DiaboLIIc system description

I. Authors (in alphabetical order)

Hervé Bredin, Wenda Chen, Sangeeta Ghangam, Anthony Larcher, Antoine Laurent, Martin Lebourdais, Meysam Shamsi

II. Abstract

Our system is based on off-the-shelf publicly available diarization baselines, combined using off-the-shelf fusion approaches.

III. Notable highlights

None.

IV. Data resources

We ended up not training anything ourselves since the whole pipeline is based on pretrained baselines.

V. Detailed description of algorithm

Despite trying many things (that did not work), we ended up running and combining various pretrained baselines in the following order:

1) pyannote.audio `dia` pretrained pipeline [1]
2) official DIHARD III baseline (including VB-HMM re-segmentation) [2]
3) official DIHARD III baseline (without VB-HMM resegmentation) [2]
4) DOVER-Lap fusion of 1, 2, and 3 with default parameters [3]

This last line is our final submission to track 2.

VI. Results on the development set

TABLE I
Results on development set, track 2, full condition

Stage	DER	JER
1	28.7	52.0
2	21.7	43.7
3	24.1	45.6
4	21.3	44.3

VII. Hardware requirements

We only used pretrained models and baselines in our final submission. Therefore, no additional training was performed.

Regarding hardware requirements at test time, we unfortunately did not keep track of this information.

References

[1] H. Bredin, R. Yin, J. M. Coria, G. Gelly, P. Korshunov, M. Lavechin, D. Fustes, H. Titeux, W. Bouaziz, and M.-P. Gill, "pyannote.audio: neural building blocks for speaker diarization," in *ICASSP 2020, IEEE International Conference on Acoustics, Speech, and Signal Processing*, Barcelona, Spain, May 2020.

[2] P. Singh, H. V. M.A., S. Ganapathy, and A. Kanagasundaram, "LEAP Diarization System for the Second DIHARD Challenge," in *Proc. Interspeech 2019*, 2019, pp. 983–987. [Online]. Available: http://dx.doi.org/10.21437/Interspeech.2019-2716

[3] D.Raj, P.Garcia, Z.Huang, S.Watanabe, D.Povey, A.Stolcke, and S.Khudanpur, "DOVER-Lap: A method for combining overlap-aware diarization outputs," *2021 IEEE Spoken Language Technology Workshop (SLT)*, 2021.