of sound source localization

- Simulation of the iCub robot moving its head towards the sound source.

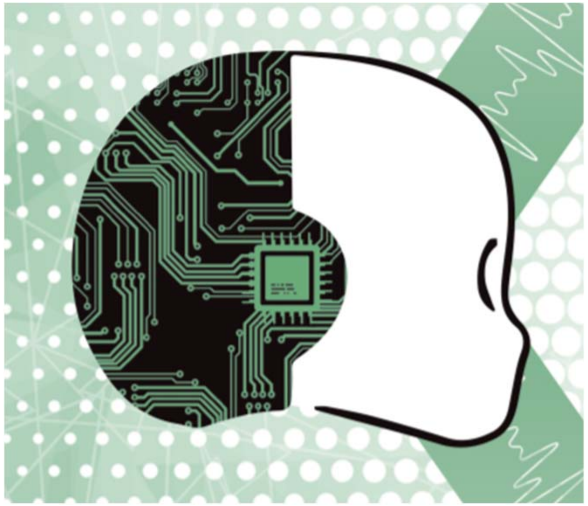


Preliminary results extracted from the tests

- **Localization accuracy**:
  - Theoretically: 180º / N_NEURONS
  - Measured: depending on the distance.
- **Maximum ITD**:
  - Theoretically: 700 μs.
  - Measured: 400 μs, since the distance between ears in iCub is 13.6 cm.
- **Sound recognition accuracy**:
  - Less than 65% due to the iCub's fan noise.

Outline

- Motivation
- Ascending auditory pathway
- The iCub robot
- Giving iCub the sense of hearing
- Preliminary results
- **Improvements & future works**
- ACKs and questions

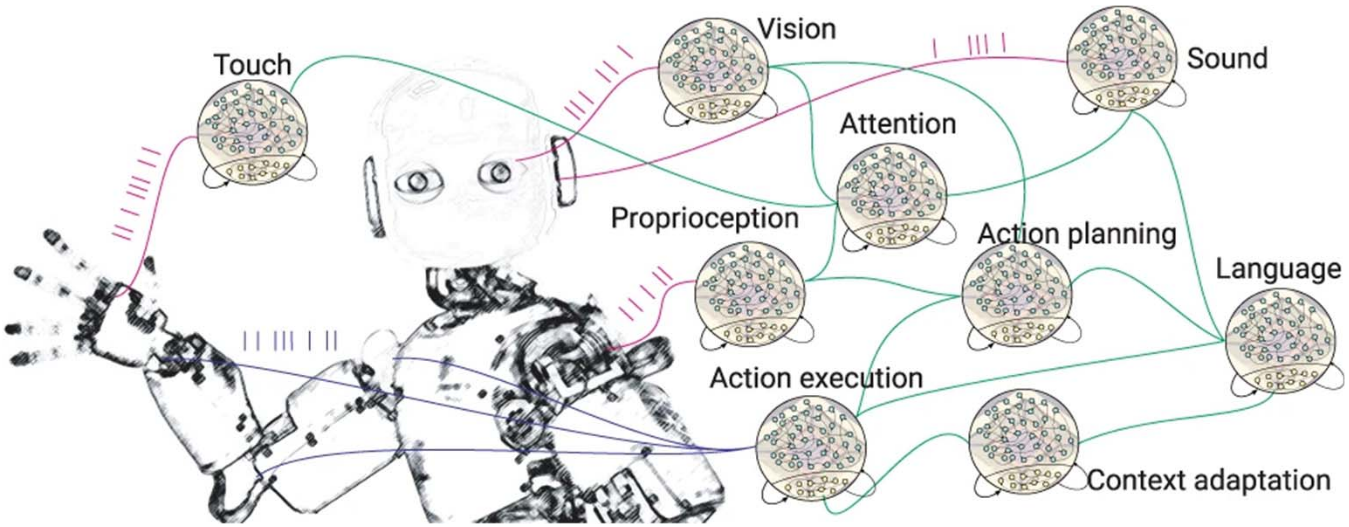


What could be better?

- To make the NAC adaptative.
- To improve the SNN models.
- To remove the iCub's fan noise:
  - It has 3 fans located close to the microphones.
- To improve the iCub's movement algorithm.

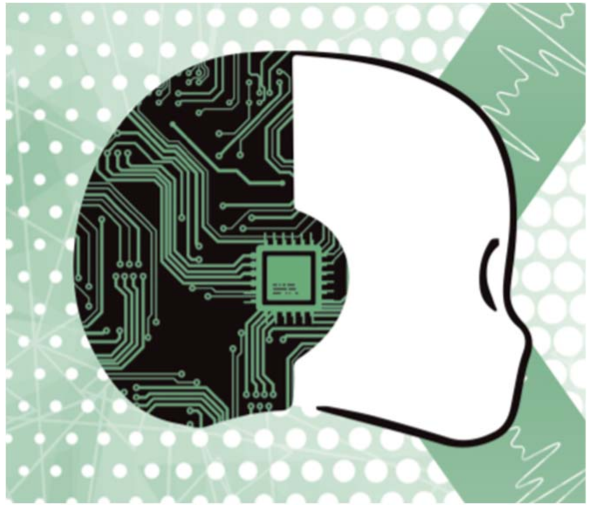
And what is next?

- Sensory integration of the neuromorphic sensors:
  - – Combining vision and hearing for attention.
  - – Cocktail party problem.



Outline

- Motivation
- Ascending auditory pathway
- The iCub robot
- Giving iCub the sense of hearing
- Preliminary results
- Improvements & future works
- **ACKs and questions**



Acknowledgements

- Thanks to:
  - Maurizio Casti, Arren Glover, Marco Monforte, Massimiliano Iacono, Simon Muller-Cleve, Ander Arriandiaga, Giulia D’Angelo Marta Caracalli.
  - COFNET project (TEC2016-77785-P)
  - MINDROB project (PID2019- 105556GB-C33/ AEI/10.13039/501100011033)

GOBIERNO DE ESPAÑA

MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

Thank you all for your attention! It has been a pleasure to let you all know more about our work.