Development of Bengali Automatic Speech Recognizer and Analysis of Error Pattern

Farzana Noshin Choudhury, Tasneem Maksud Shamma, Umana Rafiq, Hasan Rahman Shuvo, Shahnewaz alam

Abstract—The very first step for ASR was taken in 1932 and still this speech recognizing technology is constantly evolving. Although Bengali is one of the largest spoken languages in the world, a few works on Bengali speech recognition has been found in different literature reviews. There are still many areas on this field that are yet to be explored for improving the performance of ASR system in Bengali. To analyze the error pattern, a speech corpus was developed and an HMM based speech recognizer was built. Audio recordings were collected from different persons in different environment. The voice recordings were used for training and testing data. Then they were auto-checked and cross-checked with each other. Finally, the result shows the percentage of speech recognition success based on comparison.

Index Terms— ASR, HMM, HTK, SPEECH, RECOGNITION, DICTIONARY.

1 INTRODUCTION

Information exchange is the key to human development. Human can communicate with machine in different ways. But it is limited when it comes to physically handicapped people. However, Speech Recognizer has emerged as a wonderful solution to this problem. Automatic Speech Recognizer (ASR) captures acoustic signal representative of speech and determines the words that were spoken by pattern matching. It can understand spoken words and converts it into text. The Error pattern can be analyzed which occur in the detection process. Voice recording, word boundary detection, feature extraction and recognition – these are the steps involved in a typical speech recognition process. Factors like accent, time duration of pause given between the words also affect it.

The general objective of this research work is to show how to build a speech recognizer using HTK toolkit which can recognize Bengali words. Training the HTK toolkit can build Bengali speech recognizer and it can recognize any word presented in the dictionary after acoustic analysis of speech signal waves.

At present, speech recognition is one of the most widely researched topics around the world. Research on ASR technology has attracted much attention over last five decades specially speech recognition by machine. The main focus has been always on English Language. Commercially developed systems include Microsoft SAPL, Dragon –Naturally-Speech, IBM via voice which are available now. Worldwide speech recognition is now involving other different languages. But Bengali is not enriched with speech recognizer and no significant research is done so far regarding Bengali speech recognizer. So this research work is just a little attempt to enrich Bengali language with a speech recognizer.

The "literature review" section gives a brief about the ASR technology. Then "Generalized Model" section tells about the generalized speech recognition process. The methodology involving Bengali speech recognizing system is discussed in the next section called "Bengali ASR". Then finally the results of this research are presented in the later part.

- *Farzana Noshin Choudhury, Electrical and Electronic Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh, E-mail: noshin.eee@gmail.com*
- *Tasneem Maksud Shamma, Electrical and Electronic Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh E-mail: tasneemmaksud@gmail.com*
- *Umana Rafiq, Electrical and Electronic Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh, E-mail: umana.rafiq@gmail.com*
- *Hasan Rahman Shuvo, Electrical and Electronic Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh E-mail: unitedbd786@gmail.com*
- *Shahnewaz Alam, Electrical and Electronic Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh E-mail: sm.shamim89@gmail.com*

2 LITERATURE REVIEW

The current ASR technology utilizes the user interface design for speech based applications. What an ASR actually does is taking speech signals as inputs, captured by microphone or telephone and then converting them into text as close as possible to the spoken data [1]. Environmental noise and different speaking ways of different persons create the difficulties for ASR system. Applications of the modern ASR technology mainly include dictation, controlling, travel information system, weather information system etc.

The type of the ASR system design mainly depends on the type of speech input the system can handle. Word spotting and the ability to support more complex grammars opens up additional flexibility in the design, but can make the design more difficult by allowing a more diverse set of responses from the user [4]. A limited form of natural language is allowed in some current ASR systems but within a specific domain.

The ASR technology is not so easy to implement. Various types of problems regarding ASR have been identified over the past few decades. Number of speakers, varying accents, nature of utterance, vocabulary size, difference between speakers these are the factors that are related to the problems occurring for ASR technology.

