Line gain for long term environmental survey

Adrien Deverin[1 2 3]

Valentin Gies [1 3], Sébastian Marzetti [1] & Hervé Glotin [2 3]

[1] Université de Toulon, Aix Marseille Univ, CNRS, IM2NP, Toulon, France
[2] Université de Toulon, Aix Marseille Univ, CNRS, LIS, Toulon, France
[3] CIAN, int. Ct. on AI for Natural Acoustics, https://cian.univ-tln.fr

Challenge & Solution

No embedded ultra-low-power AI solution currently exists in this field of acoustic recognition!

The reason : The high-frequency **digital processing** required to compute spectrograms (FFT) and execute neural networks (CNNs) is **too energy-consuming**.

YET : living beings perform this kid of task continuously without getting exhausted.

-> They operate in analog (less energy-consuming at high fréquency)!

Explanations :

In biology, the brain is not the only element involved. Many organs, operting in an **analog** manner (eyes, ears, etc.), process the most important parts of the signal (selected through natural selection) to extract moments of interest, which are then transmitted to the brain if necessary (we don't analyze the noise).

The energy-intensive processing occurs at low frequency, but the entire system remains operational at high frequency.

This very principle led to the invention of CNNs — and we're taking it further: by extracting essential primitives directly at the acquisition stage, through a biomimetic process modeled on natural selection!

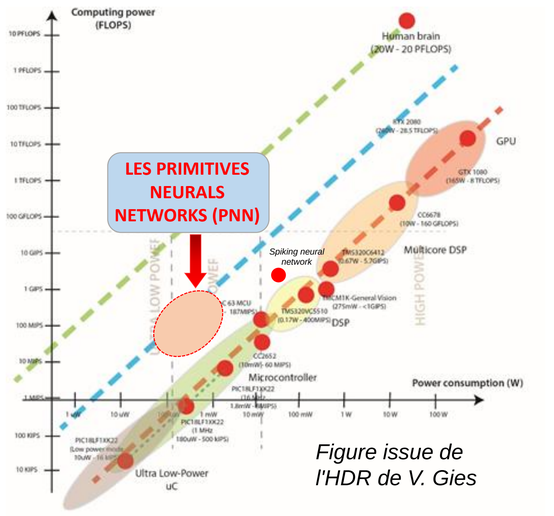
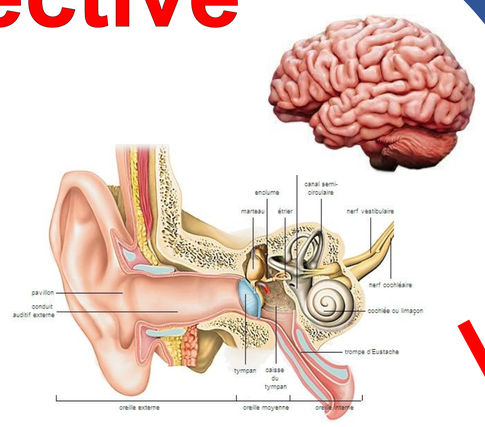
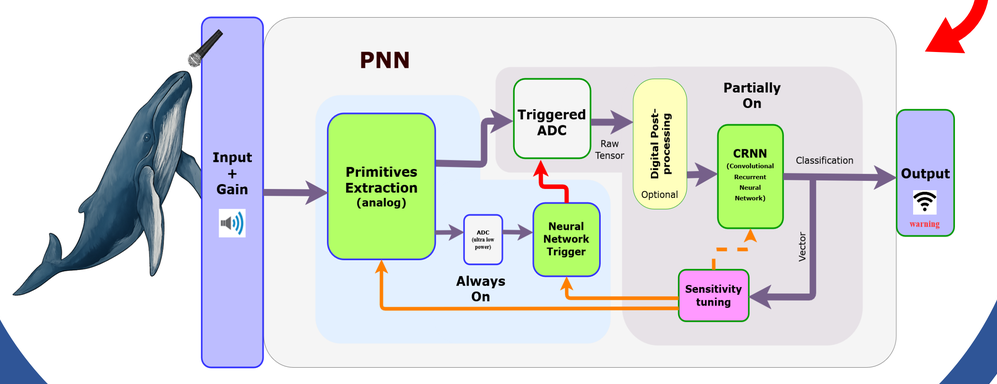


Figure issue de l'HDR de V. Gies

Objective

Designing sensors based on **Primitive Neural Networks**: a biomimetic architecture inspired by the brain-ear system, developed to **autonomously and continuously identify** marine mammals, with an energy **autonomy of up to one year**.



