Disclaimer

- The material provided in this document is not my original work and is a summary of some one else's work(s).
- A simple Google search of the title of the document will direct you to the original source of the material.
- I do not guarantee the accuracy, completeness, timeliness, validity, non-omission, merchantability or fitness of the contents of this document for any particular purpose.
- Downloaded from najeebkhan.github.io

Chapter # 6: Transcriptions and Label Files

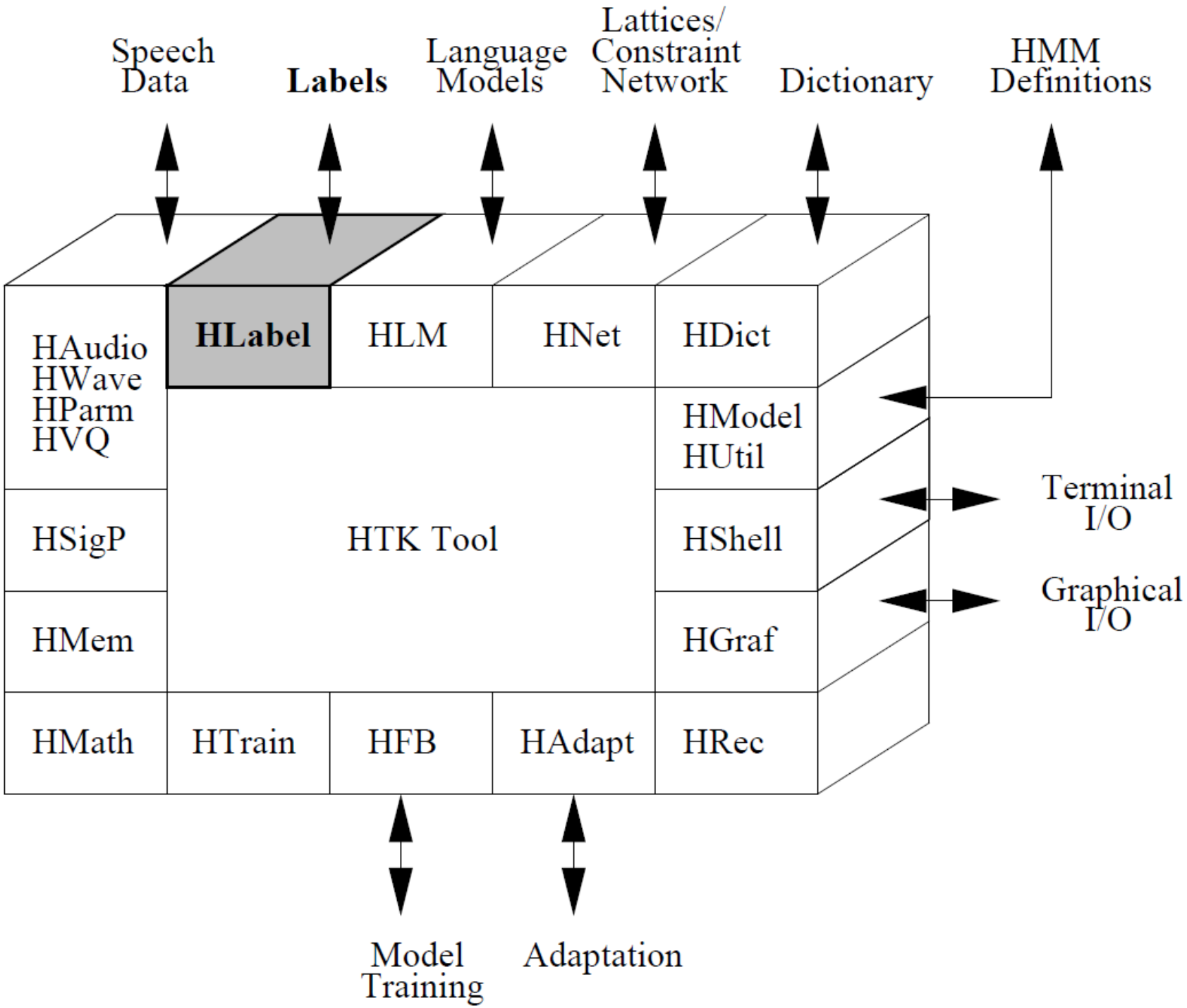
HMM TOOL KIT HTK

Outline

- Introduction
- Label File Structure
- Label File Formats
- Master Label Files
- Editing Label Files

Introduction

- Many of the operations performed by HTK which involve speech data files assume that the speech is divided into segments and each segment has a name or label
- When very large numbers of files are being processing, label file access can be greatly facilitated by using Master Label Files (MLFs)