But these problems can be overcome quickly with further research. Despite of these problems, ASR is still popular due to its unique usefulness.

3 GENERALIZED MODEL

As different languages originated, ASR technology gained popularity day by day. Like English and other European languages, Hindi, Punjabi, Sanskrit etc have their own speech recognizer system. So a generalized model of speech recognition can be derived from all the languages possessing ASR. The basic steps involved in any speech recognition process involve Corpora, HMM, Feature Extraction and HTK toolkit.

3.1 Copora

For any speech recognition system corpus (plural-corpora) is the basic building block. It includes collection of text or speech data. The speech corpus can be enriched by including diversity, for example it can be mono or multi lingual, tagged or untagged, balanced or specialized. Example of such corpus is the British national corpus; it is the largest and best known corpora, consisting of variety of styles and subjects. Bengali corpus is developed using isolated Bengali words for Bengali speech synthesis. And it is an ever evolving Corpora.

3.2 HMM

Modern speech recognizer systems are based on Hidden Markov Model (HMM). HMMs lie at the heart of virtually all modern speech recognition systems and although the basic framework has not changed significantly in the last decade or more [3]. Firstly, it can model the short time stationary speech signal as Markov model. Secondly, it can be trained automatically. And thirdly, they are simple. In any speech recognition process, the HMM model will output a sequence of n-dimensions.

3.3 Feature Extraction

The most important part of speech recognition technique is Feature extraction. It means conversion of speech signal in its raw form to another form where characteristic information is present [2]. Speech of 'WAV" format can be converted into "MFCC" format for further use in training and testing. Different types of feature extraction techniques help to distinguish between speeches. The most used feature extraction methods in ASR technology are – Mel Frequency Cepstral Co-efficient (MFCC), Linear Prediction Coding (LPC), Perceptual Linear Prediciton (PLP), Linear Discriminant Analysis (LDA). Parameters like amplitude and energy of the signal is also extracted in feature extraction. The techniques can be different but all the extracted features should fulfill some criteria like easily measurability, no repetition, should be balanced over time, natural occurrence etc.

3.4 HTK Toolkit

HTK refers to Hidden Markov Model Toolkit which is a portable software. It builds and manipulates systems that use continuous density Hidden Markov Models. It was developed by Speech group at Cambridge University engineering Development. Manipulating, transcriptions, coding data, HMM training, Viterbi decoding, results analysis etc are performed by HTK.

In HTK construction process, at first a training database is created where each element is recorded and labeled with a corresponding word. Then in the acoustic analysis step, the trained waveforms are converted into a series of co-efficient vectors. In the third step, a prototype of HMM is defined for each element of the task vocabulary. In the next step, the HMMs are initialized and trained. After that the grammar is recognized and finally the unknown input signal is recognized.

The result is presented by HResult in a form of percentage of correctly recognized words. The sentence level accuracy is presented in the first line and the second line represents the numbers concerning the word accuracy of the transcriptions generated.

4 BENGALI ASR

This research work is done by building recognizer with HTK where MATLAB has been used for ASR. A speech recognition system basically needs a microphone for the person speak into, speech recognition software, a computer to interpret the speech, a good quality sound card and proper pronunciation. Different Gui like "WAV to VEC file converter", "Create Monophone Model", "Create Triphone Model" are used to do it. There are some basic steps involved in building Bengali ASR like building dictionary, encoding the data, training, recognition and running the recognizer live.

4.1 Building Dictionary

As for Bengali ASR, at first a dictionary is formed from Bengali phonemes. Later the dictionary is used to make sentences. Bengali phonology is mostly similar to Eastern Indo-Aryan language. As the language is alphasyllabary in nature it has two types of symbols that are vowel and consonant. Bengali consists of 7 vowels and 29 consonants.

Vowels are most important for any recognition system. It is termed as the voice sounds forming in a way where air issues in continuous stream through the pharynx and mouth without obstruction and is long in duration.

