

# Engineer position - Real time speaker verification in real conditions for a mobile robot

**Labo:** Loria / Inria Nancy - Grand Est, Nancy (<http://www.loria.fr/>) and LIA (Laboratoire d'informatique d'Avignon - <http://lia.univ-avignon.fr/>)

**Supervisors:** Romain Serizel (LORIA), Driss Matrouf (LIA), and Denis Jouvét (LORIA)

**Start:** end 2021

**Duration:** 18 months (about 50% of the time at Loria / Inria, Nancy, and 50% at LIA, Avignon)

note: the position is cofunded by Inria and LIA.

## Context

This Engineer position fits within the scope of the ANR project "ROBOVOX" which involves the Multispeech team from Inria Nancy - Grand Est (<https://team.inria.fr/multispeech/>), the speech processing team from Laboratoire Informatique d'Avignon (<http://lia.univ-avignon.fr/>), and A.I. Mergence (<http://www.aimergence.com/fr/>).

## Assignment

Speaker identification has recently been deployed in several real-world applications including secured access to bank services via telephone or internet. However, identification based solely on voice remains a modality with a limited reliability under real conditions including several acoustic perturbations (noise, reverberation...), which are getting important when distant speech recording is used. Recent works indicate that multichannel speech enhancement of the test signal results in improved performance for speaker identification systems in noisy environments, especially as it enables controlling the distortion introduced on the speech signal. Additionally, the usage of deep learning for multichannel speech enhancement has recently allowed for a large performance improvement.

In Robovox, we are investigating several approaches for robust speaker verification. One approach studies the compensation of speech acoustic perturbations through deep-learning based x-vector denoising. Another approach focuses on applying a multichannel speech enhancement processing before the core speaker verification step. Extensions of these approaches are currently under investigation. In order to evaluate performance on real condition speech data, a dedicated speech corpus is under recording using a mobile robot.

## Main activities

The main activities will concern the optimization of the implementation, its adaptation for real-time operation in real conditions, and its evaluation.

As the current approaches studied rely on several programs and toolkits, a first step will consist removing dependencies to complex toolkits and implementing the whole pipeline in a standalone python code, in order to allow for an embedded implementation on the robot. The second point will be to optimize the code and the models to be compliant with real-time processing using the resources available on the robot. In addition, as the current approaches under study were developed for batch evaluations, it will be needed to reorganize some processing to match with real-time operation, and provide the result with a limited latency, whatever the duration of the incoming speech signal is.

The last aspect concerns the usage of the speech corpus currently being recorded, and its integration within the evaluation process.

## **Skills**

MSc in computer science, machine learning, signal processing

Experience with programming language Python

Experience with deep learning toolkits is a plus, as well as experience with real-time processing

## **CONTACT**

Driss Matrouf : [driss.matrouf@univ-avignon.fr](mailto:driss.matrouf@univ-avignon.fr)

Romain Serizel : [romain.serizel@loria.fr](mailto:romain.serizel@loria.fr)