A Low-cost Wireless Acoustic Sensor Network for the Classification of Urban Sounds

David Salvo, **Gema Piñero**, Pau Arce, Alberto Gonzalez
Inst. Telecommunications and Multimedia Applications (iTEAM) – UPV - Spain

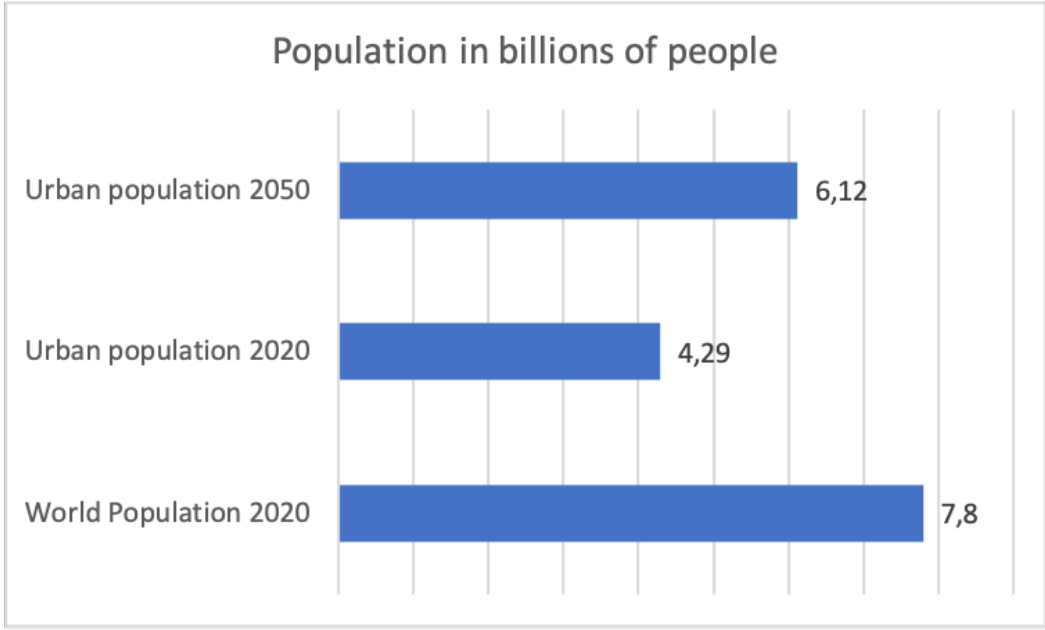
PE-WASUN 2020

OUTLINE

- Motivation
- WASN for Urban Sound Classification
- Pre-detection stage
- Conclusions

Motivation

According to United Nations* prospects:



7.8 billions*

World Population 2020

4.29 billions (55%)

Live in urban areas (2020)

6.12 billions (68% of 9B)

Will live in urban areas (2050)

*UN 2018. The World's Cities in 2018. Retrieved September 17, 2020 from https://www.un.org/en/events/citiesday/assets/pdf/the_worlds_cities_in_2018_data_booklet.pdf

Motivation

- Inspired by SONYC Project

- Combatting Acoustic Pollution:

9 of 10 adults in NYC

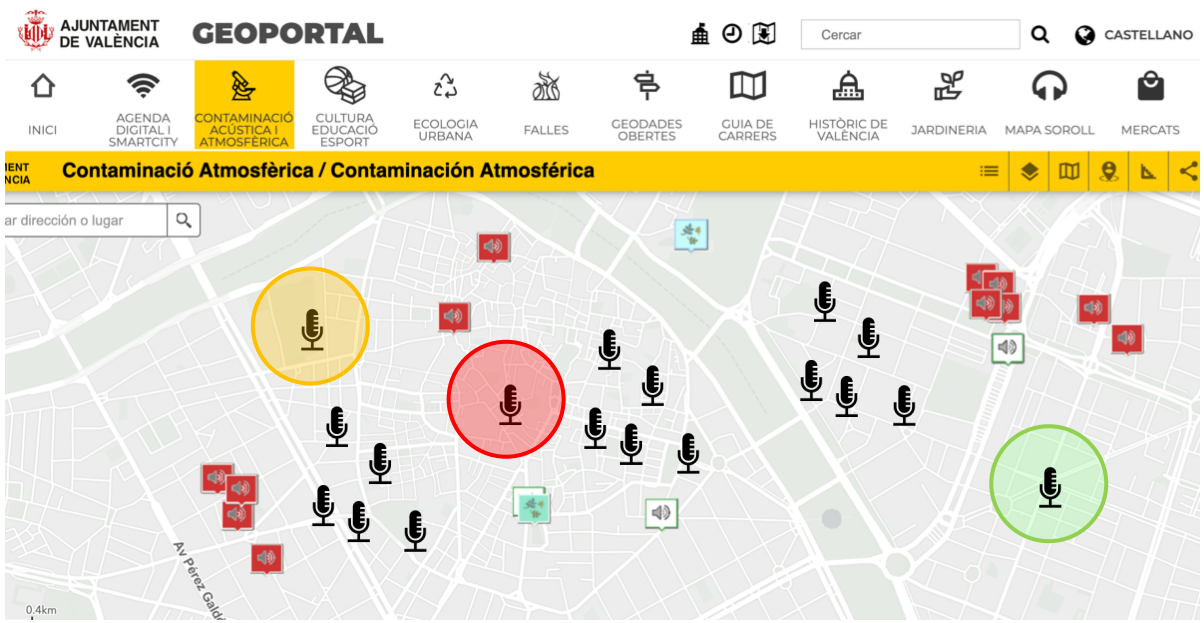
Are exposed to Sound Pressure Levels (SPLs) higher than 70 dBs

630.000 inhabitants

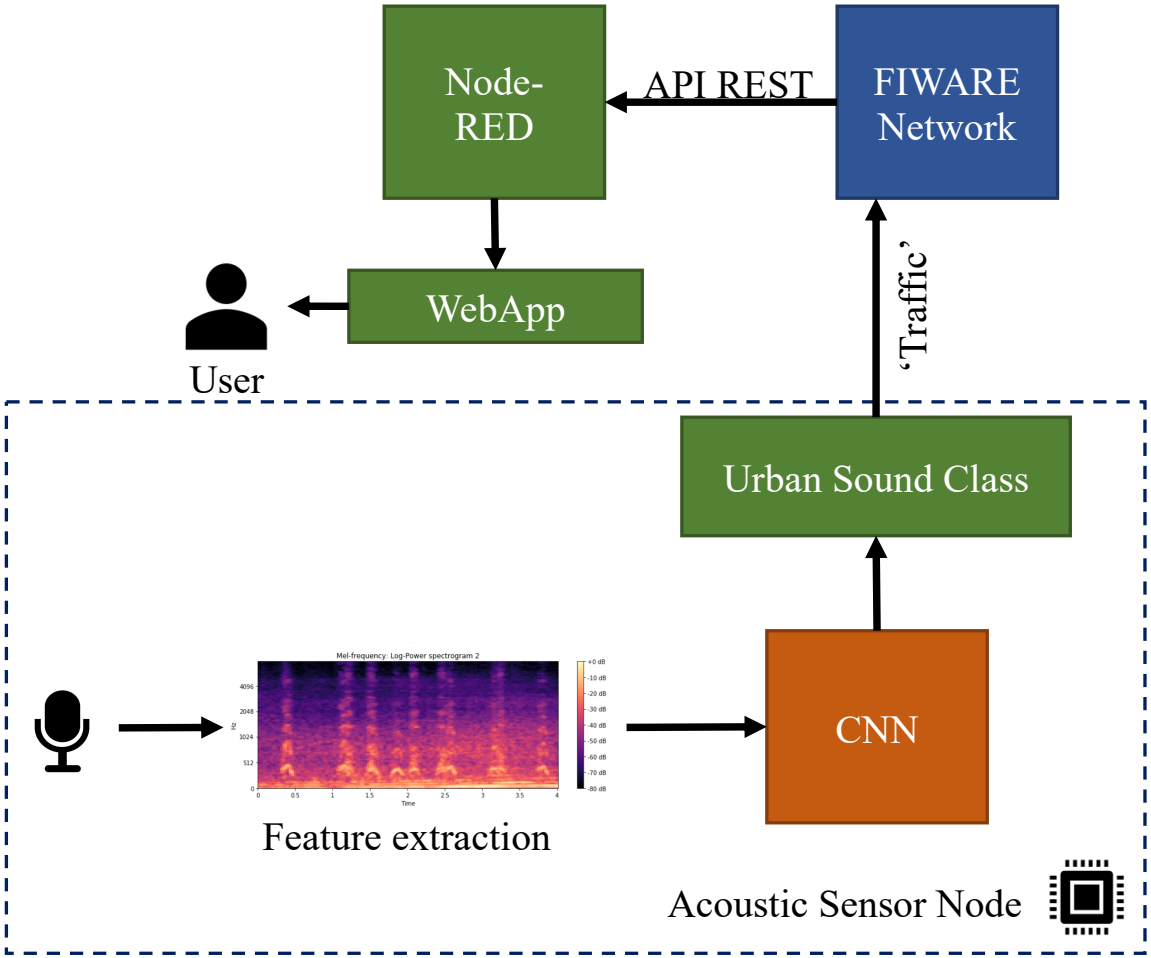
of Valencia city are exposed to SPLs higher than 70 dBs