Unlike vowels, consonants are pronounced due to the obstruction by mouth and narrowing the sounds by dental, alveolar of human. They are short in duration.

This research work firstly includes making 100 sentences using Bengali dictionary. There are 40 sentences consisting 2 words, another 40 sentences consisting 3 words and finally 20 sentences consisting 4 words. These sentences are recorded by 5 different speakers through "audacity" software and then they are converted into "WAV" files for training and testing. These files actually represent speech signal waveforms.

4.2 Encoding Data

The "encoding data" step is of high significance. The speech signal waveforms are transformed into such forms that can be

handled by recognition tool. Speech waveform is a time varying linear system which is processed to apply a short time DFT or "stDFT". The speech signal contains slow and fast variations which are separated by Cepstral co-efficient. The typical duration for successive frames for the waveform is 20 to 40 ms. Same frame samples are multiplied by Hamming Function. It's a windowing function and then a vector acoustical co-efficient is extracted from each window frame. The whole conversion process is accomplished by HCopy HTK tool which creates target list file.

4.3 Training the Model

At first, initialization is needed to train the model. HInit tool is used to initialize the HMM parameters.
The command line that initializes the HMM by time alignment of the training data with a Viterbi algorithm is given below:

HInit -A -D –T 1 -S trainlist.txt -M model/hmm0 –H model/proto/protohmm -l label -L label_dir hmm_name

Table 1: Initialization of HMM Parameters

Parameter Name	Significance
hmm_name	Name of the HMM to initialize
protohmm	Description file containing the prototype of the HMM
trainlist.txt	Complete list of the .mfcc files forming the training corpus stored in (directory data/train/mfcc)
label_dir	Directory(data/train/lab) where the label files (.lab) are saved
label	Indicates which labeled segment must be used within the training corpus
model/hmm0	The directory(must be created before) where the resulting initialized HMM description will be output

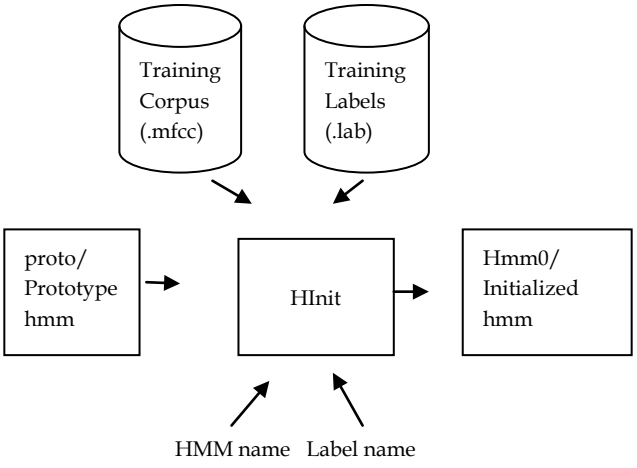


Figure 1: Initialization from a Prototype

HTK tool HRest does the training procedure. The following command line performs one re-estimation iteration with the tool mentioned above, estimating the optimal values for the HMM parameters (transition probabilities, plus mean and variance vectors of each observation function):

HRest -A -D -T 1 -S trainlist.txt -M model/hmmi
-H model/hmmi-1/hmmfile -l label -L label_dir hmm_name

Table 2: Training with HRest

Parameter Name	Significance
hmm_name	Name of the HMM to train
hmmfile	Description file of the hmm called hmm_name. It is stored in a directory whose name indicates the index of the last iteration.
trainlist.txt	The complete list of the .mfcc files forming the training corpus (stored in directory data/train/mfcc/).
label_dir	Directory where the label files (.lab) corresponding to the training corpus are stored (here: data/train/lab/).
label	Indicates the label to use within the training data.
model/hmmi	Output directory, indicates the index of the current iteration i.