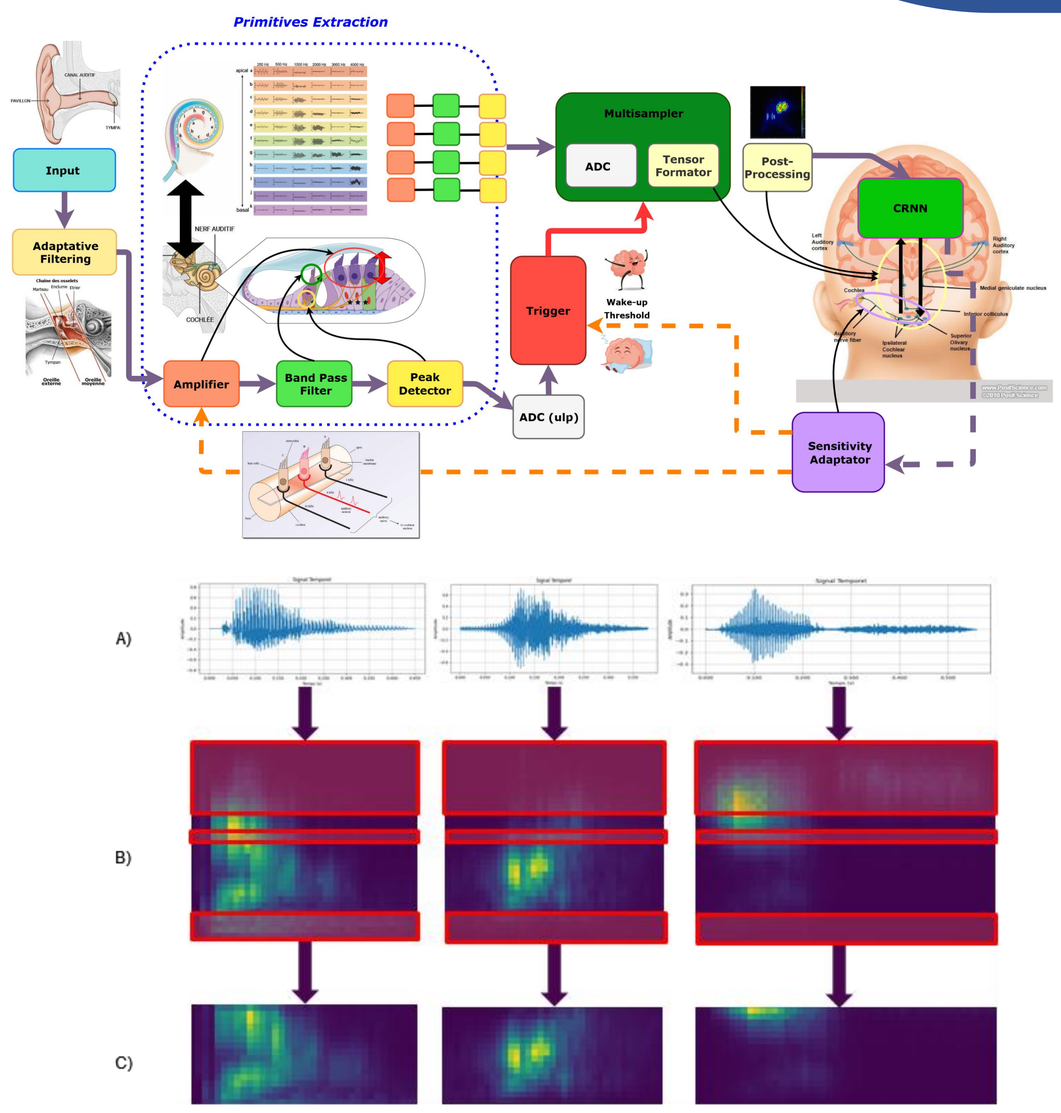


Architectural Diagram of PNNs

Implementation

Primitive Extraction : Extraction of moments of interest using a bank of band-pass filters combined with peak detectors — analogous to the function of inner hair cells.

- A spectrogram-like representation of the signal, achieved with 50x less energy than a traditional FFT (**Figure B**).
- The filter bank is optimized through end-to-end CRNN training, emulating the evolutionary specialization of the cochlea in different species, thereby further reducing energy consumption (**Figure C**)



The final filter selection is performed through a new neural network layer: the **Line Gain.**

TABLE I: CNN Architecture

Layer	Output Shape	Activation	Nb of Variables
Input	(None, 64, 90, 1)	-	-
Line Gain	(None, 64, 90, 1)	ReLU	64
Conv2D	(None, 62, 88, 8)	ReLU	80
Conv2D	(None, 60, 86, 8)	ReLU	584
Conv2D	(None, 58, 84, 8)	ReLU	584
Flatten	(None, 38976)	-	-
Dense	(None, 128)	ReLU	4,989,056
Dropout (0.2)	(None, 128)	-	-
Dense	(None, 32)	ReLU	4,128
Output (Dense)	(None, 4)	Softmax	132
Total parameters			4,994,564

Line Gain : Each row of the image is multiplied by a weight w_i , followed by $L1$ norm regularization.

$$New\_Loss = Original\_Loss + \lambda \sum_{i=1}^{n} |w_i|$$

This regularization drives the weights w_i toward zero. The lower a weight is relative to the others, the less its associated primitive contributes to the classification task.

