Study on Bangla Phonetic Features

Md. All-Mosabbir Rakib
Student Id: 011132050
Sumaia Aktar
Student Id: 011132161
Md. Tarikuzzaman Tarik
Student Id: 011132043
Ikbal Hossain Chowdhury
Student Id: 011132039

A thesis in the Department of Computer Science and Engineering presented
in partial fulfillment of the requirements for the Degree of
Bachelor of Science in Computer Science and Engineering

United International University

Dhaka, Bangladesh

May 2018

Declaration

This is to certify that the work entitled **Study on Bangla Phonetic Features** is the outcome of the investigation and research carried out by following students under the supervision of Dr. Mohammad Nurul Huda, Professor, United International University (UIU), Dhaka, Bangladesh.

_____

Md. All-Mosabbir Rakib
011132050
Department of Computer Science & Engineering (CSE)

_____

Sumaia Aktar
011132161
Department of Computer Science & Engineering (CSE)

_____

Md. Tarikuzzaman Tarik
011132043
Department of Computer Science & Engineering (CSE)

_____

Ikbal Hossain Chowdhury
011132039
Department of Computer Science & Engineering (CSE)

Certificate

This thesis titled "**Study on Bangla Phonetic Features**" submitted by Md. All-Mosabbir Rakib, Sumaia Aktar, Md. Tarikuzzaman Tarik, Ikbal Hossain Chowdhury. Student ID: "011132050", "011132161", "011132043", "011132039", has been accepted as Satisfactory in fulfillment of the requirement for the degree of Bachelor of Science in Computer Science and Engineering on April 2018.

_____

Dr. Mohammad Nurul Huda
Professor and MSCSE Coordinator
Department of Computer Science & Engineering (CSE)
United International University (UIU)
Dhaka, Bangladesh

Abstract

Speech recognition is one of the most important computer technologies for Automatic Speech Recognition (ASR) model. ASR model have to enables device for recognize and very better known as Bangla word. ASR model check sound and digitizing and also matching the word pattern for stored patterns. We have to see after recognize normal speech not better than discrete speech. Now, we can detect continuous speech (Normal Speech) for used a system. Its helps people about working for disabilities.

We used a system which is new approach for buildup in top position Bangla ASR model. Its MFCC based system used for better word recognition performance. Maximum cases used in small no. of speaker, but Bangladesh most of the wide area are used in native language. We can experiments about this language as a inputs are Mel-Frequency Cepstral Coefficients (MFCCs) and result will be based on Hidden Markov Model (HMM) classifier for word recognition performance.

We can experiments about other ASR models, firstly we used MFCC-39 based classifier and now we can used Articulatory features AF-22 based classifier for male and female voice. Then, we can see word correct rate and accuracy will be much better than MFCC-39 based classifier. That's why our suggest system help for detect gender voice and independent fact.

Acknowledgement

At first, we are highly grateful to Almighty Allah, the sustainer of the Universe, who has provided us with all kinds of opportunities to complete the thesis program.

Individual efforts alone can never contribute in totality to a successful completion of any venture. We would be failing in our duty if we did not state our gratitude and appreciate to the following that who have made valuable contribution to this report.

We are ever grateful to our supervisor, Dr. Mohammad Nurul Huda, United International University (UIU), Dhaka, Bangladesh, for his wonderful supervision, great motivation, great guidance and his valuable time. Also, he helps us choose good topic, helps for research and completing thesis.

Now, we would like to our families and also friends for appreciable their assistance and great suggestions during this thesis.

Dhaka

May 2018

Md. All-Mosabbir Rakib

Sumaia Aktar

Md. Tarikuzzaman Tarik

Ikbal Hossain Chowdhury

Table of Contents

DECLARATION……i

CERTIFICATE……ii

ABSTRACT……iii

ACKNOWLEDGEMENT……...iv

LIST OF TABLES...……...vii

LIST OF FIGURES…….....viii

1. Introduction.....1

1.1 Contribution……1

1.2 Focused on Bangla speech Recognition……...…….…1

2. Automatic Speech Recognition 2

2.1 History…… ……2

2.2 Basics of Speech Recognition……...……...…….…3

2.3 Varies of Speech Recognition 5

3. Bangla Phonetic Features 6

3.1 Bangla Phonemes Scheme ……6

3.2 Bangla IPA Table……..9

3.3 Bangla Phoneme Schemes……9

4. HMM-based Classifier.....11

4.1 About HMM……11

4.2 Modeling of HMM……...11

4.3 About HMM basic problem…….……12

4.4 Solutions of the HMM problems..……...12

5. Experiments and Results.....16

5.1 Performance Figures……16

5.2 Speech Database……..16

5.3 Experimental Setup……..……17

5.4 Experimental Results and Discussion…….……20

6. Conclusion 23

7. References.....24

LIST OF TABLES

Table 3.1: Consonants Bangla IPA Table.....8

Table 3.2: Vowels Bangla IPA Table.....9

Table 5.1: Phoneme Correct Rate 21

Table 5.2: Phoneme Accuracy 21

Table 5.3: HTK Results Analysis for Train 22

Table 5.4: HTK Results Analysis for Test.....22

LIST OF FIGURES

Figure 2.1: MFCC file Extractor 3

Figure 2.2: MFCC 39Extractor.....4

Figure 2.3: Schematic architecture for simplifies speech recognition.....5

Figure 3.1: Bangla IPA table 9

Figure 5.1: MFCC Feature Extraction.....16

Figure 5.2: MFCC Feature Extractor with Neural Network.....17

Figure 5.3: Train Model10.....18

Figure 5.4: Mix2 Model10 in Train.....18

Figure 5.5: Mix4 Model10 in Train.....19

Figure 5.6: Mix8 Model10 in Train.....19

Figure 5.7: Mix16 Model10 in Train.....20

Figure 5.8: Mix32 Model10 in Train.....20

Chapter 1

Introduction

In this world, all language has Automatic Speech Recognition (ASR) systems. Bangla language are too slowly and too little performed for kind of any other language. Its performance is too little because Bangla language lacking is it's not a proper way for speech accumulation. That's why Bangla language is very slow for word recognition performance. This issue solved or developed by used MFCC based system. MFCC based system generate text to speech. Then solved Bangla recognize lacking. This language develops efforts by Indian because their country one part West Bengal spoken by Bangla language. Also, Bangladesh full country spoken by Bangla language. They are two countries spoken, written are same Bangla language, same sound as voice, same characters but sometime uses some different variables, different pronunciations. They are uses standards Bangla language. So, they would need more research about Bangla ASR system like Bangladesh spoken Bangla language.

Bangla Automatic Speech Recognition (ASR) model uses some procedure or following some system which is works on speech recognition. We use a Mel-Frequency Cepstral Coefficients (MFCCs) based classifier for as a input and use Hidden Markov Model (HMM) based classifier for get better word recognition performance.

We also used Articulatory Features (AF) after MFCC use. Articulatory Features (AF) uses by neural network for 3000 train and 1000 test data for male and female voice. Then we can see after AF-22 used we have get better word correct rate and word accuracy.

