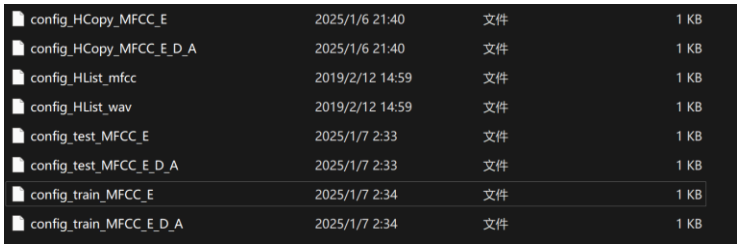
Task 1

According to the requirements of task 1, three files named after the end of MFCC_E_D_A need to be created in the config folder, as shown in the following figure.



In each file, modify the TARGETKIND parameter to MFCC_E_D_A, set the window parameters for Delta and delta-delta, DELTAWINDOW = 3, ACCWINDOW = 2, and make sure after changing the parameters. The MFCC file is 39 dimensional. The modification result is shown in the figure below.

```
TARGETKIND = MFCC_E_D_A
DELTAWINDOW    = 3
ACCWINDOW      = 2
HNET:TRACE  = 2
# next line to read data machine dependent
NATURALREADORDER = TRUE
```

```
TARGETKIND = MFCC_E_D_A
DELTAWINDOW    = 3
ACCWINDOW      = 2
HNET:TRACE  = 2
# next line to read data machine dependent
NATURALREADORDER = TRUE
```

```
SOURCEKIND     = WAVEFORM
SOURCEFORMAT   = WAV
SAVECOMPRESSED = FALSE
SAVEWITHCRC    = FALSE
TARGETKIND     = MFCC_E_D_A
TARGETRATE     = 100000
SOURCERATE     = 1250
WINDOWSIZE     = 250000.0
PREEMCOEF      = 0.97
ZMEANSOURCE    = FALSE
USEHAMMING     = TRUE
CEPLIFTER      = 22
NUMCHANS       = 23 # Number of filter bank channels
NUMCEPS        = 12 # Number of cepstal coefficients
ENORMALISE     = FALSE
DELTAWINDOW    = 3
ACCWINDOW      = 2
LOFREQ         = 64
HIFREQ         = 4000
NATURALWRITEORDER = TRUE
```

