



**Offer #2019-01675**

## **PhD Position F/M Robust audio event detection**

**Contract type** : Fixed-term contract

**Level of qualifications required** : Graduate degree or equivalent

**Fonction** : PhD Position

### **Context**

This PhD is funded by the ANR project "LEAUDS" involving the [Multispeech](#) team at Inria Nancy - Grand Est, the machine learning team at INSA Rouen, and [Netatmo](#). It will be co-supervised by [Emmanuel Vincent](#) and [Gilles Gasso](#). The successful applicant will have the opportunity to visit the machine learning team at INSA Rouen for extended periods of time, in order to benefit from the complementary scientific environment offered.

### **Assignment**

We are constantly surrounded by a complex audio stream carrying information about our environment. Hearing is a privileged way to detect and identify events that may require quick action (ambulance siren, baby cries...). Indeed, audition offers several advantages compared to vision: it allows for omnidirectional detection, up to a few tens of meters and independently of the lighting conditions. For these reasons, automatic ambient audio analysis has become increasingly popular over the past five years [1, 2].

One of the main degradations encountered when moving from lab conditions to the real world is due to the fact that ambient audio scenes are not composed of isolated audio events but of multiple events occurring simultaneously. Differences between training and test conditions also typically arise due to distant microphone capture, to the intrinsic variability of audio events, and to different acquisition hardware and settings. These problems have gained interest in the past few years, yet they remain an obstacle towards the deployment of audio event detection systems in real-world settings.

The goal of this PhD is to design an automatic audio event detection system robust to the variabilities and degradations encountered in real conditions.

[1] T. Virtanen, M. D. Plumbley and D. Ellis. Computational Analysis of audio Scenes and Events, Springer, 2017.

[2] A. Mesaros, A. Diment, B. Elizalde, T. Heittola, E. Vincent, B. Raj and T. Virtanen. Sound event detection in the DCASE 2017 Challenge. IEEE/ACM Transactions on Audio, Speech and Language Processing, 27(6), 2019, pp. 992-1006.

### **Main activities**

Starting from the existing deep learning based system developed at Inria [3], the following complementary directions may be explored.

1. Design an audio event detection system that takes the complex temporal structure (temporal coherence, duration, co-occurrence) of audio events in the scene into account. One approach to move beyond the simple model in [4] is to train an adversarial network to discriminate estimated vs. real structures and to optimize a decision function that accounts both for the predicted class(es) in each time frame and for the global structure.
2. Following the temporal attention-based algorithm in [5], develop a multiple-pass detection algorithm based on a spectro-temporal attention model. The attention model will iteratively discard the time-frequency zones corresponding to the detected events and focus on the remaining time-frequency zones. In addition, the detected events may be removed from the mixture signal by means of source separation [6]. The challenge will be to train a single neural network based system able to separate hundreds of audio classes and to exploit long-range contextual information. Robust integration and interaction between the source separation system and the above audio event detection system will also be studied.
3. Augment and transform the training data in order to increase its size and its similarity with the test domain. Heuristic approaches based on signal transformations [7] and/or generative adversarial neural networks are often used but they poorly account for the temporal structure of ambient audio scenes and they lack theoretical guarantees. The challenge will be to develop a principled data augmentation/transformation method, e.g., inspired from [8,9], that maximizes performance on the test data.

[3] N. Turpault, R. Serizel and E. Vincent. Semi-supervised triplet loss based learning of ambient audio embeddings. In Proc. ICASSP, 2019.

- [4] E. Benetos, G. Lafay, M. Lagrange and M. D. Plumbley. Detection of overlapping acoustic events using a temporally-constrained probabilistic model. In Proc. ICASSP, 2016, pp. 6450–6454.
- [5] Y. Xu, Q. Kong, W. Wang and M. D. Plumbley. A joint detection-classification model for audio tagging of weakly labelled data. In Proc. ICASSP, 2017, pp.641-645.
- [6] E. Vincent, T. Virtanen and S. Gannot. Audio source separation and speech enhancement. Wiley, 2018.
- [7] J. Salamon, D. MacConnell, M. Cartwright, P. Li and J. P. Bello. Scaper: A library for soundscape synthesis and augmentation. In Proc. WASPAA, 2017, pp. 344-348.
- [8] N. Courty, R. Flamary, D. Tuia and A. Rakotomamonjy. Optimal transport for domain adaptation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(9), 2016, pp. 1853-1865.
- [9] S. Sivasankaran, E. Vincent and I. Illina. Discriminative importance weighting of augmented training data for acoustic model training. In Proc. ICASSP, 2017, pp. 4885-4889.

## Skills

Master degree in computer science, machine learning, or audio signal processing  
Experience with programming in Python  
Experience with PyTorch is a plus

## Benefits package

- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- Flexible working hours
- Sports facilities

## Remuneration

Gross Salary per month: 1982€ brut per month (year 1 & 2) and 2085€ brut/month (year 3)

## General Information

- **Town/city** : Villers-lès-Nancy
- **Inria Center** : [Centre Inria de l'Université de Lorraine](#)
- **Starting date** : 2019-09-01
- **Duration of contract** : 3 years
- **Deadline to apply** : 2019-06-30

## Contacts

- **Inria Team** : MULTISPEECH
- **PhD Supervisor** :  
Vincent Emmanuel / [emmanuel.vincent@inria.fr](mailto:emmanuel.vincent@inria.fr)

## About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

**Warning** : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

## Instruction to apply

### Defence Security :

This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

### Recruitment Policy :

As part of its diversity policy, all Inria positions are accessible to people with disabilities.

