

ETAPE CHALLENGE 2012

The EURECOM Speaker Diarization System

Simon Bozonnet, Ravichander Vipperla, Nick Evans

Category: Segmentation (S)

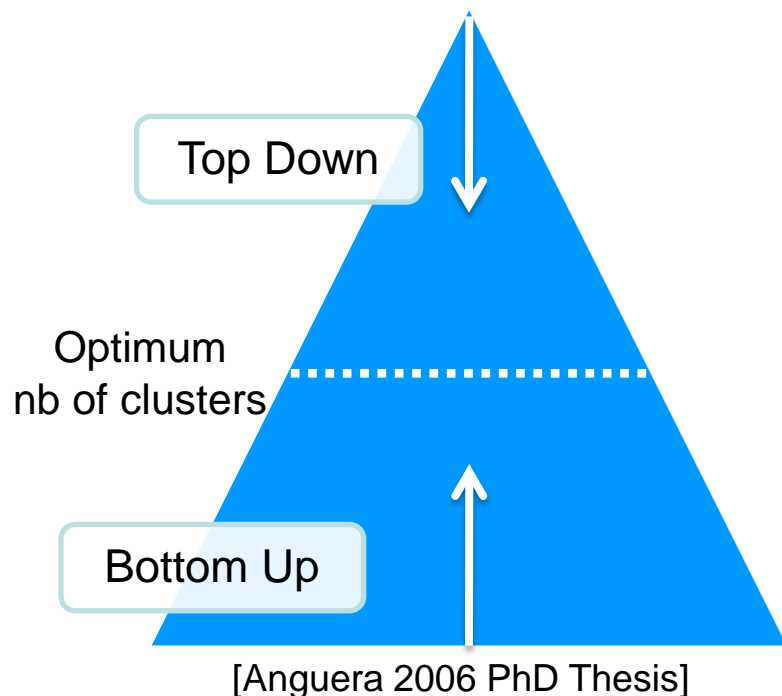
Task: SRL speaker turn segmentation

Overview

- **System Description**
- **Oracle Experiment**
- **Results**
- **Analysis**

Top-Down/Bottom-Up Approaches

- **2 main approaches:**



- **Bottom-up:**

- The most popular
[Sun et al. 2010, Wooters et al. 2007]
- Best performance in the NIST
RT evaluation [NIST 2009]

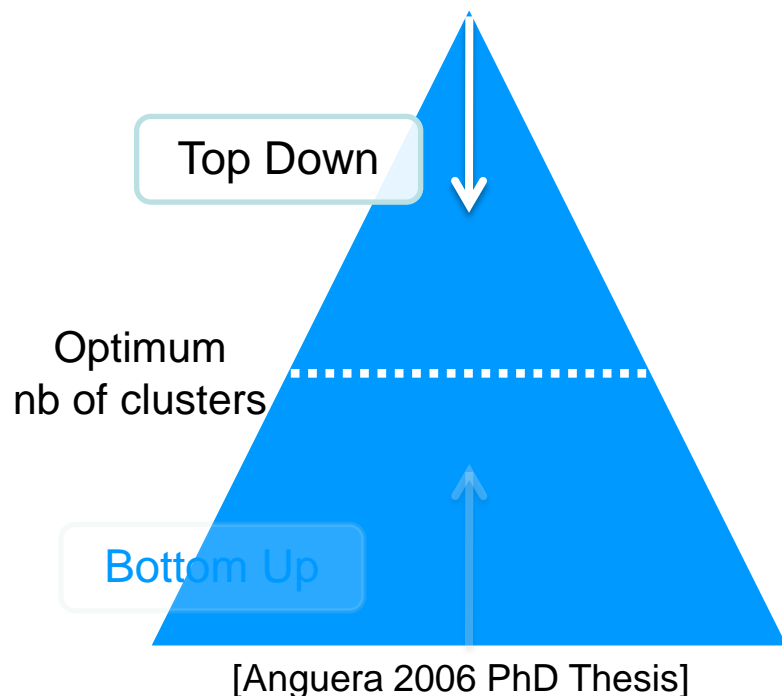
- **Top Down:**

- Less commun
- Achieved competitive results

EURECOM's system: Top-Down

Top-Down/Bottom-Up Approaches

- 2 main approaches:



- **EURECOM system:**

- Top-Down

or

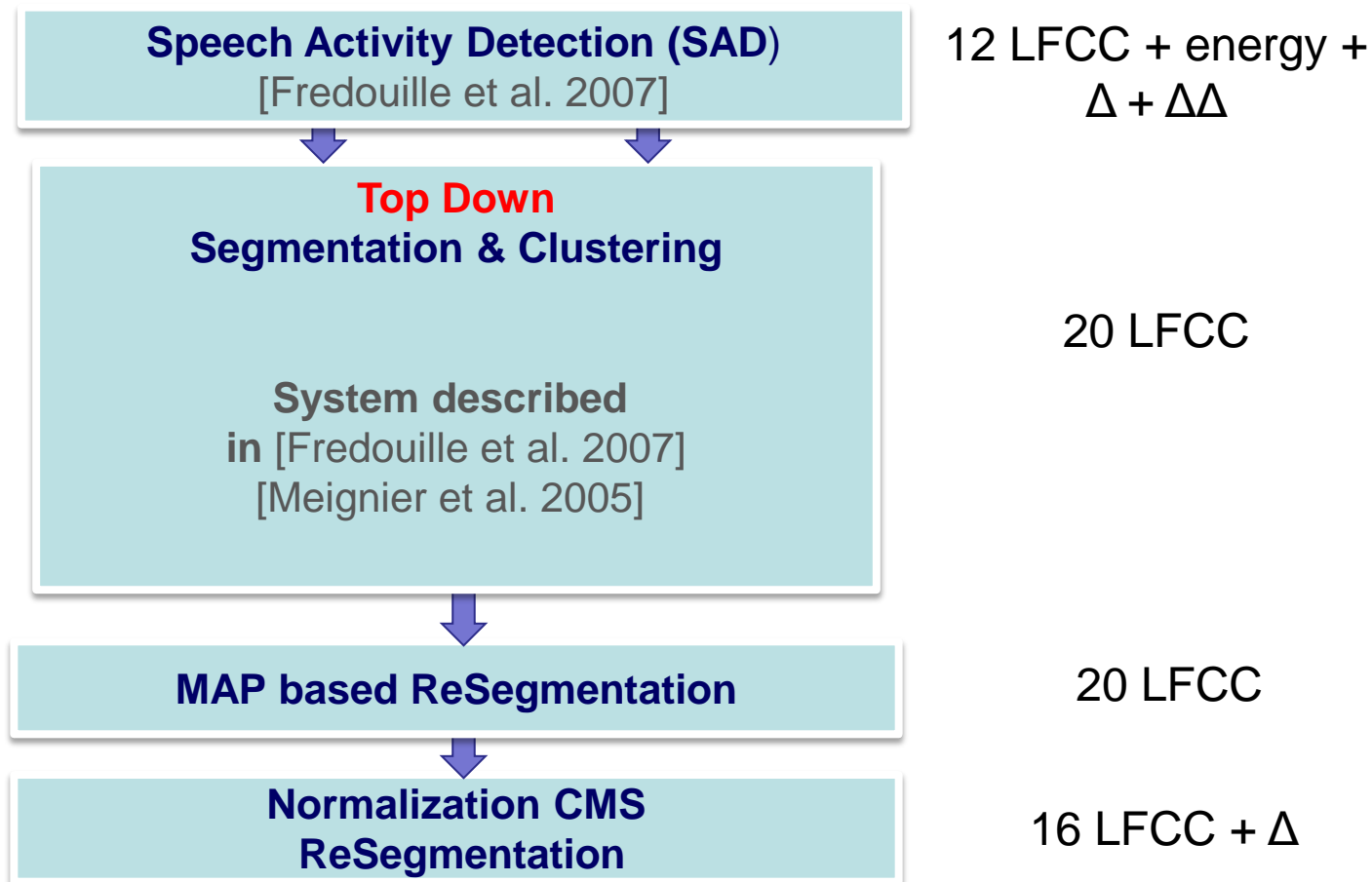
Divisive Hierarchical Clustering

- Optimized for meeting data
(**NIST** RT Corpus)

