AI Rush 2라운드
- 화자분리 -

이봉진
Clova Speech

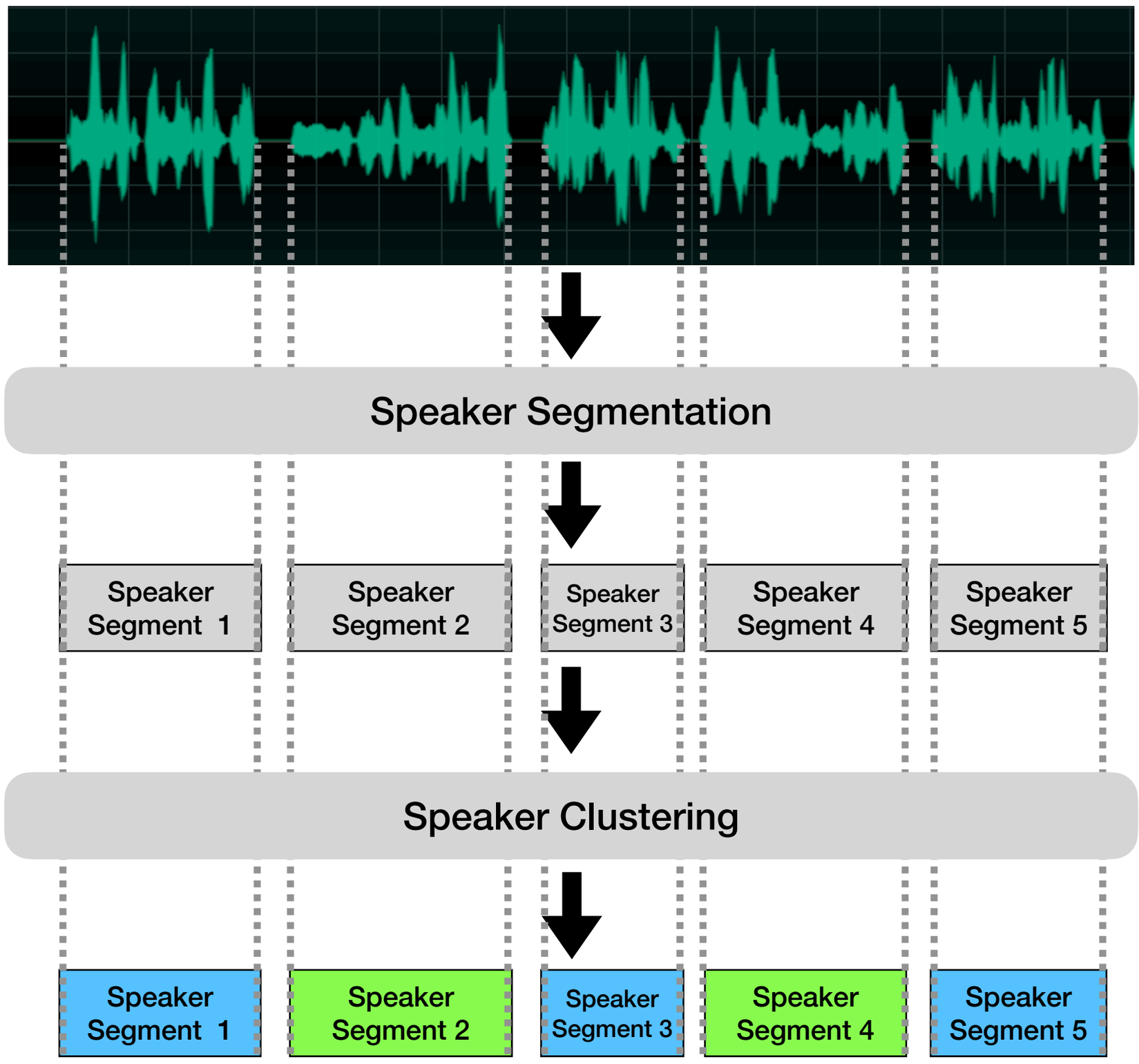
개요