ASR system main focuses on distribution frequency of separately constants and vowels for get a better result about independent male and female voices. So, we are research mainly about Bangla ASR system for give us better performance.

1.1 Contribution

This researches are lots of contributes about Bangla ASR model. We have used two types of system. This two system helps for us better accuracy for speech recognition. Hidden Markov Model (HMM) based classifier which is convert text by MFCC-39 based system. Second purpose is Articulatory Features (AF) based classifier. AF-22 works by neural network. This two purpose generate for a better word correct rate and word accuracy.

1.2 Focused on Bangla speech Recognition

Basically we have to focused on some main point that the weakness for Bangla speech recognition. Firstly, we focused on some vowel and consonants which is lack in Bangla standards language. Then, we focused on word correction rate and performance rate. Finally, we some mixture components are reduction for a better result.

Chapter 2

Automatic Speech Recognition

Automatic Speech Recognition (ASR) system is used in computer. It's also used in computer hardware based system and software based system which is located or identifies any kind of voice. Automatic Speech Recognition (ASR) also used in text to voice conversion. ASR system also works on voice to speech recognition. Automatic Voice Recognition (AVR) uses in Automatic Speech Recognition.

Automatic Speech Recognition (ASR) system basically works many type of speech recognition. ASR system uses in many kind of technology which is works on speech to voice recognition, voice to speech recognition etc. Basically, ASR is just a system human train; authenticate speech with some Bangla phoneme language. Human uses character for speech recognition. Automatic voice recognition very helpful for voice to speech recognition. Automatic speech recognition actually working on ready or saved voice which is early collected for user. Then ASR train 3000 data and test 1000 data with MFCC-39. We have to see word correct rate and word accuracy. Now, we train 3000 and test 1000 data with Articulatory Features (AF-22) which was works in neural network. So, we can saw that much better result word correct rate and word accuracy.

2.1 History

Speech Recognition basically thinking many years earlier. In 1940 first research goes to about speech recognition, when nothing to do anything about speech input. Then researchers are three Bell Labs, they are thinking speech recognition are need to develop. They are works on firstly one digit in Speech recognition and they developed first digit in 1952 World's Fair at New York.

In 1960s, many researchers are focused on speech recognition. They can be starts a small Goals then they can do the larger develop in speech recognition system. This time they create a device for test in speech recognition but that is works on small cases. In 1970s established continuous speech recognition. It's not working perfectly for user to speech recognize. This research continuous for more develop and this is something functionally working in 1980s.

Now successfully speech recognition introduces by 1990s and this time research more effetely automatic speech recognition because human can do his/her voice convert to speech or speech to voice conversion also. Automatic speech recognition is must be developed and identity for real human speech.

In 2000s Automatic Speech Recognition (ASR) system successfully by Hidden Markov Model (HMM) and then voice to speech recognition is possible now. But, as more research and developed are needed for more accurately identify.

2.2 Basics Speech Recognition

Speech recognition is the main process to communicate with machine or computer system for identify human voice. Human voice perfectly identify or known as only used by speech recognition system. Speech recognition uses Hidden Markov Model (HMM) for human voice to word or speech recognition. Computer is the main process for accurately speech recognize. We need basically how to recognize for human voice to speech or word. The main solution are following figures 2.1 showing,

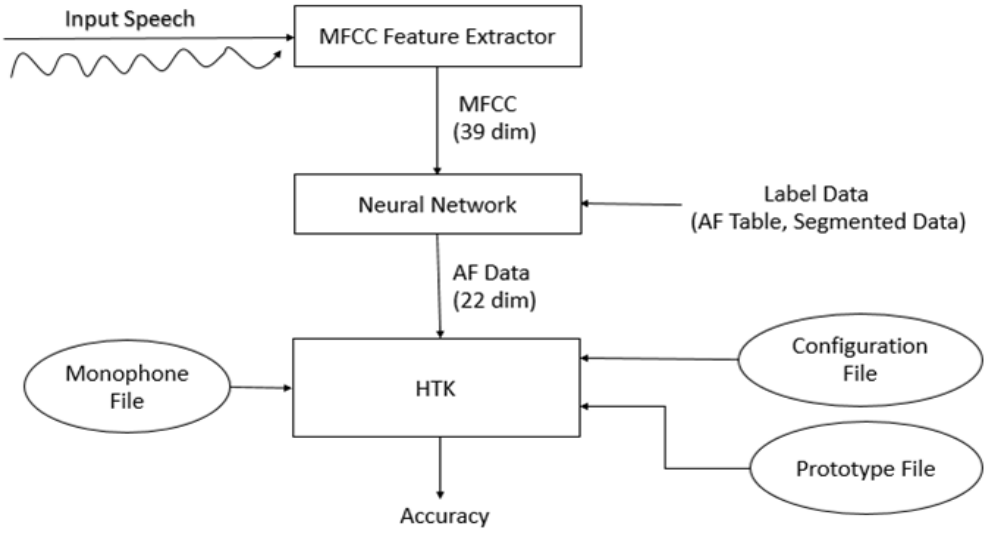

Figure 2.1 MFCC file Extractor

Firstly, we can take input as well then Mel-Frequency Cepstral Coefficients features will be extracts here. Now we can take label data MFCC-39 dim with neural network then AF-22 dim will be experiments with monophone file, configuration file and prototype file to HTK. Finally, we can saw that HTK give us accuracy. We can see MFCC file extractor in Figure 2.2,

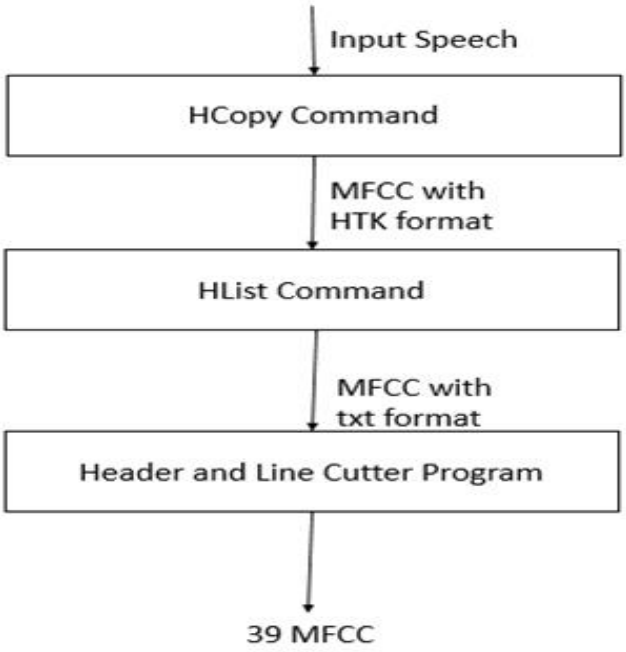

Figure 2.2 MFCC 39Extractor

Accuracy

The experiments will be giving us accuracy and it will be a long process for this research about Bangla speech recognition. Accuracy depends on how accurately voice to word or word to voice converts. System gives us accurate performance than accuracy will be better. We experiments by MFCC based system and AF based system then we can saw that AF based system will much better result than MFCC based system also AF gives better word correct rate.