- 2nd best results in SDM
conditions for **NIST** RT`09
challenge

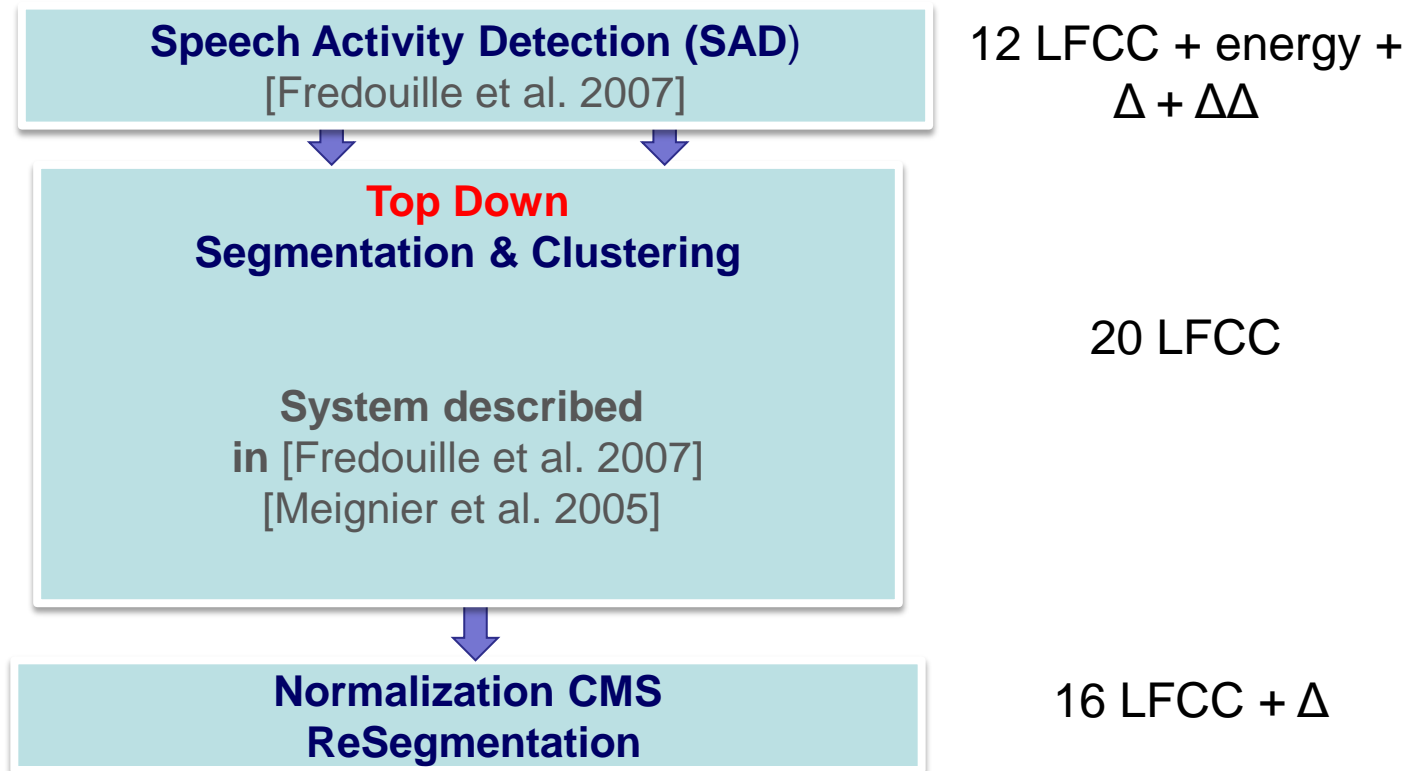
System Description: 3 submissions

- 1st submission: baseline system
(same that for NIST RT'09 evaluation)