Urban Sound Classification (USC)

- Noise monitoring system in the city: Geoportal
- Advanced noise monitoring system providing USC:
  - Efficient
  - Low-cost
  - Easy to deploy
  - Scalable
  - Open standards

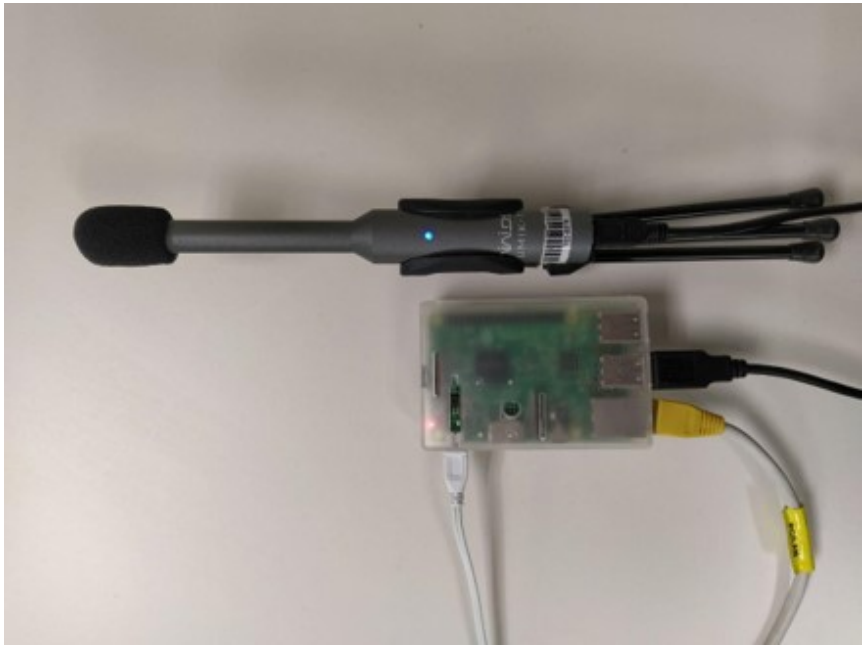


WASN for Urban Sound Classification



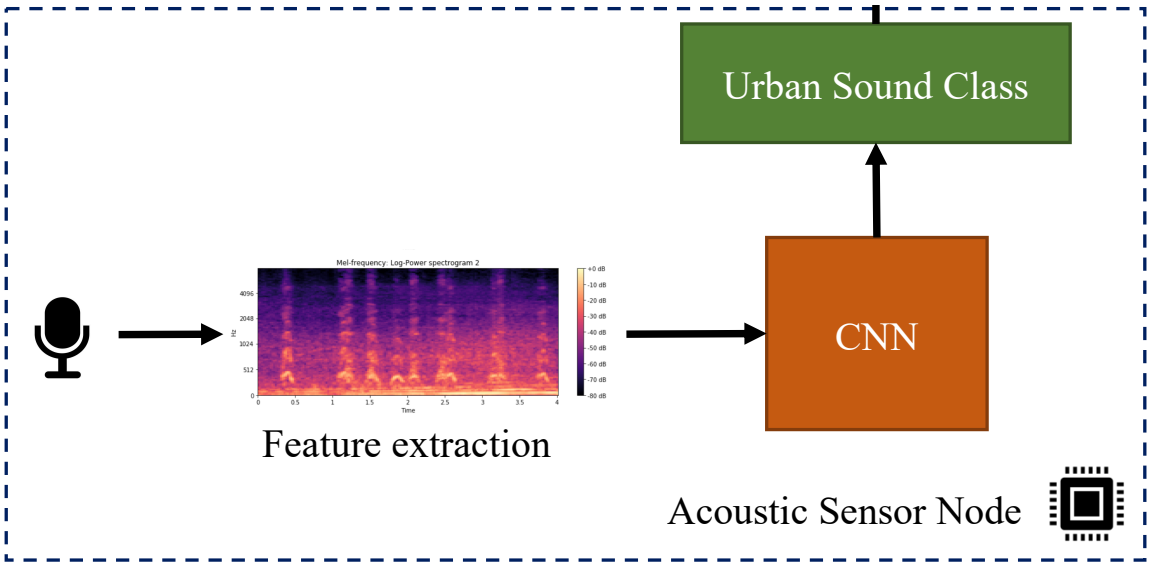
Acoustic sensor node

- Raspberry Pi 4 Model B
- miniDSP UMIK-1 microphone (outdoor)



Acoustic sensor node

- Raspberry Pi 4 Model B
- miniDSP UMIK-1 microphone (outdoor)
- The USC is performed in the node