In task 1, modify the contents of the ASR_LabDMML_MFCC_E_noCopy.pl file to change all MFCC_E paths to MFCC_E_D_A. In addition, select the proto file named s1d39 and change the NUM_COEF value to 39. Finally, save and run the modified script. The specific modification content is shown in the figure below.

```
#!C:/Strawberry/perl/bin/perl

################################################################################
#
# This script extract features from Aurora2 'Train clean' data and performs training and testing.
# Usage:  perl ASR_LabDMML.pl
# Author: Munevver Kokuer
# Date: Nov. 2018
################################################################################


# Global Variables

$REC_DIR      = "C:/Users/Jingp/Desktop/datamining/LabASR/LabASR";
$BINDIR       = "$REC_DIR/HTK3.2bin";

$HMM_DIR          = "$REC_DIR/hmmsTrained";

$LIST_train     = "$REC_DIR/list/listTrainFullPath_LabDMML_CLEAN1.scp";
$LIST_train_HCopy  = "$REC_DIR/list/listTrainHCopy_LabDMML_CLEAN1.scp";

$LIST_test      = "$REC_DIR/list/listTestFullPath_LabDMML_CLEAN1.scp";
$LIST_test_HCopy= "$REC_DIR/list/listTestHCopy_LabDMML_CLEAN1.scp";

$CONFIG_HCopy   = "$REC_DIR/config/config_HCopy_MFCC_E_D_A";
$CONFIG_train   = "$REC_DIR/config/config_train_MFCC_E_D_A";
$CONFIG_test      = "$REC_DIR/config/config_test_MFCC_E_D_A";

$WORD_LISTSP      = "$REC_DIR/lib/wordList_withSP";
$WORD_LIST        = "$REC_DIR/lib/wordList_noSP";
$LABELSP      = "$REC_DIR/label/label_LabDMML_withSP.mlf";
$LABELS       = "$REC_DIR/label/label_LabDMML_noSP.mlf";
$LABELSTEST       = "$REC_DIR/label/labelTest_LabDMML.mlf";

$ED_CMDFILE1      = "$REC_DIR/lib/tieSILandSP_LabDMML.hed";

$Proto          = "$REC_DIR/lib/proto_s1d39_st8m1_LabDMML_MFCC_E_D_A";
$mod_file       = "$HMM_DIR/hmm7/models";
```

```
$LABELS       = "$REC_DIR/label/label_LabDMML_noSP.mlf";
$LABELSTEST       = "$REC_DIR/label/labelTest_LabDMML.mlf";

$ED_CMDFILE1      = "$REC_DIR/lib/tieSILandSP_LabDMML.hed";

$Proto          = "$REC_DIR/lib/proto_s1d39_st8m1_LabDMML_MFCC_E_D_A";
$mod_file       = "$HMM_DIR/hmm7/models";
$mac_file       = "$HMM_DIR/hmm7/macros";

$flags          = "-p 0 -s 0.0";
$NET            = "$REC_DIR\\lib/wordNetwork";
$DICT           = "$REC_DIR\\lib/wordDict";

$NUM_COEF         = 39;
$PAR_TYPE     = "MFCC_E_D_A";

$RESULT       = "$REC_DIR\\result\\recognitionFinalResult.res";


#open(STDOUT, ">$REC_DIR\\result\\LogASR_LabDMML.log") or die "Can't write to STDOUT $!";


#------------------------------------------------------------------------
# Feature Extraction for mfcc_e_d_a
# HCopy: Calls HCopy to convert AURORA TESTA data files to HTK parameterised 'mfcc_e_d_a'
#------------------------------------------------------------------------

print "Coding data\n";

system("$BINDIR/HCopy -C $CONFIG_HCopy -S $LIST_train_HCopy");
system("$BINDIR/HCopy -C $CONFIG_HCopy -S $LIST_test_HCopy");

print "Coding complete\n";

#------------------------------------------------------------------------
# Training
#------------------------------------------------------------------------
```

```
================== HTK Results Analysis ==================
Date: Tue Jan 07 00:49:22 2025
Ref: >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
Rec: C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
----------------------- Overall Results ----------------------
SENT: %Correct=87.50 [H=168, S=24, N=192]
WORD: %Corr=97.92, Acc=96.99 [H=837, D=2, S=16, I=8, N=865]
============================================================
================== HTK Results Analysis ==================
Date: Tue Jan 07 00:49:42 2025
Ref: >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
Rec: C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
----------------------- Overall Results ----------------------
SENT: %Correct=88.02 [H=169, S=23, N=192]
WORD: %Corr=97.92, Acc=97.11 [H=847, D=3, S=15, I=7, N=865]
============================================================
================== HTK Results Analysis ==================
Date: Tue Jan 07 00:49:55 2025
Ref: >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
Rec: C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
----------------------- Overall Results ----------------------
SENT: %Correct=88.02 [H=169, S=23, N=192]
WORD: %Corr=97.80, Acc=96.99 [H=846, D=5, S=14, I=7, N=865]
============================================================
================== HTK Results Analysis ==================
Date: Tue Jan 07 00:50:012025
Ref: >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
Rec: C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
----------------------- Overall Results ----------------------
SENT: %Correct=88.02 [H=169, S=23, N=192]
WORD: %Corr=97.57, Acc=96.88 [H=844, D=6, S=15, I=6, N=865]
============================================================
```

```
================== HTK Results Analysis ==================
Date: Tue Jan 07 01:50:012025
Ref: >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
Rec: C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
----------------------- Overall Results ----------------------
SENT: %Correct=63.54 [H=122, S=70, N=192]
WORD: %Corr=92.02, Acc=88.67 [H=796, D=9, S=60, I=29, N=865]
============================================================
================== HTK Results Analysis ==================
Date: Tue Jan 07 01:52:112025
Ref: >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
Rec: C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
----------------------- Overall Results ----------------------
SENT: %Correct=68.23 [H=131, S=61, N=192]
WORD: %Corr=91.91, Acc=89.94 [H=795, D=11, S=59, I=17, N=865]
============================================================
================== HTK Results Analysis ==================
Date: Tue Jan 07 01:53:00 2025
Ref: >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
Rec: C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
----------------------- Overall Results ----------------------
SENT: %Correct=68.75 [H=132, S=60, N=192]
WORD: %Corr=91.68, Acc=90.17 [H=793, D=13, S=59, I=13, N=865]
============================================================
```

