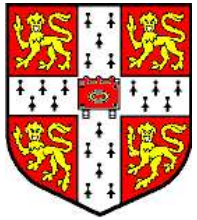
HTK – Future Developments

Gunnar Evermann & Phil Woodland
Cambridge University

May 18th 2004



HTK3 Development Team

HTK users meeting ICASSP'04

Overview

- Background
- HTK at Cambridge University
- HTK3 project
- Future plans

HTK Background

- The HTK (Hidden Markov Model Toolkit) has been at the core of ASR research at Cambridge for more than ten years
- Very flexible, modular toolkit written in ANSI C
- HTK3: available for free download since Sep 2000

  `http://htk.eng.cam.ac.uk`

  More than 20,000 registered users, heavily used in teaching, active support mailing lists

History

- Development started in 1989 at Cambridge University
- Entropic Cambridge Research Labs

entropic

- Entropic was acquired by Microsoft
- CUED arranged license agreement with Microsoft
- HTK3 made available free of charge

htk3

How is HTK used at Cambridge?

- Teaching (one year MPhil course)
- Research (staff & PhD research students)
- Building evaluation systems, e.g. 2003 CTS eval:
  - 6 sets of acoustic models: HLDA SAT MPE + Tri-/Quinphone
  - trained on 360 hours of audio data
  - 4-gram language model + 3-gram class model
  - trained on 500 million words of text
  - 58k vocabulary
  - PLP frontend, HLDA, VTLN, lattice MLLR, confidence scores
  - 190x real time, 20% word error rate

Project Aims

- Lower barrier of entering ASR research
- Give (small) research groups access to state-of-the-art research system, to work on part of problem (e.g. LM)
- Provide tool for ASR teaching
- Establish baselines for standard tasks
- Build a community of ASR researchers/students

HTK in DARPA EARS project

- Goal of the EARS (Effective, Affordable, Reusable Speech-to-Text) project is:

  "to produce powerful new speech-to-text (automatic transcription) technology [..] The intent is to create core enabling technology suitable for a wide range of advanced applications"

- CUED is one of the 4 teams working on "Rich Transcription" for 5 years
- CU-HTK will be the basis for further development at CUED
- CUED will make available more components of a large vocabulary system
- HTK-based infrastructure for standard tasks will be provided

HTK3 additions in EARS

- Fully integrated language modelling tools (word/class-based n-grams)
- Comprehensive set of lattice processing tools
- Improved adaptation code to provide more general framework to support different schemes, incl. adaptive training
- Large Vocabulary Decoder (cross-word triphones, n-gram, lattice generation and rescoring)
- Discriminative training tools to support lattice-based MMI/MPE estimation

Progress in EARS

- Internally at CUED: grand code unification. All EARS-related work now uses a single unified, version-controlled source tree.
- Snapshots of (large parts of) this source tree are regularly released as public HTK3 versions
- First release from this tree was HTK3.2, December '02, new features include:
  - HLM language modelling toolkit
  - support for global feature space transforms (e.g. used for HLDA)
  - Lattice processing tool allows lattice pruning, search and expansion (critical for multi-stage systems)
  - *2-model re-estimation* (important for state clustering)

Progress in EARS

- Release of 3.3 alpha version March '04. new features:
  - redesigned adaptation framework
  - support for constrained/unconstrained MLLR
  - full-variance transforms
  - hierarchies of transforms
  - Speaker Adaptive Training (SAT)
  - updated RM recipe
  - many bug fixes
- hardly any feedback received – we assume everyone is happily using it?!

Summary

- HTK3 project continues to be a success
- Advanced features from CU-HTK version are being released
- As part of EARS more tools and infrastructure will be made available
- Many exciting job opportunities to work on HTK at CUED.