Training

At the recognize process train is the most important things. If the train is not good then test is never gives better result. So, find out train data should be perfect. Our research will be 3000 train data and we can start experiments with MFCC based system with HMM. This data will give us better train result now we can test side with 1000 data and result will be perfect but some cases error. Then we write normalize code and firstly data normalize and then again train with Articulatory feature (AF) which is works by neural network. AF system train with 3000 data and test with 1000 data. We can saw that AF gives result much more better than MFCC based system. ASR system human voice to word recognition is perfectly.

Speech recognition is the process for human voice perfectly convert into word or speech. Bangla characters will be added in maintenance accurate sequence. When system are get right characters then system gives accurate voice and it’s decide main user voice. ASR system are not get sequent voice or characters and its doesn’t give us correct training result.

2.3 Varies of Speech Recognition

Speech Recognition is following some schematic architecture for simplified speech recognition. So, we shown figure 2.3 schematic architecture,

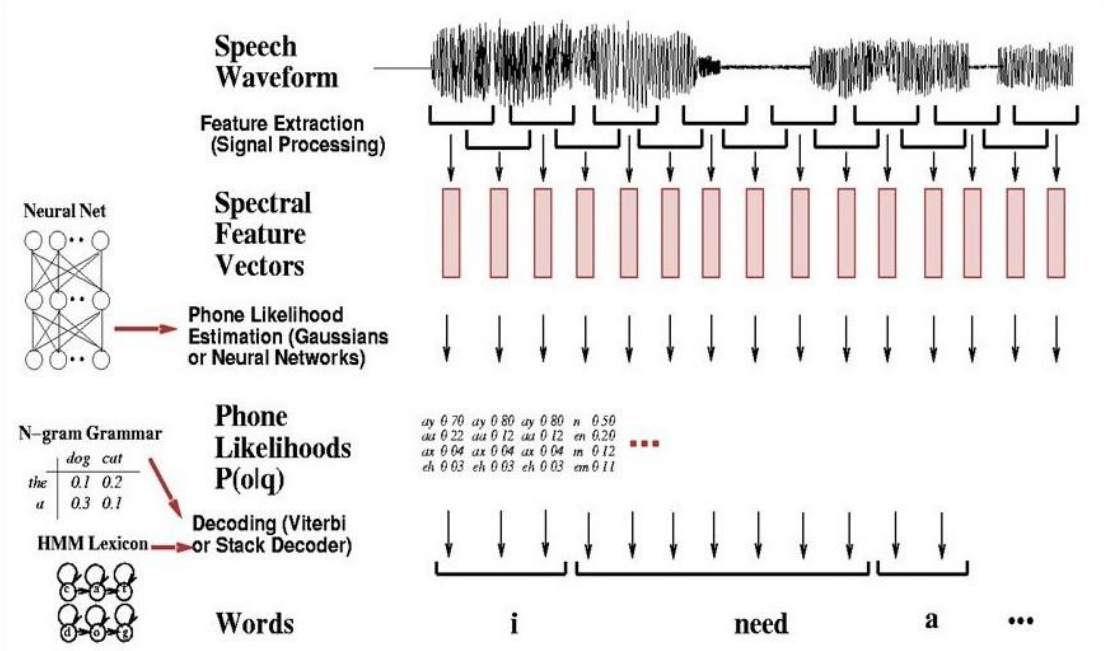

Figure 2.3: Schematic architecture for simplifies speech recognition

Schematic architecture basically works on some speech waveform and by signal processing feature extraction.

Speech recognition way to two types, i) Speaker-dependent ii) Speaker-independent

Speaker-dependent

Speaker-dependent system only works and learn a single human person and collected characters for single user. When other user come and gives voice then he firstly training then computer takes characters. It's more accurate because this system only taken a real user voice. So, result will be better and accurate.

Speaker-independent

Speaker-independent system works on anyone's characters and its system means only for real user directly can't ask to the caller before read text. Speaker-independent system are slow process because user can't get any voice when system can read and process the characters. So, its not's a better option.

So, We have research and saw that speaker-dependent software system much better than speaker-independent software system.

Chapter 3

Bangla Phonetic Features

3.1 Bangla Phonemes scheme

This Bangla Phonetic consists of 14 vowels and 7 nasal vowels and 29 consonants. This Bangla Phonetic Constructive Scheme given an IPA table, in this IPA table only 7 vowels sound are showed. There exist two more long sound of /i/ and /u/ on this sound denoted by /i: / and /u:/.The two long vowels are pronounced short fellow in modern Bangla.

Autochthonous Bangla Phonetic Features words do not permit basic clusters; the maximum syllable structure is CVC. Bangla phonetic words wide range of clusters expanding minimum composition process. English or other foreign response more clusters type into Bangla phonetic features item.

Consonants

IPA	Bengali	Examples	Approximate English Equivalent
b	ব	**বালি**	A**b**ash
bʱ	ভ	**ভী**রু	A**bh**or
d̪	দ	**দা**ম	**Th**e
d̪ʱ	ধ	**ধা**রা	Wi**th**out
ɖ	ড	**ডু**মুর	Guar**d**
ɖʱ	ঢ	**ঢো**ল	Guar**dh**ouse
dʒ	জ~য	**জল যব**ন	He**dg**e
dʒʱ	ঝ	**ঝা**উ	He**dgeh**og

f	ফ	**ফা**রাক	**Ph**antom
g	গ	**গ**ম	A**g**ate
gʱ	ঘ	**বা**ঘা	Pi**gh**eaded
ɦ	হ~ঃ	**ম**হান	**H**ead
k	ক	**ক**লম	S**c**an
kʰ	খ	**শা**খা	**C**an
l	ল	**মা**লিনী	**L**eaf
m	ম	**মা**নুষ	**M**uch
n	ন~ণ~ঞ	**নী**লা বা**ণ**	**N**ot
ŋ	ঙ~ং	বা**ঙা**ল	Ba**n**k
p	প	**পা**কা	S**p**an
pʰ	ফ	**ফ**ল	na**ph**tha (in this case, ph. is not pronounced as [f])
r	র	বি**রা**ম	trilled *r*
ɽ	ড়~ঢ়	বে**ড়া** দৃ**ঢ়**	US: lar*d*er
s	স	**সে**লাম	**S**ue
ʃ	শ~ষ~স	**ম**শা **ষা**ট **স**ময়	**Sh**oe

t̪	ত~ৎ	**ত**মাল উৎস	**Th**ink
t̪ʰ	থ	**থা**বা	**Th**ought
ʈ	ট	ব**ট**	A**rt**
ʈʰ	ঠ	কা**ঠু**র	Ar**t**ist
tʃ	চ	মু**চি**	Ca**tch**
tʃʰ	ছ	**ছা**গল	**Ch**oose
z	ज़	নমা**জ়**	**Z**ip