It can be seen from the result data that the results using Delta and delta-delta features are significantly higher than the results using only static features in terms of sentence accuracy and word accuracy. E_D_A, sentence accuracy = 88.02%, word accuracy = 97.57%; E: Sentence accuracy = 68.75%, word accuracy = 91.68%. By modifying the p-value, the gap between deletion (D) and insertion (I) can be narrowed. It can be seen that the D and I values of Delta and delta-delta features can be controlled in a low range, while the static feature model is difficult to achieve the overall balance. Therefore, it is concluded that the performance of the models using Delta and delta-delta features is much better than that using only static features, especially in terms of word recognition accuracy, reaching a very high level. Moreover, the adjustment of P-values is more effective in Delta and delta-delta feature models, and it is easier to achieve the balance between D and I values, especially in word recognition accuracy, which has reached a very high level. In addition, in Delta and delta-delta feature models, adjusting p-values is more efficient, allowing for more flexibility in achieving a balance between D-values and I-values.

Task 2

In Task 2, use the ASR_LabDMML_3mix_MFCC_E_noCopy.pl file and modify the variables according to the requirements of Task 2. The variables to be modified are $NUM_COEF = 39, $PAR_TYPE = "MFCC_E_D_A", and $num_coef = "MFCC_E_D_A". Select the config script file ending in E_D_A and the proto file of version 39.

```
# Global Variables

$REC_DIR      = "C:/Users/Jingp/Desktop/datamining/LabASR/LabASR/";
$BINDIR       = "$REC_DIR/HTK3.2bin";

$HMM_DIR          = "$REC_DIR/hmmsTrained";

$LIST_train    = "$REC_DIR/list/listTrainFullPath_LabDMML_N1_SNR10.scp";
$LIST_train_HCopy  = "$REC_DIR/list/listTrainHCopy_LabDMML_N1_SNR10.scp";

$LIST_test     = "$REC_DIR/list/listTestFullPath_LabDMML_N1_SNR10.scp";
$LIST_test_HCopy= "$REC_DIR/list/listTestHCopy_LabDMML_N1_SNR10.scp";

$CONFIG_HCopy   = "$REC_DIR/config/config_HCopy_MFCC_E_D_A";
$CONFIG_train   = "$REC_DIR/config/config_train_MFCC_E_D_A";
$CONFIG_test      = "$REC_DIR/config/config_test_MFCC_E_D_A";
```

```
================ HTK Results Analysis ================
Date: Tue Jan 07 01:49:20 2025
Ref: >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
Rec: C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
---------------------- Overall Results ----------------------
SENT: %Correct=85.42 [H=164, S=28, N=192]
WORD: %Corr=98.03, Acc=96.42 [H=848, D=2, S=15, I=14, N=865]
============================================================
=
================ HTK Results Analysis ================
Date: Tue Jan 07 1:49:42 2025
Ref: >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
Rec: C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
---------------------- Overall Results ----------------------
SENT: %Correct=87.50 [H=168, S=24, N=192]
WORD: %Corr=97.92, Acc=96.99 [H=847, D=2, S=16, I=8, N=865]
============================================================
=
================ HTK Results Analysis ================
Date: Tue Jan 07 01:49:53 2025
Ref: >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
Rec: C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
---------------------- Overall Results ----------------------
SENT: %Correct=88.02 [H=169, S=23, N=192]
WORD: %Corr=97.92, Acc=97.11 [H=847, D=3, S=15, I=7, N=865]
============================================================
================ HTK Results Analysis ================
Date: Tue Jan 07 01:50:112025
Ref: >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
Rec: C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
---------------------- Overall Results ----------------------
SENT: %Correct=88.02 [H=169, S=23, N=192]
WORD: %Corr=97.80, Acc=96.99 [H=846, D=5, S=14, I=7, N=865]
============================================================
```

```
================ HTK Results Analysis ================
Date: Tue Jan 07 01:50:012025
Ref: >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
Rec: C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
---------------------- Overall Results ----------------------
SENT: %Correct=89.06 [H=171, S=21, N=192]
WORD: %Corr=98.84, Acc=96.99 [H=855, D=0, S=10, I=16, N=865]
============================================================
================ HTK Results Analysis ================
Date: Tue Jan 07 01:52:112025
Ref: >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
Rec: C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
---------------------- Overall Results ----------------------
SENT: %Correct=89.06 [H=171, S=21, N=192]
WORD: %Corr=98.84, Acc=96.99 [H=855, D=0, S=10, I=16, N=865]
============================================================
================ HTK Results Analysis ================
Date: Tue Jan 07 01:53:00 2025
Ref: >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
Rec: C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
---------------------- Overall Results ----------------------
SENT: %Correct=91.15 [H=175, S=17, N=192]
WORD: %Corr=98.27, Acc=97.92 [H=850, D=4, S=11, I=3, N=865]
============================================================
================ HTK Results Analysis ================
Date: Tue Jan 07 01:53:00 2025
Ref: >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
Rec: C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
---------------------- Overall Results ----------------------
SENT: %Correct=91.67 [H=176, S=16, N=192]
WORD: %Corr=98.38, Acc=98.03 [H=851, D=3, S=11, I=3, N=865]
============================================================
```