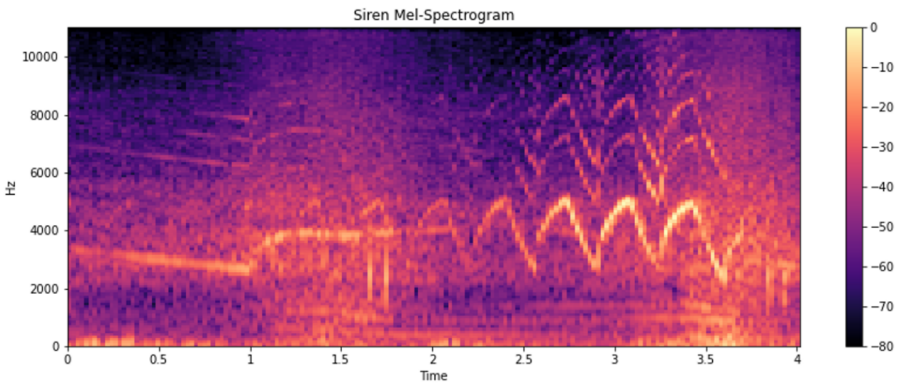


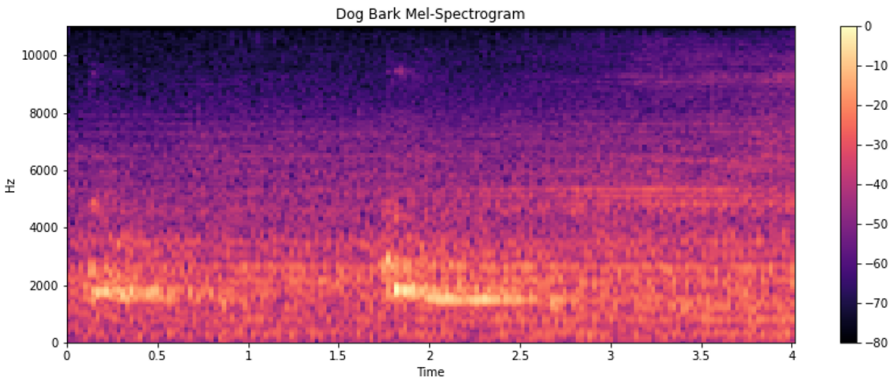
Dataset

- UrbanSound8K (SONYC project): https://urbansounddataset.weebly.com/
- 8732 labelled sound excerpts (<=4s) of urban sounds
- All excerpts are taken from field recordings uploaded to www.freesound.org
- 10 classes: air conditioner, car horn, children playing, dog bark, drilling, engine idling, gun shot, jackhammer, siren, and street music

Feature extraction

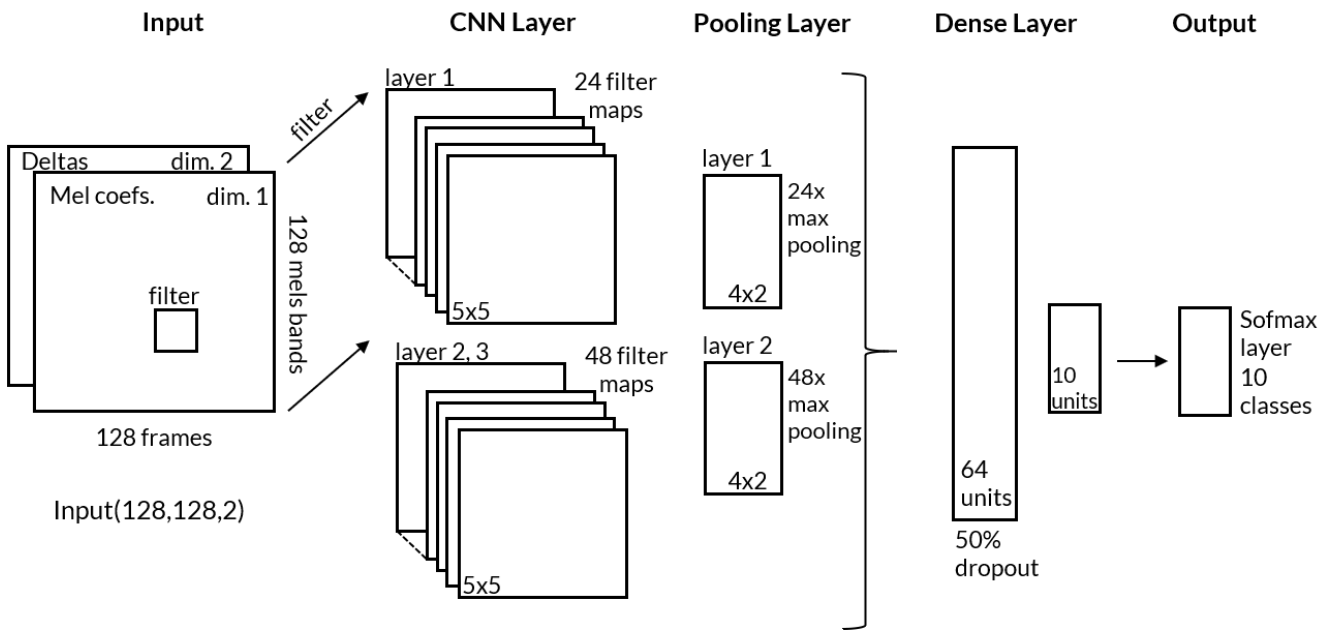
- Well-known features in the field of Environmental Sound Classification (ESC)
- 3-seconds excerpts -> 128 time frames
- Mel frequency spectral coefficients (MFCCs) plus their DeltaMFCC for 128 bands covering 0-11025 Hz
- 128x2 coefficients per 128 frames: network input formed by a [128 x 128 x 2] matrix





Convolutional Neural Network

- Based on SONYC: 3 convolutional layers, 2 dense layers, 1 dropout layer

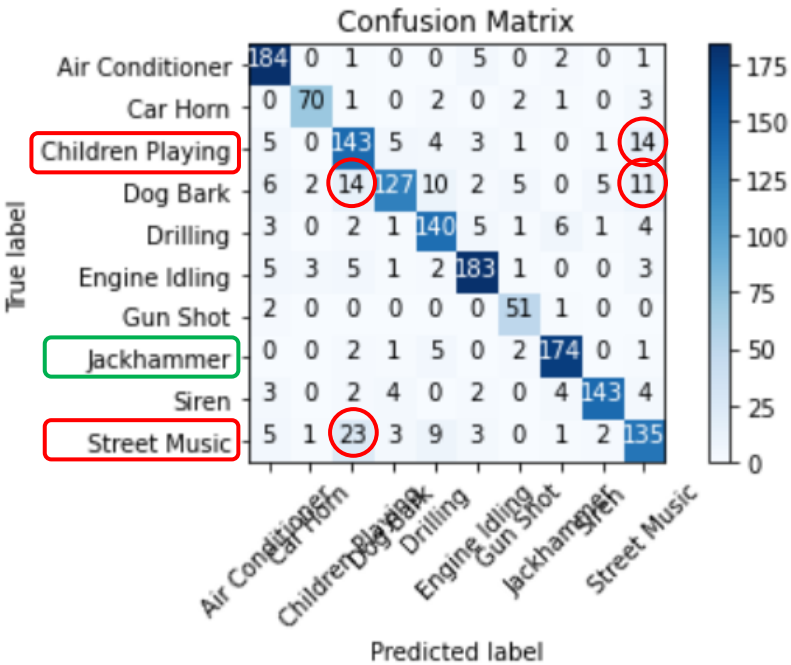


- 80% train, 10% validation, 10% test
- Stochastic gradient descent, learning rate 0,001
- ReLu for all the layers except the last dense layer (Softmax)

