Diaboliic @ DIHARD 3

Third Dihard Challenge Workshop

2021

Team Structure & Approach

- Team: LIUM, (Herve Bredin) IRIT, Wenda Chen & Sangeeta Ghangam

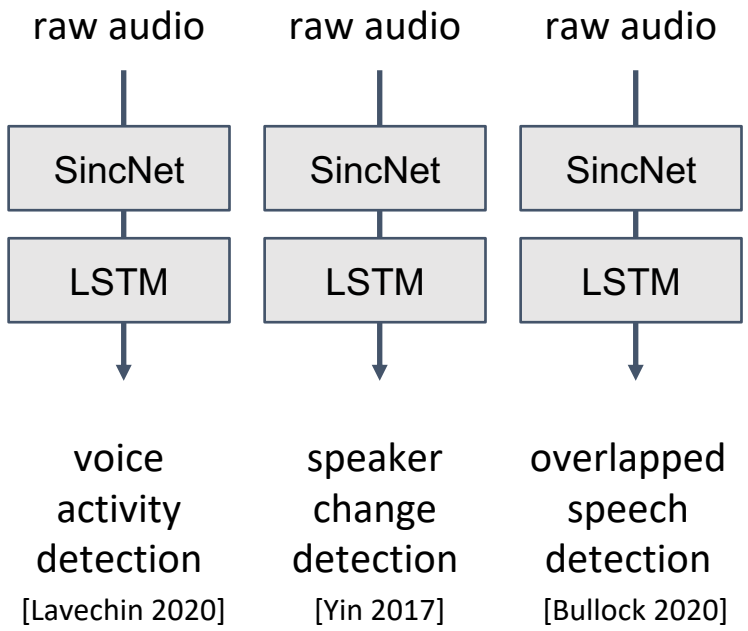
- Summary –

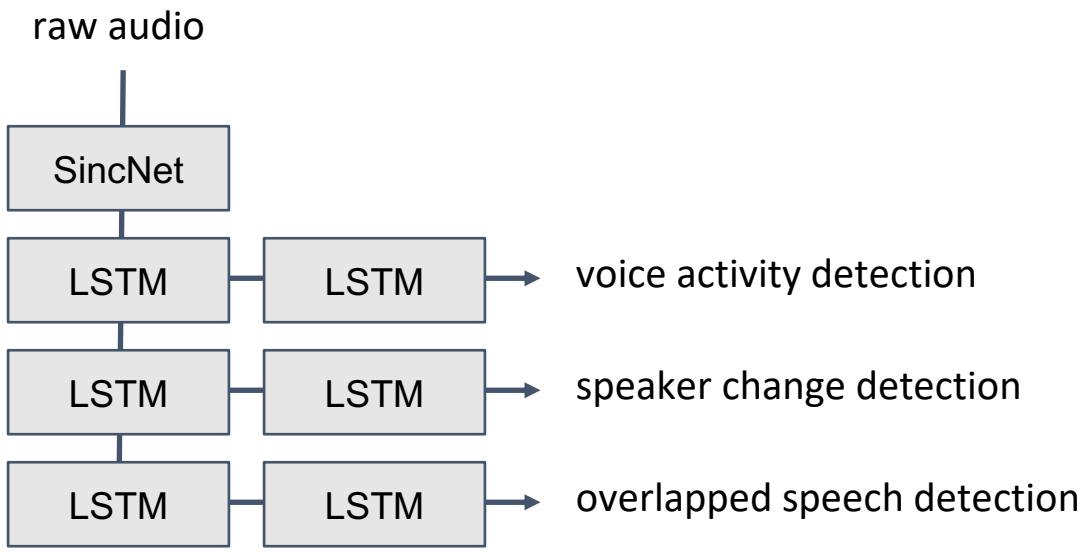
  - ❖ Team created different systems, each optimized for specific modules - Segmentation, Embedding, Re-segmentation
  - ❖ An ensemble approach was used for both the tracks as part of the final results submission

Segmentation

Based on `pyannote.audio` [Bredin 2020]