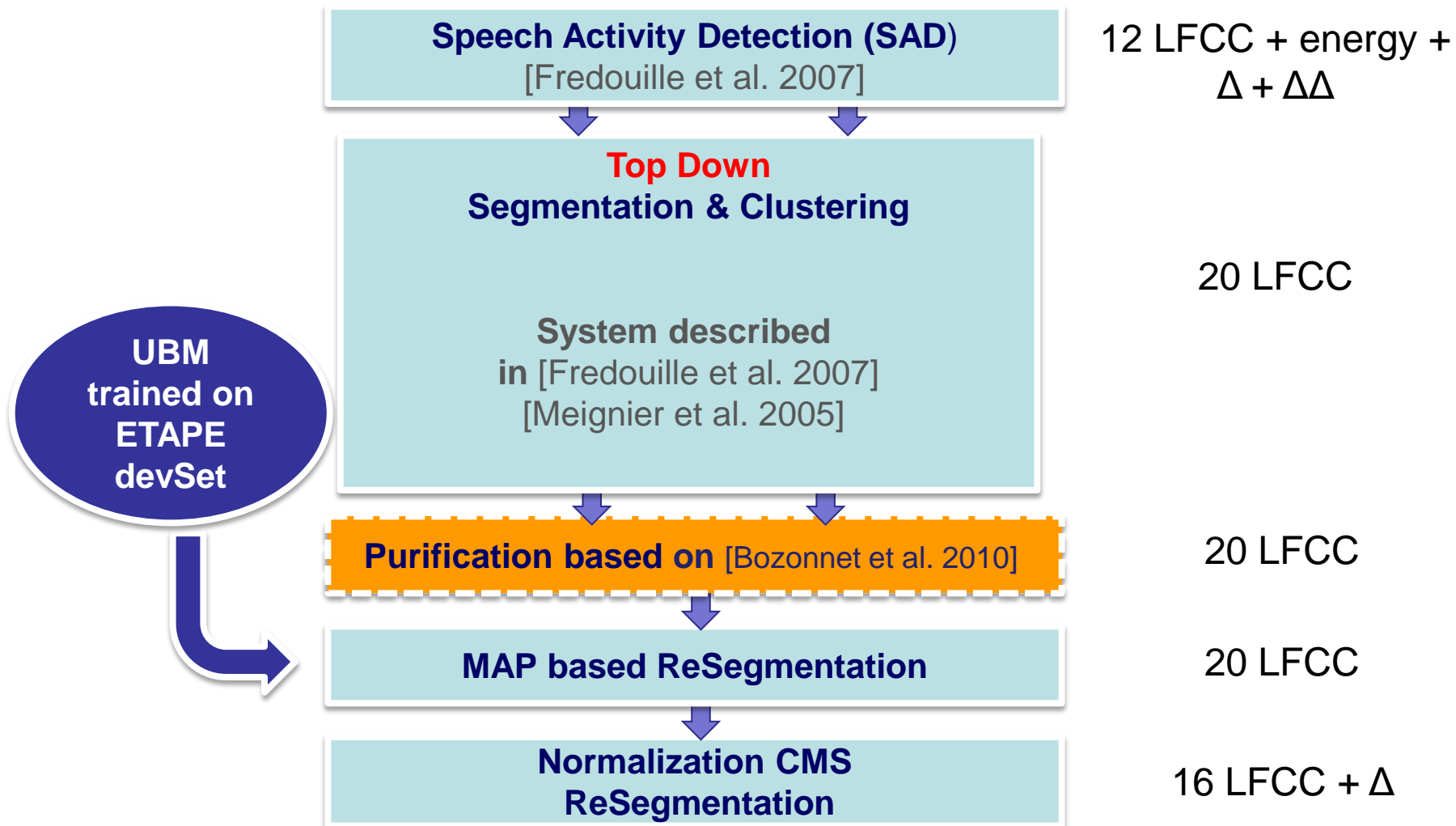
System Description: 3 submissions

- 2nd submission: remove the use of UBM (specific to meeting)