Table 3.1: Consonants Bangla IPA Table

Vowels

IPA	Bengali	Examples	Approximate English equivalent
a	আ—পা	**আ**ঙিনা	F**a**ther
ɛ	অ্যা—প্যা ~ এ—পে	**ত্যা**গ **খে**লা	B**e**t
e	এ—পে	**মে**দিনী	B**e**d
i	ই—পি ~ ঈ—পী	**দি**ন **দী**ন	Ver**y**
o	ও—পো ~ অ—প	**ও**জন **অ**ঙ্কুর	S**o**le

ɔ	অ—প	**অ**মর	S**o**rt
u	উ—পু ~ ঊ—পূ	**কু**ল **কূ**ল	L**oo**k
~	ঁ	**ইঁ**দুর	Nasal vowel ([ã], [õ], etc.)
◌̯	ই~উ~ও~য়	ঝা**উ** খা**য়**	semivowel ([i̯], [u̯], [o̯] and [e̯])

Table 3.2: Vowels Bangla IPA Table

3.2 Bangla IPA Table

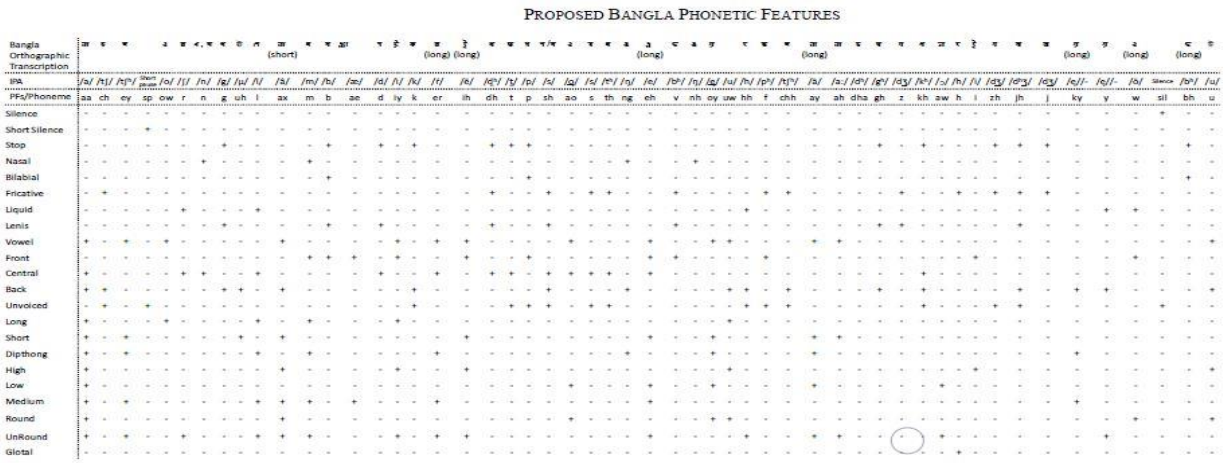

Figure 3.1: Bangla IPA table.

3.3. Bangla Phoneme Schemes

Bangla Phonetic features and IPA (International Phonetic Alphabet) table for Bangla is related. Bangla words using by tri phone model based on HMM features to resolve in Bangla words.

At present the real problem to do Bangla Phoneme is need to proper Bangla speech corpse. In fact, such a corpse is not obtainable at least one referenced in any of the existing literature. Bangla Speech remains medium size of evolved.

The male training remains 30*100 sentences. On the other hand, 3000 male voice same sentence uttered by 30 female voice are training corpse. Hundred interrogations from the daily newspaper and under the thirty male pedagogue of dissimilar district of Bangladesh to find the male training corpse.

On the other hand, 100 sentence from the same newspaper spoken 10 different male lecturers and 10 different female lecturers are used to male test and female test lectures are from Bangladeshi nationals and local lectures. This Bangladeshi lectures ages are ranges from 20 to 40 years.
We have chosen a wide area of Bangladesh from Dhaka central region. There are different region from chosen the areas West region, North region, North-West region, North-East region, East region to speak in standard of Bangla.

We found invalidate place of voice where ceiling fan and air cardiopulmonary-exercise were switch on some level and noise could on the heard. Software was used to document to voice checked and former voices.

Speech Signal are extended to Acoustic Feature Extraction 39 dim and 12 MFCC and 39 dim P. 12 del MFCC and del P are conventionality progress of ASR uses where P queue for raw energy input speech signal and method of MFCC feature extraction for HMM features. The value of pre emphasis factor is 0.97.

Chapter 4

HMM-based Classifier

4.1 About HMM

HMM this features model basically phonemes structures of three left to right states and a simple left-to-right topology as that features illustrated. To move joining of the models unitedly the entry and exit states are toggle. The exit state of one phoneme model can be connected with the entry state of another to form a structure for this phonemes are connected to the features. This phonemes models to be joined together to form words and words to be joined together to cover complete features and this particular features predication.

4.2 Modeling of HMM

Generally, this phoneme is specified by a five-tuple of this features:

(S, O, η, A, B)

(1) S = {1,2, …,N}, Set of clandestine states

N: the number of states of this feature.

Start the number of states to begin the process and mid line that five tuples on number of states and standard HMM phoneme model.

And that the features to start become the first states to standard honor HMM phonemes model to same process using the second process to Gaussian mixtures.

Si: the state at time t,

(2) O = {o_1,o_2,…..o M}, set of observation symbols

M: the number of observation systems to that features

(3) η =$\{ \eta_i \}$ η_i = P ($S_{0 = I}$) the elemental state distribution

(4) A = $\{a_{ij}\}$ a_{ij} = P (Si – jI Si = j) elemental states to distribution

Observance symbol probability distribution in state.

To revenue up, absolute specification of an HMM exclude:

(i) Two constant-size parameters: N and M (representing the total number of states the size of observation symbols)

Three sets of probability distribution: φ = (A, B, ɳ)

4.3 About HMM basic problem

Hidden Markov Model (HMM) is very important for Bangla speech recognition. HMM works really good all the time when it's works on MFCC and AF. But, something is wrong sometime actually HMM facing three major problems,

i. Firstly, facing problem evaluation when its evaluate probability for human voice then we can saw that result will never positive.

ii. Secondly, sometime it's facing decoding problem. It never is control characters state sequence.

iii. The third one is problem on learning and that's why some parameters are maximization. So, no adjustments are possible.

4.4 Solutions of the HMM problems

We can saw that HMM system facing three major problems for speech recognition. So, those problem are in solutions is normalization. Than we learn and write a normalization code which is shown as down some lines. After normalize the three problem is solved.