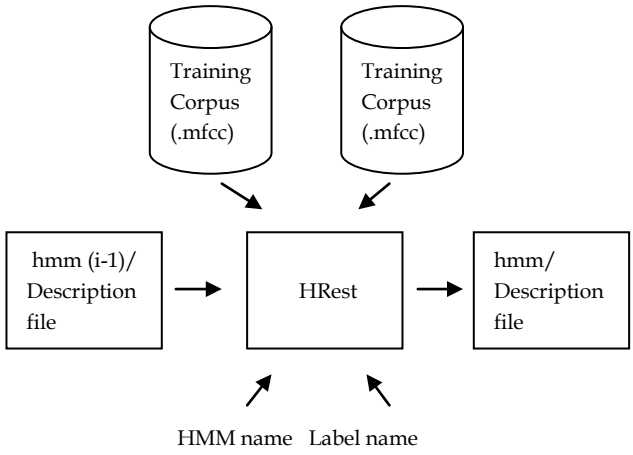
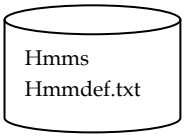


Figure 2: A Re-estimation Process

4.4 Recognition

In the recognition step all that required is to run the recognizer. Now Bengali ASR is ready for recognition.



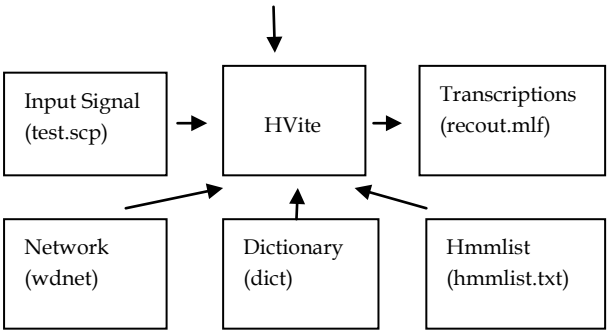


Figure 3: Speech Recognition Process [4]

HTK tool HCopy transforms the input speech signal into a series of "acoustical vectors". The result is stored into a file called test.scp. Viterbi algorithm processes the input observation. It matches it against the recognizer's markov model using the HTK tool HVite. The output is stored in a file (recout.mlf) which contains the transcription of the input. Here, recoult.mlf is the output recognition transcription file, Wdnet is the task network, dict is the task dictionary, hmmlist.txt list the name of the models to use and Test.scp is the input data to be recognized.

The recognizer is then run with live input. The source here is direct audio with sample period of 62.5 sec. The silence detector is enabled and a measurement of the background speech/silence levels was made at start-up [4]. A warning is printed when the silence measurement is being made. After setting up the configuration file for direct input, the HTK tool HVite was again used to recognize the live in put using a microphone.

The steps discussed in this section are essential for the process and any error in a single step can be the reason for failure of the mechanism.

5 Results

HTK tool HResult analyzes the overall performances of the ASR Speech recognition is performed by HVite and it obtains transcriptions for testing. HResult computes recognized statistics such as percentage of correctly recognized words. In the recognition statistics, the first line represents the sentence level accuracy. The second line represents numbers concerning the word accuracy of the transcriptions generated.

Here, H=Number of Correct Words
D=Number of Deletions
S=Number of Substitutions
I=Number of Insertions
N=Total number of words in reference transcriptions

So, the percentage of correctly recognized words: Correct=H/N ×100%

& the word recognition accuracy is given by: Accuracy= H-I/N×100%

Successful completion of the steps leads to the results where error pattern can be detected and compared. Voice signal samples of five users have been used. The tests for checking error pattern are of two types – self test (where 'train' and 'test' samples are from the same person) and cross test (where 'train' and 'test' samples are from different person). Errors found in cross tests are larger compared to the self tests. Error types mainly include spelling errors and wrong sentence detection. The errors mainly occurred for complicated words and conjugated letters. The summarized result is presented below:

Table 3: Combined Error Analysis (Train-sels, Test-self)