System Description: 3 submissions

- 3rd submission: specific UBM and purification



Results – Development Set

Metric:	Evaluation Set - DER (%)		
	NIST no ovlp	NIST with ovlp	ETAPE ovlp
Baseline System	30,72		
System No MAP	27,31		
System + Purif + MAP (UBM Etape)	26,51		

Results – Eval Set

Metric:	Evaluation Set - DER (%)		
	NIST no ovlp	NIST with ovlp	ETAPE ovlp
Baseline System	26,88	30,73	30,88
System No MAP	28,13	31,87	31,67
System + Purif + MAP (UBM Etape)	25,36	29,14	29,32

Results – Eval Set / Category

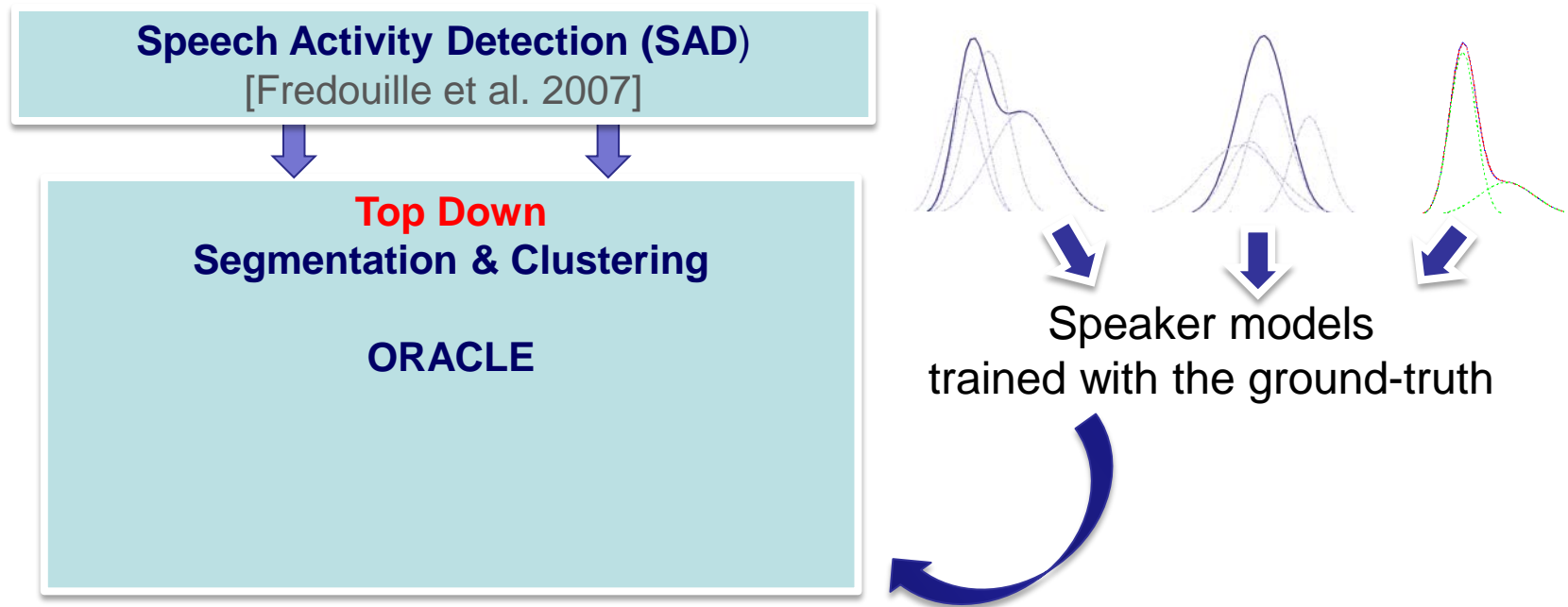
Metric:	Evaluation Set - DER (%)		
	NIST no ovlp	NIST with ovlp	ETAPE ovlp
Baseline System	26,88	30,73	30,88
System No MAP	28,13	31,87	31,67
System + Purif + MAP (UBM Etape)	25,36	29,14	29,32