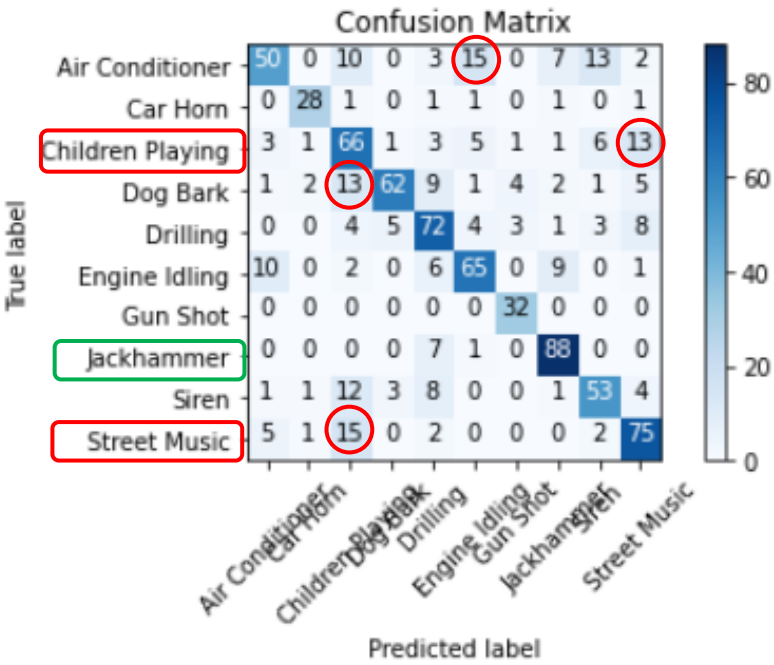
CNN Performance

- Accuracy: 71%
- We are now working on hyperparameter optimization

Validation Dataset, 1579 events



Test Dataset, 837 events



USC on the Raspberry Pi

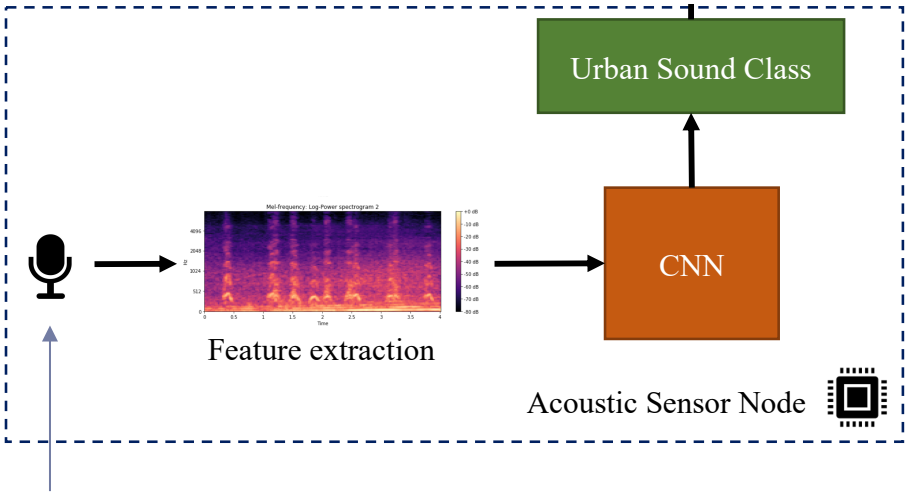
- Python libraries (librosa)
- Tensorflow and Keras for training in the PC, Tensorflow lite for copying the model to the Raspberry Pi
- Computation times of the different processes in the tables

Table 1: Hardware Specifications and Total Processing Time

Device	Processor	RAM	Time
Dell XPS 15 (2015)	I7-6700HQ 2.60 GHz	16GB	0.195 s
Raspberry Pi 3 Model B	Quad Core 1.2 GHz	1GB	2.147 s

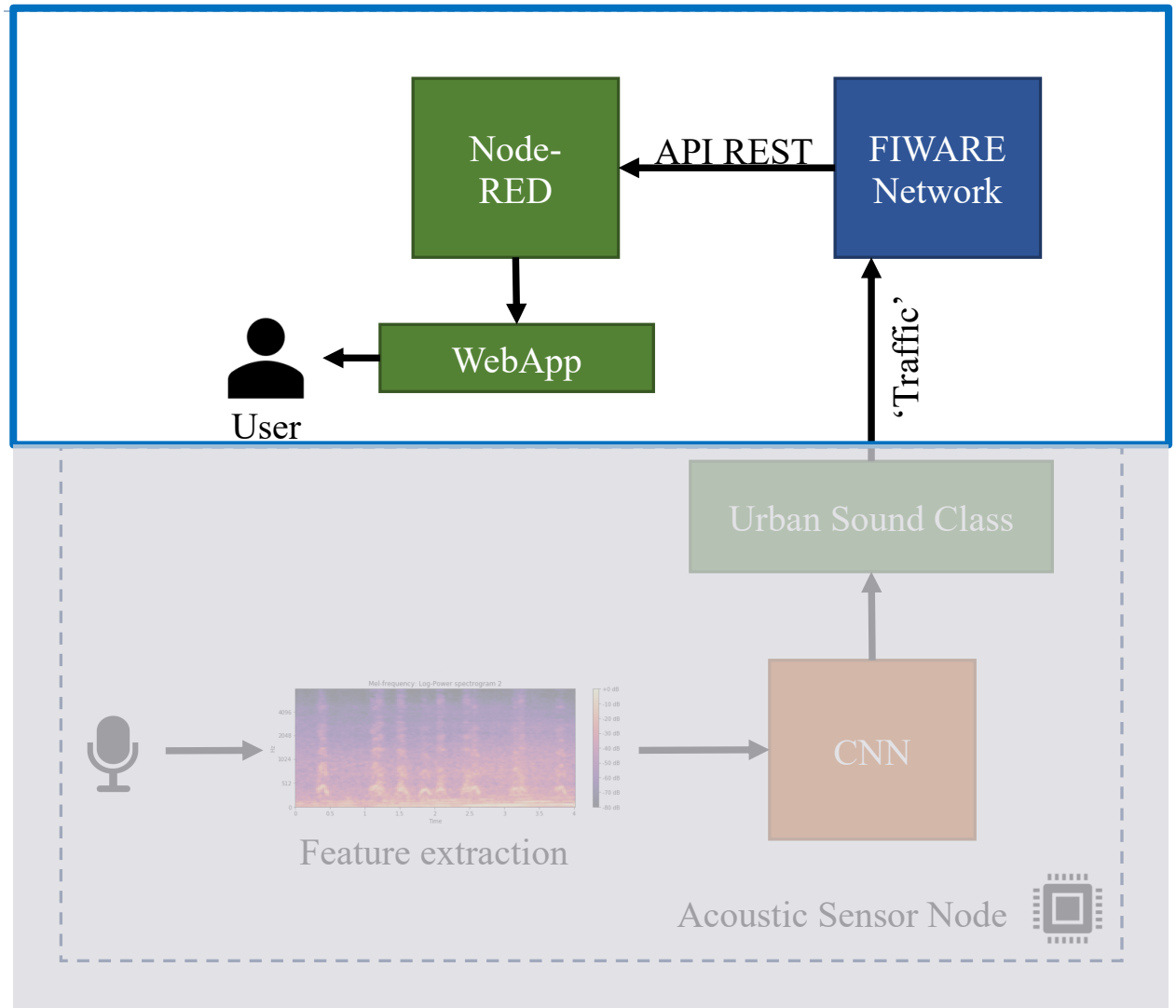
Table 2: Processing Time per Specific Task

Device	*wav* file Loading	Feature Extraction	CNN Classification
Dell XPS 15 (2015)	0.124 s	0.012 s	0.059 s
Raspberry Pi 3 Model B	2 s	0.052 s	0.095 s



