NT Project Group

SoSe22
Meeting diarization for natural conversations
Meeting diarization for natural conversations

● “Who speaks when?”

● Meeting scenario:
  ○ Multiple people talking over a long duration
  ○ Large amounts of silence
  ○ Speakers can be inactive for a long time

● Challenges:
  ○ Re-identification of speakers
  ○ Handling of long meetings
  ○ Overlapping speech

● Current Systems:
  ○ Extract speaker embeddings for short windows
  ○ Assign each segment to a speaker with global clustering
  ○ Combination of neural networks and model-based techniques
General Information

- **Duration:** 1 or 2 semester (9 or 18 ECTS)
- **Goals:**
  - Refine an existing Diarization model (previous project group)
    - Modification of the model for realistic data (e.g. overlap)
    - Incorporation of Multi-Channel data
  - Test model on real recordings
  - Visualization of the model / Development of a demonstrator
- **Workload:**
  - Extending the existing model with new techniques
  - Systematic evaluation and comparison of extensions
  - Holding an intermediate presentation (only 2 semester projects)
  - Writing the project report
  - Holding the final presentation
Requirements for a fast start in the project

1. Experience with Python coding (https://github.com/fgnt/python_crashkurs)
2. Familiarity with Linux (https://pc2.uni-paderborn.de/de/teaching/trainings/hpc-user-trainings/selftests/selftest-linux)
3. Knowledge of statistical signal processing
4. Basic understanding of neural networks (beneficial)

First meeting: Tuesday, April 5 (Look at our website for more details)

We will have a short Python test at the beginning to check whether you have the necessary basics