	% DER
BFMTV	39.39%
EST2BC	34.49%
LCP	13.70%
TV8	27.51%


Oracle Experiment

- **Each speaker model:**
 - Introduced iteratively
 - Trained on the ground-truth
 - Use all the available data for all the speakers



Oracle Experiment: Results – Eval Set

Metric:	Evaluation Set - DER (%)		
	NIST no ovlp	NIST with ovlp	ETAPE ovlp
Oracle Setup	11,44	15,66	15,82
Baseline System	26,88	30,73	30,88
System No MAP	28,13	31,87	31,67
System + Purif + MAP (UBM Etape)	25,36	29,14	29,32



	% DER
BFMTV	14.89
EST2BC	15.23
LCP	12.71
TV8	13.28

Oracle Experiment: Results – Eval Set

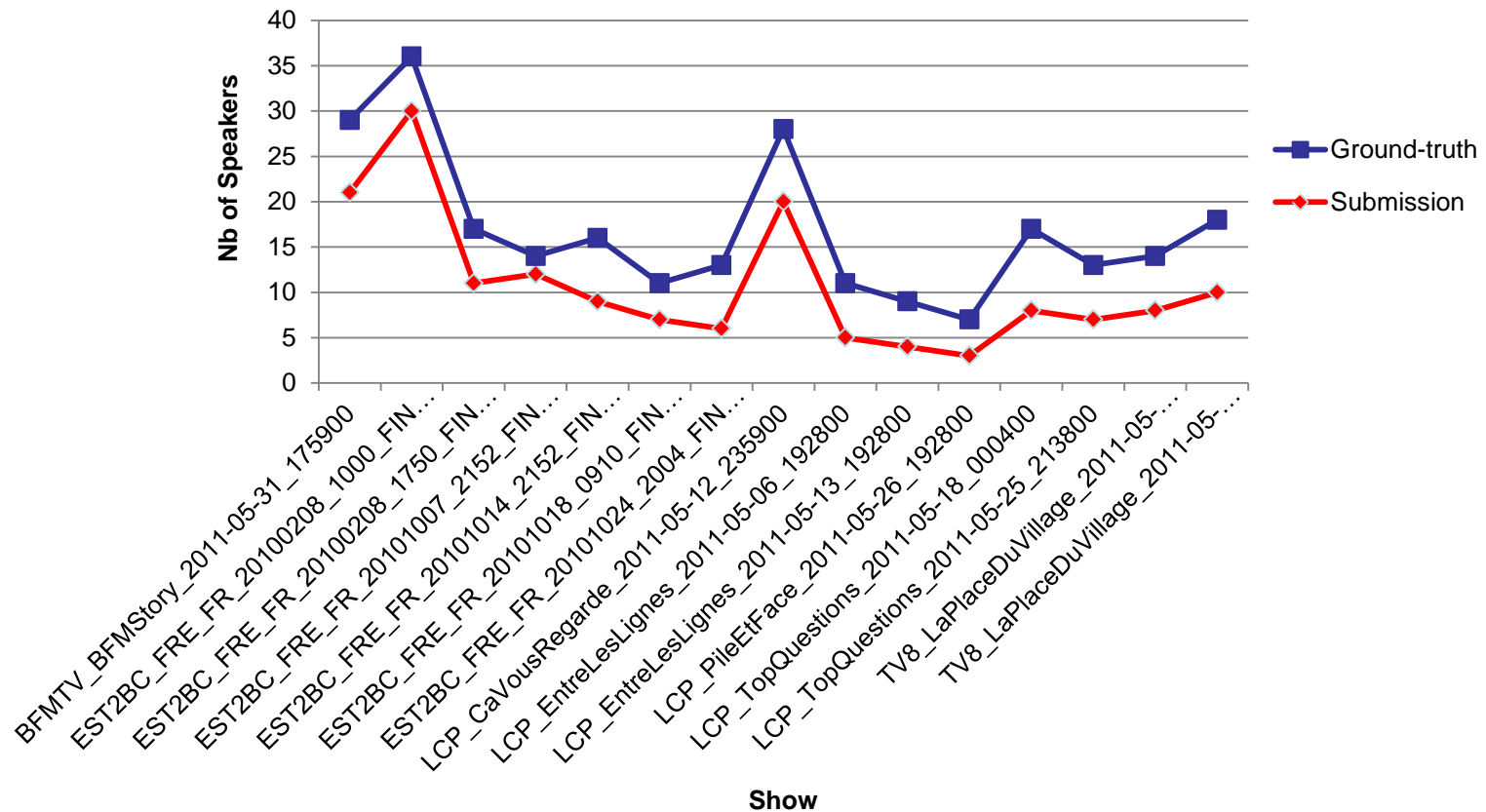
Show	% DER (No ovlp scored)
BFMTV_BFMStory_2011-05-31_175900	14.89
EST2BC_FRE_FR_20100208_1000_FINTER_DEBATE	4.97
EST2BC_FRE_FR_20100208_1750_FINTER_DEBATE	27.47
EST2BC_FRE_FR_20101007_2152_FINTER_DEBATE	20.52
EST2BC_FRE_FR_20101014_2152_FINTER_DEBATE	19.77
EST2BC_FRE_FR_20101018_0910_FINTER_DEBATE	7.08
EST2BC_FRE_FR_20101024_2004_FINTER_DEBATE	11.59
LCP_CaVousRegarde_2011-05-12_235900	3.66
LCP_EntreLesLignes_2011-05-06_192800	23.63
LCP_EntreLesLignes_2011-05-13_192800	25.78
LCP_PileEtFace_2011-05-26_192800	4.22
LCP_TopQuestions_2011-05-18_000400	16.95
LCP_TopQuestions_2011-05-25_213800	2.04
TV8_LaPlaceDuVillage_2011-05-03_201300	17.88
TV8_LaPlaceDuVillage_2011-05-12_172800	8.68