From the data comparison, we can see that compared with the single Gaussian model, the third-order Gaussian model has improved the accuracy of all aspects, among which the sentence accuracy is improved by about 1.5%, the word recognition accuracy is improved by about 0.6%, and the overall word-level accuracy is improved by about 0.9%. In addition, both insertion errors (I) and deletion errors (D) were reduced to 3, resulting in a more balanced overall performance. Although the single Gaussian model is simpler, shorter in time and lower in resource consumption, its accuracy is lower than that of the third-order Gaussian model. The third-order Gaussian model performs better in recognition accuracy.

Task 3a.

```
$LIST_train    = "$REC_DIR/list/listTrainFullPath_LabDMML_N1_SNR10.scp";
$LIST_train_HCopy  = "$REC_DIR/list/listTrainHCopy_LabDMML_N1_SNR10.scp";

$LIST_test     = "$REC_DIR/list/listTestFullPath_LabDMML_N1_SNR10.scp";
$LIST_test_HCopy= "$REC_DIR/list/listTestHCopy_LabDMML_N1_SNR10.scp";
```

```
$LIST_train     = "$REC_DIR/list/listTrainFullPath_LabDMML_N1_SNR15.scp";
$LIST_train_HCopy  = "$REC_DIR/list/listTrainHCopy_LabDMML_N1_SNR15.scp";

$LIST_test      = "$REC_DIR/list/listTestFullPath_LabDMML_N1_SNR15.scp";
$LIST_test_HCopy= "$REC_DIR/list/listTestHCopy_LabDMML_N1_SNR15.scp";
```

After performing the above operations, the data obtained is as follows:

```
====================== HTK Results Analysis =======================
  Date: Tue Jan 07 00:39:14 2025
  Ref : >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
  Rec : C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
------------------------ Overall Results --------------------------
SENT: %Correct=68.75 [H=132, S=60, N=192]
WORD: %Corr=93.99, Acc=90.75 [H=813, D=4, S=48, I=28, N=865]
===================================================================
```

```
====================== HTK Results Analysis =======================
  Date: Tue Jan 07 00:49:22 2025
  Ref : >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
  Rec : C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
------------------------ Overall Results --------------------------
SENT: %Correct=81.25 [H=156, S=36, N=192]
WORD: %Corr=96.42, Acc=94.45 [H=834, D=1, S=30, I=17, N=865]
===================================================================
```

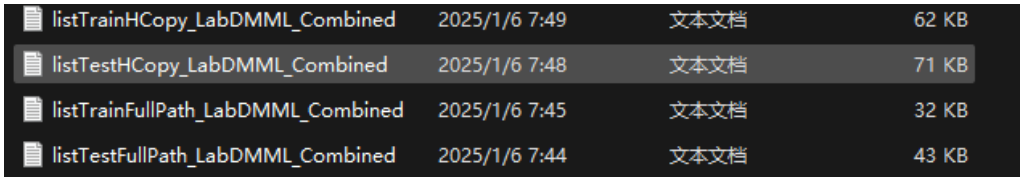
Data analysis in Figures 1 and 3:

Results under N1_SNR10 (signal-to-noise ratio 10dB) : SENT %Correct = 68.75%, WORD %Corr = 93.99%, WORD %Acc = 90.75%. Error analysis: Replace error (S) = 48, insert error (I) = 28, Delete error (D) = 4. Performance: Under the influence of large noise interference, the sentence accuracy decreased significantly, but the word-level accuracy remained at a high level. Results under N1_SNR15 (signal-to-noise ratio 15dB) : SENT %Correct = 81.25%, WORD %Corr = 96.42%, WORD %Acc = 94.45%. Error analysis: Replace error (S) = 30, insert error (I) = 17, delete error (D) = 1. Performance: With the increase of SNR, the accuracy of sentence and word level is significantly improved, and the number of errors is significantly reduced.