Slightly improved overlapped speech detection thanks to multi-task training

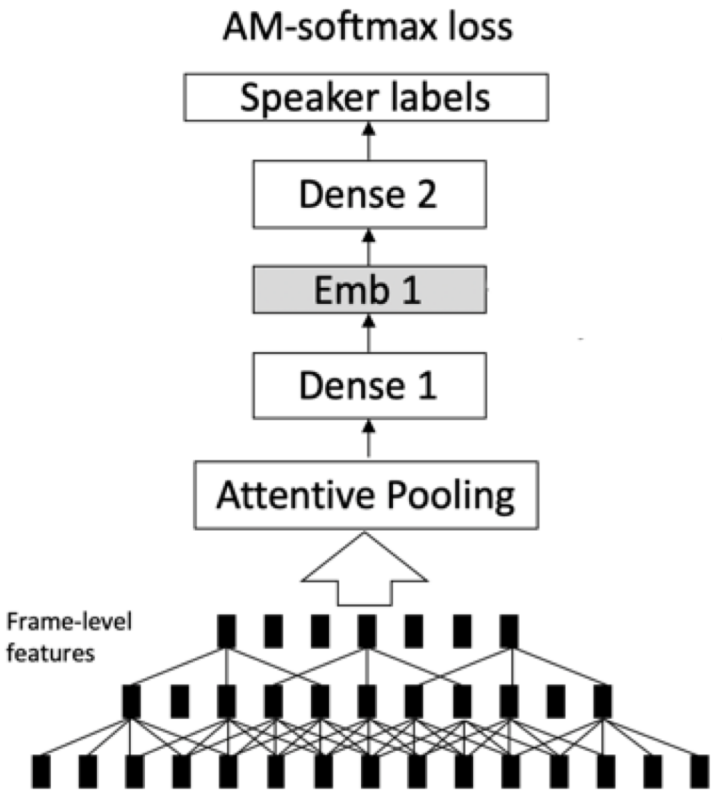




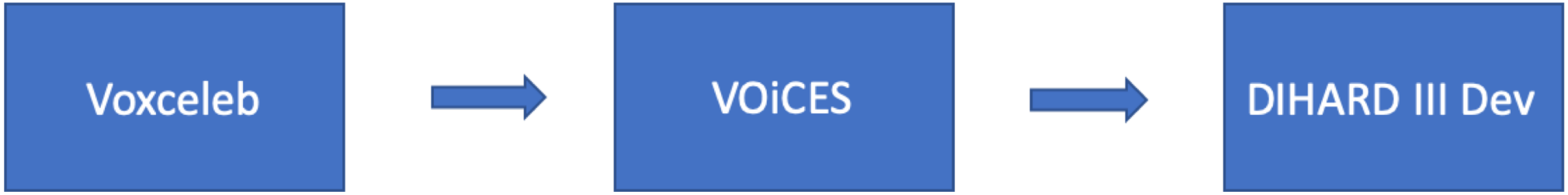
Efficient Embeddings

- X-vector model efficiencies for short-duration speech segments [Chen 2020]
- Distance: model_1 0.941; model_2 0.953

Models	1s	2s	4s	Full
Model1: AM-Softmax	12.74	6.70	3.99	1.90
Model2: AM-Softmax-IRL	13.67	7.13	3.69	1.49



- SID to diarization



Resegmentation

- Based on VBx system [Landini 2020]
- Optimized Parameters - Ploop (0.40), Interpolation Alpha (0.75)
- 4% improvement in the DER when the resegmentation system was combined with the baseline pyannote system

Final Results Summary

- Track1 - Ensemble of following systems
  - Pyannote (Second DIHARD)
  - Baseline Third DIHARD system

These were combined used Dover-lap [Raj 2021]

- Track2 - Ensemble of following systems
  - Pyannote (Updated with segmentation/embeddings)
  - Resegmentation output
  - VBx Baseline System [Landini 2020v2]

These were combined using Dover [Stolcke]

Results Summary - 1

Track	System	Dataset	DER	JER
Track 1	Pyannote (Second DIHARD)	Third DIHARD Dev	20.70	
	Baseline Third DIHARD system	Third DIHARD Dev	19.4	
	Final - Full	Third DIHARD Eval	18.89	42.98
	Final - Core	Third DIHARD Eval	20.61	48.38

These were combined used Dover-lap [Raj 2021]

Results Summary - 2

Track	System	Dataset	DER	JER
Track 2	Pyannote Fusion (Updated segmentation/Embedding)	Third DIHARD Dev	28.40	45.63
	Resegmentation	Third DIHARD Dev	24.32	42.72
	VBx Baseline [Landini 2020v2]	Third DIHARD Dev	15.71	34.48
	Final - Full	Third DIHARD Eval	21.18	42.58
	Final - Core	Third DIHARD Eval	22.96	47.17

These were combined using Dover [Stolcke]

References

- [Bredin 2020] *“pyannote.audio: neural building blocks for speaker diarization”*. ICASSP 2020
- [Bullock 2020] *“Overlap-aware diarization: resegmentation using neural end-to-end overlapped speech detection”*. ICASSP 2020
- [Yin 2017] *“Speaker change detection in broadcast TV using bidirectional long short-term memory networks”*. InterSpeech 2017
- [Lavechin 2020] “End-to-end domain-adversarial voice activity detection”. InterSpeech 2020
- [Chen 2020] “Length- and noise-aware training techniques for short-utterance speaker recognition”. InterSpeech 2020.
- [Landini 2020] *“BUT system for the Second DIHARD Speech Diarization Challenge”* ICASSP 2020
- [Stolcke]*”Improving Diarization Robustness using Diversification, Randomization and the DOVER Algorithm”*
- [Raj 2021]*”DOVER-Lap: A Method for Combining Overlap-aware Diarization Outputs”*
- [Landini 2020v2]*”Bayesian HMM clustering of x-vector sequences (VBx) in speaker diarization: theory, implementation and analysis on standard tasks”*