Introduction to HTK (Hidden Markov Model Toolkit)
HTK History

1. Development and Releases of HTK

• In 1989, The first version of the HTK Hidden Markov Model Toolkit, by Steve Young at the Speech Vision and Robotics Group of the Cambridge University Engineering Department (CUED)

• In early 1992, HTK V1.3, Phil Woodland joined Steve as co-developer of HTK. HTK V1.3 were sold by the University (via the University company Lynxvale) at a cost of 450/950 pounds for a source site license for academic sites/companies.

• In 1993, Entropic Research Laboratories (ERL) to take over distribution and maintenance of HTK.

• HTK V1.4D, The first ERL HTK release. included a license manager, and the ability to purchase support. At this time the cost of the software was significantly increased.

• HTK V1.5, in October 1993, Steve and Phil and Bill Byrne

• In 1995, the Entropic Cambridge Research Laboratory (ECRL) was established which was a joint venture between ERL and Cambridge University.
HTK History

1. Development and Releases of HTK (cont.)

- In October 1955, HTK V2.0, Steve and Phil, and Julian Odell, Valtcho Valtchev and Dave Ollason.
- By 1999, HTK V2.2
- Microsoft purchased Entropic in November 1999. Microsoft decided to make the core HTK toolkit available again and licensed the software back to CUED so that it could distribute and develop the software.

- From September 2000, HTK became available in source form at no cost from a CUED Web site (htk.eng.cam.ac.uk) with the intention of further developing it as a speech recognition research platform. The initial release of HTK 3 was based on the final Entropic release, but contained a few minor bug fixes. However the intention is to develop HTK further and provide infrastructure support for investigating state-of-the-art speech recognition, and other sites are encouraged to make available additions to the core HTK functions.
HTK History

2. Major Features of Various HTK Releases

- **Version 1.0**: Initial CUED-internal release. Small amount of documentation. Initial definition of libraries and tools. Support for diagonal and full covariance Gaussian mixture HMMs.
- **Version 1.2**: Added automatic parameter coercion and byte swapping; pruning added to HERest and HVite; support for tied output distributions.
- **Version 1.3**: Arbitrary HMM parameter tying; multiple data streams; MFCC analysis; qualifiers for delta coefs; logical to physical HMM mapping via HMM lists; extensive tracing options added; HHEd created. User, reference and programmer manuals created.
- **Version 1.4**: Sub-word based word recognition supported; faster/smaller word-pair grammars in HParse; tee models for inter-word silence; acceleration coefs; robust state clustering; variance floor macro; X-windows version of HGraf and HSLab added; error codes improved and documented. Support for all features of 1992 CUED Resource Management evaluation system. V1.4D added support for ESPS FEA files and included a license manager.
HTK History

2. Major Features of Various HTK Releases (cont.)

• **Version 1.5:** Master model files (MMFs) and master label files (MLFs) added essential for large-scale systems; forced alignment; parameter file compression; cepstral mean normalisation; addition of the RM recipe. Total documentation ran to 286 pages.

• **Version 2.0:** Major redesign of many library and tools. Documentation via the HTK Book. Support for discrete density HMMs. Complete rewrite of recognition tools using new lattice-based grammar format (HNet/HRec modules). Support for cross-word triphones; lattice and N-best recognition output; and back-off bigram language models. Decision-tree state clustering. Redesigned speech and audio input (HWave/HParm) to support coercion from waveform and real-time audio input. Configuration files.

• **Version 2.1:** HParm partially re-designed and an energy based silence detector included. HNet optimised. Pronunciation probabilities in HVite. Automatic byte swapping for all binary file formats. Support for Microsoft WAV format.

• **Version 2.2:** HEAdapt included for MLLR (mean and variance) and MAP adaptation. HVite also supports adaptation.
2. Major Features of Various HTK Releases (cont.)

- **Version 3.0:** Code based on 2.2 release with minor bug fixes and C++ compatibility. The major changes are the new licensing and distribution arrangements.
- **Version 3.1:** Perceptual Linear Prediction (PLP) frontend implemented; support for Vocal Tract Length Normalisation (VTLN) and cluster-based cepstral mean and variance normalisation added.
- **Version 3.2:** HLM language modelling toolkit integrated. HLRescore lattice post-processing tool added. Support for global feature space transforms. 2-model re-estimation in HERest.
- **Version 3.3:** Adaptation code rewritten and extended, supports MLLR, Constrained MLLR and variance transforms. In addition Speaker Adaptive Training with COnstrained MLLR added, HERest replaces HEAdapt as the tool to generate linear transforms.
- **Version 3.4:** Discriminative training, both MPE and MMI, using HMMIRest added. Code for estimating Semi-Tied and HLDA transform added to HERest. A large vocabulary decoder (HDecode) that supports trigram decoding with cross-wrod triphone models added as an extension to HTK V3.4. HDecode is distributed under a more restrictive license that the main code-base.
3. Other HTK-Related Software

- Entropic produced and sold various products that were related to HTK. These included HAPI (HTK API) which was bundled with HTK in later versions. Other Entropic-produced software included Graphvite (a graphical grammar builder/tester) and Transcriber (a large vocabulary recognition engine and toolkit). None of these are included in the HTK 3 release. At various points in time other HTK software has been produced at CUED and released on a restricted basis. This has included the Lattice Toolkit and a large vocabulary decoder called JRlx written by Julian Odell. None of this software is part of the HTK 3 release although HLRescore and HDecode support much of the functionality.
Structure of HTK for developing a speech recognition

- HLED
- HLSTATS
  - Transcriptions
  - Speech
    - mfcc
    - HSLAB
      - HCOPY
      - HLIST
      - HQUANT

- Data Preparation
  - Training
  - Testing
  - Analysis

- HDMAN
- Dictionary
- Networks
- HBUILD
- HPARSE
- Transcriptions
- HVITE
- HMMs
- HRESULTS
Procedure to generate a speech recognition system

• Prepare the speech data for
  – Define a set of words to recognize
  – Define and record a set of training words - wave file
• Compute Feature Vectors (Hcopy.exe) - mfcc file
• Generate a set of initial hmms for all words to recognize
  – Define a general hmm model(prototype) for a word
  – Compute variances (HCompV.exe) for all the training data
  – Generate a set of initial Hmms for all words to recognize
• Train a set of Hmm(Herest.exe)
• Test recognition rate (Hvite.exe)
• Analyze Result (Hresults.exe)