Speech Data
Labels
Language Models
Lattices/ Constraint Network
Dictionary
HMM Definitions
HAudio
HWave
HParm
HVQ
HLabel
HLM
HNet
HDict
HModel
HUtil
HSigP
HTK Tool
HShell
Terminal I/O
HMem
HGraf
Graphical I/O
HMath
HTrain
HFB
HAdapt
HRec
Model Training
Adaptation

Label File Structure

- Most transcriptions are single-alternative and single-level, that is to say, the associated speech file is described by a single sequence of labelled segments
- In training a HMM system it is useful to have both the word level transcriptions and the phone level transcriptions side-by-side
- All non-HTK formats are limited to single-level single-alternative transcriptions

Label File Structure

ice	cream

(a) 1-alternative, 1-level

ice		cream			
ay	s	k	r	iy	m

(b) 1-alternative, 2-level

I	scream
ice	cream
eyes	cream

(c) 3-alternative, 1-level

Label File Structure

- As with speech data files, HTK not only defines its own format for label files but also supports a number of external formats other than HTK (SOURCELABEL, TARGETLABEL)

HTK Label Files

- The HTK label format is text based
- A single label file can contain multiple alternatives and multiple-levels

```
[start  [end] ] name [score] { auxname [auxscore] } [comment]
```

HTK Label Files

```
0000000 3600000 ice
3600000 8200000 cream
```

```
0000000 2200000 ay
2200000 3600000 s
3600000 4300000 k
4300000 5000000 r
5000000 7400000 iy
7400000 8200000 m
```

```
ice
cream
```

```
0000000 2200000 I
2200000 8200000 scream
///
0000000 3600000 ice
3600000 8200000 cream
///
0000000 3600000 eyes
3600000 8200000 cream
```

Label File Structure

- ESPS Label Files
  - An ESPS/waves+ label file is a text file with one label stored per line
  - Each label indicates a segment boundary

```
time   ccode   name
```

- TIMIT Label Files
  - TIMIT label files are identical to single-alternative single-level HTK label files without scores
  - Start and end times are given as sample numbers rather than absolute times

Label File Structure

- SCRIBE Label Files
  - The HTK SCRIBE format recognizes just three label types

LBA	– acoustic label
LBB	– broad class label
UTS	– utterance

  - The LBA and LBB types have 4 fields: start sample, center sample, end sample and label
  - The UTS type has 3 fields: start sample, end sample and label
  - For use in HTK word blanks are converted to underscore characters

Master Label Files

- To use a training tool with isolated word data may require the generation of hundreds or thousands of label files each having just one label entry
- Each label file must be stored in the same directory as the data file it transcribes, or all label files must be stored in the same directory
- Every HTK tool which uses label files has a -I option which can be used to specify the name of an MLF file

Master Label Files

- MLFs can do two things
  - They can contain embedded label definitions so that many or all of the needed label definitions can be stored in the same file
  - They can contain the names of sub-directories to search for label files

Master Label Files

MLF = "#!MLF!#"
MLFDef { MLFDef }

MLFDef = ImmediateTranscription | SubDirDef

ImmediateTranscription =
Pattern
Transcription
"."

SubDirDef = Pattern SearchMode String
SearchMode = "->" | "=>"

Master Label Files

a.lab contains

```
000000   590000   sil
600000  2090000   a
2100000 4500000   sil
```

b.lab contains

```
000000   990000   sil
1000000  3090000  b
3100000  4200000  sil
```

```
#!MLF!#
"*/a.lab"
 000000   590000   sil
 600000  2090000   a
2100000  4500000   sil
.
"*/b.lab"
 000000   990000   sil
1000000  3090000   b
3100000  4200000   sil
.
```

