HTK (Hidden Markov Model Toolkit)

HTK:

- an open-source speech recognition toolkit

- htk.eng.cam.ac.uk

- source language is C

- written for linux, but also compiles under Windows

- it consists of many little executables that can be called from scripts

- manual: https://www.inf.u-szeged.hu/~tothl/speech/htkbook.pdf

- you won't have to compile it, the most important executables are in   
 https://www.inf.u-szeged.hu/~tothl/speech/Numbers-Demo.zip

The main modules of HTK:  
 - preprocessing: HCopy

- training HMM acoustic models: HInit, HRest, HeRest

- creating language models: HParse, HBuild

- evaluation: HResults

We will have to create the following components to perform speech recognition:

- acoustic model (learned from training data)

- language model (given as grammar or learned from training data)

- pronunciation dictionary (prepared manually)

File formats of HTK:

- HTK prefers text file formats, so we can look into the files

- input features: .mfc format (this is binary)

- label files: text format

- pronunciation dictionary: text format

- acoustic model: text (or binary)

- language model: text (or binary)

Preprocessing:

- HCopy.exe -C config input.wav output.mfc

- To preprocess many files, you can pass a file list to it using -S filelist

- example config file: preprocess.config --> MFCC features

Model training:

- if phonetic segmentation is available> HInit.exe, HRest.exe

- see examples runhinit.bat and runhrest.bat

- if only word-level transcripts are available: HeRest.exe (see later!)

Recognition:

- HVite.exe or HDecode.exe

- see examples runhvite.bat and runhvitep.bat

Evaluation:

- HResults.exe

- see examples runhresults.bat and runhresultsp.bat