SPEAKER DIARIZATION

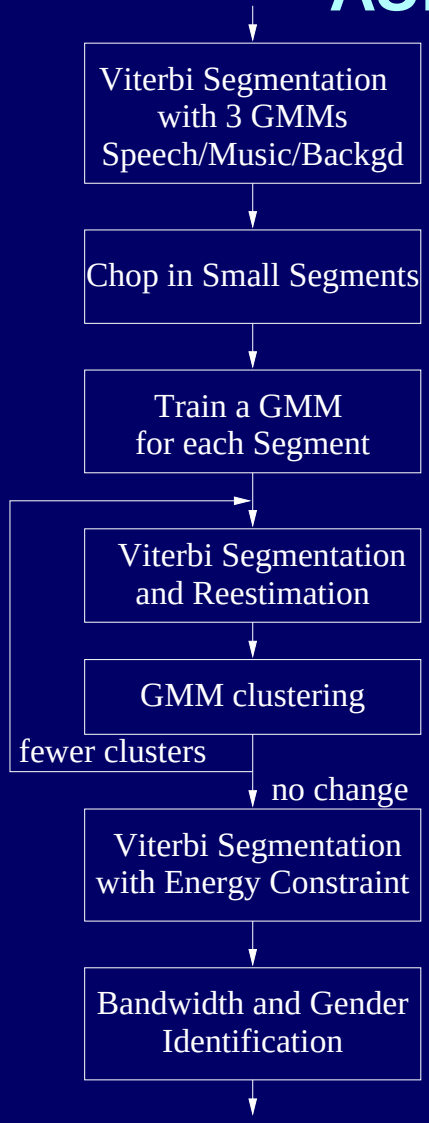
J.L. Gauvain, C. Barras

RT03 meeting
Boston, MA
May 20, 2003

TALK OUTLINE

- Speaker diarization for the BNEWS task
- Audio partitioner used in the BN STT system since 1998
- Two parameters tuned for better speaker segmentation accuracy
- Post STT filtering
- Interaction with STT performance

AUDIO PARTITIONING FOR BN DATA



- Audio stream mixture model, where each speaker/environment condition is modeled by a GMM
- Maximum likelihood segmentation/clustering iterative procedure
- Objective function is a penalized log-likelihood

$$\sum_{i=1}^{N} \log f(s_i|\lambda_{c_i}) - \alpha N - \beta K$$

- Designed for STT not MDE

PARAMETER TUNING

Data from a single speaker is divided in 2 or more clusters, sometimes with different background acoustic conditions (good to minimize STT WER, not good for diarization)

- Minimize time error
- α: the maximum log-likelihood loss for a merge
- β: the segment boundary penalty

Dev03	*Time Error*	*Word Error*
STT tuning α= β=160	37.31%	28.90%
MDE tuning α= β=230	26.79%	18.79%

POST STT FILTERING

Use recognizer hypothesis to remove interword segments above 30ms

Dev03	*Time Error*	*Word Error*
MDE tuning α= β=230	26.79%	18.79%
MDE tuning α= β=230,+filter	24.38%	20.05%

Eval03 summary

Eval03	*Time Error*	*Word Error*
STT tuning α= β=160	33.97%	27.96%
MDE tuning α= β=230	26.26%	19.49%
MDE tuning α= β=230,+filter	24.47%	20.92%

INTERACTION WITH STT

Eval03	*Word Error Rate*						
	ABC	CNN	MNB	NBC	PRI	VOA	Avg.
STT segments	10.3	12.9	10.2	11.5	10.8	11.4	11.2
MDE segments	10.5	13.2	10.4	12.4	11.3	11.9	11.6
STT segments, notel	10.5	12.9	9.4	11.5	10.9	11.2	11.1
LL segments, notel	11.1	13.1	9.8	13.0	11.6	12.1	11.8

CONCLUSIONS

- Standard LIMSI STT partitioner slightly tuned for MDE task
- Filtering with STT output gives mixed results
- Optimizing MDE error rate requires different criteria than optimizing STT error rate
- Word error rate is lower with original tuning (11.2% vs 11.6%)
- MDE substitutions are not a problem for STT
- Identifying the correct number of speakers is a key problem for the speaker MDE task