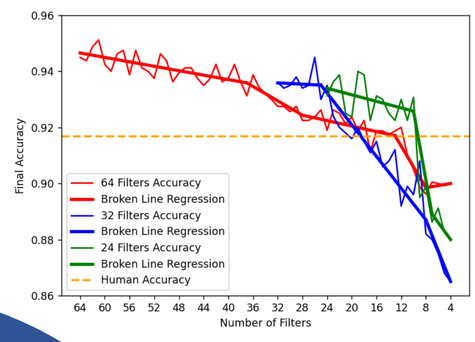
The remaining filters are then used to train the following components:

- **The Trigger block,** which detects when activation is necessary,
- **The CRNN block,** which performs final signal classification in ULP mode,
- **The Sensitivity Adaptator block,** which enables self-adaptive attention to cope with environmental variability.

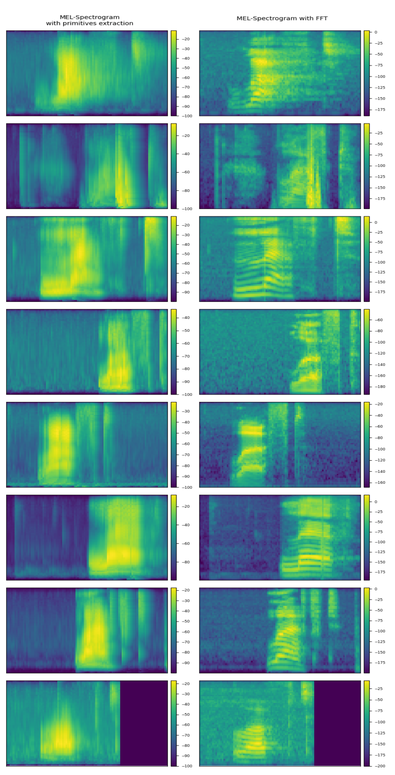
Results

Empirical results show that 4th-order band-pass filters are required to achieve performance comparable to FFT-based acquisition. An approach using around a dozen filters is sufficient to match human-level performance*, while enabling an average power consumption as low as **150 µW** in 'Always-On' mode.

Impact of Filter Prunning on CRNN Classification Accuracy over 8 Target Classes



** Performance may, of course, vary as a function of the number and complexity of the target classes.*



Analog vs Digital Comparison of Spectrograms Obtained with 64 Bands

Conclusion : We have demonstrated that a mixed analog-digital approach can achieve accuracy comparable to state-of-the-art purely digital methods (based on FFT followed by a CNN), while consuming significantly less energy. This makes it suitable for deployment on microcontrollers and enables operation for several months on battery power.

***Presence of species : 1/60 of the time; cost without microphone.*

Video

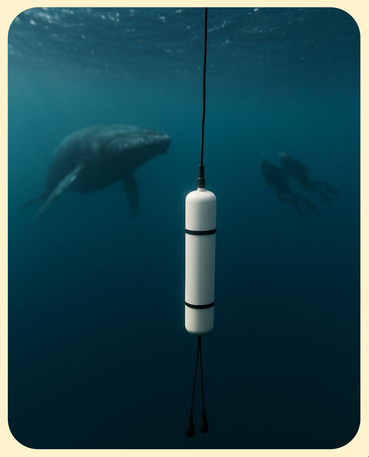
Primitive Neural Networks (PNNs) embody an ecological and scientifically grounded strategy, by closely mimicking biological processes to enable efficient, embedded artificial intelligence systems.

Explanatory video

Development Perspective

An immersible construction and real-world testing of the entire system are desired by the LIS and IM2NP laboratories.

An upcoming PhD thesis will focus on the silicon implementation of the architecture, with the goal of reducing the energy cost of the filters and breaking the symbolic threshold of ultra-low power consumption (<100 µW).



Reference : Deverin A., Gies V., Marzetti S., Barchasz V., Glotin H (2024) Eusipco. Reducing AI complexity using ear bio-inspired primitives.

We thank ANR for the grant of this Phd
ANR-21-CE04-0019 SylvanIA (SYnchronized Low power Versatile Acoustic Network with embedded AI)