



# CRIM'S SPEAKER DIARIZATION SYSTEM FOR ETAPE 2011

Themos Stafylakis, Vishwa Gupta, Pierre  
Ouellet, and Gilles Boulianne

18 June 2012





## WITHIN-FILE DIARIZATION

- Divide audio into homogeneous segments
- System1: Multi-stage Segmentation and clustering
- System2: Agglomerative Clustering using single Gaussians



## DIVIDE AUDIO INTO HOMOGENEOUS SEGMENTS

- Use change-point detector:
  - Symmetric KL2 metric over sliding 1.7 sec windows
- Iterative Viterbi re-segmentation:
  - model each segment by its mean and variance
  - minimum 1 sec duration of each segment
- Remove music and noise segments:
  - use 5 GMMs: speech, music, music+speech, noise, noise+speech



## Multi-Stage Segmentation and Clustering

- BIC agglomerative clustering: 12 MFCCs + energy
- Iterative Viterbi re-segmentation
- Repeat BIC and Viterbi re-segmentation
- Male/Female Labeling of each cluster
- SID clustering of male/female clusters:
  - 256 mixture GMMs, Gaussianized MFCCs, cross-likelihood ratio
- Viterbi re-segmentation using GMMs



## Agglomerative Clustering using single Gaussians

- Bayesian Modeling
- Single Gaussian to model speaker distributions
- Closed-form expressions using the Normal-Inverse Wishart prior
- Avoids BIC-like approximations: uses exact marginal likelihood of the model.

## WITHIN FILE DIARIZATION RESULTS



System	DER	System	DER
Telecom paristech	11.67	Orange 2	20.77
Lium primary	17.05	Orange 4	22.23
Lium bic ilp	17.52	Orange 3	22.31
Lium clr sr flt2 jfa	17.64	Orange primary	22.36
Lium clr sr	17.80	Orange 6	22.38
Lium bic ilp flt2	18.01	Lia primary	23.75
Crim 2: single Gauss	18.08	Lia secondary	24.81
Lium clr sr flt2	18.48	Eurecom pur ma cm	26.79
Limsi primary	19.35	Eurecom primary	28.36
Crim primary	19.77	Eurecom no map	29.52
Orange 5	20.06		



## CROSS-FILE DIARIZATION

### THREE DIFFERENT SYSTEMS:

- Renumber within file speakers to make them unique (crim x primary)
- Cross-file agglomerative clustering of output from multi-stage segmentation and clustering (using GMMs) (crim x 2)
- Constrained cross-file clustering to avoid any further within-file clustering (crim x 3)
  - cannot combine two cross-file clusters with a common file.



## CROSS-FILE CLUSTERING RESULTS

- Note: crim x 3 is restricted cross-file diarization

System	DER
limsi x primary	18.12
lium x primary bic ilp flt2 jfa clr	18.43
lium x secondary bic sr clr flt2 jfa clr	19.98
crim x 3 (not submitted)	20.97
lium x bic sr clr flt2 clr	21.41
crim x 2	22.61
crim x primary	24.17
lia x primary	26.26





## CROSS-FILE DIARIZATION CPU TIME

- Unrestricted cross-file diarization: 8 hours (22.61% DER)
- Restricted cross-file diarization: 15 minutes (20.97% DER)



## CONCLUSION

- Agglomerative clustering using single Gaussians is very fast and gives lower DER than multi-stage segmentation and clustering. (18.08% versus 19.77%)
- Restricted cross-file clustering is very fast and increases DER compared to within-file clustering by only 1.2% (19.77% versus 20.97%).