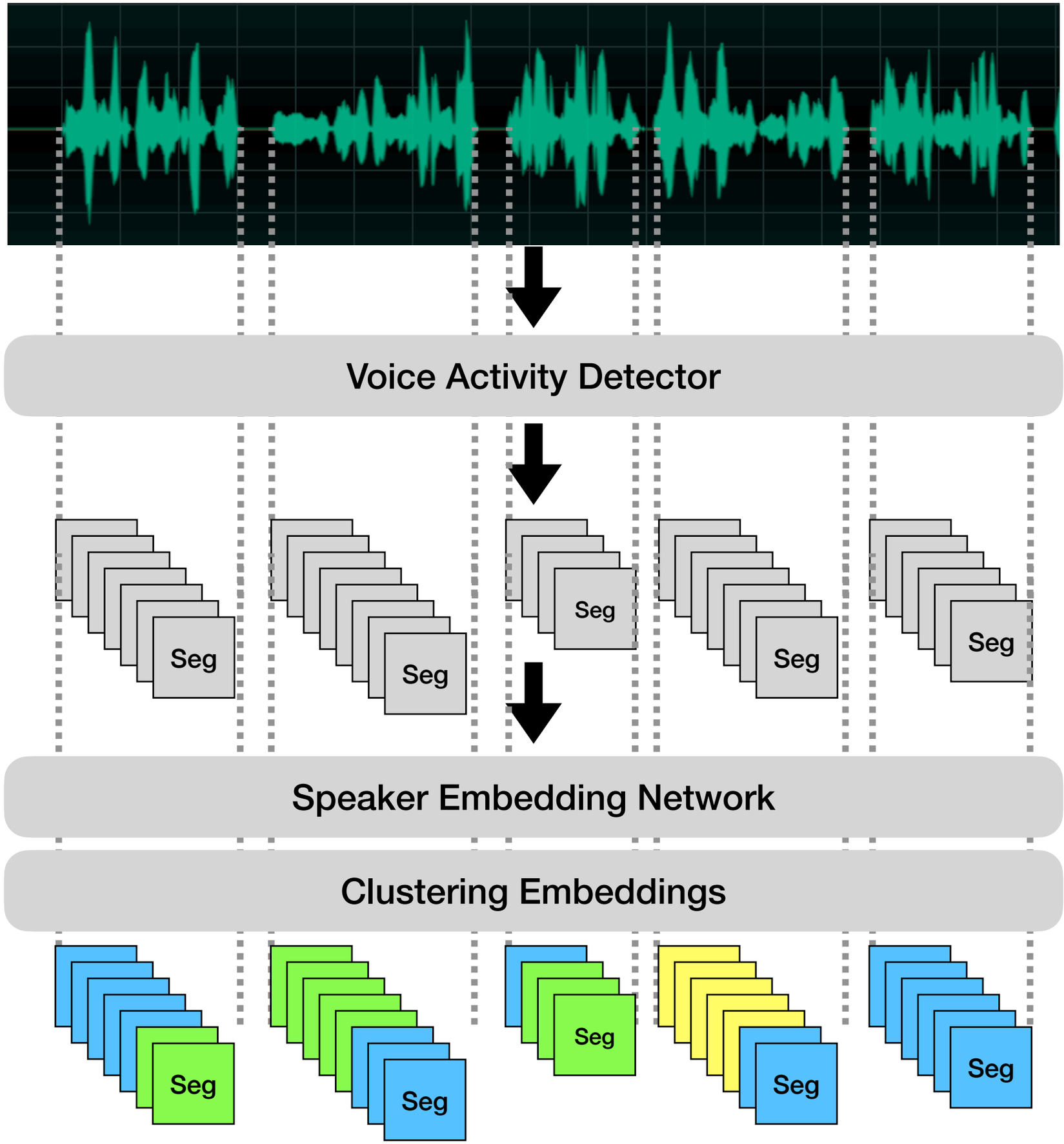
- 화자분리란?
- 성능 평가
- 개선 방향

화자분리

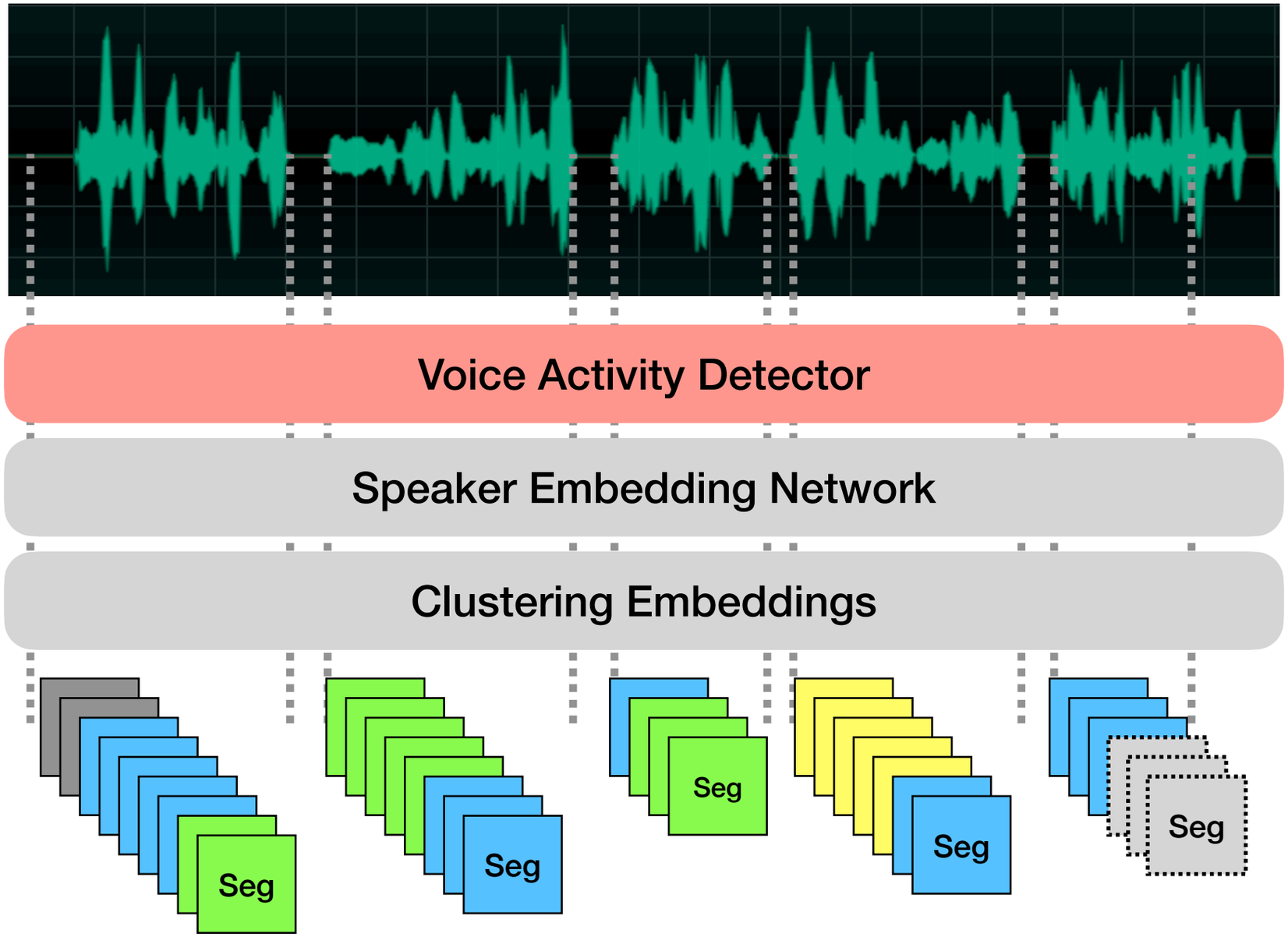
- Speaker Diarization
- "Who Spoke When"
- Speaker Segmentation + Speaker Clustering

