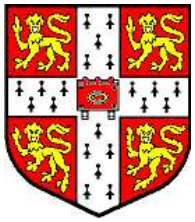
EARS STT Overview

Phil Woodland

February 4th 2004



Cambridge University Engineering Department

Outline

- STT in EARS
  - Teams & the Super Team
  - Targets & Measuring Progress
  - Data & Algorithms
- Where We Started This Year
  - RT03 tasks and outcome
- What We Have Been Doing
  - New Data
  - New Technical Work & Workshops
- Workshop Highlights

  - Martigny September 2003
  - St. Thomas December 2003

- Where We Are Going: towards RT04

- Key Issues

  - Data for 2004
  - Accounting for WER Floor

STT in EARS

- Aggressive Targets on WER reduction in English + run-time constraints: :
  - 3 fold reduction in WER for broadcasts over 3.5 years
  - 5 fold reduction in WER for conversational telephone speech over 5 years
  - Average 27% relative reduction in WER per year required
  - Go/No-Go program decisions based on achieving these targets
  - Move from unlimited times real time to 1xRT while reducing error rate!
- Also Arabic and Mandarin Chinese for both broadcasts and conversations
- Teams:
  - BBN, LIMSI, U Pittsburgh, U Washington
  - Cambridge U
  - IBM [STT only]
  - SRI, ICSI, U Washington
- Inter-team collaboration strongly encouraged

Measuring STT Progress

Current Test Sets : Traditional NIST speech recognition evaluation approach

- Evaluates English/Arabic/Mandarin for both broadcasts and conversations
- Test sets available in time via LDC and for other sites taking part in RT0x
- Evaluation data becomes development data for future years
- Data specification may evolve (broadcast epoch, conversation topic set etc ...)

Progress Test Set : Fixed test set to eliminate test-set variability for tracking progress towards program goals

- Broadcasts from Feb'2001
- CTS Fisher data
- Detailed results & reference transcripts NOT released
- Only available to EARS STT sites
- Baselines for progress from BBN systems existing in 2002

Targets & RT03 Progress

Progress Test Set (targets and best systems):

	Broadcast			
	Target	Achieved		
Baseline	18.0@UL	—		
RT03	14.4@20x	12.3@20x	12.7@10x	16.4@1x
RT04	9.6@10x	—		
RT05	6.0@1x	—		

	CTS			
	Target	Achieved		
Baseline	27.8@UL	—		
RT03	22.2@UL	17.4@UL	18.2@10x	25.9@1x
RT04	15.6@20x	—		
RT05	10.0@10x	—		
RT07	5.6@1x	—		

Where We Started (RT03)

- Achieved first year targets!
- 1xRT RT03 performance better than baselines
- Roughly similar picture for current and progress test sets:
  - BN English results 2-3 abs% harder progress than current
  - CTS Fisher progress very similar to Fisher component of current

For non-English

- Improvements for all languages / tasks
- Relatively little data for non-English
- Work focused on English and porting to non-English

Whats Been Happening: Data

- Large quantities of new training data released in 2003:
- Allow the use of *an order of magnitude* more training data for building STT systems
- Low(er) quality transcriptions than conventionally used

Fisher Collection

- 2000 hours of telephone conversations collected (8 mins each)
- Quick transcription at Wordwave and LDC (typically 5-8xRT)
- BBN provided automatic segmentations for WordWave data
- STT sites now starting to experiment with this data

Broadcast Data

- STT sites started to look at large quantities with only closed captions: initially TDT data (TDT2,3,4)
- LDC are releasing large amounts of new data: (TDT sources March-July 2001), new recordings from 2003 with closed captions
- Lending library model for distribution of new audio data
- Semi-automatic correction of closed captions for TDT4 training set will result in 250 hours of corrected transcripts (LDC/LIMSI)

Whats Been Happening: Workshops

- RT02/RT03 didn't allow STT sites to report in detail on experiences
- Hence sites organised a series of EARS STT workshops

  Martigny Sep'03 Post-Eurospeech'03
  St. Thomas Dec'03 Post-ASRU'03
  Montreal May'04 (planned) Pre-ICASSP'04

- EARS STT sites plus a few others: fairly small numbers typically 30-40
- Give results on current EARS tasks
- Detailed presentations and in-depth discussion: complements other meetings
- Real-time creation of web-site of presentations (& wi-fi)
- Forum for building the **super team**!