Conclusion: Higher SNR is helpful to improve the performance of speech recognition. Under the condition of low SNR, deletion, replacement and insertion errors increase significantly, which indicates that the speech signal is seriously interfered with, and it is difficult to achieve accurate recognition even with the three-Gaussian mixture model.

Task 3b.

Generate two HCopy.combined files for train and test according to the problem requirements. In addition, two fullpath.combined files are generated, each containing paths to all train and test files. As shown in the figure.



Do the following,

```
$LIST_train    = "$REC_DIR/list/listTrainFullPath_LabDMML_Combined.scp";
$LIST_train_HCopy  = "$REC_DIR/list/listTrainHCopy_LabDMML_Combined.scp"

$LIST_test     = "$REC_DIR/list/listTestFullPath_LabDMML_Combined.scp";
$LIST_test_HCopy= "$REC_DIR/list/listTestHCopy_LabDMML_Combined.scp";
```

```
================== HTK Results Analysis ==================
Date: Tue Jan 07 02:53:00 2025
Ref: >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
Rec: C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
------------------------ Overall Results ----------------------
SENT: %Correct=81.20 [H=553, S=128, N=681]
WORD: %Corr=97.18, Acc=94.85 [H=3000, D=5, S=82, I=72, N=3087]
============================================================
```

```
================== HTK Results Analysis ==================
Date: Tue Jan 07 02:53:50 2025
Ref: >s/Jingp/Desktop/datamining/LabASR/LabASR//label/labelTest_LabDMML.mlf
Rec: C:/Users/Jingp/Desktop/datamining/LabASR/LabASR//result/result.mlf
------------------------ Overall Results ----------------------
SENT: %Correct=90.10 [H=173, S=19, N=192]
WORD: %Corr=98.15, Acc=97.69 [H=849, D=4, S=12, I=4, N=865]
============================================================
```

Compared to clean training, mixed training achieves lower sentence accuracy (81.2% vs. 90.1%) and word accuracy (97.18% vs. 98.15%) in a clean environment, indicating that mixing clean and noisy data affects performance in clean conditions. However, mixed training outperforms pure noise training in low SNR (68.75% vs. 81.2% for SNR10), though the gap narrows at higher SNR (81.25% vs. 81.2% for SNR15), showing that low SNR significantly impacts recognition accuracy.

Task 4

```
$LIST_train    = "$REC_DIR/list/listTrainFullPath_LabDMML_Combined.scp";
$LIST_train_HCopy  = "$REC_DIR/list/listTrainHCopy_LabDMML_Combined.scp";

$LIST_test     = "$REC_DIR/list/listTestFullPath_F_N1_SNR10.scp";
$LIST_test_HCopy= "$REC_DIR/list/listTestHCopy_LabDMML_Combined.scp";
```

```
------------------------ Overall Results ----------------------
SENT: %Correct=64.58 [H=62, S=34, N=96]
WORD: %Corr=93.53, Acc=89.73 [H=419, D=0, S=29, I=17, N=448]
============================================================
```

```
$LIST_train    = "$REC_DIR/list/listTrainFullPath_LabDMML_F.scp";
$LIST_train_HCopy  = "$REC_DIR/list/listTrainHCopy_LabDMML_Combined.scp";

$LIST_test     = "$REC_DIR/list/listTestFullPath_F_N1_SNR10.scp";
$LIST_test_HCopy= "$REC_DIR/list/listTestHCopy_LabDMML_Combined.scp";
```

```
---------------------- Overall Results -------------------------
SENT: %Correct=72.92 [H=70, S=26, N=96]
WORD: %Corr=95.76, Acc=92.86 [H=429, D=1, S=18, I=13, N=448]
================================================================
```

Compared to the previous system, the female training model and the hybrid model demonstrated significant improvements in recognizing female voices, increasing sentence recognition accuracy from 64.58% to 72.92% and word recognition accuracy from 93.53% to 95.76%. Additionally, the D and I error values were relatively low, making this new system model more effective for female voice recognition.