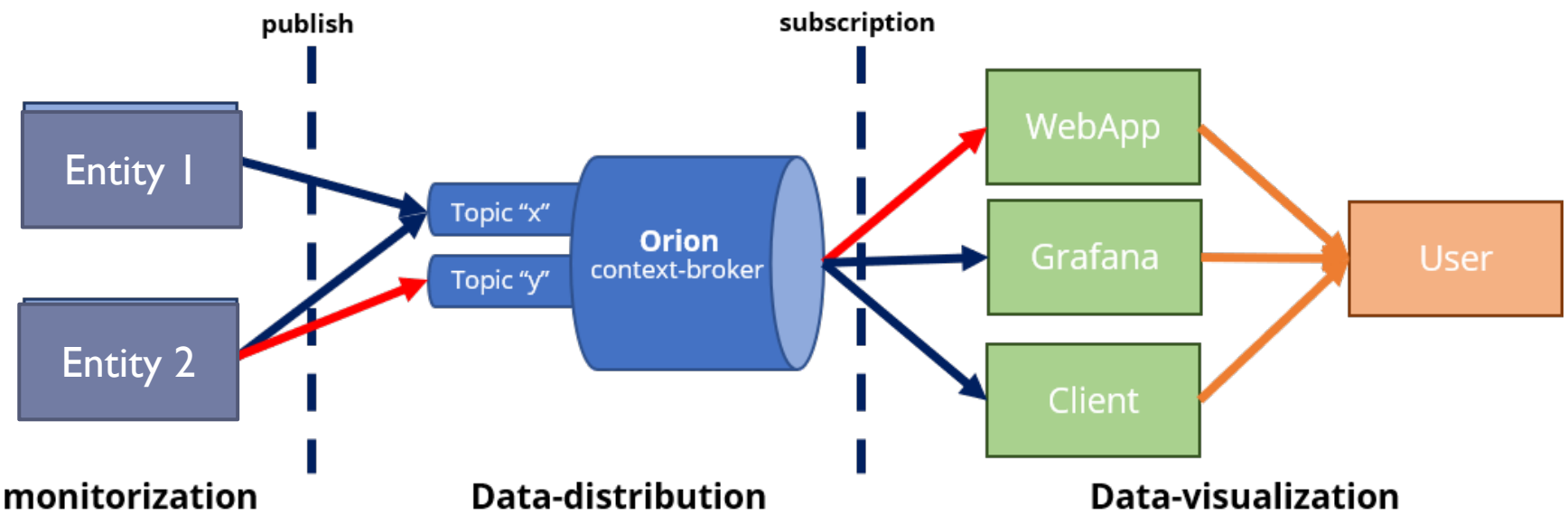
Common audio record time: 3.089s

WASN for Urban Sound Classification

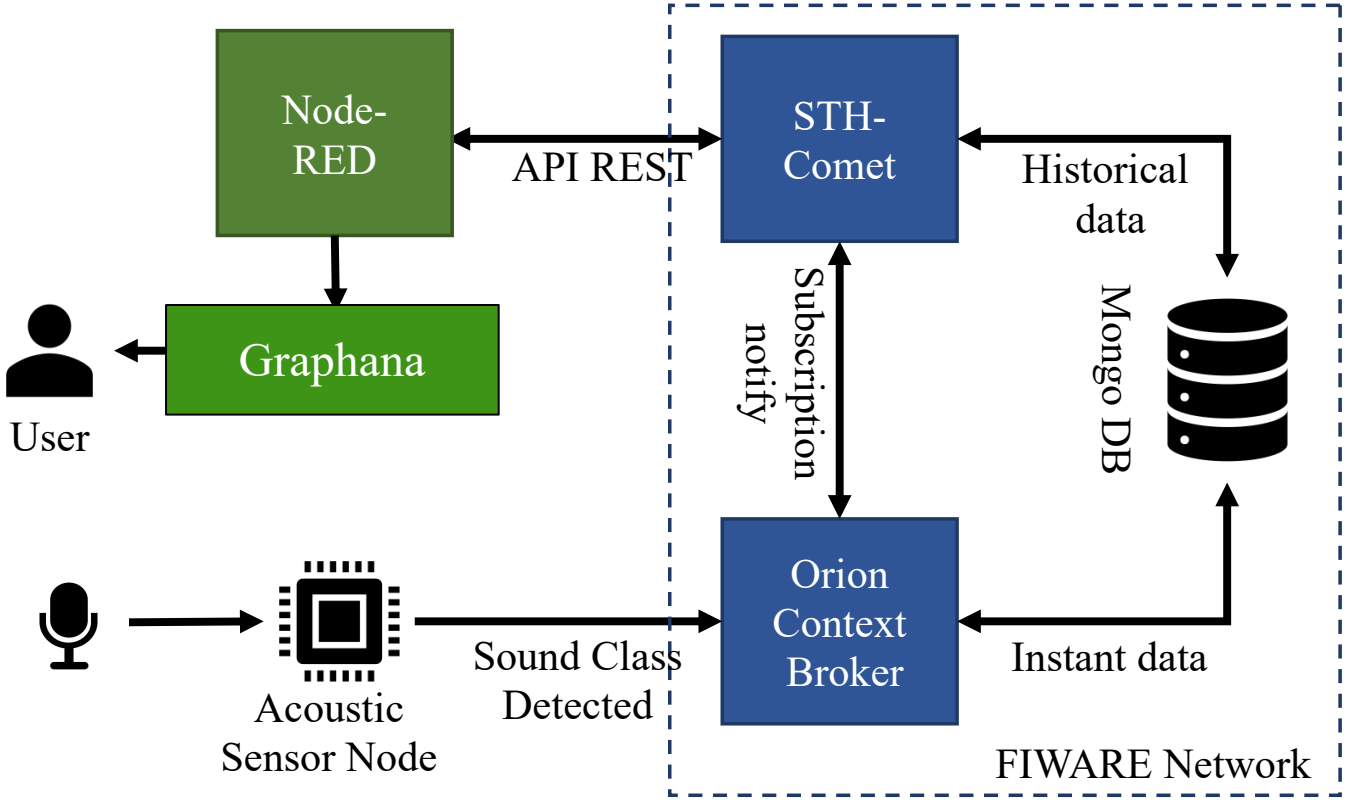


Wireless network: IoT FIWARE

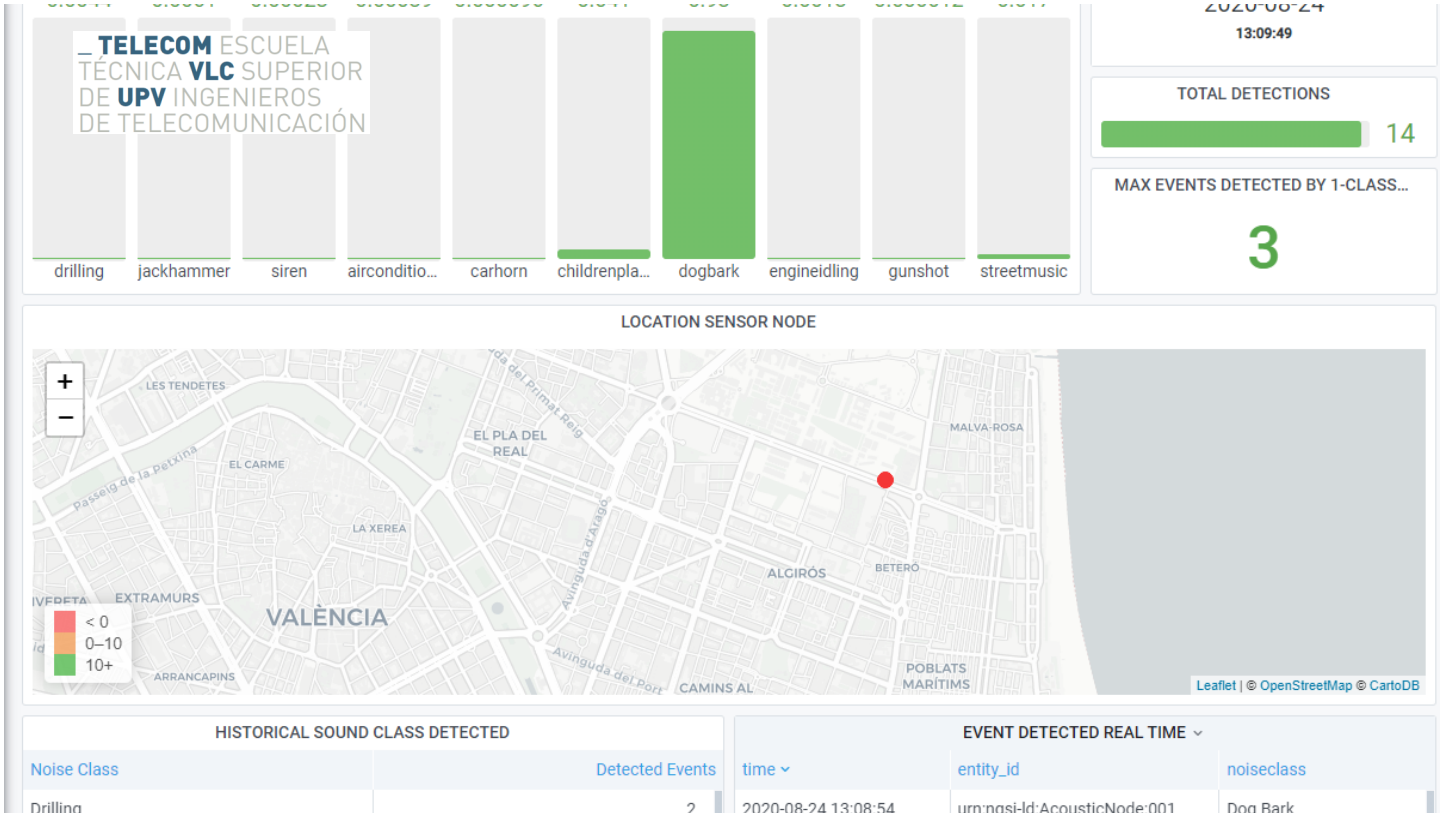
Agents in the IoT FIWARE network architecture:



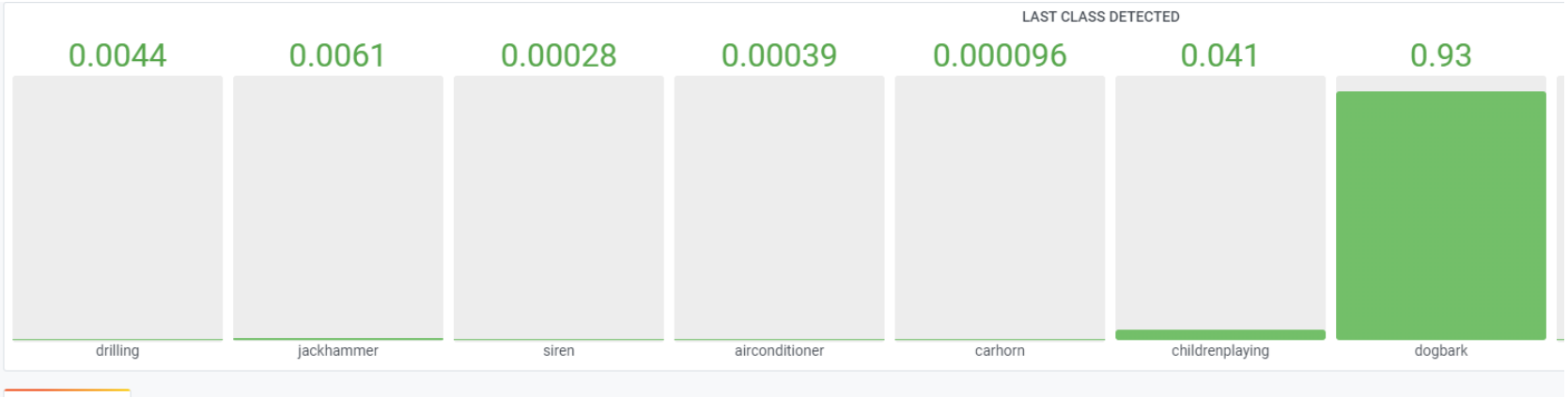
Proposed FIWARE network



Dashboard: graphana