Normalization code

```
Void WriteWave(double **vec, char *fname, int fnum, int ch);

void main(void){

        int n;

        HISTOGRAM*hist;

        char fname[MAX_FNAME]={'\0'};

        char Path[MAX_FNAME]={'\0'};

        //char LPath[MAX_FNAME]={'\0'};

        char temp[6]={'\0'};

        char tmp[10]={'\0'};

        //char phone[6]={'\0'};

        int fnum=0;

        FILE *fp,*fp1,*fp3;

        double data;
```

```
        int k,i;
        //long x,y;
        k=1;
        printf("Program is running.....\n");
       fp1=fopen("Out_Regenerate_Vec.test","r");
        if(fp1==NULL){
                                printf("\nCan not open input file");
                                exit(1);
        }
        rewind(fp1);
        fp3=fopen("Test.scp","w");
        if(fp3==NULL){
                                printf("\nCan not open script file");
        exit(1);
        }
        rewind(fp3);
        if((hist=(HISTOGRAM*)malloc(sizeof(HISTOGRAM)))==NULL){
                        printf("\nCan not allocate memory for histogram");
                        exit(1);
                }
        memset(hist,0,sizeof(HISTOGRAM));
        if ((hist->peaks = (double **)calloc(MAX_FRAME_NUM, sizeof(double *))) ==
NULL)
{
                                printf("Cant allocate memory for hist->peaks\n");
                                exit(1);
        }
```

```
for(n = 0; n < MAX_FRAME_NUM; n++){
        if ((hist->peaks[n] = (double *)calloc(BIN, sizeof(double ))) == NULL){
                                printf("Cant allocate memory for hist->peaks\n");
                                exit(1);
                }
}
while((!feof(fp1))){
        strcpy(fname,"");
        strcpy(temp,"");
        strcpy(fname,"Train\\file");
        itoa(k,temp,10);
        strcat(fname,temp);
        strcat(fname,".vec");
      putc('\n',fp3);
        fprintf(fp3,"%s",fname);
        fnum=0;
        strcpy(Path,"");
        fscanf(fp1,"%s",Path);
        printf("\nInput File::%s",Path);

        fp=fopen(Path,"r");
        if(fp==NULL){
           printf("\nCan not open input Data");
          exit(1);
        }
        rewind(fp);
        n=0;
```

```
            while(!feof(fp)){
                for(i=0;i<BIN;++i){
                  fscanf(fp,"%lf",&data);
                  hist->peaks[n][i]=data;
                }
                ++n;
            }
            fnum=n-1;
            WriteWave(hist->peaks,fname, fnum, BIN);
            fclose(fp);
            ++k;
        }
        fclose(fp1);
        fclose(fp3);
        return;
}
```

Chapter 5

Experiments and Results

HMM Performance of different types the results were analyzed benefit features of results. First we overall insertion deletion and substitution fault were accuracy and correctness from these.

5.1 Performance Figures

Progress of ASR substructure uses of MFCC conventionality of breadth 12 MFCC where input speech signal and method of MFCC feature extraction HMM window of 0.25 ms is used extraction features. The value of P emphasis factor 0.97.

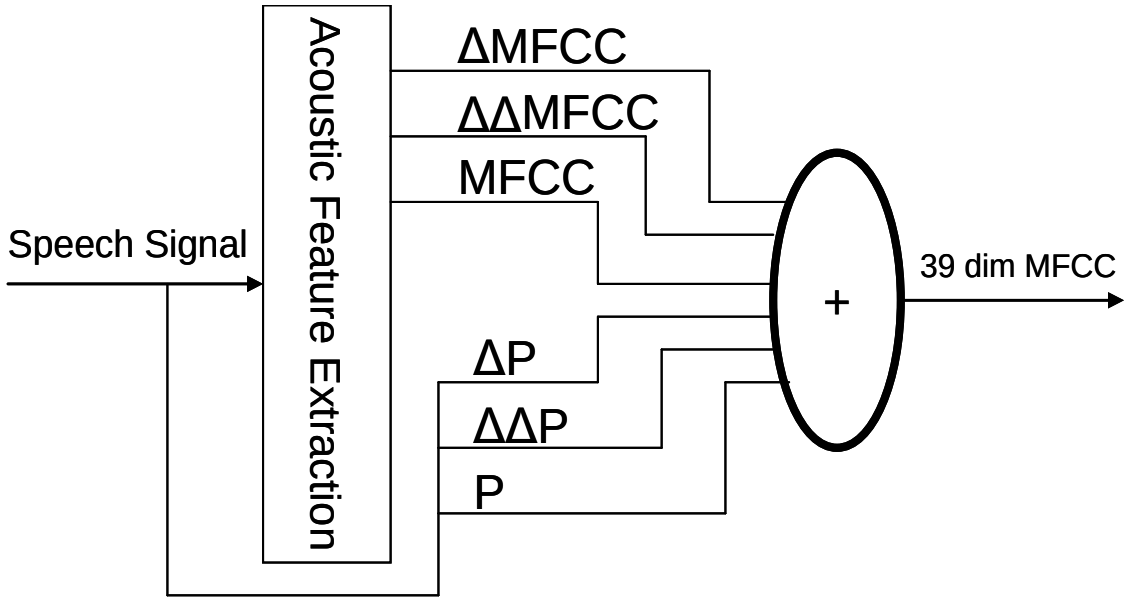

Figure 5.1: MFCC Feature Extraction.

5.2 Speech Database

At the feature extraction stage, the input function of first converted into AFs perform spectrum along the time and frequency axes. Linear defence to applying three-point along the frequency. That frequency to time capability pattern example to LFs to input features. After compressing these two AFs with 24 dimensions and 12 dimensions use to DCT to P stands for long power of raw speech signal vector called AF. After that linear function using discrete cosine transform to 25 dimensions.

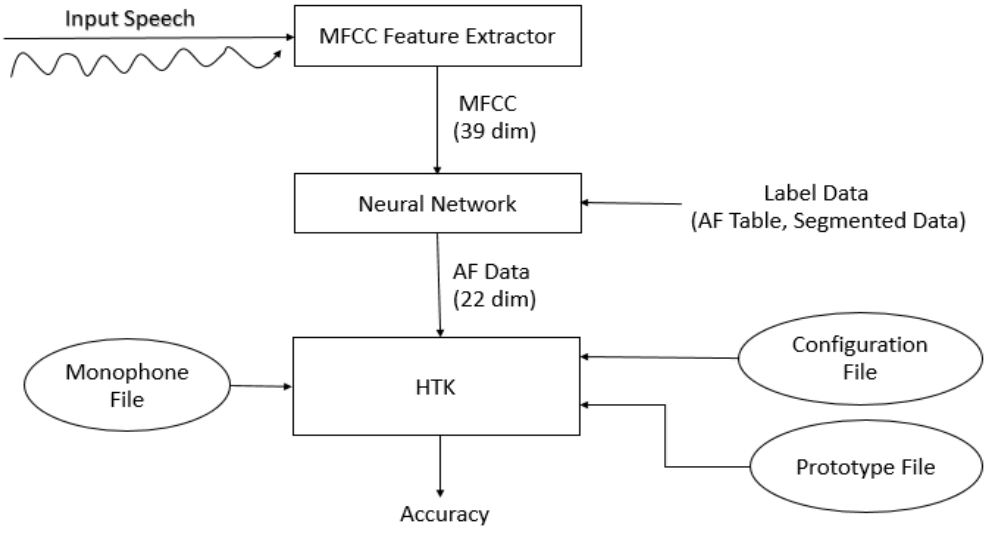

Figure 5.2: MFCC Feature Extractor with Neural Network.