Speakers	Sent (%Corr)	H=	S=	Word (%Corr)	Acc	Error Types
Speaker_1	100	100	0	100	100	Spelling Errors:13
Speaker_2	99	99	1	100	99.64	Spelling Errors:13 Wrong Sentence:1
Speaker_3	100	100	0	100	100	Spelling Errors:12
Speaker_4	87	87	13	88.978	87.90	Spelling Errors:17 Wrong Sentences: 9
Speaker_5	95	95	5	96.80	96.44	Spelling Errors:13 Wrong Sentences:4

Table 4: Combined Error Analysis (Cross-test)

SPEAKERS	HTK RESULT ANALYSIS
SPEAKER_2_TRAIN_SPEAKER_1_TEST	SENT: %CORR =77, H=77, S=2, WORD: %CORR=76.16, ACC=75.44
SPEAKER_1_TRAIN_SPEAKER_2_TEST	SENT: %CORR =59, H=59, S=41, WORD: %CORR=59.07, ACC=56.94
SPEAKER_3_TRAIN_SPEAKER_1_TEST	SENT: %CORR =19, H=19, S=81, WORD: %CORR=23.13, ACC=21.71
SPEAKER_1_TRAIN_SPEAKER_3_TEST	SENT: %CORR =7, H=7, S=93, WORD: %CORR=7.83, ACC=4.98
SPEAKER_5_TRAIN_SPEAKER_1_TEST	SENT: %CORR =2, H=2, S=98, WORD: %CORR=3.91, ACC=1.78
SPEAKER_1_TRAIN_SPEAKER_5_TEST	SENT: %CORR =1, H=2, S=99, WORD: %CORR=2.14, ACC=11.74
SPEAKER_4_TRAIN_SPEAKER_1_TEST	SENT: %CORR =1, H=1, S=99, WORD:%CORR=1.42, ACC=0.0
SPEAKER_1_TRAIN_SPEAKER_4_TEST	SENT: %CORR=1, H=1, S=99, WORD: %CORR=1.42,ACC=0.36
SPEAKER_2_TRAIN_SPEAKER_4_TEST	SENT; %CORR =1, H=1, S=99, WORD: %CORR=2.14, ACC=11.74
SPEAKER_4_TRAIN_SPEAKER_2_TEST	SENT: %CORR =0, H=0, S=100, WORD: %CORR=1.78, ACC=1.42
SPEAKER_2_TRAIN_SPEAKER_5_TEST	SENT: %CORR =1, H=1, S=99, WORD: %CORR=2.14, ACC=12.1
SPEAKER_5_TRAIN_SPEAKER_2_TEST	SENT: %CORR =1, H=1, S=99, WORD: %CORR=2.14, ACC=1.78
SPEAKER_2_TRAIN_SPEAKER_3_TEST	SENT: %CORR =5, H=5, S=95, WORD: %CORR=6.41, ACC=3.56
SPEAKER_3_TRAIN_SPEAKER_2_TEST	SENT: %CORR =16, H=16, S=84, WORD: %CORR=17.08, ACC=15.3
SPEAKER_3_TRAIN_SPEAKER_5_TEST	SENT: %CORR =1, H=1, S=99, WORD: %CORR=0.71, ACC=0.71
SPEAKER_5_TRAIN_SPEAKER_3_TEST	SENT: %CORR =2, H=2, S=98, WORD: %CORR=3.91, ACC=7.83
SPEAKER_3_TRAIN_SPEAKER_4_TEST	SENT: %CORR =3, H=3, S=97, WORD: %CORR=3.2, ACC=1.07
SPEAKER_4_TRAIN_SPEAKER_3_TEST	SENT: %CORR =1, H=1, S=99, WORD: %CORR=2.49, ACC=1.49
SPEAKER_5_TRAIN_SPEAKER_4_TEST	SENT: %CORR =1, H=1, S=99, WORD: %CORR=2.49, ACC=6.05
SPEAKER_4_TRAIN_SPEAKER_5_TEST	SENT: %CORR =2, H=2, S=98, WORD: %CORR=1.78, ACC=1.42

6 CONCLUSION

In this research the main task was to develop an automatic speech recognizer for Bengali and analyze the error pattern. Total 500 voice samples from 5 different speakers were taken as input data. They were compared with each other by training and testing. The results summarize that there is a huge difference in the same voice waveform used for training and testing versus two different voices used for training and testing. The system performance is optimum when the same voice input is used as training data and testing data. The error occurs only when the sentences are spoken incorrectly or if they are not spoken sequentially. Even if the word recognition process has a 100% success still some errors are found to be persistent. The errors are mainly because of the spelling mistakes in the words in the .mlf file. The recognizer cannot find similarities between the words in the .mlf file to which it is comparing with and the original words in the dictionary.

