Université de Lorraine - IDMC
Masters SC & TAL – M1
2020-2021

Fiche de projet tutoré / Project form

Titre du projet / Title of the project

Speaker diarization with overlapped speech

Encadrement / Supervisors
1. équipe, laboratoire / team, lab
MULTISPEECH

1. encadrant·e principal·e (nom, email) / main supervisor (name, email)
Md Sahidullah, md.sahidullah@inria.fr

2. autres encadrant·es / other supervisors
-

Description / Description
1. projet global/global project

Speech is the most convenient means for human communication. Automatic *speaker diarization* is a computer-based method of determining **"who spoke when"** in a multi-talker speech conversation. This technology has applications in many important practical problems-- for example, automatic captioning of videos, creating text transcriptions of meetings & interviews, etc.

The speaker diarization system includes different sub-systems such as *speech activity detector* (SAD), *automatic speaker verification* (ASV), *speaker clustering* (SC). The SAD module identifies the speech regions from the entire audio recording whereas ASV module computes the corresponding speaker similarity of different speech segments. Finally, SC groups different speech segments according to the speakers present in the given audio recording. The state-of-the-art speaker diarization system with deep neural network technology has shown promising results for good quality audio data. However, the performance severely degrades in the realistic conditions. For example, recent studies on DIHARD 2019 challenge dataset show reasonably good performance can be achieved for audio recordings from audio books, interviews in the studio. But the performance severely degrades for the audio-data collected in restaurants or extracted from web videos. Speaker diarization becomes very challenging mainly due to the poor ASV performance in degraded acoustic conditions and due to the presence of overlapped speech.

One possible way to improve the ASV performance is to accurately detect the speakers in overlapped speech regions. This can be achieved with *overlapped speech detection* followed by *detection* of the speakers.

With this background, the overall aim of this project is to improve the speaker diarization performance in challenging scenarios by addressing the overlapped speech in an audio

recording. The students will explore different overlap speech detection methods with DIHARD 2019 (single channel) dataset consisting of 11 different acoustic conditions. The acoustic conditions have different numbers of speakers and background noise. This will help the students to explore the robustness of overlap detection methods across diverse conditions. Finally, the students will explore speaker detection methods with overlapped speech for improved speaker diarization.

2. biblio. UE 705 (semestre 7)

For the semester 7, the students will get familiar with the speaker diarization and overlapped speech detection literature, dataset and state-of-the-art methods. The students are also expected to analyze the impact of overlapped speech on speaker diarization performance. The following papers/tech reports are suggested readings.

- Anguera, X., Bozonnet, S., Evans, N., Fredouille, C., Friedland, G. and Vinyals, O., 2012. Speaker diarization: A review of recent research. IEEE Transactions on Audio, Speech, and Language Processing, 20 (2), pp.356-370.
- Ryant, N., Church, K., Cieri, C., Cristia, A., Du, J., Ganapathy, S., Liberman, M. (2019) The Second DIHARD Diarization Challenge: Dataset, Task, and Baselines. Proc. Interspeech 2019, 978-982.
- Ryant, N., Church, K., Cieri, C., Cristia, A., Du, J., Ganapathy, S. and Liberman, M., 2019. Second DIHARD challenge evaluation plan. Linguistic Data Consortium, Tech. Rep.
- Bullock, L., Bredin, H. and Garcia-Perera, L.P., 2020. Overlap-aware diarization: Resegmentation using neural end-to-end overlapped speech detection. Proc. ICASSP ASSP 2020, 7114-7118.
- Baseline codes for speaker diarization on DIHARD 2019: https://github.com/iiscleap/DIHARD_2019_baseline_alltracks
- Link to the papers published in DIHARD 2019 challenge: https://www.isca-speech.org/archive/Interspeech_2019/

3. réalisation. UE 805 (semestre 8)

For the next semester, the students will experiment with baseline systems in Section 2. During this period, the evaluation of the integrated system will be made on DIHARD 2019 dataset.

Informations diverses : matériel nécessaire, contexte de réalisation /
Various information: material, context of realization

The study will be done on the second DIHARD challenge (DIHARD) 2019 dataset. The preliminary study will be made to analyze the impact of overlapped speech on speaker diarization performance. The students will then explore deep neural networks for the classification of overlapped and non-overlapped speech. In the following studies, the student will develop methods to detect the speakers in the overlapped speech. Towards this, the non-speech regions can be utilized for creating speaker models.

Livrables et échéancier / Deliverable and schedule

Semester 7:

- Familiarization with state-of-the-art speaker diarization
- Familiarization with state-of-the-art overlapped speech detection
- Understanding baseline codes
- Presentation of the work done in Semester 7

Semester 8:

- Experiments with DIHARD dataset
- Writing project report
- Writing article for a conference
- Creating reproducible research repository

Bibliographie /References (max. 4-5)

[*il ne s'agit pas de la bibliographie complète qui sera fournie aux étudiants au début du projet mais d'une bibliographie indicative pour aider à cerner le sujet*]

1. Anguera, X., Bozonnet, S., Evans, N., Fredouille, C., Friedland, G. and Vinyals, O., 2012. Speaker diarization: A review of recent research. IEEE Transactions on Audio, Speech, and Language Processing, 20 (2), pp.356-370.
2. Cornell, S., Omologo, M., Squartini, S. and Vincent, E., 2020, October. Detecting and counting overlapping speakers in distant speech scenarios. Proc. INTERSPEECH 2020.
3. Sell, G., Snyder, D., McCree, A., Garcia-Romero, D., Villalba, J., Maciejewski, M., Manohar, V., Dehak, N., Povey, D., Watanabe, S. and Khudanpur, S., 2018. Diarization is Hard: Some Experiences and Lessons Learned for the JHU Team in the Inaugural DIHARD Challenge. In Interspeech (pp. 2808-2812).
4. Moattar, M.H. and Homayounpour, M.M., 2012. A review on speaker diarization systems and approaches. Speech Communication, 54(10), pp.1065-1103.
5. Andrei, V., Cucu, H. and Burileanu, C., 2019. Overlapped Speech Detection and Competing Speaker Counting–-Humans Versus Deep Learning. IEEE Journal of Selected Topics in Signal Processing, 13(4), pp.850-862.