FEATURE EXTRACTION PROGRAMS FOR SPEECH RECOGNITION

Philip Loizou
University of Arkansas at Little Rock

Installation instructions

After uncompressing the file 'features.zip' using pkunzip, you will see the file 'features.tar'. Un-tar the file 'features.tar' in a Unix workstation by typing:

tar xvf features.tar

This will create a new directory called 'feature'.

Compilation instructions

There is a makefile in the directory 'feature'. To compile the main program 'cparam', just type:

make **cparam**

The makefile is using the standard C compiler 'cc'.

In addition to the 'cparam' program, you may also compile a few other utility programs, such as: **adheader, wav2htk** and **cview** by typing for example: make wav2htk

Running the program 'cparam' for feature extraction

The program **cparam** creates HTK-compatible feature files.

Usage: **cparam [options] infile outfile**

Option		Default
-c	Output cepstral coefficients	LPC
-m	Output mel-freq cepstra coeffs	LPC
-r	Output reflection coefficients	LPC
-t	Output bilinear cepstrum coeffs	LPC
-B	Output FBANK energ(-m should be set)	LPC
-M	Output normalized mel-freq cepstra	LPC
-d	Append Delta Coefficients	Off
-e	Append Log Speech Energy	Off
-g	Append Delta Energy	Off
-a	Use LPC spectrum in mel analysis	Magnitude
-b	Plot LPC and FFT spectrums	Off
-h	Do not output header in code file	On
-q	Do not apply hamming window	On
-j	Compute LFCC coefficients	LPC
-O	Perform octave-bank analysis	LPC

```
-R      Remove global mean from feature vec Off
-A N    Octave (N=0),1/3 octave(N=1)spacing  1
-I F    Initial offset (Hz) in octave anals 0.0
-u N    Process N frames only               All
-n N    Set number of parameters to N       12
-p N    Set analysis  order to N            12
-o N    Set bilinear transform order to N   12
-v N    Set LPC_MEL order to N (with -a opt)12
-i F    High pass freq in mel analysis (Hz) 0.0
-l N    Set cepstrum liftering window to N  24
-x N    Skip the first N frames              0
-f T    Set frame period to T (msecs)      10.0
-w T    Set window duration to T (msecs)   20.0
-s F    Set preemphasis factor to F        0.97
-H frm  Read file format frm (TIMIT,NIST,ISO)HTK
-P      Process frames from a label file     All
```

The ‘infile’ is the input speech file (assumed to be in HTK format), and the ‘outfile’ is the output feature file in HTK format. The program ‘cparam’ also supports other file formats using the –H option, such as the old TIMIT (.adc) files, NIST SPHERE (.wav) format, and ISOlet format (OGI’s).

If you have a speech file in other file format, you may use the utility program ‘**adheader’** to convert the file in HTK format.

Examples:

For illustration purposes, I have included the file ‘speech.wav’ taken from OGI’s Alphadigits corpus. Say for instance that you want to parameterize this speech file into Mel-frequency cepstrum coefficients (MFCC), appended by delta MFCC coefficients assuming a feature base dimension of 12. The command to do that is as follows:

cparam -m –l 0 –d –H NIST speech.wav speech.mfc

The –m option is for computing MFCC features, the –l 0 option is to avoid liftering, the -d option is to create Delta MFCC, and the –H NIST option is for indicating that the file ‘speech.wav’ is an NIST SPHERE file (1024-byte long header). The default feature dimension is 12. You can change that by using the –n P option.
The resulting feature file ‘speech.mfc’ contains the HTK header, and can therefore be used by HTK for training (HInit, HRest, etc.) or recognition (HVite).

The **cparam** program creates nearly identical MFCC coefficients as HTK’s **HCode** program. If you were to use Hcode with the following options, then you would get nearly identical MFCC coefficients:

HCode –m –d –h –k 0.97 –s 1.0 –w 20.0 speech.htk speech-htk.mfc

The 'speech.htk' file is the file 'speech.wav' converted to HTK format using the utility program 'wav2htk' (see description below). The –h option is used for applying a hamming window, the –k 0.97 is for applying a pre-emphasis filter of the form $H(z) = 1 - 097z^{-1}$, the –w 20.0 is for changing the frame size to 20 msecs, and the -s 1.0 is for not scaling the energy term (default scaling = 0.1).

Finally, if you want to create MFCC coefficients (using 24 mel-spaced filters), with delta MFCC coefficients, normalized energy, and delta energy, then type the following:

cparam –m –p 24 –d –g –e –H NIST speech.wav speech.mfcc

The –p 24 option denotes the number of mel-spaced filters to be used in MFCC computation, the –g option denotes the delta energy, and the –e option denotes the speech energy. The resulting feature vector has a dimension of 26 (=12 MFCC + 12 Delta MFCC + Energy + Delta Energy). This feature vector is probably the most popular feature vector used for speech recognition.

UTILITY PROGRAMS

In addition to the **cparam** program, I have also included three other utility programs:

- cview - lists feature files
- wav2htk - converts NIST .wav files to HTK format
- adheader - converts any file to HTK format (waveform type)

CVIEW

This program can be used for listing (or viewing) HTK feature files. It is equivalent to HList in HTK.

```
Usage: cview [options] infile

 Option                                    Default

 -s N    Start at frame N                    1
 -e N    Stop at frame N                     eof
 -n N    Read N numbers at a time            10
 -d      Display 16-bit numbers              Float
 -x      Display in Hex format               Float
 -h      Skip the header                     Off
 -k N    Skip N bytes                         0
 -c      Skip a 4-byte header                Off
```

Example:
If you want to see for instance the feature vectors of the first two frames of the file

'speech.mfc' (created above by cparam using the –m and –d options), the type the following:

cview –h –e 2 –n 24 speech.mfc

The –h option is for skipping the HTK header, the –e 2 option is for listing only up to the 2nd frame, and the –n 24 option is for indicating the feature dimension which is 24 (12 MFCC + 12 Delta MFCC).

If you type the above command you will see the following:

```
Frame [ 1] :
    1. -3.92100   2.  0.32934   3.  0.11431   4. -0.45338   5. -0.66613
    6. -0.53214   7. -0.10043   8. -0.27172   9. -0.12697  10.  0.23583
   11. -0.08274  12.  0.00000  13.  0.54272  14. -0.04120  15. -0.79489
   16. -0.07456  17.  0.31415  18.  0.34557  19.  0.09667  20. -0.03021
   21.  0.18084  22. -0.14627  23.  0.05524  24.  0.00000
Frame [ 2] :
    1. -3.37828   2.  0.28813   3. -0.68058   4. -0.52794   5. -0.35198
    6. -0.18657   7. -0.00376   8. -0.30194   9.  0.05387  10.  0.08956
   11. -0.02750  12.  0.00000  13. -0.07816  14.  0.45321  15.  0.64058
   16.  0.83789  17. -0.04612  18. -0.51061  19. -0.20720  20. -0.23168
   21. -0.10837  22. -0.13441  23. -0.04334  24.  0.00000
```

WAV2HTK

This program converts NIST SPHERE .wav files, which have a 1024-byte long header, to HTK waveform type files.

```
Usage: wav2htk [options] infile outfile

Options                                 Default
-t     Add an ISOLET header               HTK
-l     Add no header                      HTK
```

ADHEADER

This program converts any speech file format to HTK format (waveform type), assuming you know the sampling frequency (-r option) and the size of the header (-s option) of the file.

```
Usage: adheader [options] infile outfile

Options                                 Default
-r F   Set sampling frequency to F     8000 Hz
-t     Add an ISOLET header               HTK
-s K   Skip K bytes                       0
```