NAVER CLOVA SUBMISSION TO THE THIRD DIHARD CHALLENGE

Hee-Soo Heo[1], *Jee-weon Jung*[1,2], *Youngki Kwon*[1], *You Jin Kim*[1],
Jaesung Huh[3], *Joon Son Chung*[1], *Bong-Jin Lee*[1]

[1]Naver Corporation, South Korea
[2]School of Computer Science, University of Seoul, South Korea
[3]Visual Geometry Group, Department of Engineering Science, University of Oxford, UK

ABSTRACT

This report describes Naver Clova's submission to the third DIHARD speech diarization challenge. Our system consists of following subsystems: speech activity detection, overlapped speech detection, speaker embedding extraction, feature enhancement, and clustering. Main improvements of our submitted system over existing diarization systems are feature enhancement based on the dimensionality reduction and overlap detection system. We reduce the dimensionality using an auto-encoder for each utterance to adapt speaker representations for the clustering system, and then perform attention-based aggregation, which we developed to enhance the clustering result. We also use some variants of CRNN based overlapped speech detection networks and their ensemble to further reduce the missed detection of overlapped speech regions. The submitted system achieves competitive performance in the `DIHARDIII_Task1_CORE`, which shows 14.96%, and 15.40% diarization error rate for the development and the evaluation sets, respectively.

Index Terms— Speaker diarisation.

1. REFERENCES