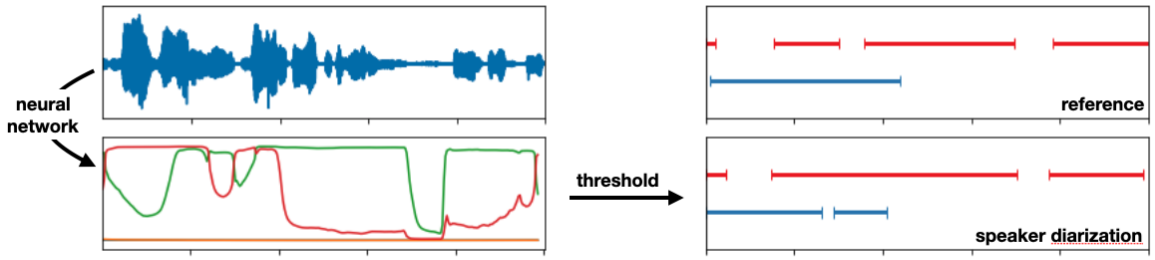
Projets de fin d'étude // Master internships
Several internships around speaker diarization

Speaker diarization is the task of partitioning an audio stream into homogeneous temporal segments according to the identity of the speaker. Most dependable diarization approaches consist of a cascade of several steps: voice activity detection to discard *non-speech* regions, speaker embedding to obtain discriminative speaker representations, and clustering to group speech segments by speaker identity.

A new family of approaches have recently emerged, rethinking speaker diarization completely. Dubbed end-to-end diarization (EEND), the main idea of this approach is to train a single neural network – in a permutation-invariant manner – that ingests the audio recording and directly outputs the (overlap-aware) diarization output.



Such a model has recently been integrated into the pyannote.audio open-source speaker diarization library (github.com/pyannote/pyannote-audio, based on pytorch) and pretrained pipelines based on such models can be tested online (hf.co/spaces/pyannote/pretrained-pipelines). We propose several internships around this type of models and pipelines.

- In (research) project #1, we will investigate the training of multi-task models capable of both speaker diarization and speaker separation (understand: hf.co/pyannote/segmentation and github.com/asteroid-team/asteroid in the same model)
- In (research) project #2, we will investigate the training of multi-task models capable of both speaker diarization and speaker embedding (understand: hf.co/pyannote/segmentation and hf.co/pyannote/embedding in the same model)
- In (research) project #3, we will investigate constrained clustering approaches (e.g. with must link and cannot link constraints) and their application to speaker diarization.

- In (engineering) project #4, we will implement a streaming speaker diarization pipeline with adjustable latency, and aim at making it run in real-time on small devices (such a mobile phone, a Raspberry Pi, or an NVIDIA Jetson Nano)
- In (engineering) project #5, we will implement a Jupyter widget for interactive annotation of audio recording (with pretrained models in the loop).
- In (engineering) project #6, we will implement a tool to automatically transcribe and diarize collections of audio documents (e.g. podcasts, or TV shows).

In practice, those projects will be based on (and contributed to) pyannote.audio open source toolkit and experiments will make good use of *Jean Zay* supercomputer.

Please send your application (CV, grades, references) arguing about your prefered project

- to herve.bredin@irit.fr
- with subject "Internship about speaker diarization"

Location: France, Toulouse, IRIT, SAMoVA team
Date and duration : 5 to 6 months, possibly leading to a PhD (research projects) or research engineer (engineering projects) position.
Allowance: around 600 euros monthly