Comparison RT Meeting data/ ETAPE data

Attribute	NIST RT'09 meeting	ETAPE eval Set
<i>Nb Of Shows</i>	7	15
<i>Evaluation Time</i>	25 min	4170 min
<i>Total Speech</i>	13 min	378 min
<i>Avg Nb Of Segments</i>	882	294
<i>Avg Segment length</i>	1 sec	6 sec
<i>Ovlp</i>	3 min	1.5 min
<i>Avg Nb Spk</i>	5	11
<i>most active</i>	535 sec	834 sec
<i>least active</i>	146 sec	0.25 sec

	# meetings	Avg spk	Avg turn duration	% silence	% ovlp
BFMTV	1	21	10 sec	32%	2%
EST2BC	6	12.5	5 sec	35%	3%
LCP	6	7.8	8 sec	22%	7%
TV8	2	9	2 sec	28%	4%

Analysis of the Number of Detected Speakers



References

- **[Fredouille et al. 2009] The LIA-EURECOM RT`09 Speaker Diarization System.** In RT`09, NIST Rich Transcription Workshop, 2009, Melbourne, Florida.
- **[Bozonnet et al. 2010] The LIA-EURECOM RT`09 Speaker Diarization System: enhancements in speaker modeling and cluster purification.** In Proc. ICASSP, Dallas, Texas, USA, March 14-19 2010.
- **[Anguera 2006 PhD Thesis]**

Thank you!!

Statistics over the eval dataset

■ Number of speakers:

	real number of spk	Baseline	eurecom_no_map		eurecom_purif_mapetape_cms		
		nb clusters		nb clusters		nb clusters	
BFMTV_BFMStory_2011-05-31_175900	21	8	13	9	12	8	13
EST2BC_FRE_FR_20100208_1000_FINTER_DEBATE	30	5	25	6	24	6	24
EST2BC_FRE_FR_20100208_1750_FINTER_DEBATE	11	3	8	7	4	6	5
EST2BC_FRE_FR_20101007_2152_FINTER_DEBATE	12	2	10	3	9	2	10
EST2BC_FRE_FR_20101014_2152_FINTER_DEBATE	9	6	3	7	2	7	2
EST2BC_FRE_FR_20101018_0910_FINTER_DEBATE	7	4	3	4	3	4	3
EST2BC_FRE_FR_20101024_2004_FINTER_DEBATE	6	7	1	7	1	7	1
LCP_CaVousRegarde_2011-05-12_235900	20	8	12	9	11	8	12
LCP_EntreLesLignes_2011-05-06_192800	5	5	0	6	1	6	1
LCP_EntreLesLignes_2011-05-13_192800	4	5	1	5	1	5	1
LCP_PileEtFace_2011-05-26_192800	3	4	1	4	1	4	1
LCP_TopQuestions_2011-05-18_000400	8	9	1	9	1	9	1
LCP_TopQuestions_2011-05-25_213800	7	6	1	6	1	6	1
TV8_LaPlaceDuVillage_2011-05-03_201300	8	6	2	6	2	6	2
TV8_LaPlaceDuVillage_2011-05-12_172800	10	8	2	8	2	8	2
	161	86	83	96	75	92	79

System Description

Speech Activity
Detection
(SAD)

Segmentation
&
Clustering

ReSegmentation

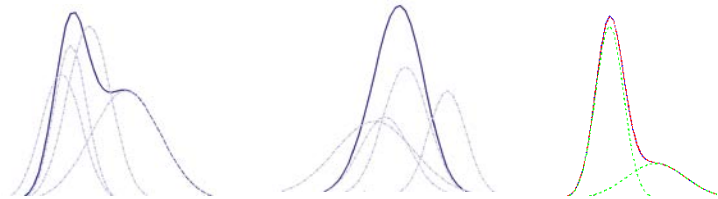
Normalization
&
Resegmentation

Cluster 1

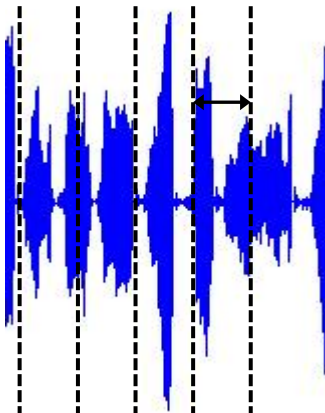
Cluster 3

Cluster 2

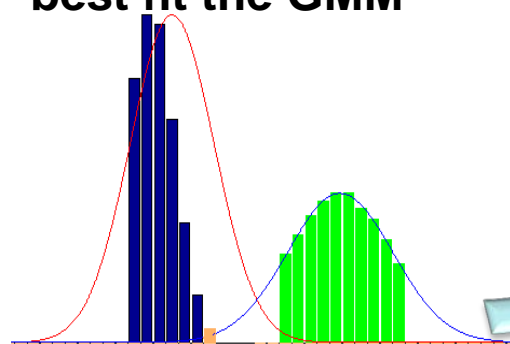
1. Train
16G-GMM



2. Make 500ms segments for
each cluster

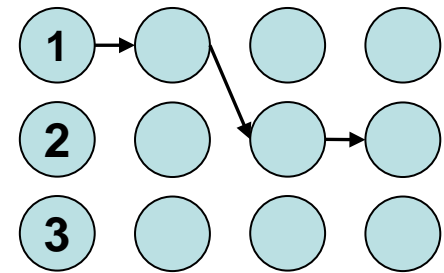


3. Keep the
segments which
best fit the GMM

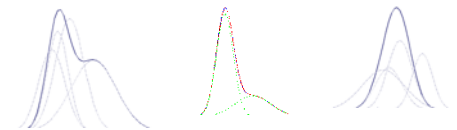


X 10

5. Viterbi Decoding



4. Train
16G-GMM



Experimental Setups

■ 3 submissions:

➤ Baseline system

- ☞ Originally optimized for **meeting domain**

➤ Baseline without using MAP adaptation

- ☞ Segmentation (EM)
 - ☞ Normalization & ReSeg (EM)
- } No more need for UBM

➤ With purification, MAP use a UBM trained on ETAPE dev set, CMS normalization

- ☞ Segmentation (EM)
- ☞ Purification (EM)
- ☞ ReSegmentation MAP (UBM trained on ETAPE data)
- ☞ Normalization and ReSegmentation