5.3 Experimental Setup

The shortness of frame are rate of set 25 ms and 10 ms to this frame shift between two disordered frames according to MFCC features to extract acoustic features. Articulatory features from an input features. MFCC and AFs compare to 39 and 25 dimensional vector.

For the perfect sentence recognizer and word correct rates and word exactitude and sentence correct poor rates for data set are revaluated using HMM based classifier 39 dimension MFCC and 25 dimension AFs to research. That matrices are used diagonal. The Bangla features male and female data set are design for tri phones HMMs with five states, three loops, left to right models. Input features of that data set are correct rate extraction.

For evaluating data set to performance to different method to incorporate the functional method we have to designed,

Experiment-I

(a) MFCC (Train: 3000 Male, Test: 1000 male + 1000 female).
(b) AF (Train: 3000 Male, Test: 1000 male + 1000 female).

Experiment-II

(a) MFCC (3000 Female, Test: 1000 male + 1000 female).
(b) AF (3000 Female, Test: 1000 male + 1000 female).

Experiment-III

(a) MFCC (Train: 3000 male + 3000 Female, Test: 1000 male + 1000 female).

(b) AF (Train: 3000 male + 3000 Female, Test: 1000 male + 1000 female).

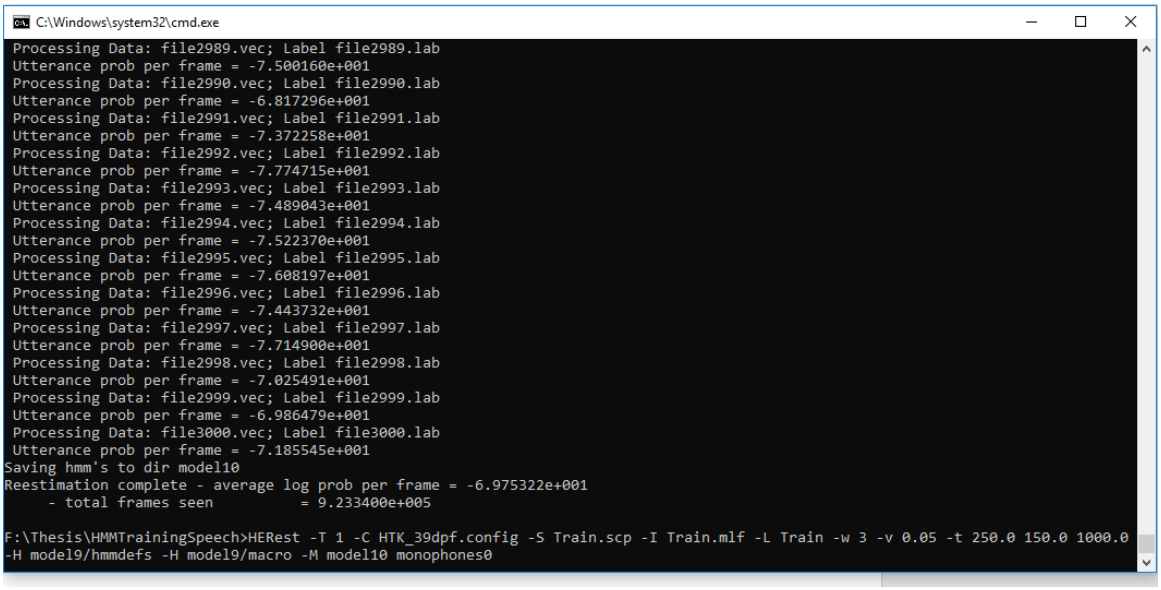

Figure 5.3: Train Model10.

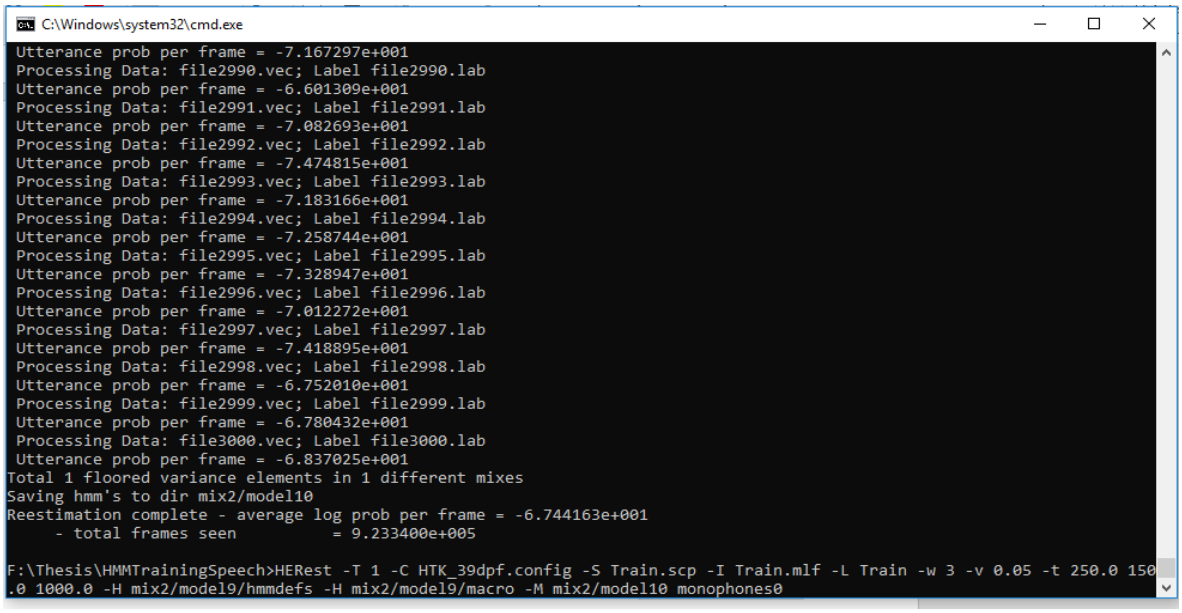

Figure 5.4: Mix2 Model10 in Train.

