RT-04 MDE Evaluation Systems at CUED

M. Tomalin, S. Tranter, M.J.F. Gales, R. Sinha, S. Umesh, and P.C. Woodland

8th November 2004

Cambridge University

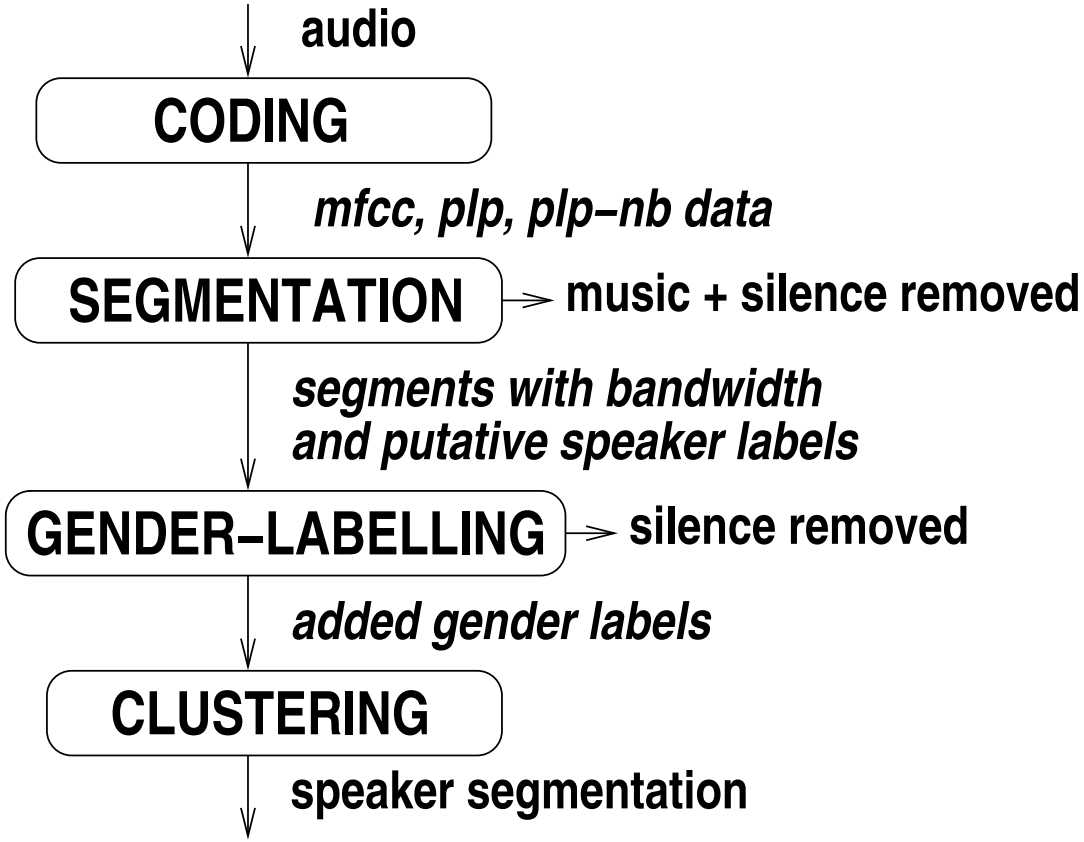
Overview

- CUED Diarisation systems for RT-04.
  - Modification since RT-03s.
  - Progress since RT-03s.
  - Conclusions.

- CUED Structural MetaData Extraction (SMD) systems for RT-04.
  - Large data sets for Slash Unit Boundary Detection.
  - Filler Word Detection system.
  - Interruption Point Detection system.
  - Conclusions.

Diarisation System : General Architecture

3 stage process : Segmentation →Gender Labelling →Clustering



Diarisation System : Modifications since RT-03s

- **Segmentation** : LIMSI-style iterative scheme
  - Audio type/ bandwidth detection using GMM classifier.
  - Oversegmentation using divergence based change detector.
  - Iterative segmentation-clustering.
  - Segmenter also produces putative speaker labels.
- **Clustering** :
  - top-down, AHS distance, full-correlation matrix of PLPs.
  - Uses BIC-based stopping criterion.
  - Segments are sorted before clustering to help initialisation .