https://en.wikipedia.org/wiki/Speaker_diarisation

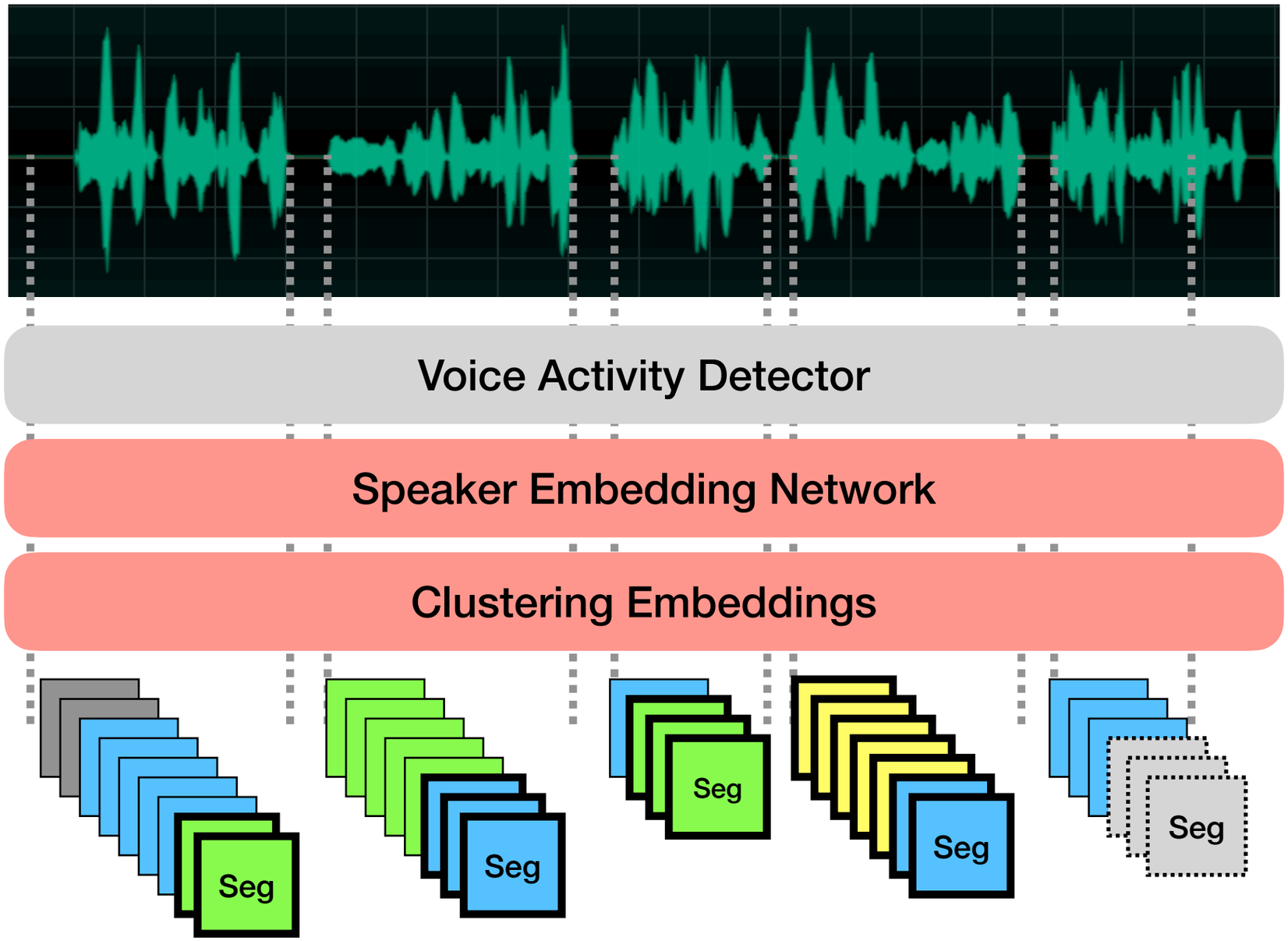

Speaker Segmentation
Speaker Segment 1
Speaker Segment 2
Speaker Segment 3
Speaker Segment 4
Speaker Segment 5
Speaker Clustering
Speaker Segment 1
Speaker Segment 2
Speaker Segment 3
Speaker Segment 4
Speaker Segment 5


Voice Activity Detector
Seg
Seg
Seg
Seg
Seg
Speaker Embedding Network
Clustering Embeddings
Seg
Seg
Seg
Seg
Seg

FA & Miss Error



Confusion Error



성능 평가

- Diarization Error Rate (DER)
  - FA + Miss + Confusion
- Jaccard Error Rate (JER)

$$Error = \frac{DER + JER}{2}$$

https://github.com/nryant/dscore#iii-metrics

개선 방향

- VAD 성능
- Speaker Embedding Network
- Clustering Algorithm

개선방향

VAD

- Baseline: webrtcvad
- 어떤 알고리즘, 모델이든 상관없음

개선방향

Speaker Embedding Network

- Baseline: ResNetSE34L
  - 학습데이터: Voxceleb2
  - Pretrained Model 제공
  - 학습 모델 성능 평가: Equal Error Rate(EER)
    - 화자분리 성능(DER,JER)과 완벽하게 일치하지는 않음
- 어떤 알고리즘, 모델이든 상관없음

개선방향

Clustering Algorithm

- Baseline: Agglomerative Hierarchical Clustering (AHC)
- 어떤 알고리즘, 모델이든 상관없음

평가 방법 & 기타

- 1주차: 전체 DB의 1/3 로 평가
- 2주차: 전체 DB의 2/3 로 평가
- 3주차: 전체 DB로 평가
- 당연히 Error가 낮을 수록 좋음
- 더 자세한 내용은 README 에 작성해 두었습니다!