Master Label Files

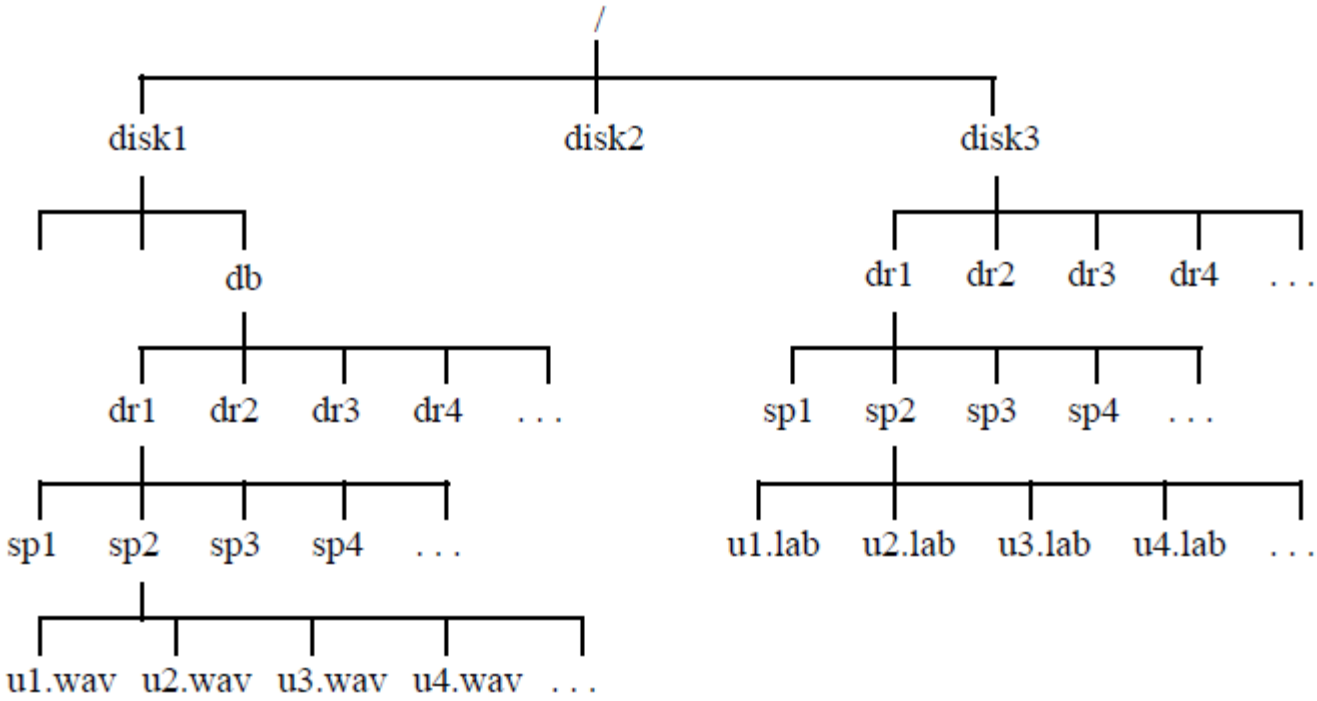
```
#!MLF!#
"*/one.*.lab"
one
.
"*/two.*.lab"
two
.
"*/three.*.lab"
three
.
<etc.>
```

Master Label Files

```
#!MLF!#
"*" -> "/db/dr1/labs"
"*" -> "/db/dr2/labs"
...
"*" -> "/db/dr7/labs"
"*" -> "/db/dr8/labs"
```

```
#!MLF!#
"*/dr1_*" -> "/db/dr1/labs"
"*/dr2_*" -> "/db/dr2/labs"
...
"*/dr7_*" -> "/db/dr7/labs"
"*/dr8_*" -> "/db/dr8/labs"
```

Master Label Files




```
#!MLF!#
"*" => "/disk3"
```

Editing Label Files

- HTK training tools typically expect the labels used in transcription files to correspond directly to the names of the HMMs chosen to build an application
- Hence, the label files supplied with a speech database will often need modifying
- Groups of labels corresponding to a sequence of acoustic events (e.g. pcl p') might need converting to some simpler form (e.g. p) which is more suitable for being represented by a HMM
- The labels in the transcription must be converted to show the required contexts explicitly.

Editing Label Files

- HTK supplies a tool called HLEd for rapidly and efficiently converting label files
- The HLEd command invocation specifies the names of the files to be converted and the name of a script file holding the actual HLEd commands

```
HLEd -l newlabs edfile.led l1 l2 l3
```

- Each edit command stored in an edit file is identified by a mnemonic consisting of two letters

Editing Label Files

```
# Map 61 Phone Timit Set -> 48 Phones
SO
DE q
RE cl pcl tcl kcl qcl
RE vcl bcl dcl gcl
RE sil h# #h pau
```

```
0000 2241 h#
2241 2715 w
2715 4360 ow
4360 5478 bcl
5478 5643 b
5643 6360 iy
6360 7269 tcl
7269 8313 t
8313 11400 ay
11400 12950 dcl
12950 14360 dh
14360 14640 h#
```

```
      0 1400625 sil
1400625 1696875 w
1696875 2725000 ow
2725000 3423750 vcl
3423750 3526875 b
3526875 3975000 iy
3975000 4543125 cl
4543125 5195625 t
5195625 7125000 ay
7125000 8093750 vcl
8093750 8975000 dh
8975000 9150000 sil
```