Diarisation System : Progress since RT-03s

Results on RT-04 evaluation data

Coding	Segmentation	Clustering	DER
RT-03s	RT-03s	RT-03s	**36.33**
RT-03s	RT-03s	RT-04	27.90
RT-03s	RT-04	RT-04	**22.48**
† RT-04	RT-04	RT-04	23.86

† Official evaluation submission

- Modifications since RT-03s resulted in 14% absolute drop in DER.
- Slight degradation in DER due to switching compilers in coding.

Diarisation System: Conclusions

- RT-04 system gave 34% relative improvement in DER over RT-03s system.
- Over-sensitivity of system to small changes in input could be due to top-down clustering approach used.
- Integrating segmenter speaker information could help improve performance further.
- Possible future work includes:
  - exploiting speaker labels from segmenter in clusterer.
  - cluster voting of segmenter and clusterer outputs.
  - investigating the use of proxy speaker models.

CUED SMD: from RT-03f to RT-04

One SMD task attempted for RT-03f:

- CTS Slash Unit Boundary Detection (SUBD)

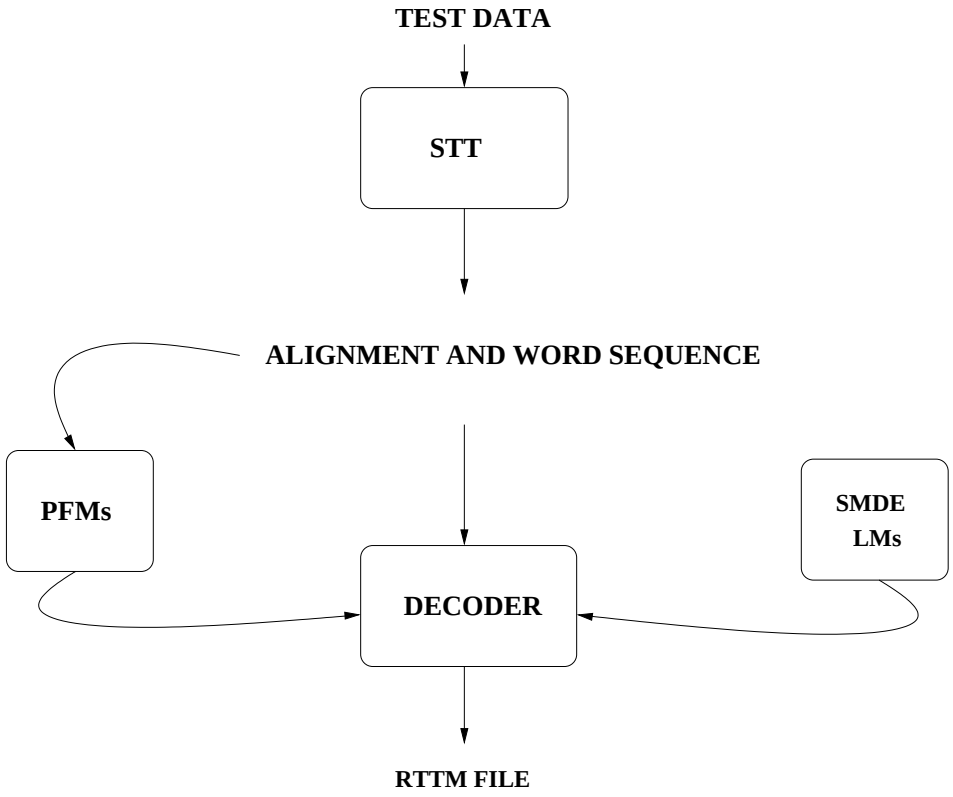
Four SMD tasks attempted for RT-04:

- CTS SUBD.
- BN SUBD.
- CTS Filler Word Detection.
- CTS Interruption Point Detection.

Three of the CUED SMD systems were developed for RT-04.

General System Architecture

The SMD systems used same generic architecture:



General System Architecture

CTS SMD systems:

- input: audio files, CUED 20xRT CTS STT output.
- task-specific Language Models (LMs).
- task-specific Prosodic Feature Models (PFMs).
- 1-Best lattice-based Viterbi Decoder.

BN SMD systems:

- input: audio files, CUED 20xRT BN STT output
- task-specific LMs.
- task-specific PFMs.
- 1-Best lattice-based Viterbi Decoder.