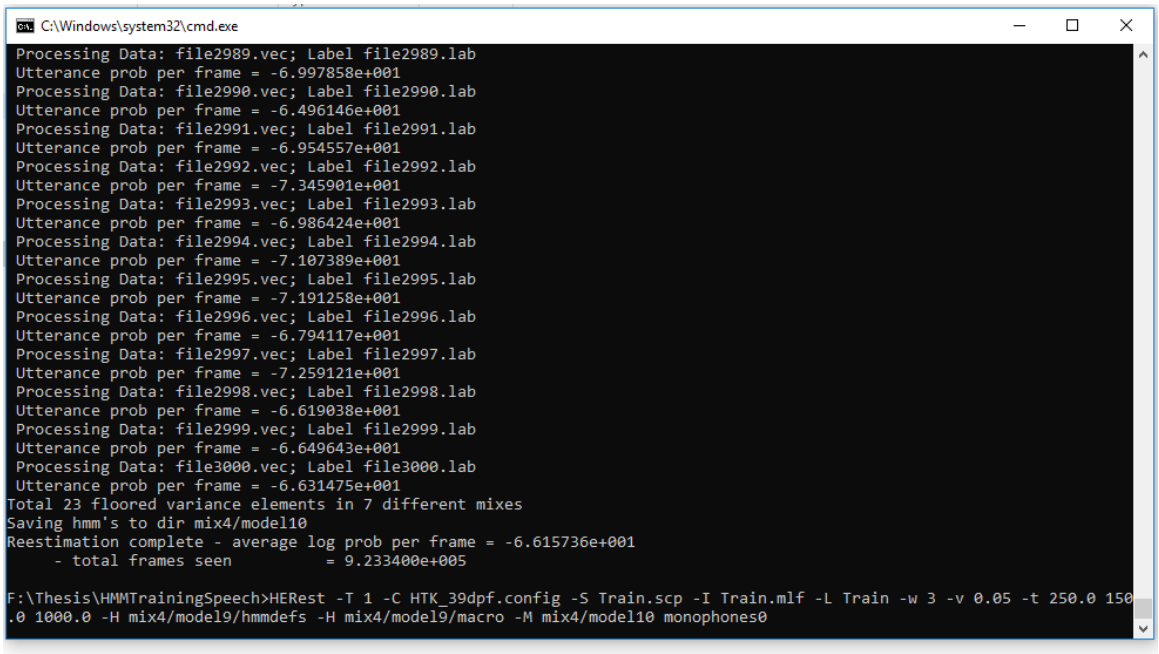

Figure 5.5: Mix4 Model10 in Train.

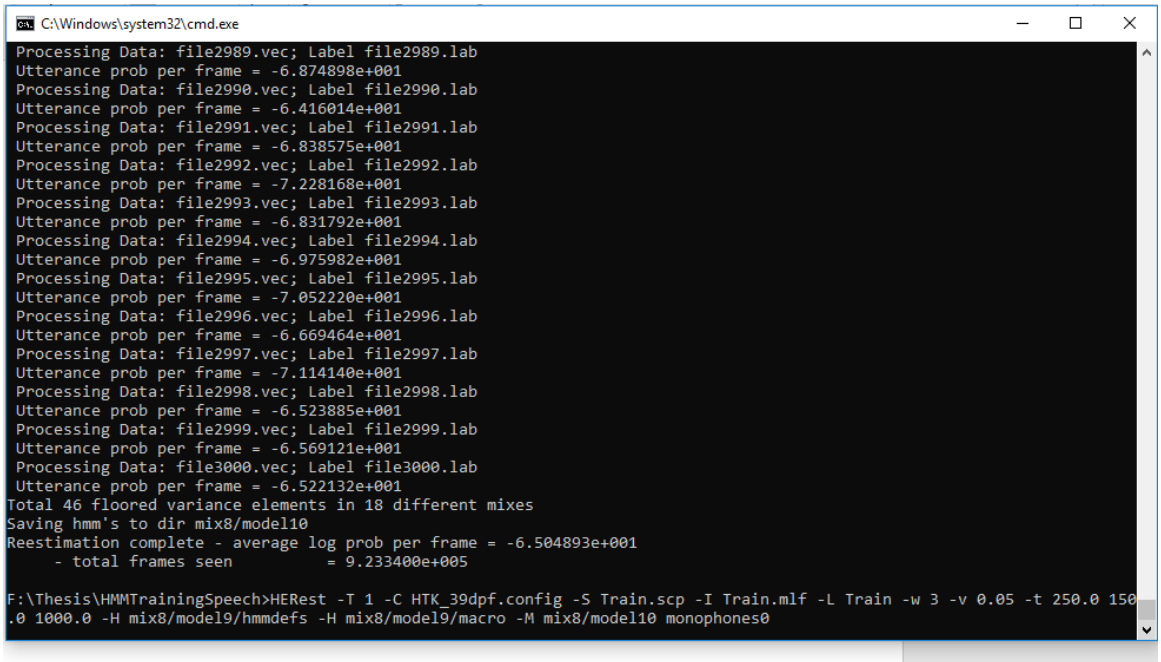

Figure 5.6: Mix8 Model10 in Train.

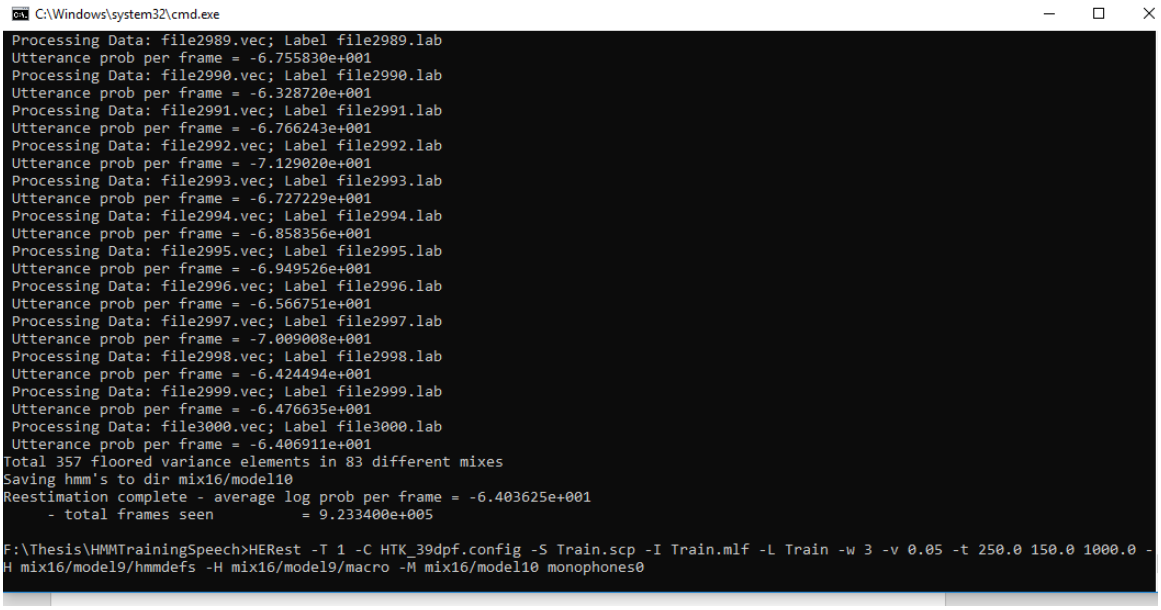

Figure 5.7: Mix16 Model10 in Train.