In the past, speech recognizers were only used for dictating or performing command. But now its applications are not li-

mited. It can likely be used to send instant messages, to annotate and comment, to keep real time transcripts during conversations, to instruct and answer computers in a hands-free environment like driving cars and eventually for general computer iteration e.g. the Linguistic User Interface (LUI). So, in future further research in ASR technology will introduce wider range of applications.

APPENDIX: 100 SENTENCES

1. AALOCHONA AACHE
2. AAMRA BANGLADESHI
3. BIDDUT ACHE
4. AUSTRELIAY GECHHE
5. SONDHAN DAU
6. DHAKAY AACHE
7. NOTUN NATTO
8. DURGHOTONAY NIHOTO
9. MONOROM DRISHSHO
10. SHANTI AACHE
11. LALON SONGIT
12. AAGUN AACHE
13. MACHH UTPADON
14. BORO DESH
15. NIRBHORJOGGO MEYE
16. AMDANI BONDHO
17. UTTOR DAU
18. KOTO DORSHOK
19. EKSHATHE GAIBE
20. SUJOG AACHE
21. JELA WOPOJELAI
22. BINODON BAABOWSTHAY
23. WONNISSHO EKATTOR
24. BOITHOK SOBHA
25. BANGLA VASHA
26. SOMODHIKAR BASTOBAYONE
27. BIDDUT PROKOLPO
28. AMDANI ROPTANI
29. MORICHIKAR MOTO
30. NIRBHORJOGGO SHOMORTHON
31. BANGLADESHER DRISHSHO
32. BISSHO SHANTI
33. CHADER AALO
34. NIRBACHONER SHOPOCKHE
35. KHULNA SHOHOR
36. GREPTAR AIN
37. JHUDDHE JABO
38. HOSTOKKEPER JONNO
39. GORIB CHHELER
40. PARBOTTO CHOTTOGRAM
41. ADHIKAR BASTOBAYONER ABEDON
42. ADHIKAR BASTOBAYONER SHOPNO
43. NARI SHIKKHA NITI
44. NARI SAMMELON SHURU
45. TUMI NIRBHORJOGGO NOU
46. AUSTRALIAY SHANTI AACHHE
47. BIBHINNO BIDDALOYER BOSTU
48. BANGLADESH AMAR DESH
49. BIBHINNO TOTTHOU SONDHAN
50. BIBHINNO MUKHI AALOCHONA
51. SORASORI SONGIT SOMPROCHAR
52. BONGER SOINIK SOMMELONE
53. SROMIKER ADHIKARER DABI
54. BANGLADESHER MUKTIR DABI
55. NARI SOMODHIKAR SONGGOTHON
56. NOTUN BABSHA SHURU
57. SHABEK BIRODHIDER SOBHA
58. BOSONTER UTSAB SURU
59. SHANTIR SOMMELON SHURU
60. APONAR BORO DAITTO
61. SHIKHKHA BABOWSTHAY EGIYE
62. BIBHINNO KARJOKROM BASTOBAYONE
63. JHUDDHE JABE SHENA
64. AMDANI ROPTANI KARI
65. BISSHO EBONG PROJUKTI
66. ROBINDRO SONGIT SHAROWN
67. PONNO UTPADON SHURU
68. CHITRO MILIYE DEBE
69. PORIKHHAR JONNO SHOMORTHON
70. AIN SONGGOTHON NIONTRONE
71. CHITRO NIRMATA JABE
72. ESID DOGDHO MEYE
73. POTRIKAY NOTUN PROTIBEDON
74. SHARA DESHE BRISTI
75. PLASTIK PONNO BIKROY
76. PURONO TOTTHOU WODDHAR
77. BANGLA AMAR VASHA
78. WOPOJELAI TODONTO HOBE
79. SORASORI SOMPROCHAR HOCHCHHE
80. BOSHONTER LAL RONGE
81. AUSTRELIAY BAARLO BANGLADESHER BABSHA
82. DHAKAY BIDROHIDER BISHAL SOMMELONE
83. CHOTTOGRAM SOHOR KOTO BORO
84. BIDESHI BONDUK AMDANI BONDHO
85. NARI NITI BASTOBAYONER DABI
86. LALON MELA SURER UTSAB
87. DHAKA DINAJPUR THEKE BORO
88. TUMI AMAR MEYE NA
89. MILI AMAR CHELER MAA
90. BIDROHIDER DABI NIROBICHCHHINO BIDDUT
91. ESID DURGHOTONAY NARI NIHOTO
92. AMAR DESH EGIYE GECHHE
93. NIRAPOD SHETU BANGLADESHER DABI
94. NIRBHORJOGGO NIRBACHON MOORICHIKAR MOTO
95. SUNDORBONE O MODHUMOTI BORO MONOROM
96. DESHE DHORMOGHOT DITE BOITHOK
97. TARA EKSATHE CHOTTOGRAM GECHHE
98. MATRO SHURJO MILIYE GELO