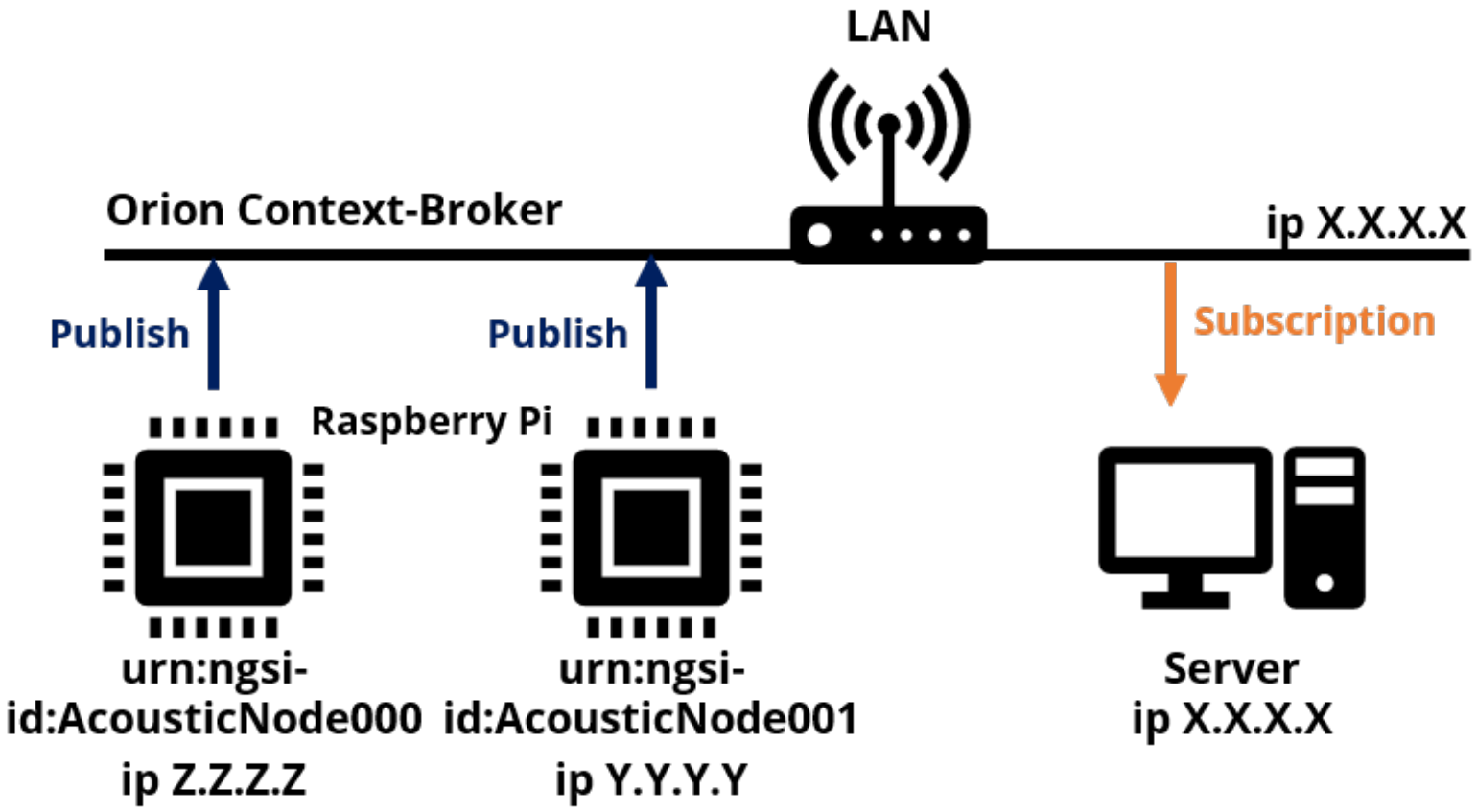
Dashboard: graphana



EVENT DETECTED REAL TIME

time	entity_id	noiseclass
2020-08-24 12:51:47	urn:ngsi-ld:AcousticNode:001	Children Playing
2020-08-24 12:52:16	urn:ngsi-ld:AcousticNode:001	Children Playing
2020-08-24 12:52:56	urn:ngsi-ld:AcousticNode:001	Air Conditioner
2020-08-24 12:57:01	urn:ngsi-ld:AcousticNode:001	Drilling