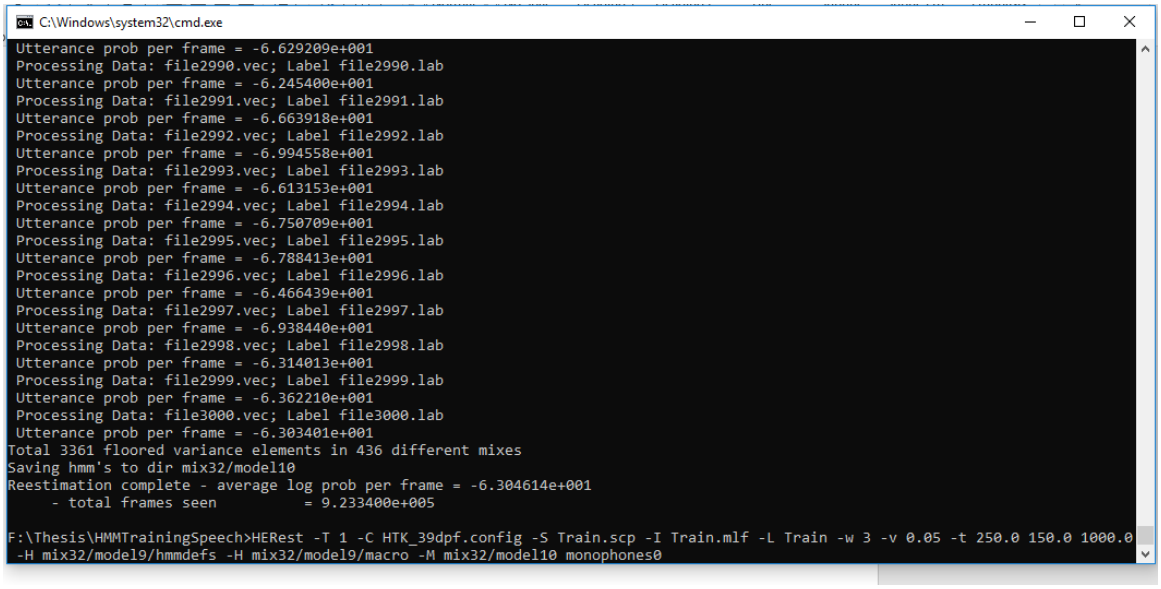

Figure 5.8: Mix32 Model10 in Train.

5.4 Experimental Results and Discussion

Segmentation of this features are silence, short silence, stop, nasal, bilabial, fricative, liquid, lenis, vowel, front, interior, back, unvoiced, long, short, diphthong, high, low, medium, round, unground and glottal AFs are depicted for ideal and real cases for utterance, From both the figures, it is observed that features segments of nasal, liquid, vowel and front are more precise nasal and unvoiced features, long, diphthong, high, low, medium, unground and glottal exhibit better segments with importance to ideal

segmentation in Again, Figure 5 shows correct rates for each of the AFs using the test utterances in data set, where AF correct rates for the corresponding AFs are 97.83%, 52.88%, 75.15%, 75.88%, 64.30%, 84.68%, 49.20%, 84.67%%, 95.72%, 87.83%, 88.22%, 78.42%, 93.79%, 87.49%, 86.65%, 82.97%, 77.82%, 70.75%, 92.62%, 86.15%, 89.00%, and 100.00%, respectively.

Phoneme Correct Rate:

Method	**Mix1**	**Mix2**	**Mix4**	**Mix8**
AF22 (Train)	10.70%	48.14%	50.21%	52.96%
AF22 (Test)	10.96%	47.27%	49.48%	51.68%

Table 5.1: Phoneme Correct Rate.

Phoneme Accuracy:

Method	**Mix1**	**Mix2**	**Mix4**	**Mix8**	**Mix16**
Close Test AF22(Train)	6.86%	12.84%	17.75%	26.22%	32.25%
Close Test AF22(Test)	7.15%	10.39%	15.17%	21.42%	25.38%

Table 5.2: Phoneme Accuracy.

HTK Results Analysis for Train:

Method	**Mix1**	**Mix2**	**Mix4**	**Mix8**
Insertion, I	3105	28550	26252	21623
Deletion, D	29790	6183	6160	6352
Substitution, S	42414	35765	34108	31697
Total, N	80856	80880	80880	80880

Table 5.3: HTK Results Analysis for Train.

HTK Results Analysis for Test:

Method	**Mix1**	**Mix2**	**Mix4**	**Mix8**
Insertion, I	919	8888	8269	7294
Deletion, D	8692	1623	1643	1698
Substitution, S	12767	11086	10532	9946
Total, N	24100	24100	24100	24100

Table 5.4: HTK Results Analysis for Test.

Chapter 6

Conclusion

We have to learn about Bangla Phonetic Features. Actually, Bangla phonetic features depend on some kind of system. We use two classifier methods with Automatic Speech Recognition (ASR) system. Bangla speech recognition need some improvement because have to some lacking. We have experiments for Bangla speech recognition for better result. We are following some criteria:

i. Mel-Frequency Cepstral Coefficients (MFCCs) system provide some improvement word correct rate and word accuracy for both of the training and test data.
ii. We can uses MFCC-39 based classifier for highly improvement for word correct rate and word accuracy for both of the training and test data.
iii. Now, we have to write normalize code and train (for 3000 data) and test (for 1000 data) by neural network with Articulatory Features (AF-22). Then, we can saw that AF-22 based system give us much better word correct rate and word accuracy then MFCC-39 based system.
iv. We have learned about how accurately text to voice or voice to speech recognition.
v. We have learned how to better speech recognition and better word accuracy also word correct rate.

So, we can do the research about Bangla Automatic Speech Recognition (ASR) system and we saw that many improvements are requirement in this research. Because, many characters are problems for text to speech recognition. We are experiments Bangla ASR system with MFCC and AF base system and we get better result give us AF then MFCC. That's why we get better word accuracy and also word correct rate. So, our research is much positive but, many improvements are possible for Bangla ASR system. It can be helps future research about Bangla ASR system for more improvement.

References

[1] M. J. Alam, N. Uzzaman, and M. Khan. N-gram based statistical grammar checker for Bangla and English. Proc. Of 9th International Conference on Computer and Information Technology (ICCIT 2006), 2006.

[2] L.R. Bhal, P. F. Brown, P. V. De Souza, and R. L. Mercer. Acoustical markov models used in the tangora speech recognition system. In Proc. IEEE Int. Conf. on Acousticals, Speech, and Signal Processing, 1988.

[3] K. H. Davies, R. Biddulph, and S. Balashek. Automatic speech recognition of spoken digits. J.Acoust. Soc. Am., 1952.

[4] L.Deng, M. Lennnnig, V.N. Gupta, and P.Mermelstein. Modeling acousticalphonetic detail in an hmm-based large vocabulary speech recognizer. IEEE Int. Conf. on Acousticals, Speech and Signal Processing, 1988.

[5] A.M. Derouault. Context- dependent phonetic markov models for large vocabulary speech recognition. IEEE Int. Conf. One Acousticals, Speech and Signal processing, 1987.

[6] N. R. Dixon and H. F. Silverman. The 1976 modular acoustical processior (map0. IEEE Trans, Acousticals, Speech and Signal Processing, 1977.