Large Training Data Sets for CTS and BN SUBD

An attempt to use large STT training data sets for SUBD task:

- c.1500 hrs of STT CTS WordWave (WW) training data was mapped to approximate SU annotations
- c.4100 hrs of STT BN training data (db98, bn2003) was mapped to approximate SU annotations

The mapping rules:

- full-stop $\longrightarrow$ statement boundary
- comma $\longrightarrow$ statement boundary
- question mark $\longrightarrow$ question boundary

Word-based and class-based SULMs were built using mapped training data.

SUBD for CTS and BN

SUBD results for CTS and BN:

SYSTEM	%Err (DEL/INS/ERR)		
	dev03	**eval03**	**dev04**
CTS PFM+MDE_SULMs	31.8/14.8/57.0	31.3/13.8/56.1	29.1/14.7/54.4
+ WW_SULM	30.7/15.4/**56.7**	30.4/14.3/**55.8**	28.1/15.3/**54.2**
BN PFM+MDE_SULMs	46.1/14.8/63.4	45.4/15.3/63.9	53.7/21.7/78.8
+ db98_SULMs	42.4/16.6/61.7	42.9/16.7/63.1	52.0/22.4/77.9
+ bn2003_SULMs	41.0/17.2/**61.0**	42.1/16.8/**62.5**	51.5/22.8/**77.8**

WW_SULM and MDE_SULMs interpolated: Err falls slightly for eval03 and dev04.

mapped BN SULMs and MDE_SULMs interpolated: Err falls by 4% abs on average.

FWD for CTS

The FWD system consisted of:

- trigram Filler Word Language Models (FWLMs).
- 1-Best lattice-based Viterbi Decoder.

The IPD system consisted of:

- trigram and 40 class class-based Interruption Point Language Models (IPLMs).
- An Interruption Point PFM.
- 1-Best lattice-based Viterbi Decoder.

RT-04 Eval Results

Results for CUED SMD RT-04 Evaluation Systems:

SYSTEM	%Err (ERR only)			
	dev03	eval03	dev04	eval04
CTS FMD (spch)	52.2	55.2	44.3	45.8
CTS FMD (ref)	25.3	25.4	25.5	27.4
CTS IPD (spch)	61.4	62.2	58.1	63.5
CTS IPD (ref)	42.8	42.1	44.5	47.2
CTS SUBD (spch)	56.7	55.8	54.2	56.5
CTS SUBD (ref)	52.0	50.6	45.2	46.2
BN SUBD (spch)	61.0	62.5	77.8	72.2
BN SUBD (ref)	57.5	60.6	75.1	71.1

CTS eval04 performance in line with dev set performance for all tasks.

dev04 and eval04 sets for BN SUBD harder than dev03 and eval03 sets.

CTS SUBD: from RT-03f to RT-04

Difficult to compare RT-03f and RT-04 system performance:

- RT-03f: SUB errors not scored; V5 MDE annotation spec.
- RT-04: SUB errors scored; V6.2 MDE annotation spec.

Results using V5 and V6.2 versions of the eval03 scoring ref files:

SYSTEM	DEL	INS	SUBS	%Err (DEL/INS)
RT-03f_sys/V5_ref	33.1	19.3	11.7	64.1 (52.4)
RT-03f_sys/V6.2_ref	34.1	21.2	10.9	66.1 (55.2)
RT-04_sys/V5_ref	32.0	15.1	13.9	61.0 (47.1)
RT-04_sys/V6.2_ref	30.4	14.3	11.2	55.8 (44.7)

RT-04 sys ERR rates between c.5% and c.11% abs lower than RT-03f sys ERR rates.

SMD Conclusions

- RT-04 CTS SUBD system produces ERR rates c.8% lower than RT-03f system.
- CTS FWD, IPD, and BN SUBD systems constructed for RT-04.
- Same basic system architecture can be used for all SMD tasks.
- FWD system does not use PFM data.
- Mapping large STT training data sets for MDE worth exploring.

References

For more information about CUED RT-04 SMD systems:

Tomalin and Woodland, 'The RT-04 Evaluation Structural Metadata Systems at CUED', Proc. Fall 2004 RT-04 Workshop, to appear

For more information about CUED RT-04 Diarisation systems:

Tranter et al., 'The Development of the Cambridge University RT-04 Diarisation System', Proc. Fall 2004 RT-04 Workshop, to appear