99. SANGBADIKDER SOMMELONE SOBHA SHURU
100. MOBAIL NETWORKING NOTUN NOY

ACKNOWLEDGMENT

This research work was supervised by Mr. Jakaria Rahimi, Assistant Professor, Department of Electrical and Electronic Engineering, Ahsanullah University of Science and Technology. Without his help it could not have been possible to accomplish this research.

REFERENCES

[1] Mohit Dua, R.K.Aggarwal, Virender Kadyan, Shelza Dua, Punjabi Automatic Speech Recognition Using HTK, International Journal of Computer Science Issues, Vol. 9, Issue 4, No 1, July 2012
[2] Jitendra Singh Pokhariya and Dr. Sanjay Mathur, Sanskrit Speech Recognition using Hidden Markov Model Toolkit, International Journal of Engineering Research and Technology, Vol. 3, Issue 10, October 2014
[3] Mark Gales and Steve Young, The Application of Hidden Markov Models in Speech Recognition, Foundation and Trends in Signal Processing, Vol. 1, No. 3, 2007, DOI: 10.1561/2000000004
[4] Muhirwe Jackson, Automatic Speech Recognition: Human Computer Interface for Kinyarwanda Language, A project report submitted in partial fulfillment of the requirements for the award of the degree of Master of Science in Computer Science., Makerere University, August 2005.
[5] Historical Development and Future Directions in Speech Recognition and Understanding by Janet M. Baker, Li Deng, Sanjeev Khudanpur, Chin-Hui Lee, James Glass, and Nelson Morgan.
[6] Baum, L.E., and Petrie, T., (1966). Statistical Inference for Probabilistic functions of Finite-State Markov Chains, Annotated Mathematical Statistics,37:1554-1563.
[7] Bridle, J., Deng, L., Picone, J., Richards, H., Ma, J., Kamm, T., Schuster, M., Pike, S., Reagan, R., 1998. An investigation of segmental hidden dynamic models of speech coarticulation for automatic speech recognition. Final Report for the 1998 Workshop on Language Engineering, Center for Language and Speech Processing at Johns Hopkins University, pp. 161.
[8] Cole, R. Noel, M. Burnet, D.C., Fanty,M., Lander, T., Oshika, B., Sutton ,S., 1994 Corpus development activities at the center for spoken language understanding. Human Language Technology Conference archive, Proceedings of the workshop on Human Language Technology. Pages: 31 - 36 .
[9] R. Cole, K. Roginski, and M. Fanty.,1992 A telephone speech database of spelled and spoken names. In ICSLP'92, volume 2, pages 891–895.
[10] Deshmukh, N., Ganapathiraju, A, Picone J., (1999), Hierarchical Search for Large Vocabulary Conversational Speech Recognition. IEEE Signal Processing Magazine, 1(5):84-107
[11] Liu, F.H., Liang G., Yuqing G. AND Picheny, M, (2004). Applications of Language Modeling in Speech-To-Speech Translation INTERNATIONAL JOURNAL OF SPEECH TECHNOLOGY (7):221-229.
[12] Ma, J., Deng, L., 2004. Target-directed mixture linear dynamic models for spontaneous speech recognition. IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 12, NO. 1, JANUARY 2004
