HTK – Past, Present & Future

Gunnar Evermann & Phil Woodland
Cambridge University

May 10th 2001

HTK3 Development Team

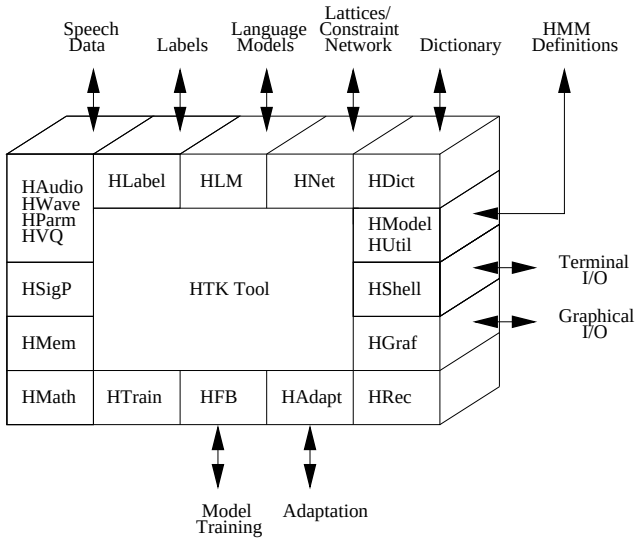
HTK users meeting ICASSP'01

Overview

- Background
- HTK at Cambridge University
- HTK3 project
- Version 3.1
- Future plans

What is HTK?

- Hidden Markov Model Toolkit
- Set of tools for training and evaluating HMMs
- Primarily used in automatic speech recognition
- Also applied to speech synthesis, DNA sequencing and economic modelling
- Modular implementation, (relatively) easy to extend
- Includes 300 page manual



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

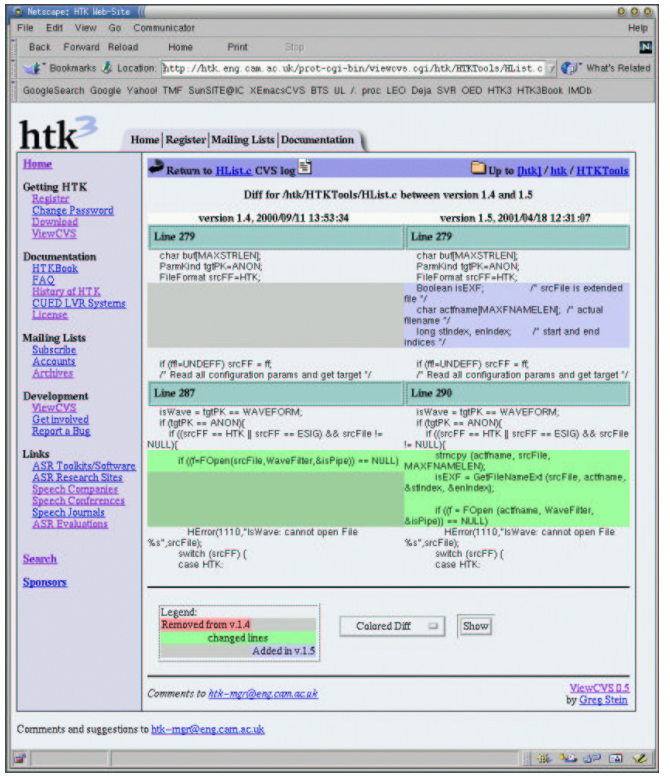
How is HTK used at Cambridge?

- Teaching (one year MPhil course)
- Research (staff & PhD reserch students)
- Building evaluation systems, e.g. 2001 Hub5 (Switchboard) eval:
  - 4 sets of acoustic models: MLE/MMIE + Tri-/Quinphone
  - trained on 265 hours of audio data
  - 4-gram language model + 3-gram class model
  - trained on 200 million words of text
  - 54k vocabulary
  - PLP frontend, VTL normalisation, MLLR adaptation, confidence scores
  - 300x real time, 24.6% word error rate

HTK3

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

- Available since September 2000
- Free of charge, includes full source & HTK book
- Web site got hits from 25k unique IP addresses
- 5000 registered users
- Active mailing lists for users and developers



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
- Build a community of ASR researchers/students

Version 3.1

- Perceptual Linear Prediction (PLP) frontend
- Vocal Tract Length Normalisation (VTLN)
- Improved cepstral mean/variance normalisation
- Better support for unsegmented data
- Many bug fixes

Future Plans

- Tools for word lattice manipulation (rescoring, LM expansion, pruning)
- N-gram Language Modelling tools ($N > 2$)
- Large vocabulary decoder
- Maximum mutual information estimation training
- Inside HTK – Programming Manual

Future Plans II

- Automated regression tests (e.g. RM Recipe)
- WSJ Recipe
- Acoustic models, dicts, LM for standard task (WSJ, Hub5?)
- Word lattices for standard test set (WSJ eval?)
- *your suggestions?*

Summary

- Successful start of HTK3 project
- Advanced features from CU-HTK version get merged into HTK3
- Contribute!