Workshop Highlights: Martigny Sept'03

- 2 day meeting hosted by IDIAP
- Topics covered included:
  - large training sets with broadcast data (BBN, CU, LIMSI, Panasonic)
  - CTS data (BBN, CU)
  - Acoustic modelling (BBN, CUx2, ICSI-NA)
  - Language modelling (BBN, IBM, LIMSI, SRI, UW)
  - Decoding (SRI)
  - System improvements (SRI)
- Key event was interaction in discussions, breaks and elsewhere
- Real feeling of collaboration

Workshop Highlights: St. Thomas Dec'03

- 1.5 day meeting after ASRU (another good location!)
- Topics covered included:
  - Use of large amounts of Fisher (CU, LIMSI)
  - Fisher segmentation (BBN)
  - Use of lots of broadcast data (BBN, SRI)
  - Work on Arabic and mandarin (LIMSI, UW/SRI)
  - Acoustic modelling (BBN, CUx2, IBM, ISL, SRI)
  - Language modelling (LIMSI, SRI, UW/SRI)
  - Novel approaches update (ICSI/SRI, Microsoft)
  - Error analysis (BBN, UW)
  - Decoding (CU, IBM, SRI)
- Probably not enough discussion time in only 1.5 days: very full schedule

STT Collaboration Examples

- Sharing & Preparing Data
  - Wordwave segmentations (BBN)
  - Common broadcast news development sets (LIMSI + BBN,CU,SRI)
  - Shared TDT transcriptions (all sites working on broadcasts)
  - Shared CTS transcriptions (CU)
  - Shared lattices (CU, LIMSI)
  - Shared CTS and Broadcast segmentations (all)
- Comparing results/techniques
  - SRI with BBN on almost parsing LM
- Sharing Ideas
  - Workshops

  - Inter-team meetings (e.g. BBN+LIMSI visit to CU)

- Distribution of code (& cross usage)
  - SRI LM toolkit
  - HTK
  - Draft revised scoring (IBM)

Key Issues: 2004 Data

- New data planned for collection in 2004 includes
  - more Fisher English CTS data
  - 200 hours of Levantine Arabic CTS data
  - 200 hours of Mandarin CTS data
  - English dev test: BN (by STT sites and by LDC) & Fisher
  - Non-English dev test: broadcast and CTS
- STT data budget didn't have room for
  - as much Fisher data as we would like
  - any new broadcast collections (but there will be for TIDES)
  - any more closed caption correction
- Key issues:
  - delivery schedules for corpora needed for RT04
  - whether current plan can be refined/improved further

Key Issues: Accounting for WER Floor

- Key issue in measuring word error rates is dealing with inherent uncertainty in reference
- Human transcribers would often disagree in such regions
- Proposal from STT sites on how this **reference noise** which causes **a non-zero floor in WER** can be accounted for in practice:
  - Only score words in regions where two independently derived human references agree
  - Better estimate of true error rate
  - Tends to remove approx constant value from WERs
  - Details presented in later talk
- Improved scoring procedure would need to be applied to the Progress Test set for both conversational and broadcast data

Conclusions

- STT in EARS is working well: much progress has been made
- Year 1 targets met
- Future targets (esp after RT04) extremely challenging!
- New data sources becoming available for RT04
- Finding out how to make most effective use of new data
- Also investigating many new algorithms
- Collaboration is working very well across teams: workshops, data sharing etc.
- Need to finalise issues concerning
  - scoring and the WER floor
  - data collection/transcription for 2004