WASN in the iTEAM laboratory



USC on the Raspberry Pi

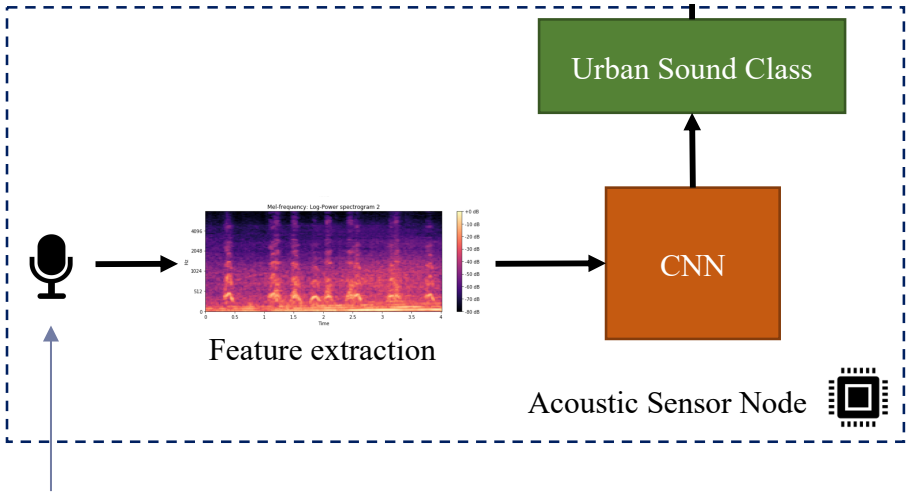
- The classification process takes 97% of the CPU capacity and the 30% of the RAM
- USC runs 4.9% of the excerpt time (3s)
- ¿Is it possible to reduce the USC energy consumption?

Table 1: Hardware Specifications and Total Processing Time

Device	Processor	RAM	Time
Dell XPS 15 (2015)	I7-6700HQ 2.60 GHz	16GB	0.195 s
Raspberry Pi 3 Model B	Quad Core 1.2 GHz	1GB	2.147 s

Table 2: Processing Time per Specific Task

Device	*wav* file Loading	Feature Extraction	CNN Classification
Dell XPS 15 (2015)	0.124 s	0.012 s	0.059 s
Raspberry Pi 3 Model B	2 s	0.052 s	0.095 s



Common audio record time: 3.089s

Pre-detection stage for traffic

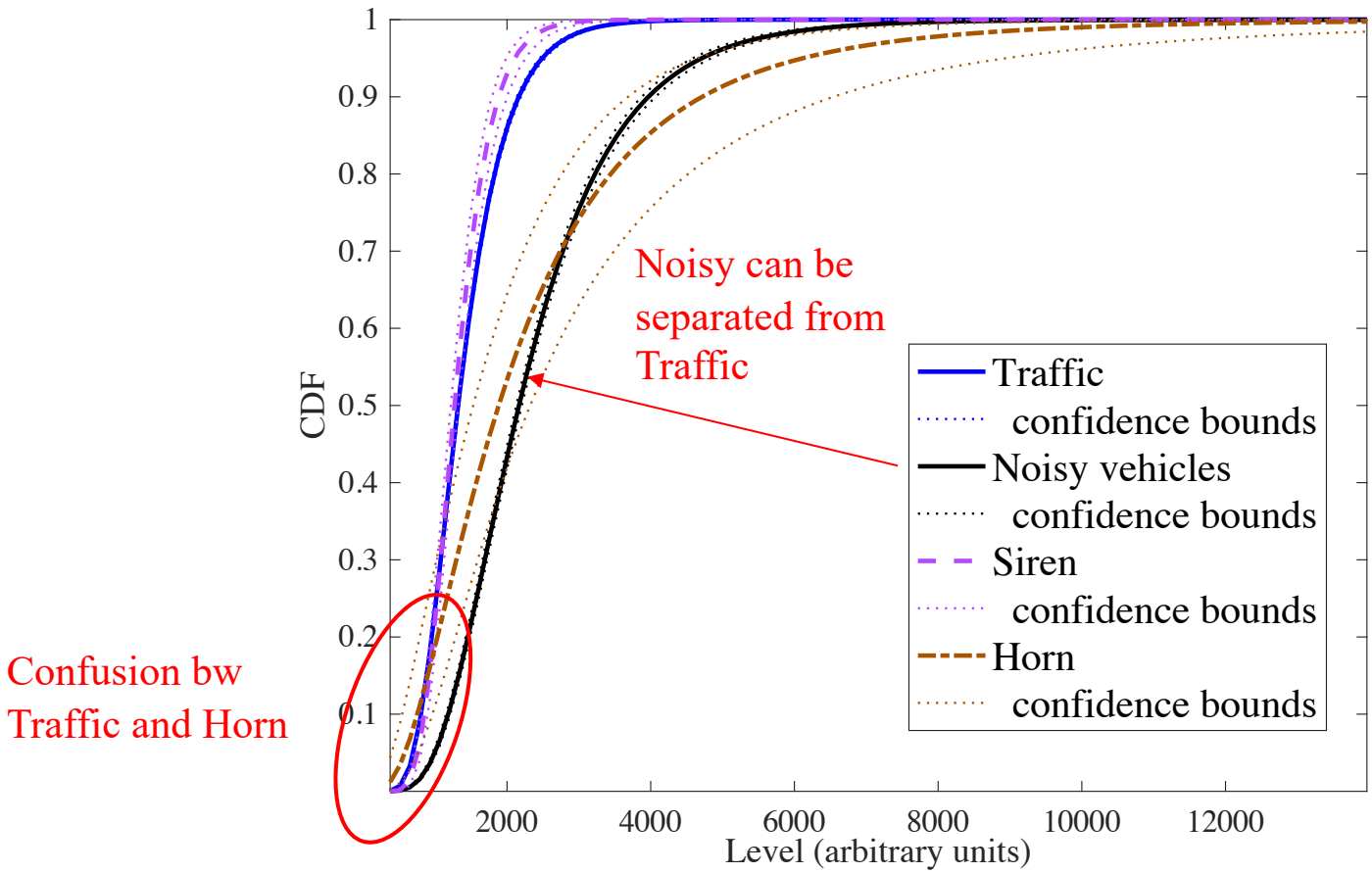
- The main noise sources in big avenues are:
  - Traffic: background traffic noise
  - Noisy vehicles as motorbikes and sport cars
  - Sirens
  - Car horns
- "Traffic" is always in the background. The other three are considered as *sound events* to classify
- New audio records have been taken at two big avenues of Valencia. One of them close to a big public hospital:

Class	*wav Files*	*Observation time*	
Traffic	23	775 s	**66.9%**
Noisy Vehicles	16	339 s	**29.3%**
Siren	3	34 s	**2.9%**
Horn	8	10 s	**0.9%**

sound events: 33.1%

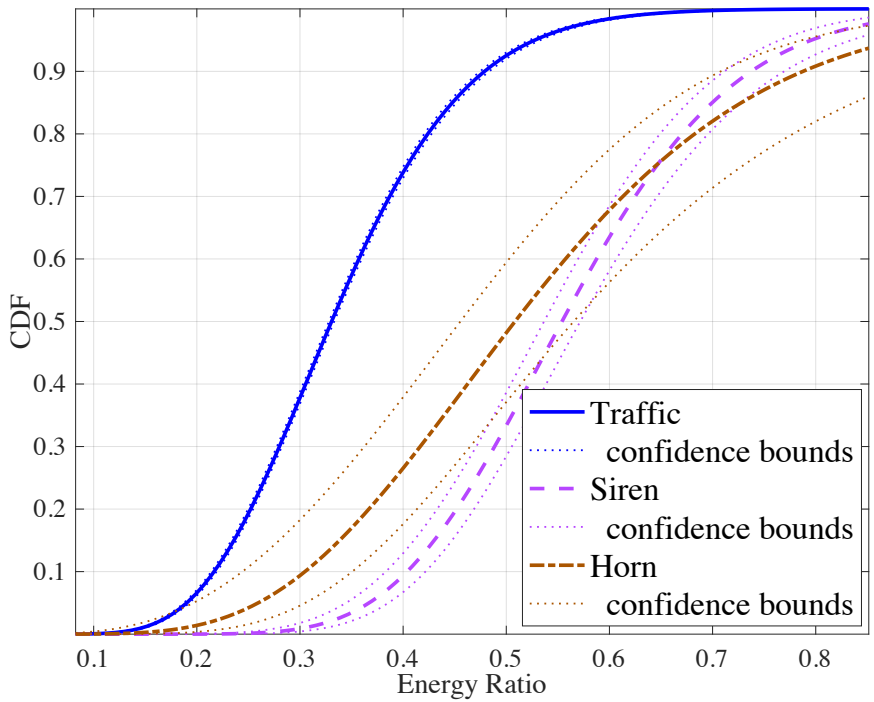
Statistical model (1)

- We compute the equivalent continuous sound pressure level L(*m*) where *m* is the time frame

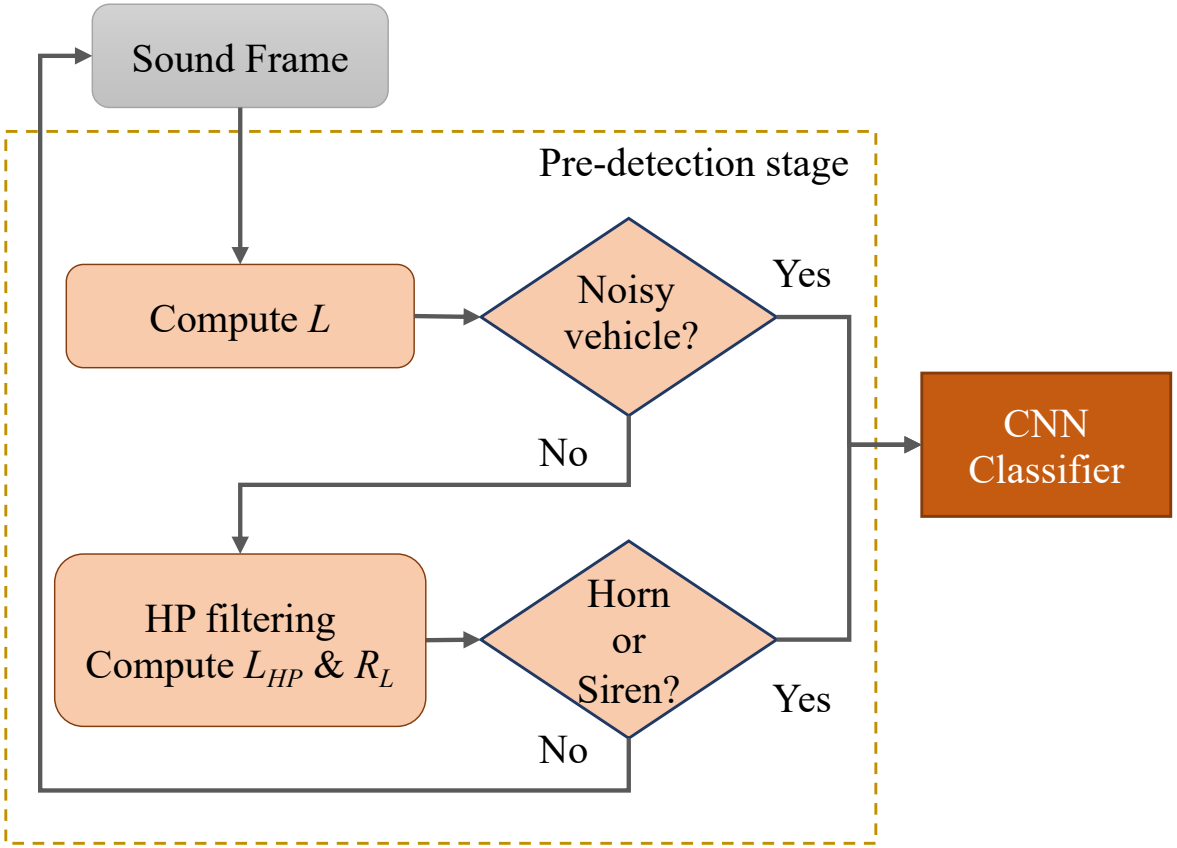


Statistical model (2)

- Horns are high-level sounds, but they also present high-frequency content as well as the siren sounds.
- We filter the recorded audio by a high-pass filter and compute their energy as $L_{HP}(m)$, obtaining the energy ratio $R_L(m) = L(m)/L_{HP}(m)$.



Pre-detection stage



Energy saving

Raspberry Pi 4 Model B State*	Power Consumption
Idle	540 mA (2.7 W)
400% CPU load (stress --cpu 4)	1280 mA (6.4 W)

* Power Consumption Benchmarks, https://www.pidramble.com/wiki/benchmarks/power-consumption

- Without pre-detection stage:
  - USC working 0.147s every 3s = 4.9% of the CPU time
  - CPU working at 97% -> almost 400% quad-core
  - **Energy consumption** = 0.049*1280mA = **62.72 mAh**
- With pre-detection stage:
  - 66.9% is traffic:.
  - CPU working at 15% -> close to idle
  - Energy consumption only pre-detection = 0.049*540mA = 26.46mAh
  - **Total energy consumption** = 0.669*26.46+ 0.331*62.72 = **38.46 mAh**

Energy saving: 38.7%

Conclusions

- A low-cost wireless acoustic sensor network has been built for USC
- Sound classification is carried out in the acoustic node
- The network is based on open standard FIWARE
- A pre-detection stage has been designed in order to reduce the energy conssumption
- Future work:
  - Improve USC performance
  - Other platforms: NVIDIA Jetson nano, Google coral
  - Expand the audio recording dataset with local recording from Valencia city

Thank you!

Questions?

A Low-cost Wireless Acoustic Sensor Network for the Classification of Urban Sounds

gpinyero@iteam.upv.es,