Editing Label Files

```
ME b bcl b
ME d dcl dh
ME t tcl t
```

Editing Label Files

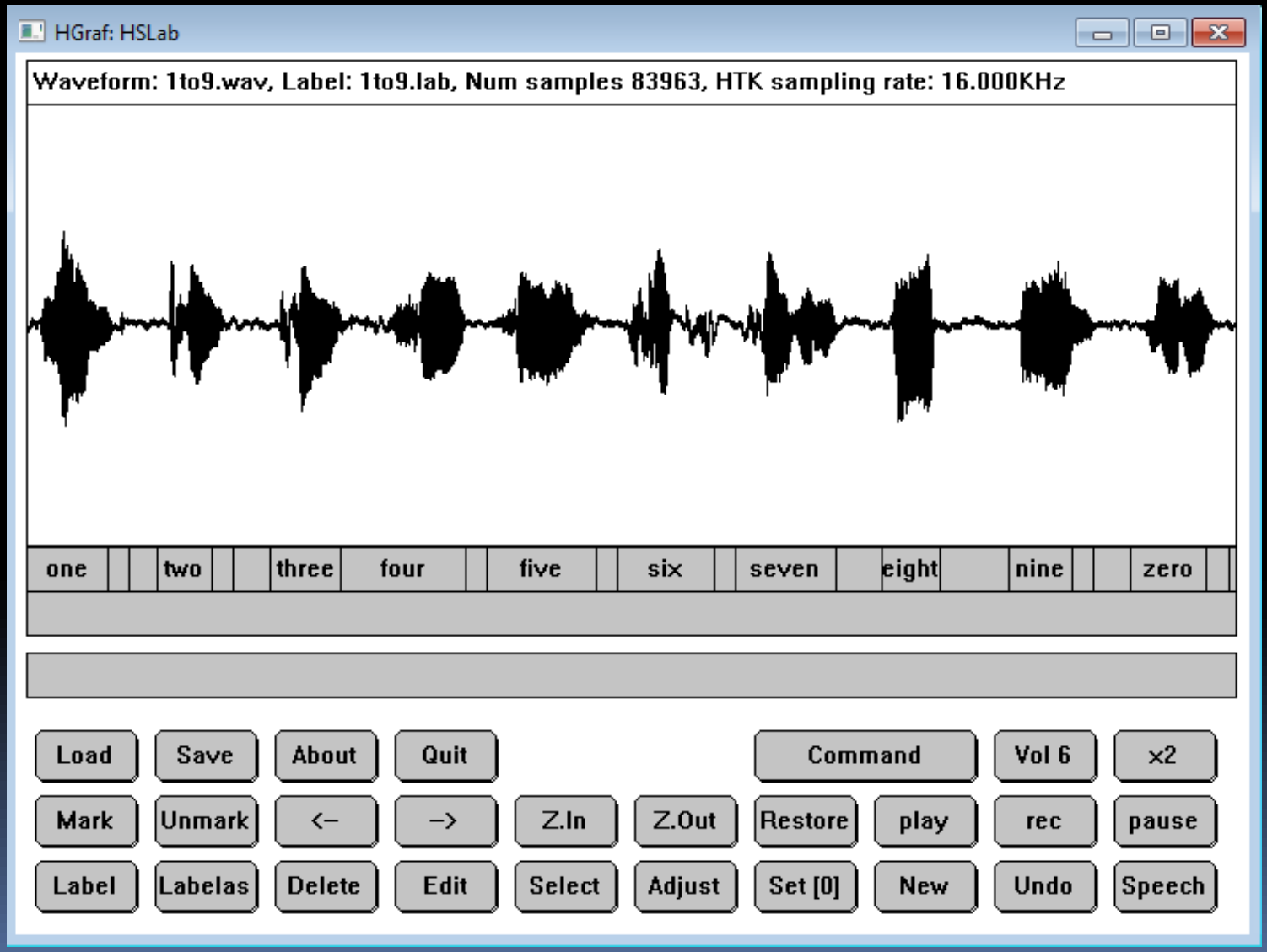
```
DC V iy ah ae eh ix ...
DC C t k d k g dh ...
DC L l r w j ...
DC N n m ng ...
DC S h# #h epi ...
```

```
CH V-ah+V V ah V
CH V-ah+C V ah C
CH V-ah+N V ah N
CH V-ah+L V ah L
 ...
 etc
```

Recognition Output

```
"'*'/1to9.rec"
0 3600000 one -2852.590088
3600000 4500000 six -661.472961
4500000 5700000 silence -526.582642
5700000 8100000 two -1732.673218
8100000 9000000 six -661.718262
9000000 10600000 silence -823.747559
10600000 13700000 three -2177.576904
13700000 19100000 four -3675.707275
19100000 20000000 silence -535.357056
20000000 24800000 five -3246.559570
24800000 25700000 silence -460.714935
25700000 29900000 six -3184.442871
29900000 30800000 silence -488.392578
30800000 35200000 seven -3200.251709
35200000 37200000 silence -1066.845337
37200000 39700000 eight -1636.414307
39700000 42700000 silence -1624.928955
42700000 45500000 nine -2064.114014
45500000 46400000 six -656.805420
46400000 48000000 silence -805.105774
48000000 51300000 zero -2384.838867
51300000 52300000 six -708.313904
```

Recognition Output



Thank You