[13] Mane, A., Boyce,S., Karis,D.,Yankelovich,N., (1996) Designing the User Interface for Speech Recognition Applications SIGCHI Bulletin 28(4):29-34.
[14] Mengjie, Z., (2001) Overview of speech recognition and related machine learning techniques, Technical report. retrieved December 10, 2004 from http://www.mcs.vuw.ac.nz/comp/Publications/archive/CS-TR-01/CS-TR-01-15.pdf
[15] [Mori R.D, Lam L., and Gilloux M. (1987). Learning and plan refinement in a knowledgebased system for automatic speech recognition. IEEE Transaction on Pattern Analysis Machine Intelligence, 9(2):289-305.
[16] Picheny, M., (2002). Large vocabulary speech recognition, IEEE Computer, 35(4):42-50.
[17] Pinker, S., (1994), The Language Instinct, Harper Collins, New York City, New York, USA.
[18] Rabiner L.R., S.E.L.evinson: (1981) "Isolated and connected word recognition – Theory and selected applications", IEEE Trans. COM-29, pp.621-629
[19] Rabiner, L., R., and Wilpon, J. G., (1979). Considerations in applying clustering techniques to speaker-independent word recognition.Journal of Acoustic Society of America.66(3):663-673.
[20] Nidhi Desai, Prof. Kinnal Dhameliya & Prof. Vijayendra Desai, "Feature Extraction and Classification Techniques for Speech Recognition : A Review", International Journal of Emerging Technology and Advanced Engineering(IJETAE) on December 2013
[21] Shanti Therese S. & Chelpha Lingam, "Review of Feature Extraction Techniques in Automatic Speech Recognition", International Journal of Scientific Engineering and Technology, 1 June 2013
[22] Mark Gales & Steve Young, "The Application of Hidden Markov Models in Speech Recognition", Fundamental and Trends in Signal Processing
[23] Urmilla Shrawankar & Dr. Vilas Thakare, "Techniques for Feature Extraction in Speech Recognition System : A Comperative Study"
[24] Deller, J.,J. Proakis & & J. Hansen, "Discrete-Time Processing of Speech Signals", New York, Macmillan Publishing, 1993
[25] Jeinek, F. , "Statistical Methods for Speech Processing", Language,Speech & Communication Series, Cambridge, MA : MIT Press, 1997
[26] Rabiner, L., "A Tutorial on Hidden Markov Models & Selected Applications in Speech Recognition", Proceedings of the IEEE 77, 1989
[27] B. –H. Juang., "Fundamentals of Speech Recognition" ,Prentice Hall Signal Processing Series, Englewood Cliffs, NJ: Prentice Hall,1993
[28] Smyth, P., D. Heckerman & M.I. Jordan, "Probabilistic Independence Networks for Hidden Markov Probability Models", Neutral Computation 9.227-270, 1997
[29] Morgan, N., & B. Gold., "Speech and Audio Signal Processing", Wiley Press
[30] Robertson, J., Wong, Y.T., Chung, C., and Kim, D.K., (1998) Automatic Speech Recognition for Generalised Time Based Media Retrieval and Indexing, Proceedings of the sixth ACM International Conference on Multimedia(pp 241-246) Bristol, England.